Paul Muller
Yusuf Leblebici

ACSP
Analog Circuits and Signal Processing

CMOS Multichannel Single-Chip Receivers for Multi-Gigabit Optical Data Communications

Springer

CMOS MULTICHANNEL SINGLE-CHIP RECEIVERS FOR MULTI-GIGABIT OPTICAL DATA COMMUNICATIONS

ANALOG CIRCUITS AND SIGNAL PROCESSING SERIES

Consulting Editor: Mohammed Ismail. *Ohio State University*

Titles in Series:

ULTRA LOW POWER CAPACITIVE SENSOR INTERFACES
Bracke, W., Puers, R. (et al.)
ISBN: 978-1-4020-6231-5

BROADBAND OPTO-ELECTRICAL RECEIVERS IN STANDARD CMOS
Hermans, C., Steyaert, M.
ISBN: 978-1-4020-6221-6

CMOS MULTICHANNEL SINGLE-CHIP RECEIVERS FOR MULTI-GIGABIT OPTICAL DATA COMMUNICATIONS
Muller, P., Leblebici, Y.
ISBN: 978-1-4020-5911-7

SWITCHED-CAPACITOR TECHNIQUES FOR HIGH-ACCURACY FILTER AND ADC...
Quinn, P.J., Roermund, A.H.M.v.
ISBN: 978-14020-6257-5

LOW-FREQUENCY NOISE IN ADVANCED MOS DEVICES
von Haartman, M., Östling, M.
ISBN: 978-1-4020-5909-4

CMOS SINGLE CHIP FAST FREQUENCY HOPPING SYNTHESIZERS FOR WIRELESS MULTI-GIGAHERTZ APPLICATIONS
Bourdi, Taoufik, Kale, Izzet
ISBN: 978-1-4020-5927-8

ANALOG CIRCUIT DESIGN TECHNIQUES AT 0.5V
Chatterjee, S., Kinget, P., Tsividis, Y., Pun, K.P.
ISBN-10: 0-387-69953-8

IQ CALIBRATION TECHNIQUES FOR CMOS RADIO TRANCEIVERS
Chen, Sao-Jie, Hsieh, Yong-Hsiang
ISBN-10: 1-4020-5082-8

FULL-CHIP NANOMETER ROUTING TECHNIQUES
Ho, Tsung-Yi, Chang, Yao-Wen, Chen, Sao-Jie
ISBN: 978-1-4020-6194-3

THE GM//ID DESIGN METHODOLOGY FOR CMOS ANALOG LOW POWER INTEGRATED CIRCUITS
Jespers, Paul G.A.
ISBN-10: 0-387-47100-6

PRECISION TEMPERATURE SENSORS IN CMOS TECHNOLOGY
Pertijs, Michiel A.P., Huijsing, Johan H.
ISBN-10: 1-4020-5257-X

CMOS CURRENT-MODE CIRCUITS FOR DATA COMMUNICATIONS
Yuan, Fei
ISBN: 0-387-29758-8

RF POWER AMPLIFIERS FOR MOBILE COMMUNICATIONS
Reynaert, Patrick, Steyaert, Michiel
ISBN: 1-4020-5116-6

ADVANCED DESIGN TECHNIQUES FOR RF POWER AMPLIFIERS
Rudiakova, A.N., Krizhanovski, V.
ISBN: 1-4020-4638-3

CMOS CASCADE SIGMA-DELTA MODULATORS FOR SENSORS AND TELECOM
del Río, R., Medeiro, F., Pérez-Verdú, B., de la Rosa, J.M., Rodríguez-Vázquez, A.
ISBN: 1-4020-4775-4

SIGMA DELTA A/D CONVERSION FOR SIGNAL CONDITIONING
Philips, K., van Roermund, A.H.M.
Vol. 874, ISBN: 1-4020-4679-0

CALIBRATION TECHNIQUES IN NYQUIST A/D CONVERTERS
van der Ploeg, H., Nauta, B.
Vol. 873, ISBN: 1-4020-4634-0

ADAPTIVE TECHNIQUES FOR MIXED SIGNAL SYSTEM ON CHIP
Fayed, A., Ismail, M.
Vol. 872, ISBN: 0-387-32154-3

WIDE-BANDWIDTH HIGH-DYNAMIC RANGE D/A CONVERTERS
Doris, Konstantinos, van Roermund, Arthur, Leenaerts, Domine
Vol. 871, ISBN: 0-387-30415-0

METHODOLOGY FOR THE DIGITAL CALIBRATION OF ANALOG CIRCUITS AND SYSTEMS: WITH CASE STUDIES
Pastre, Marc, Kayal, Maher
Vol. 870, ISBN: 1-4020-4252-3

HIGH-SPEED PHOTODIODES IN STANDARD CMOS TECHNOLOGY
Radovanovic, Sasa, Annema, Anne-Johan, Nauta, Bram
Vol. 869, ISBN: 0-387-28591-1

CMOS Multichannel Single-Chip Receivers for Multi-Gigabit Optical Data Communications

By

PAUL MULLER
Marvell Semiconductor, Etoy, Switzerland

and

YUSUF LEBLEBICI
École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

Springer

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN 978-1-4020-5911-7 (HB)
ISBN 978-1-4020-5912-4 (e-book)

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

www.springer.com

Printed on acid-free paper

About the Authors

Paul Muller received the engineering degree (diploma) in electrical engineering from the École Polytechnique Fédérale de Lausanne (EPFL) in 1999. From 1999 to 2002, he worked as a mixed-signal design engineer at XEMICS (now part of Semtech Corp.), where he contributed to several sensing and data-acquisition circuit designs. In 2002, he joined the Microelectronic Systems Laboratory (LSM) at EPFL as a research assistant, where he obtained his Dr. Sc. degree in electrical engineering in July 2006. His thesis research was on the modeling and design of multichannel gigabit receivers for short-distance optical communication interfaces, including design, modeling and test of transimpedance amplifiers, limiting amplifiers and clock and data recovery circuits. Since 2006, he is with Marvell Switzerland Sàrl., a subsidiary of Marvell Technology Group Ltd. He has published a number of papers in international conferences in the field and is a member of the IEEE.

Yusuf Leblebici received the B.S. and M.S. degrees in electrical engineering from Istanbul Technical University (ITU) in 1984 and 1986, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Illinois at Urbana-Champaign in 1990. Between 1991 and 2001, he worked as a faculty member at the University of Illinois at Urbana-Champaign, at Istanbul Technical University, and at Worcester Polytechnic Institute (WPI) in Massachusetts. From 2000 to 2001, he took the responsibility of developing the microelectronics degree program at Sabanci University, while also continuing his academic involvement at WPI during this time, as an affiliate associate professor. Since January 2002, Dr. Leblebici is a full (chair) professor at the École Polytechnique Fédérale de Lausanne (EPFL) and director of the Microelectronic Systems Laboratory.

Dr. Leblebici is the author or coauthor of more than 150 scientific articles published in international journals and conferences, and two textbooks. He has served on the organizing and steering committees of several international conferences in the field of integrated circuits. He is a senior member of the IEEE, served as associate editor of IEEE Transactions on Circuits and Systems II and associate editor of IEEE Transactions on VLSI.

Foreword

The intention of this book is to address a number of timely, performance-critical issues within the field of short-distance optical communications, from a circuit designer's perspective. It discusses the major trade-offs the designer has to deal with in the development of monolithically integrated receivers in CMOS technologies. As such, it is based on Dr. Muller's doctoral dissertation entitled "A Standard CMOS Multi-Channel Single-Chip Receiver for Multi-Gigabit Optical Data Communications", submitted to the School of Engineering of the École Polytechnique Fédérale de Lausanne (EPFL) in May 2006. The dissertation material has been enhanced by the presentation of a number of alternative design approaches and circuit topologies, providing exhaustive coverage of the state of the art in optical short-distance receiver circuit design.

The need for a new processor input/output (I/O) interface paradigm is dictated by ongoing technology scaling and the advent of multi-core systems. Indeed, each new generation of microprocessors and digital signal processors provides higher computing power and data throughput, whereas the available bandwidth of the I/O interfaces is subject to much slower growth. Moving beyond upcoming serial links to an optical data link paradigm for very short-distance (board-to-board and chip-to-chip communications allows for considerable I/O interface bandwidth enhancement. Fully integrated silicon CMOS receivers are considered to be the technology of choice to lead this solution to economic success, because monolithic integration results in lower volume-manufacturing cost, improved yield and reduced assembly and test expenses.

This book provides the reader with the necessary background knowledge to fully understand the trade-offs in optical short-distance communication receiver design. Thorough discussion of the presented material guides the reader in his design choices and leads to in-depth understanding of the design

trade-offs he is facing. An exhaustive list of references and suggestions for further reading complement the technical material presented in the book.

With its timely and up-to-date content, this book could be categorized as required reading for practicing engineers and researchers in the field of short-distance optical communications and optical CMOS receiver design, as well as for graduate students and photonics engineers working on high-speed photodetection systems.

This book would not have taken shape without the input and feedback of many people, which directly or indirectly influenced the research on which it is based. We are indebted to Professor M. Selim Ünlü and Dr. Matthew Emsley from Boston University, who laid the foundations of this project with their work on resonant cavity enhanced photodetectors. Particular thanks go to Dr. Armin Tajalli from the École Polytechnique Fédérale de Lausanne (EPFL) for his contribution to the clock recovery analysis and design. We would also like to express our profound gratitude to Dr. Thomas Morf and Dr. Martin Schmatz, from IBM Zurich Research Laboratory for their very valuable and regular feedback on this work, as well as Mr. Larry DeVito from Analog Devices for the tremendous review of the dissertation manuscript. We greatly appreciated the collaboration and discussion with the staff of IBM Research in Zurich, researchers from the École Centrale de Lyon and fellow researchers from the École Polytechnique Fédérale de Lausanne.

Finally, we would like to gratefully acknowledge Mark de Jongh at Springer Science + Business Media for giving us the opportunity to publish this work. Our thanks also go to Mrs. Cindy Zitter, Editorial Assistant, supporting us in the publishing process and the whole Springer staff committed to this work.

Paul Muller

Etoy, May 2007

Yusuf Leblebici

Lausanne, May 2007

Table of Contents

About the Authors . . . . . *v*

Foreword . . . . . *vii*

Table of Contents . . . . . *ix*

Constants, Symbols and Acronyms . . . . . *xiii*

CHAPTER 1 *Introduction* ***1***

CHAPTER 2 *Integrated Photonic Systems* ***5***

2.1 Long-Haul Communication Links **6**
2.2 Metropolitan-Area Networks **6**
2.3 Local Area Networks and Short-Distance Interconnects **7**
2.4 Optical Backplane Technology **9**
2.5 Optical on-chip Interconnects **10**

CHAPTER 3 *Basic Concepts* ***13***

3.1 Modulation of Optical and Electrical Signals **15**
3.2 NRZ Random Data **15**
3.3 Clock Recovery Basics **18**
3.4 Bit Error Ratio **19**
3.5 System Bandwidth and Inter-Symbol Interference **19**
3.6 Amplitude Noise **21**
3.7 Jitter **22**

3.7.1 *Jitter Contributions* ... 22
3.7.2 *Jitter Specifications* ... 24
3.7.3 *Regenerators and Timing Repeaters* ... 24
3.8 Multichannel Systems ... 25
3.9 Definition of Transistor-Level Conventions ... 26
3.9.1 *Large-Signal Model* ... 27
3.9.2 *Small-Signal Model* ... 28

CHAPTER 4 *System-Level Specifications* ... 29

4.1 Technology ... 30
4.2 System-Level Requirements ... 32
4.3 Receiver System Specifications ... 33
4.4 Subblock Parameters ... 35
4.5 Transimpedance Amplifier Analysis ... 36
4.5.1 *Transfer Function* ... 37
4.5.2 *Noise Model* ... 39
4.5.3 *Specification of Design Parameters* ... 41
4.6 System Gain and Bandwidth Specifications ... 42
4.7 Bit Error Ratio Evaluation ... 43
4.7.1 *Vertical Eye Closure Contribution* ... 44
4.7.2 *Horizontal Eye Closure Contribution* ... 47
4.7.3 *Combined Horizontal–Vertical Eye Closure Estimation* ... 49
4.8 Block Specification Flow ... 53

CHAPTER 5 *Silicon Photodetectors* ... 57

5.1 Photodetection ... 57
5.2 PIN Photodiodes ... 59
5.3 Avalanche Photodiodes ... 60
5.4 Differential N-Well Detector ... 61
5.5 Equalized PIN Detector ... 63
5.6 Resonant Cavity-Enhanced Detector ... 65
5.6.1 *The Resonant Cavity Structure* ... 65
5.6.2 *Distributed Bragg Reflectors* ... 66
5.6.3 *Detector Fabrication* ... 67
5.7 Metal–Semiconductor–Metal Photodetectors ... 70
5.8 Lateral Photodetectors ... 70
5.9 Photodetector Characterization ... 71

CHAPTER 6 *Transimpedance Amplifier Design* ... 73

6.1 Principles of I-V Conversion ... 73
6.2 Transimpedance Amplifier Topologies ... 76
6.2.1 *Common-Gate Input Stage Amplifiers* ... 76
6.2.2 *Voltage Amplifier Based TIAs* ... 77
6.2.3 *Source Follower Output Stages* ... 79

6.2.4 *Topology Discussion* 80
6.3 Specifications 82
6.4 Transimpedance Amplifier Design 82
6.4.1 *Description of the TIA Amplifier* 82
6.4.2 *Design Procedure* 83
6.4.3 *DC Compensation* 85
6.4.4 *Complete Block Diagram* 87
6.4.5 *Input Noise Estimation* 89
6.5 Block Layout 91
6.5.1 *Channel Constraints* 91
6.6 Circuit Measurement 92
6.6.1 *Frequency-Domain Measurements* 92
6.6.2 *Time-Domain Measurements* 93

CHAPTER 7 *Limiting Amplifier Design* 95

7.1 Principles of Signal Limiting 95
7.2 Simple Limiting Amplifier Topologies 97
7.2.1 *Cascade of Gain Stages* 97
7.2.2 *Optimization Theory* 98
7.2.3 *Group Delay* 98
7.2.4 *Basic Gain Stage* 99
7.3 Bandwidth Enhancement in Limiting Amplifiers 100
7.3.1 *Inductive Peaking* 100
7.3.2 *Active Inductors* 101
7.3.3 *Capacitive Degeneration* 102
7.3.4 *Negative Miller Capacitance* 103
7.3.5 *The Cherry–Hooper Amplifier* 103
7.3.6 *Active Feedback Structure* 106
7.3.7 *Interstage Scaling Technique* 107
7.3.8 *Topology Discussion* 107
7.4 Specifications 108
7.5 Inductorless Limiting Amplifier Design 108
7.5.1 *Interstage Scaling Optimization* 108
7.5.2 *Transistor Dimensioning* 110
7.5.3 *Discussion* 111
7.6 Design of an Inductive Peaking Limiting Amplifier 112
7.6.1 *The Inductor* 112
7.6.2 *Magnetic Coupling Coefficient* 113
7.6.3 *Design of the Amplifier Core* 115
7.6.4 *Discussion* 116
7.7 Complete Limiting Amplifier 117
7.7.1 *Offset Compensation* 117
7.7.2 *Complete Block Diagram* 119
7.7.3 *Limiting Amplifier Noise Estimation* 119
7.8 Block Layout 120
7.9 Circuit Measurements 121
7.9.1 *Measurement Setup for Time-Domain Measurements* 121
7.9.2 *Eye Diagram Measurements* 123
7.9.3 *Magnetic Coupling Measurements* 124

7.9.4 Measurement Summary . . . 125
7.10 Discussion . . . 125

CHAPTER 8 *Clock and Data Recovery Circuit . . . 127*

8.1 Clock Recovery Principles . . . 128
8.1.1 NRZ Data Phase Detection . . . 130
8.1.2 Linear Phase Detector . . . 131
8.1.3 Binary Phase Detector . . . 133
8.2 CDR Topologies . . . 135
8.2.1 PLL-Based Topologies . . . 136
8.2.2 DLL-Based Topologies . . . 137
8.2.3 Oversampling Receivers . . . 138
8.2.4 Phase Interpolation . . . 139
8.2.5 Injection Locking . . . 141
8.2.6 Gated Oscillator Topology . . . 142
8.3 Topology Discussion . . . 143
8.4 Specifications . . . 144
8.5 The Gated Oscillator Topology . . . 144
8.5.1 The Multichannel Gated Oscillator CDR . . . 145
8.5.2 The Clock Recovery Core . . . 145
8.5.3 CDR Top-Down Design Methodology . . . 149
8.6 Statistical Modeling of the Gated Oscillator . . . 150
8.6.1 Jitter Statistics . . . 151
8.6.2 Random Jitter Components . . . 152
8.6.3 Deterministic Jitter Components . . . 153
8.6.4 Sinusoidal Jitter Model . . . 154
8.6.5 Bathtub Curves and BER Estimation . . . 158
8.7 Time-Domain Modeling . . . 161
8.7.1 CDR Model . . . 161
8.7.2 Input Data Source with Extended Jitter Model . . . 163
8.7.3 Duration of Edge Detection Pulse . . . 169
8.8 Transistor-Level Design . . . 170
8.8.1 Current-Mode Logic Cells . . . 170
8.8.2 Low-Power CDR Design . . . 172
8.8.3 Scaling Theory . . . 174
8.9 CDR Measurements . . . 176
8.9.1 Eye Diagram Measurements . . . 176
8.9.2 Bit Error Ratio Measurements . . . 177
8.10 Discussion . . . 180

CHAPTER 9 *Conclusions . . . 181*

References . . . 183

Constants, Symbols and Acronyms

Constants

Symbol	Description	Value
ε_0	Permittivity of free space	$8.8542\ 10^{-12}$ F/m
μ_0	Permeability of free space	$0.4\pi\ 10^{-6}$ H/m
h	Planck constant	$6.626\ 10^{-34}$ Js
k_B	Boltzmann constant	$1.3806\ 10^{-23}$ J/K
q	Electron charge	$1eV = 1.6022\ 10^{-19}$ C

List of Symbols

Symbol	Description	Units
β	Transconductance parameter in bulk-referenced model	A/V^2
β	Amplifier feedback coefficient in TIA configuration	–
β_{BB}	BB path gain in higher-order BB loop	–
β_{int}	Integral path gain in higher-order BB loop	–
Δf_{TR}	CDR frequency tuning range	MHz or GHz
γ	Excess thermal noise factor	–
η	Quantum efficiency	–
f_0	Channel surface potential	V
ϕ_e	Phase error	rad
ϕ_F	Fermi potential in the the substrate	V
ϕ_i	Phase of the input (data) signal	rad

ϕ_o	Phase of the output (recovered clock) signal	rad
κ	Oscillator noise parameter	–
q	Time-varying phase angle	rad
μ_e	Electron mobility	$cm^2/(Vs)$
ρ_{PD}	Phototdetector responsivity	A/W
σ_0	RMS noise amplitude on logic “0”	V
σ_1	RMS noise amplitude on logic “1”	V
σ_{CKJ}	RMS oscillator clock jitter amplitude	UI
σ_{RJ}	RMS random jitter amplitude	UI
$\tau_{transit}$	PD transit time constant	μs
υ	Optical frequency	nm
ξ	Body-effect factor in source-referenced model	–
ζ	ISI proportionality factor	–
ζ	Transfer function damping factor	–
ω_n	Transfer function natural frequency	rad/s
ω_{pi}	Open-loop bandwidth of a single LA stage	rad/s
ω_{TIA}	TIA open-loop bandwidth	rad/s
A_{vi}	Voltage gain of a single LA stage	dB
A_{vDC}	Limiting amplifier DC voltage gain	dB
A_{vDCi}	DC voltage gain of a single LA stage	dB
A_{vLA}	Limiting amplifier voltage gain	dB
A_{vOS}	Offset compensation voltage gain in LA	dB
A_{vTIA}	TIA core amplifier voltage gain	dB
A^0_{vTIA}	TIA core amplifier open-loop DC gain	dB
BER	Bit error ratio	–
BW_{LA}	Limiting amplifier bandwidth	GHz
BW_{TIA}	Transimpedance amplifier closed-loop bandwidth	GHz
C_c	AC-coupling capacitance	fF
C_d	Drain junction capacitance	fF
C_f	TIA feedback capacitance	fF
C_{gs}	Gate-source capacitance	fF
C_{inCDR}	CDR input capacitance	fF
C_{inLA}	Limiting amplifier input capacitance	fF
C_L	Load capacitance	fF
C_{PD}	Photodetector capacitance	fF

C_{ox}	Gate oxide capacitance	fF/μm^2
DJ_{chan}	Deterministic jitter contribution of the channel	UI
DJ_{PP}	Peak–peak deterministic jitter amplitude	UI
DJ_{RJ}	Deterministic jitter contribution of the receiver	UI
dSJ	Differential sinusoidal jitter	UI
E	Electric field in depletion region	kV
E_c	Critical electric field in pinch-off region	kV
E_g	Material bandgap energy	eV
$erfc$	Complementary error function	–
f_0	CDR center frequency	GHz
f_B	Raw data rate	Gb/s or GHz
f_{BB}	Frequency step in a bang-bang CDR	GHz
f_{max}	Unity current gain frequency in diode configuration	GHz
f_{mod}	Normalized sinusoidal jitter frequency	–
f_{SJ}	Sinusoidal jitter frequency	GHz
f_T	Transit frequency	GHz
GBW_{TIA}	TIA core amplifier gain-bandwidth product	GHz
g_{ds}	Small-signal output conductance	A/V
g_m	Small-signal transconductance	A/V
g_{mb}	Small-signal bulk transconductance (source-referenced model)	A/V
g_{ms}	Small-signal source transconductance (bulk-referenced model)	A/V
i_d	Drain current in small-signal model	A
I_D	Drain current in large-signal model	A
$I_{in\ min}$	Receiver input sensitivity	μA_{RMS}
I_{nPD}	Photodetector dark current	nA_{RMS}
I_{nTIA}	TIA input-referred integrated noise current	nA_{RMS}
I_{ntot}	Total receiver input-referred integrated noise current	nA_{RMS}
i_{nTIA}	TIA core amplifier small-signal current noise	nA
i_{PD}	Photodetector current	A
I_{PD}	Phase detector / charge pump current	μA
$JTOL$	CDR jitter tolerance	UI
k	Scaling factor between in LA stages	–
k_σ	Peak–peak to RMS conversion factor	–
k_{ind}	Inductive coupling coefficient	–
K_{PD}	Phase detector gain	V/rad

K_{VCO}	VCO gain	Hz/V
K'	Transconductance parameter of source-referenced model	A/V^2
L	MOS transistor drawn channel length	μm
l_a	Length of the absorption region	μm
l_d	Length of the depletion region	μm
M	Number of gain stages in LA	–
n	Slope parameter in bulk-referenced model	–
OMA	Optical modulation amplitude	dBm
P_0	Optical power corresponding to a logic “0”	mW
P_1	Optical power corresponding to a logic “1”	mW
P_{AVG}	Average received optical power	dBm
q	Electron charge	
Q_{BER}	Bit error ratio Q-factor	–
r_e	Extinction ratio	–
R_{TIA}	Transimpedance gain	dBΩ
RJ	RMS random jitter amplitude	UI
RJ_{PP}	Peak–peak random jitter amplitude	UI
RJ_{RMS}	RMS random jitter amplitude	UI
SJ_{PP}	Peak–peak sinusoidal jitter amplitude	UI
S_{niTIA}	TIA core amplifier current noise power spectral density	μV/√Hz
S_{nvTIA}	TIA core amplifier voltage noise power spectral density	μV/√Hz
T	Absolute temperature	K
T_B	Bit period	ps
$t_{rt\text{-}ft}$	Rise-time/fall-time duration	ps
TJ	Total jitter	UI
TrJ_{PP}	Tracking jitter (hunting jitter)	UI
U_T	Thermodynamic voltage	V
v_{bs}	Bulk-source voltage in source-referenced small-signal model	V
V_{cminLA}	LA common-mode input voltage	V
V_{DD}	Power supply voltage	V
v_{ds}	Drain-source voltage in small-signal model	V
V_{DS}	*Drain-source voltage in large-signal model*	*V*
v_g	Gate voltage in bulk-referenced small-signal model	V
V_G	Gate voltage in bulk-referenced large-signal model	V
v_{gs}	Gate-source voltage in source-referenced small-signal model	V

V_{GS}	Gate-source voltage in source-referenced large-signal model	V
V_{inPP}	Peak–peak input voltage amplitude	V
V_{ISI}	Intersymbol interference voltage amplitude	V
V_{minCDR}	CDR input sensitivity	mV_{PP}
V_{ni}	Input-referred integrated noise voltage of a single LA stage	μV_{RMS}
V_{nLA}	Limiting amplifier input-referred integrated noise voltage	μV_{RMS}
$V_{noutTIA}$	TIA output-referred integrated noise voltage	μV_{RMS}
v_{nR}	TIA feedback resistor voltage noise	μV
V_{nRMS}	Random noise amplitude at TIA output	μV_{RMS}
v_{nTIA}	TIA core amplifier small-signal voltage noise	μV
V_{PP}	Peak–peak voltage at TIA output	V
v_s	Gate voltage in bulk-referenced small-signal model	V
V_S	Source voltage in bulk-referenced large-signal model	V
V_{swing}	Voltage swing in current-mode logic gates	V
V_T	MOS transistor threshold voltage in source-referenced model	V
V_{T0}	MOS transistor threshold voltage in bulk-referenced model	V
W	MOS transistor drawn channel width	μm
Z_{in}	Input impedance	Ω

List of Acronyms

APD	Avalanche Photodiode
AM–PM	Amplitude Modulation to Phase Modulation
ATM	Asynchronous Transfer Mode
BB	Bang–bang
BER	Bit Error Ratio
BERT	Bit Error Ratio Tester
BUJ	Bounded Uncorrelated Jitter
C4	Controlled Collapse Chip Connection
CBR	Constant Bit Rate
CCDL	Current Controlled Delay Line
CCO	Current Controlled Oscillator
CDF	Cumulative Distribution Function
CDR	Clock and Data Recovery
CID	Consecutive Identical Digits
CJTPAT	Compliant Jitter Tolerance Pattern
CK	Clock signal

CMFB	Common-Mode Feedback
CML	Current-Mode Logic
CMOS	Complementary Metal Oxide Semiconductor
CMP	Chemical-Mechanical Polishing
DAC	Digital-to-Analog Converter
DBR	Distributed Bragg Reflector
DCD	Duty-Cycle Distortion
DDJ	Data-Dependent Jitter
DDR	Double Data Rate
DJ	Deterministic Jitter
DLL	Delay-Locked Loop
DQDB	Distributed-Queue-Dual-Bus
DSP	Digital Signal Processor
DUT	Device Under Test
DWDM	Dense Wavelength Division Multiplexing
ESD	Electrostatic Discharge
FC	Fibre Channel
FC-PH	Fibre Channel Physical Layer
FDDI	Fiber Distributed Data Interface
FEC	Forward Error Correction
FFT	Fast Fourier Transform
FIB	Focused Ion Beam
FIFO	First In First Out
FR4	Flame Resistant 4
FTOL	Frequency Tolerance–Tolerance to frequency offsets
FWHM	Full Width at Half Maximum
GaAs	Gallium-Arsenide
GD	Group delay
GO	Gated Oscillator
GCCO	Gated Current Controlled Oscillator
GVCO	Gated Voltage Controlled Oscillator
HDL	Hardware Description Language
IDM	Integrated Device Manufacturer
IL	Injection Locking
InGaAs	Indium–Gallium–Arsenide
InP	Indium-Phosphite
IP	Intellectual Property
ISI	Intersymbol Interference

I/O	Input/Output
JGEN	Jitter Generation
JTOL	Jitter Tolerance
JTPAT	Jitter Tolerance Pattern
JTRAN	Jitter Transfer
LA	Limiting Amplifier
LAN	Local Area Network
LASER	Light Amplification by Stimulated Emission of Radiation
LF	Loop Filter
LPCVD	Low Pressure Chemical Vapor Deposition
LSB	Least Significant Bit
LVDS	Low-Voltage Differential Signaling
MAN	Metropolitan Area Networks
MMF	Multimode Fiber
MSB	Most Significant Bit
NoC	Network on Chip
NRZ	Non-Return to Zero
OMA	Optical Modulation Amplitude
OOK	On–Off Keying
PCB	Printed Circuit Board
PD	Photodetector
PD	Phase Detector
PDF	Probability Density Function
PI	Phase Interpolation
PIN	P-doped - Intrinsic - N-doped
PJ	Periodic Jitter
PLL	Phase-Locked Loop
POF	Plastic Optical Fiber
POP	Point Of Presence
PRBS	Pseudo-Random Bit Sequence
PSD	Power Spectral Density
PSRR	Power Supply Rejection Ratio
ppm	Parts per million
QoS	Quality of Service
RCE	Resonant Cavity Enhanced
RF	Radio Frequency
RJ	Random Jitter
RMS	Root Mean Square

RZ	Return to Zero
SAN	Storage-Area Network
SAW	Surface Acoustic Wave
SEM	Scanning Electron Microscope
Si	Silicon
SiGe	Silicon–Germanium
SJ	Sinusoidal Jitter
SNR	Signal-to-Noise Ratio
SoC	System on Chip
SOI	Silicon-On-Insulator
TIA	Transimpedance Amplifier
UDJ	Uncompensatable Deterministic Jitter
UI	Unit Interval
VBR	Variable Bit Rate
VCDL	Voltage Controlled Delay Line
VCO	Voltage Controlled Oscillator
VCSEL	Vertical Cavity Surface Emitting Laser
VSR	Very Short Reach
WAN	Wide Area Network

CHAPTER 1 *Introduction*

Digital high-performance circuits grow in size and complexity following an exponential trend line. This trend was first recognized and announced by Gordon E. Moore in 1965 [1] (hence the name "Moore's law") and shows that the number of components in an integrated circuit doubles every 18 months. Technology scaling allows to implement more and more complex functions on a single chip, lowering at the same time the cost per function and increasing the operating frequency of digital cores. Digital cores used in microprocessors and digital signal processors (DSPs) profit from this evolution to handle an increasing quantity of data in shorter and shorter time slots.

Long-haul communication systems offer an impressive bandwidth improvement due to improvements in routing processors, which again benefit from technology scaling, and especially fiber-optic communication systems. The optical communication medium benefits of the huge signal bandwidth available on optical carriers, based on LASER (light amplification by stimulated emission of radiation) sources. While the Internet traffic has been doubling every 12 months over the past few years, the available capacity has been growing much faster between 1995 and 2000. Some consider this excess capacity as one of the reasons of the 2000 telecom crash. We can nonetheless expect the capacity to continue growing at the same rate than the traffic in the years to come. A technological hurdle is only expected to arrive at the end of this decade, when dense wavelength division multiplexing (DWDM) will occupy the full available spectrum of the fibers and the cost of additional capacity will increase. Today, local-area network (LAN) protocols are based on the existing optical technology developed for long-haul systems to achieve increasing multi-gigabit data rates at relatively low cost. While long-haul data communication still enjoys a relatively large bandwidth capacity in the foreseeable future, short-distance (chip-to-chip or board-to-board) interconnects face much tighter bottlenecks.

P. MULLER, Y. LEBLEBICI,
CMOS Multichannel Single-Chip Receivers for Multi-Gigabit Optical Data Communications,
DOI 10.1007/978-1-4020-5912-4_1, © 2007 Springer

Currently, the microprocessor peripheral bus relies on an obsolete parallel bus paradigm limiting the total system performance, which growth rate cannot compete with either the intrinsic microprocessor performance or the network interfaces (Figure 1.1). For consistent comparison, the processor performance is calculated by multiplying its clock frequency with the data bus width and expressed in megabit per second.

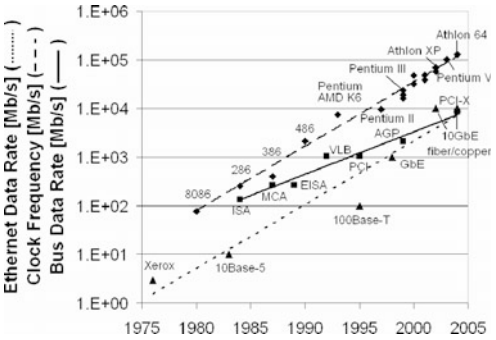

FIGURE 1.1. Performance comparison between microprocessor, peripheral bus and Ethernet

Today's microprocessors receive and send all the data to be shared with peripherals via their input/output (I/O) pads using electrical signaling. While the processor complexity can grow with technology scaling and the area of the silicon die, the number of I/Os is limited by the mechanical constraints of the packaging and the perimeter of the die. Figure 1.2 shows the respective growth rates of I/O count and gate count in state-of-the-art microprocessors.

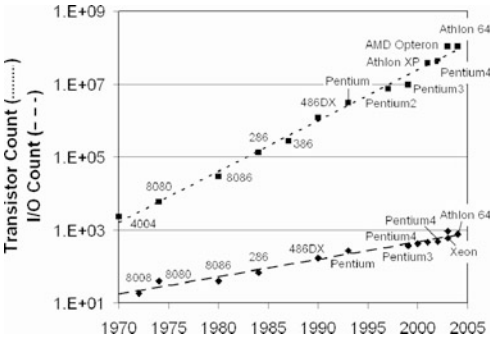

FIGURE 1.2. Comparison of I/O count to gate count

To compensate for the decreasing I/O count–processor complexity ratio in future microprocessors, the per pin I/O bandwidth must be considerably increased. Current computer buses improve the data rates by increasing the bus width and sampling several times per clock cycle (e.g. double data rate – DDR). Extending the bus width comes at the penalty of printed circuit board (PCB) area and will be ultimately limited by the number of I/Os. Data rates in electrical signaling are limited by RLC line effects, i.e. time constants of signal propagation and cross talk. Although cross talk can be somewhat improved by using low-voltage differential signaling (LVDS), the resulting bandwidth-distance trade-off is a strong limitation when targeting high data rates.

Serial data links using low-voltage electrical signals (e.g. PCI Express) operate at higher data rates, but are still limited by the intrinsic bandwidth limitation of copper interconnects. While optical signals experience hardly any attenuation in the optical fiber for the distances considered here, electrical signals suffer of diverse loss mechanisms [2]. As a result, the required transmit power increases with the distance between transmitter and receiver. Above a given distance, which depends on the data rate of transmission, optical communications can transmit the data at the same rate at lower energy.

Energy, or power consumption, is an important concern in microprocessor I/Os. State-of-the-art cores already dissipate several watts to tens of watts of power and deep submicron effects like gate and sub-threshold leakage worsen with process generations. Considering an actual microprocessor with several hundred I/O pins, the power consumption per pin must be optimized. For increasing data rates, only optical short-distance data communication can provide a complete solution to today's I/O bottleneck.

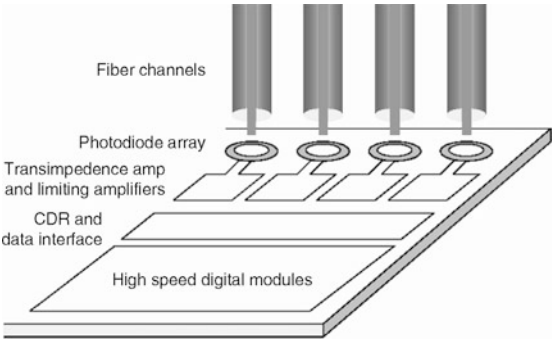

FIGURE 1.3. Single-chip digital core with optical I/Os

For obvious reasons, the optical I/Os have to be manufactured in the same silicon CMOS technology than the microprocessor or DSP core itself (Figure 1.3). The implementation may however benefit

of the move to silicon-on-insulator (SOI) substrates, which tend to become mainstream for high-performance digital circuits.

This book presents the different aspects of multichannel single-chip CMOS photoreceiver design. After motivating the work in the present chapter, Chapter 2 introduces the different fields of optical communications, ranging from long-haul systems established in the 1980s to modern optical backplanes. Chapter 3 familiarizes the reader with the basic concepts used in the design of fiber-optic receivers, like random non-return-to-zero data patterns, optical modulation principles and data encoding. The effects of receiver nonidealities, be it amplitude noise, bandwidth limitations or jitter, on the receiver bit error ratio are addressed. Clock recovery principles, a major issue in serial communication links, are also mentioned.

Chapter 4, discusses system-level considerations and receiver specifications. Starting from common specifications for short-distance communications, the requirements for each subblock are determined using a specification-driven design approach. The receiver design fundamentals are revisited in the scope of CMOS technologies and a design method for bandwidth optimization is proposed.

Chapter 5 presents various silicon photodetector topologies published in the literature. First, infrared light absorption in silicon is introduced, then different approaches to address the intrinsically low bandwidth-efficiency product of silicon are discussed. Measurement results illustrate the performance of the different designs.

The design of transimpedance amplifiers, capable of converting the high-speed photocurrent delivered by the detector, is discussed in Chapter 6. Chapter 7 proposes different limiting amplifier topologies for high data rate applications. Among other topics, the effects of spiral inductors for limiting amplifier bandwidth enhancement and the potentiality of magnetic coupling are also presented.

Conventional clock recovery topologies, based on phase-locked loops, are found not to be the optimum solution for multichannel receivers in Chapter 8. First, the different clock and data recovery architectures are discussed in the scope of multichannel short-distance communications. Then an innovative design based on a gated oscillator structure, able to achieve low area low power clock recovery, is introduced. In this chapter, we present a top-down design methodology developed to analyze and verify the specification compliancy of this topology. A statistical model of the different jitter components for jitter and frequency tolerance estimation is discussed, as well as behavioral models for time-domain verification of the system.

Finally, Chapter 9 concludes the discussion.

CHAPTER 2

Integrated Photonic Systems

Historically, fiber optic communications have experienced an impressive growth in the long-haul domain due to their inherently large bandwidth and low attenuation and dispersion. Indeed, these characteristics allow for the transmission of a large amount of data over long distances at limited latency, which makes optical fibers the medium of choice for intercontinental communication links.

In spite of the burst of the dot-com bubble at the beginning of this decade, today's fiber communications market is still very much dominated by the same paradigms and major players as 5 or 10 years ago. It appears however that the increasing processing power in computers from PCs to servers raises the bar for peripheral communication links to a level where the parallel electrical bus paradigm cannot satisfy the needs anymore. Serial links of all kinds are defined by various standardization bodies and special interest groups and optical serial links appear as an ultimate solution to the growing bandwidth needs of tomorrows systems.

Various optical communication systems apply different communication schemes like network-based routing, channel-based switching or bus-like networks providing shared services over point-to-point links. As this work focuses on the physical layer of a communication link and does not consider the higher layers of such a protocol, these details will not be further discussed.

This chapter will provide an overview of today's long-haul fiber-optic communication systems, deeper insight into a number of optical short-distance (de facto) standards and provide an outlook at a possible scenario for tomorrow's short-distance optical links.

P. MULLER, Y. LEBLEBICI,
CMOS Multichannel Single-Chip Receivers for Multi-Gigabit Optical Data Communications,
DOI 10.1007/978-1-4020-5912-4_2, © 2007 Springer

2.1 Long-Haul Communication Links

As previously mentioned, low dispersion and absorption are critical specifications when sending optical signals over distances of several hundreds or thousands of kilometers. These specifications directly determine the number of optical amplifiers and repeaters required in the link. Obviously these wide-area network (WAN) links are operated in the low–loss minimum-dispersion windows at 1,330 nm and 1,550 nm using single-mode fibers to obtain minimum dispersion. The cost of such high-performance fibers and a number of optical amplifiers with their energy sources renders the cost of transmitters and receivers almost negligible.

The fiber-optic synchronous transmission protocols SONET (North America and Japan) and SDH (Europe) have been developed to transfer constant bit rate (CBR) traffic with latency requirements, like the voice services. These heavily standardized protocols build most of the backbone of the actual voice networks. In such bandwidth and absorption-specified long-haul standards, transmitters and receivers have very stringent performance requirements to allow transcontinental communications with a minimum number of repeaters and regenerators.

The asynchronous transfer mode (ATM) was defined to replace SONET/SDH and offer as well quality-of-service (QoS) guaranteed CBR as variable bit rate (VBR) best-effort services in order to optimize the bandwidth usage. As for SONET/SDH, the physical requirements of ATM are appropriate for the long-distance transmission wavelengths (1,330 nm and 1,550 nm).

2.2 Metropolitan-Area Networks

Metropolitan-Area-Networks (MAN) cover areas of several tens of kilometers and include as well backbones of large campuses, as local network provider infrastructures and regional links between neighboring cities. Robust redundant topologies have been developed in the 1980s like the Fiber Distributed Data Interface (FDDI) and the Distributed-Queue-Dual-Bus (DQDB, IEEE 802.6). Both are dual-ring fiber topologies and essentially differ in the physical layer definition (optical wavelength and fiber type) and the queuing algorithms. They have essentially be replaced by Fast Ethernet and Gigabit Ethernet due to lower cost and higher speed. For longer distance communications and situations where robustness is important, SONET/ATM solutions have taken over. With the emergence of 10 G Ethernet, there is little place left for the development of optimized protocol solutions for the metropolitan area networks.

2.3 Local Area Networks and Short-Distance Interconnects

Although a LAN is supposed to connect computer frames as well as printers and other independent (intelligent) network devices, the birth of optical short-distance interconnects blurs the border between the latest peripheral links and computer interconnects. While protocols like Ethernet with a large software overhead (due to the TCP-IP stack) may singularly stay away from the processing cores, this is not true for all packet-based networks. Thin (low-overhead) packet-based protocols are even considered in advanced research for tomorrow's networks-on-chip (NoCs). A neat distinction between LAN protocols and peripheral solutions would be considered arbitrary and avoided here. As electrical interconnects are able to operate up to a few gigabit per second over short distances, this area is covered both by high-performance purely optical schemes as well as optical protocols that evolved from lower-speed electrical definitions. The short distances to be covered allow for the use of inexpensive multimode fibres and sometimes even plastic optical fibers (POFs). The large number of devices to be potentially connected increases the number of transceivers in such solutions. This finally inverts the cost relationship between link and end-point equipment, resulting in a need for low-cost transceiver designs.

The OC-192 VSR (OIF 99.120) standard, where VSR stands for "very short reach", offers a low-cost match to the OC-192 SONET/SDH standard (around 10 Gb/s) for short-distance intra-office and service provider points-of-presence (POPs) interconnects. POPs are the interface between high-performance long-haul equipment and MAN routers and switches and typically have a large number of short-reach OC-192 inter-router interconnects. The standard defines serial, 4-fiber parallel and 12-fiber parallel interconnects based on low-cost VCSELs and receivers.

Today's most popular LAN protocol is without any doubt Ethernet (IEEE 802.3). Due to its long history which dates back to the 1970s, components are readily available, backward-compatible and provide a low cost-performance ratio. In order to stay with its customer base in the LAN and enterprise market while trying to enter the metropolitan area, both electrical and optical standards are defined at multi-gigabit data rates (802.3ab and 802.3z are the electrical and optical standards operating at 1 Gb/s). Because of its widening LAN–MAN focus, Ethernet supports both 850 nm and 1,310 nm optical wavelengths. In opposition with the long-haul standards around 1,300 nm and 1,550 nm, the shorter wavelength window at 850 nm allows for the use of low-cost vertical cavity surface emitting lasers (VCSELs), which moreover can be fabricated in arrays suitable for coupling with multifiber ribbons in parallel optical links. This wavelength window is not used in long-haul systems due to the higher optical absorption and the resulting need for additional amplification.

Fibre Channel (FC-PH) is an optical high-performance static interconnect infrastructure operating at 1 Gb/s below an IP or SCSI protocol stack. Its dominant position in storage-area networks (SAN) and server farms. Products of the 10 Gb/s version 10 GFC using either 850 nm or 1,330 nm has been released in 2005 and offer support for various data protocols like SCSI, HIPPI, IP.

InfiniBand is a switched, channel-based I/O architecture providing mapping for various transport protocols (IP, HIPPI, SCSI,...) with a unclear positioning between fibre-channel dominated server interconnects and the computer-bus market (processor-to-peripheral interconnect). As it specifies both electrical and optical media, it could become a high-speed replacement for PCI, but it may also extend to the SAN or even LAN markets (Figure 2.1).

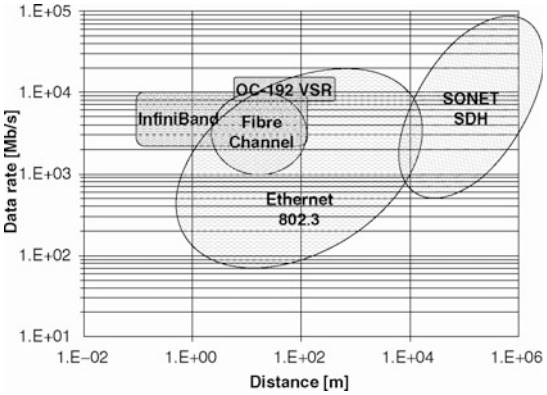

FIGURE 2.1. Comparison of bandwidth and data rates

Several processor and peripheral bus manufacturers have teamed up to define and market three protocols dedicated to the processor peripheral bus area. HyperTransport, PCI-Express and RapidIO, respectively backed by AMD, Intel and Motorola-TI, are competing in this domain. While the first two are host-centric solutions uniquely targeting the computer bus market, RapidIO is more of a communication-centric solution for embedded systems. Although none of their released specifications contains any definition of an optical physical layer, it has been unofficially mentioned for both HyperTransport and PCI-Express. The planned number of parallel communication channels in these protocols (32 for both HyperTransport and PCI-Express) shows the limitations of such schemes.

2.4 Optical Backplane Technology

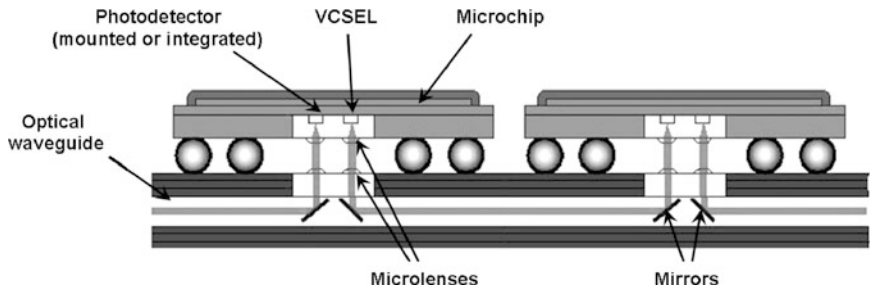

FIGURE 2.2. The optical backplane principle

At the same time, some integrated device manufacturers (IDMs) have developed their own solutions which may suit from none to several of the above discussed standards. While Infineon's Paroli line was discontinued in early 2005, Agilent used its parallel optics business as a basis to define a multi-source agreement (POP4) with Zarlink and Primarion. Most standard-independent research efforts are however spent on optical backplanes (Figure 2.2), tomorrow's technology for high-performance chip-to-chip communications. The optical waveguides are integrated in the substrate of an otherwise conventional printed circuit board. The manufacturing techniques, today ready for mass production, are currently still costly due to the limited order volume, but may drop considerable within a short timeframe. In collaboration with some smaller companies and academia, IBM has demonstrated optical backplanes

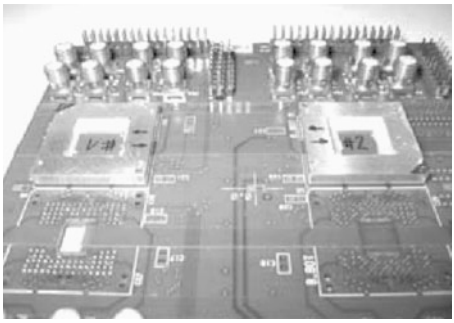

FIGURE 2.3. IBM's optical backplane demonstrator [3]

with 12.5 Gb/s/channel capabilities over 30 cm integrated waveguides on several parallel channels ([3], Figure 2.3). However, the development of improved integrated photonic components is necessary to provide the breakthrough to this solution. Indeed, the modulation rate of today's VCSEL sources is

limited to about 20 Gb/s by physical constraints [4], while copper links are close to achieve comparable performance at the expense of complex channel equalization [5]. We remain confident however that the immense research effort in this area will lead to the development of improved light sources within a couple of years, respectively of monolithically integrable optical modulators achieving comparable performance. [6].

On the detector side, several scenarios have been proposed to circumvent the bandwidth-absorption trade-off of conventional silicon photodetectors. While a detailed analysis of the domain is beyond the scope of this chapter, we would like to mention the use of horizontal detector structures [7], the compensation of lacking detector bandwidth through equalization [8] and finally the use of resonant cavities, as discussed in Chapter 5.

Silicon-based lasers have been the holy grail of photonics researchers for years. Intel has published results of a silicon Raman laser [9], which is pumped by an external III-V laser. The more interesting results come from ST Microelectronics in Italy [10] developing high-efficiency silicon light sources. Combining this latter technology with resonant cavities, a silicon "laser-like" coherent light source seems close to become reality. Finally, advances in the manufacturing of on-chip optical waveguides also present a coherent interface with optical chip-to-chip interconnects.

2.5 Optical on-chip interconnects

Although it is currently not yet a technologically feasible and economically viable alternative, the immense research efforts invested in the domain should make optical on-chip or wafer-scale interconnects become a viable alternative to power-draining clock trees [12] and wave-pipelined global on-chip interconnects [13] in the midterm future. A further step into this direction has been recently reported in the form of a new transistor structure capable of emitting coherent light, dubbed the "transistor laser" [14]. Figure 2.4 shows the evolution of optical short-distance links from rack-to-rack links to on-chip interconnects.

The first conclusion of this overview is that the move to serial communication schemes, the immense design experience available and low-cost manufacturing capacity allow well-established electrical backplane schemes to remain competitive at increasing data rates. There is little doubt however that even for the short distances in chip-to-chip communications, the physical bandwidth-distance trade-off will soon call for real breakthrough alternatives like low-cost multichannel optical links. More and more technical barriers like light coupling, mirror alignment, optical waveguide integration in backplane and especially silicon photonic integration are removed. Once fully monolithic integration

in relatively conventional processes is available at a larger scale, the impressive available bandwidth at low cost should allow for commercial market success.

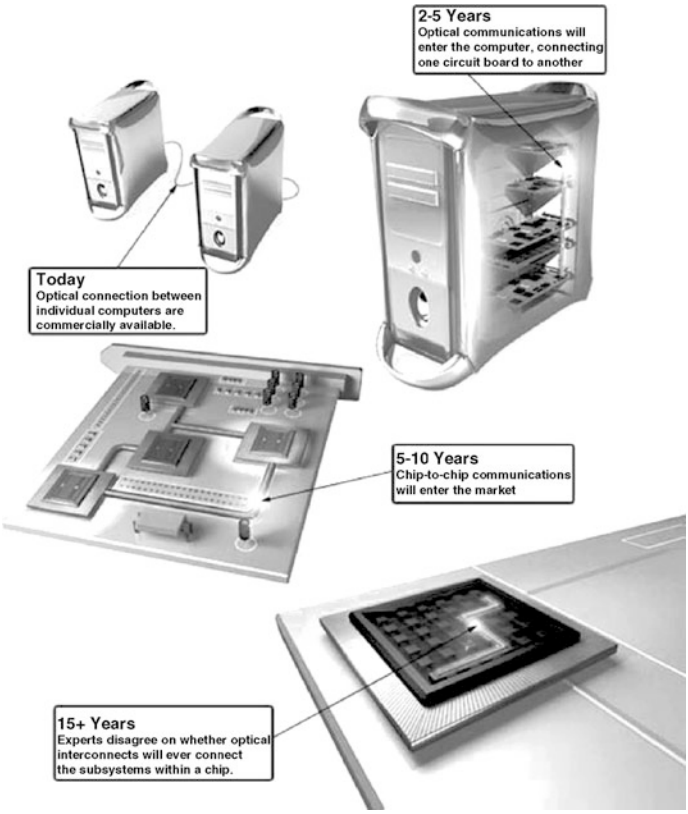

FIGURE 2.4. Evolution of short-distance optical communications, referenced to 2002 [11]

CHAPTER 3

Basic Concepts

A fiber-optic communication system is composed of one or multiple optical links using the modulation of light to transmit the data. For technical reasons, two-level (binary) intensity modulation is almost exclusively used in today's systems. We mentioned before that for best performance the system is operated in one of the low-absorption regions in the infrared domain around 850 nm, 1,310 nm or 1,550 nm (Figure 3.1). The absorption peaks are due to OH^- impurities in the fiber, nowadays mostly removed in the manufacturing process.

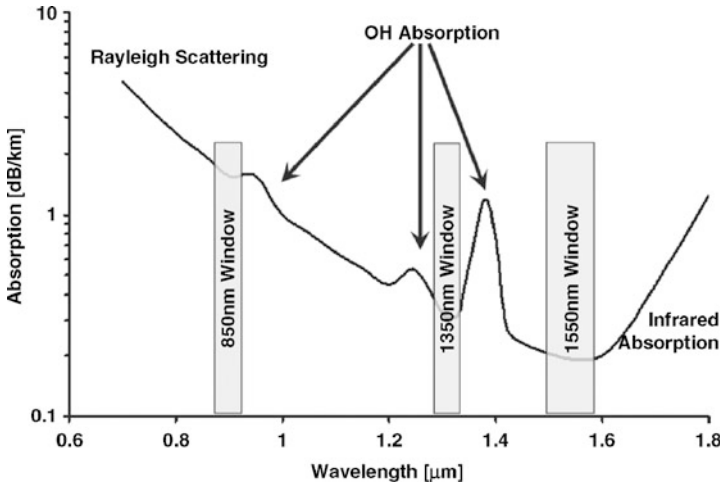

FIGURE 3.1. Low-absorption wavelength regions in optical fibers

P. MULLER, Y. LEBLEBICI,
CMOS Multichannel Single-Chip Receivers for Multi-Gigabit Optical Data Communications,
DOI 10.1007/978-1-4020-5912-4_3, © 2007 Springer

Lower system cost related to the availability of vertical cavity surface emitting laser (VCSEL) sources and multimode fibers call strengthen the operation in the 850 nm window for short-distance communication. Figure 3.2 shows one of the optical links between two processors, each including transmitters and receivers.

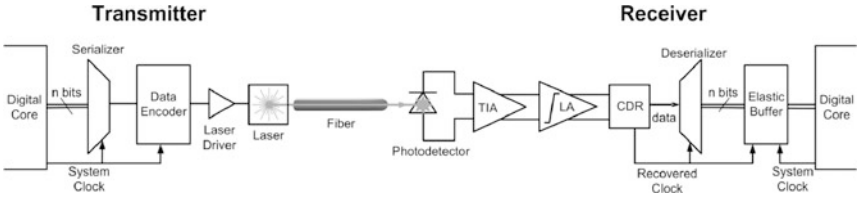

FIGURE 3.2. Single-channel fiber-optic communication link

In the transmitter, the data to be sent is delivered by the digital core to the serializer, which then delivers the 1-bit high-speed stream to the data encoder for scrambling and encoding. The commonly off-chip laser driver modulates the laser intensity according to the output of the encoder. Direct modulation of VCSELs results in lower bill of material and complexity compared to the use of separate laser modulators in long-haul links. The light beam is directly coupled into a 50/65 μm core step index multimode fiber (MMF). In absence of stringent absorption and dispersion specifications, this fiber type has been selected for its low fabrication cost and minimum coupling loss as a benefit of the large core diameter.

The photodetector receives the optical signal at the other end of the fiber. The transimpedance amplifier (TIA) converts the generated photocurrent into a voltage and amplifies the signal at the same time. The limiting amplifier (LA) provides additional gain and guarantees a fixed output swing independent of the input voltage swing. The two-level signal it delivers then enters the clock and data recovery (CDR) circuit, which extracts clock information from the incoming data stream and performs synchronization. The received data is then deserialized and the bit stream is stored in the elastic buffer. The elastic buffer is a FIFO-type structure which reads the data using the recovered clock and delivers it to the digital core of the processor using the system clock. As such, it transfers the data from one clock domain to the other.

With the present work focusing on the design of integrated fiber-optic receivers, the issues on the transmitter side will not be further addressed. An introduction to the basic concepts of the design of such fiber-optic receivers will allow the reader to fully understand the importance of each step later addressed in the modeling and design of the receiver. An overview of photonic detection principles will be given when addressing the silicon photodector used in this work in Chapter 5.

3.1 Modulation of Optical and Electrical Signals

Considerable effort has been spent in the past to develop more efficient modulation schemes compared to the currently used binary data transmission. Although this basic two-level modulation is sometimes called on–off keying (OOK), it appears impossible to obtain a 100% modulation of the laser's intensity in high-speed optical communication systems. We thus consider two levels of optical power, the higher power level P_1 corresponding to a logical "1", the lower level P_0 corresponding to a logical "0" [16]. The extinction ratio r_e is defined as the ration between both power levels:

$$r_e = \frac{P_1}{P_0} \quad \text{(EQ 3.1)}$$

The average total power in a balanced code (defined later in this chapter) is of course the average between both levels:

$$P_{AVG} = \frac{P_1 + P_0}{2} \quad \text{(EQ 3.2)}$$

while the optical modulation amplitude (OMA) is the difference between both levels.

$$OMA = P_1 - P_0 = 2 \cdot P_{AVG} \cdot \frac{r_e - 1}{r_e + 1} \quad \text{(EQ 3.3)}$$

With such a modulated optical signal, the optoelectronic conversion by the photodetector directly delivers a baseband photocurrent and requires no additional demodulation.

3.2 NRZ Random Data

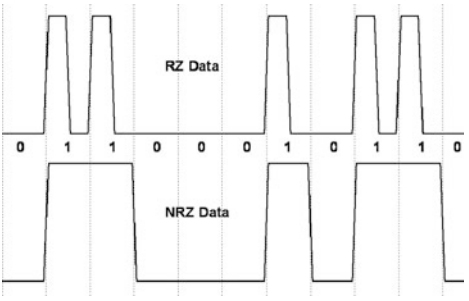

FIGURE 3.3. Time-domain representation of return-to-zero and non-return-to-zero line codes

Most frequently, two-level baseband signals are represented using a non-return-to-zero (NRZ) line code, which means that each data bit keeps its value during a whole bit period, with no neutral position.

In opposition to this, in return-to-zero (RZ) schemes the signal returns to its neutral state during the second half of each bit period (Figure 3.3).

To distinguish the pulse duration from the bit duration, the shape of the signal during the whole bit period is also called a symbol. This denomination is partially related to the fact that some modulation schemes in electrical and wireless communications allow a symbol to carry the information of more than one bit. The interested reader can find a complete discussion of these in [17]. The duration of a symbol T_B is called a unit interval (UI). In two-level modulation schemes it corresponds to the inverse of the data rate f_B. Unit intervals are a common measure for time-domain signals like jitter.

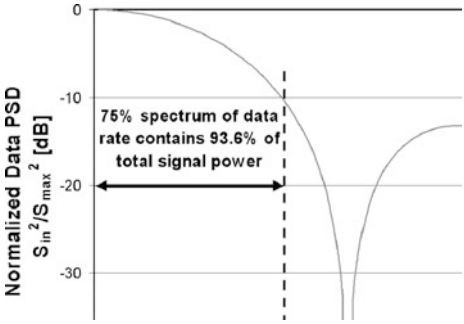

FIGURE 3.4. Frequency spectrum of an NRZ signal

As in such a modulation scheme the intensity of the optical input signal never becomes zero, one can argue whether the implemented line code uses a unipolar or bipolar representation. The fact is that the input amplifier will remove the DC component of the received signal and will result in a bipolar electrical signal. One of the reasons for choosing an NRZ representation is the reduced bandwidth occupied by NRZ signals as a consequence of the increased duration of each symbol. When addressing the clock recovery design in Chapter 8, the absence of spectral energy at the data rate will also be of importance, while the fact that a spectrum of 3/4 of the data rate contains almost 94% of the signal power simplifies the design of the receiver signal chain (Figure 3.4).

However, the strong low-frequency content of such a signal represents a serious threat to the synchronization of the receiver. The clock information is not transmitted separately from the data and must

be reconstituted from data transitions (in the so-called *clock recovery* process, discussed later). A stream containing few transitions (Figure 3.5) would require a lot of "memory" to maintain the correct clock rate during that process and "memory" (interestingly like digital memory) requires considerable chip area. While scrambling increases somewhat the number of data transitions, more benefits can be gathered from 8b/10b data encoding. The name indicates that the encoder outputs 10 bits of data for each eight incoming bits, resulting in a data rate loss of 20%. The output stream however does not only have a guaranteed transition density (0.6% or 60%) equivalent to a reduced low-frequency spectral content, it also guarantees a DC-balanced output signal (no spectral energy at DC) and an absolute maximum run length (5 bits). The run length defines the maximum number of consecutive identical digits (CID) appearing in a run of identical bits.

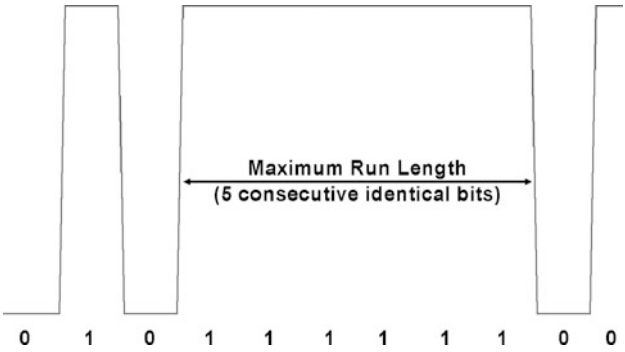

FIGURE 3.5. Run of NRZ coded identical bits

AC-coupling in the signal path and offset compensation mechanisms result in a band-pass behavior of the system. Filtering out the low-frequency component of a random NRZ data pattern results in a loss of short-term DC-balance, also called baseline wander, as shown in Figure 3.6 and more thoroughly discussed in [18]. The reduced low-frequency spectral content of an 8b/10b encoded pattern however relaxes the constraints on lower cutoff frequencies in the case of AC coupling and offset compensation. Although the impressive run lengths required for SONET compliant receivers requires a

strict minimum of four decades of passband to avoid baseline wander, we found that this requirement drops to a little more than three decades for 8b/10b encoding.

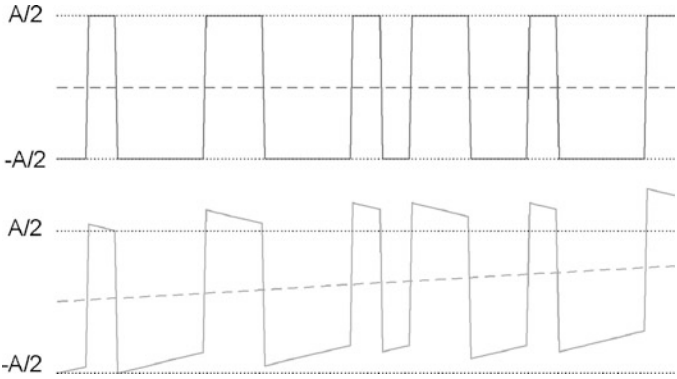

FIGURE 3.6. Base-line wander due to removal of low-frequency content

3.3 Clock Recovery Basics

A random data stream has a unique interpretation only if a time reference is provided for synchronization. The communication protocol defining the data link specifies the data rate (with some tolerance), which is a starting point for developing a synchronization scheme. As the instantaneous frequency is not perfectly known because of tolerances on the data rate and addition of jitter in the link, the clock information has to be recovered from the incoming data stream by the clock and data recovery block (Figure 3.7).

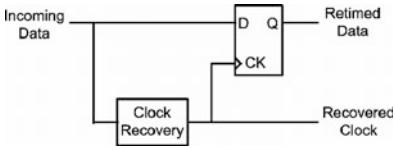

FIGURE 3.7. Clock recovery and data retiming

We have however seen in Figure 3.4 that the data spectrum does not contain any energy at the data rate. In order to allow recovery of the clock information contained in the data stream, energy must be created by a nonlinear operation, like edge detection, at the input of the CDR. Edge detection is a combination of differentiation and rectification (Figure 3.8, [19]). Several phase detector circuit topologies

for phase-locked loops (PLL) operating on random data streams have been developed. These structures, performing implicit edge detection at the input of the PLL, are discussed in more detail in [19].

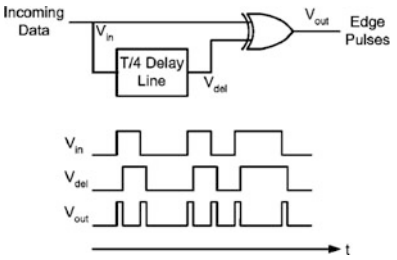

FIGURE 3.8. Explicit edge detection of NRZ data

It is important for the reader to understand that the sampling instant determined by the rising clock edge (respectively the falling clock edge in some designs) is considered as the reference point in time for measurement of jitter and bit error ratio.

3.4 Bit Error Ratio

If at the sampling instant the received data signal has, be it due to jitter or amplitude noise, the wrong logic level, the synchronized data will contain a bit error. The bit error ratio (BER) is defined as the ratio between the number of bit errors and the total number of received bits (Equation 3.4). General BER estimation techniques will be presented in Chapter 4, while detailed analysis of the jitter-BER relationship is discussed in Chapter 8.

$$BER = \frac{\#\text{ bit errors}}{\#\text{ bits received}} \qquad \textbf{(EQ 3.4)}$$

The bit error ratio and circuit performances like amplitude, noise and jitter are related through probability density functions. The representation of the receiver system performance can be simplified by introducing a new parameter Q_{BER} related to the BER by Equation 3.5, where *erfc* is the complementary error function [20].

$$\frac{1}{2} \cdot erfc\left(\frac{Q_{BER}}{\sqrt{2}}\right) = BER \qquad \textbf{(EQ 3.5)}$$

3.5 System Bandwidth and Inter-Symbol Interference

The ideal signals considered up to now were presented with infinitely sharp edges respectively zero rise-and fall-times. The finite bandwidth of the signal path can only deliver signals with nonzero rise-and

fall-times. If the bandwidth is considerably lower than the signal bandwidth, this will affect the edge timing and the amplitude at the sampling instant (Figure 3.9). Influence of preceding and succeeding symbols on the amplitude of the signal at the sampling instant is called *intersymbol interference* (ISI). This is best observed in a so-called *eye diagram*, which is obtained by superimposing all the symbols of the data pattern centered around their sampling instant. The lowest bit error ratio is achieved when the incoming data is sampled at the ideal sampling instant where the eye opening is maximal.

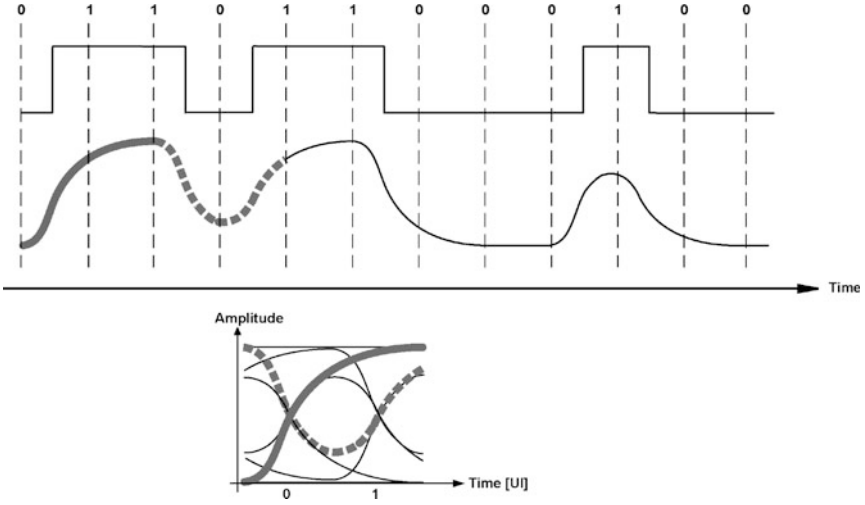

FIGURE 3.9. Effects of bandwidth limitations and construction of an eye diagram

Figure 3.10 illustrates the relationship between system bandwidth and intersymbol interference.

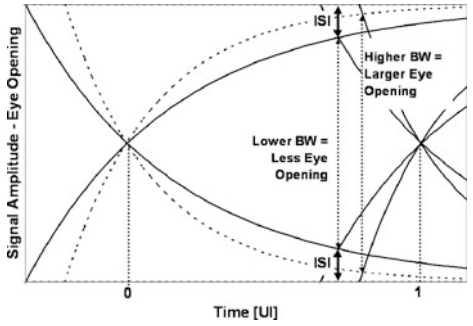

FIGURE 3.10. Resulting eye diagram for different system bandwidth and resulting intersymbol interference (ISI)

3.6 Amplitude Noise

The vertical eye opening is limited by intersymbol interference and by amplitude noise, which is superimposed on the signal (Figure 3.11). As for the device noise in most electronic signal amplification systems, the major contributors are located in the first stages of a receiver, i.e. the photodetector itself and the first stages of amplification. This device noise is a combination of thermal noise, shot noise and flicker noise. While the last component can be considerably suppressed through offset compensation in the amplifiers, the others depend directly on the system bandwidth. Without going into more details for the moment, we notice there is an obvious trade-off between intersymbol interference and noise bandwidth when defining the system bandwidth.

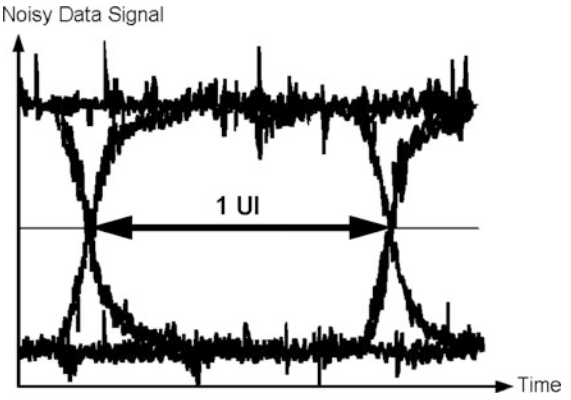

FIGURE 3.11. Effect of amplitude noise on the eye diagram

In the vertical eye closure analysis, Q_{BER} depends on the signal amplitude V_{inPP}, the random noise amplitude σ_0 and σ_1 on either binary level of the eye and the amount of intersymbol interference V_{ISI}:

$$Q_{BER} = \frac{V_{inPP} - 2V_{ISI}}{\sigma_0 + \sigma_1} \qquad \text{(EQ 3.6)}$$

3.7 Jitter

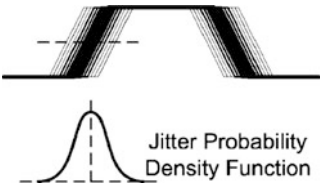

FIGURE 3.12. Illustration of jitter affecting a data edge

The horizontal eye opening on the other hand is affected by another form of noise called *jitter.* The term *jitter* is commonly used to describe any unwanted phase error appearing at the edges of the considered signal (Figure 3.12), somewhat the same way the term *noise* is used in a general fashion to define any unwanted amplitude contribution to a signal. Probability density functions (PDFs) are conveniently used to describe the different types of jitter sources, the above illustration showing a Gaussian distribution. Total jitter is typically viewed in an eye diagram or a bathtub curve, built using the cumulative distribution functions (CDFs) of both data edges (Figure 3.13).

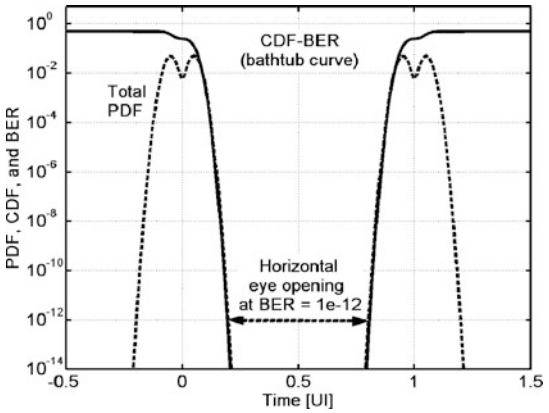

FIGURE 3.13. Bathtub curve. (After Kossel and Schmatz). The CDF has already been multiplied by the transition probability of 0.5 to show the obtained horizontal eye opening for given BER

3.7.1 Jitter Contributions

Jitter is usually decomposed into several components, as shown in Figure 3.14. Random jitter (RJ) is due to device noise sources like thermal noise, shot noise and flicker noise. As such, its probability density function (PDF) is generally defined as a Gaussian distribution, defined by the well-known

formula shown in Equation 3.7. In the rest of the text, the root-mean-square (RMS) random jitter amplitude *RJ* will be used in place of the standard deviation σ_{RJ}. In situations where both RMS and peak–peak random jitter values are used, they are explicitly differentiated by RJ_{RMS} and RJ_{pp}.

$$p_{RJ}(t) = \frac{1}{\sigma_{RJ}\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2\sigma_{RJ}^2}} = \frac{1}{RJ\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2RJ^2}} \quad \text{(EQ 3.7)}$$

Deterministic jitter (DJ) is a combination of intrinsic and extrinsic contributors. It includes design-dependent effects as data-dependent jitter (which includes duty-cycle distortion and intersymbol interference), periodic jitter resulting from electromagnetic interference, as well as bounded uncorrelated jitter due to coupling from neighboring channels or digital switching noise [22].

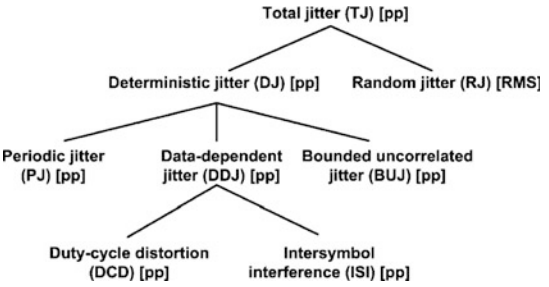

FIGURE 3.14. Jitter contributions [22]

Mathematically speaking it is obvious that jitter components from random sources are added in power domain (Equation 3.8), while jitter components from deterministic sources are added in amplitude domain (Equation 3.9). Combining deterministic and random jitter sources however requires some care. While random jitter amplitude is defined as a root-mean-square (RMS) value, the deterministic jitter amplitude, which is bounded, is measured as a peak–peak value. In order to be able to sum up both contributions, we have to translate the RMS value of the random jitter amplitude into a peak–peak value (Equation 3.10).

$$RJ_{tot} = \sqrt{RJ_1^2 + RJ_2^2} \quad \text{(EQ 3.8)}$$

$$DJ_{tot} = DJ_1 + DJ_2 \quad \text{(EQ 3.9)}$$

$$RJ_{PP} = RJ_{RMS} \cdot k_\sigma \quad \text{(EQ 3.10)}$$

Because a Gaussian distribution is not bounded, the peak value to be considered for a BER specification of $10^{-\alpha}$ corresponds to the probability that the jitter amplitude exceeds this value only once out of

10^{α} received bits. Intuitively speaking, the lower the specified BER, the larger the area under the tails of the Gaussian distribution that one must account for. The required proportionality factor k_{σ} between RMS and peak–peak values for common BER values is defined by Kossel and Schmatz [21]:

$$\frac{1}{2} \cdot erfc(\sqrt{2} \cdot k_{\sigma}) = BER \qquad \text{(EQ 3.11)}$$

where *erfc* is the complementary error function. The reader will notice that the parameter k_{σ} corresponds to $Q_{BER} / 2$. This is easily understood when considering that the RMS-to-PP conversion considers both tails of the gaussian distribution, while the BER calculation from Q_{BER} considers only one tail exceeding the sampling threshold or instant. Values for k_{σ} for common BER levels are given in Table 3.1, where most state-of-the-art transmission protocols require a BER of 10^{-12}.

TABLE 3.1. Proportionality factor for RMS–peak–peak conversion of signals with Gaussian distribution

BER	k_{σ}	BER	k_{σ}
10^{-9}	12.00	10^{-13}	14.70
10^{-10}	12.72	10^{-14}	15.30
10^{-11}	13.41	10^{-15}	15.88
10^{-12}	14.07	10^{-16}	16.44

3.7.2 Jitter Specifications

Although the bandwidth limitations and noise sources of the amplifiers may contribute to the jitter budget, the amount of tolerable jitter essentially depends on the specifications of the clock recovery circuit. Jitter transfer (JTRAN) is the amount of jitter transferred from the data input of the clock and data recovery block to its recovered clock output. This specification is important in applications where the recovered clock is used to synchronize the data on the transmit side. Jitter generation (JGEN) is the amount of jitter on the recovered clock generated inside the clock recovery circuit. Jitter tolerance (JTOL) specifies the maximum allowable jitter the clock recovery can handle before synchronization is lost, before the bit error ratio specification is exceeded.

3.7.3 Regenerators and Retiming Repeaters

In long-haul fiber optic links, optical amplifiers are regularly spaced in the link to compensate for the attenuation of the optical signal traveling over several hundreds or thousands of kilometers. Optical amplifiers however do not take care of dispersion due to the wavelength dependency of the propagation

velocity of the light in the fiber [23]. Regenerators convert the optical signal into a voltage and perform clock recovery, data retiming and finally reemit the retimed data using the recovered clock as a timing reference. As the recovered clock is subject to imperfections because of jitter transfer and generation, the regenerated signal is not a jitterless signal.

In opposition to this, retiming repeaters also recover the clock from the incoming data, but then send the data using a local timing reference, typically based on a crystal oscillator. In this way and unlike regenerators, retiming repeaters reset the jitter budget (i.e. do not retransmit jitter) thanks to the synchronization with the local clock.

In short-distance communications, where no regenerators are needed, jitter generation and transfer characteristics are not specified. Jitter tolerance is tested by superimposing, in addition to the jitter sources already present in the channel (see "Jitter Contributions" on page 22), a sinusoidal jitter component of swept amplitude and frequency. In order to be considered compliant to most short-distance protocols, a receiver must tolerate SJ amplitude exceeding the jitter tolerance specification (called jitter tolerance mask) shown in Figure 3.15.

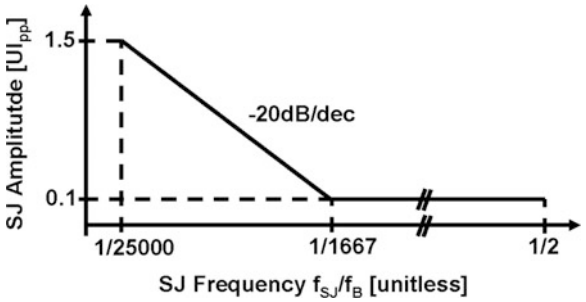

FIGURE 3.15. Jitter tolerance mask of a typical optical link standard

3.8 Multichannel Systems

In multichannel short-distance optical links, the frequency of each received data link is tightly controlled (typically to ±100 ppm) by the transmitter's local oscillator. As the length and physical conditions (temperature, bending, strain) of each fiber is not monitored, the phase difference between, thus the instantaneous frequency of, the links connected to a receiver are however unknown. Unlike parallel

buses with tight routing constraints, such *plesiochronous* (Greek: near time) links require independent clock recovery for each channel to take care of the timing variations (Figure 3.16).

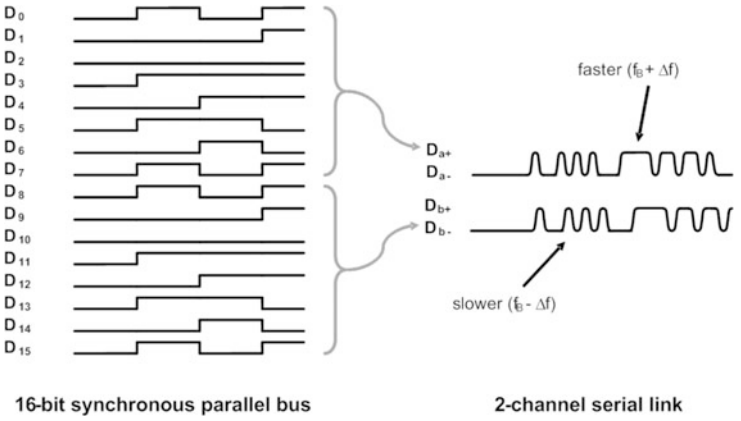

FIGURE 3.16. Comparison between a 16-bit parallel bus and a 2-channel serial link at equivalent data rates. The two serial links do not operate at exactly the same frequencies

It may again be pointed out that the economics of such applications are not comparable to those of conventional long-haul links. Compared to long-haul links, the direct modulation capability of VCSELs eliminates the need for an external (bulky) optical modulator, which has a considerable impact on cost and form factor. Operation over short distances calls for less stringent dispersion and attenuation specifications, permitting the use of 50/65 μm core multimode fibers. These fibers are not only cheaper to manufacture than single-mode fiber with core diameters of about 10 μm, especially their reduced alignment requirements for proper optical coupling has a considerable effect on the connector and system cost. As a result, the relative contribution of the transceiver units to the overall system cost in a short-distance link is much larger compared to long-haul links. In order to avoid a further increase of this ratio in presence of multiple link channels, the transceiver unit design must make use of this multiplicity to lower the cost per channel. Only in this way the overall transceiver cost, be it in terms of power, silicon area or test and verification, can result in a competitive alternative to current electrical links.

3.9 Definition of Transistor-Level Conventions

Before closing this chapter, we would like to describe some transistor-level conventions used throughout this thesis. For the reader's convenience, a comparison of the bulk-referenced framework

used in this document with the as commonly used source-referenced framework for MOS models is presented below. Additional background information on this comparison can be found in [24].

3.9.1 Large-Signal Model

Figure 3.17 shows the voltage and current conventions for an NMOS transistor in both frameworks.

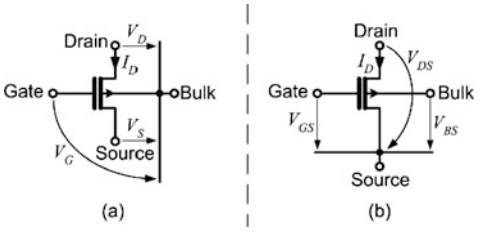

FIGURE 3.17. Voltage and current conventions in (a) a bulk-referenced model compared to (b) a source-referenced model

The large signal current-voltage characteristic of a saturated NMOS transistor is shown in Equation 3.12 for the bulk-referenced and source-referenced frameworks respectively.

$$I_D = \frac{\beta}{2n}(V_G - V_{T0} - nV_S)^2 \qquad I_D = \frac{K'}{2}\cdot\frac{W}{L}\cdot(V_{GS} - V_T)^2 \qquad \text{(EQ 3.12)}$$

While the bulk-referenced model is drain-source symmetrical and thus has an explicit formulation as a function of the source voltage, the source-referenced model represents non-zero source-bulk voltage under the form of bulk-modulated threshold voltage V_T. This behavior is implemented by the body-effect factor (called ξ instead of γ to avoid confusion with the excess thermal noise factor) in the relationship between V_T and the intrinsic threshold voltage V_{T0} (Equation 3.13). ϕ_0 is the channel surface potential in strong inversion, which can be approximated by $\phi_0 = 2\phi_F + \Delta\phi \approx 2\phi_F + 6U_T$, where ϕ_F is the Fermi potential in the substrate and $U_T = \frac{kT}{q}$ the thermodynamic voltage [25].

$$V_T = V_{T0} + \xi(\sqrt{\phi_0 + V_S} - \sqrt{\phi_0}) \qquad \text{(EQ 3.13)}$$

The slope factor n used in the bulk-referenced framework (as well as in some source-referenced representations like [26]) is determined by the ratio between the bulk-channel and the gate-channel capacitances, respectively C_{BC} and C_{GC}:

$$n = 1 + \frac{C_{BC}}{C_{GC}} \quad \text{(EQ 3.14)}$$

Both models can be related to each other by replacing

$$V_T = V_{T0} + (n-1)V_S \quad \text{(EQ 3.15)}$$

and

$$\frac{\beta}{n} = K' \cdot \frac{W}{L} \quad \text{(EQ 3.16)}$$

3.9.2 Small-Signal Model

The small-signal model of an NMOS transistor in both frameworks is presented in Figure 3.18. In the bulk-referenced framework, the saturated MOS transistor presents three conductance parameters: the gate transconductance g_m, the source transconductance g_{ms} and the output conductance g_{ds}. While the output conductance is the same in the source-referenced model, the gate transconductance $g_m' = g_m$ acts on the gate-source voltage, while the source transconductance is replaced by the bulk transconductance g_{mb}.

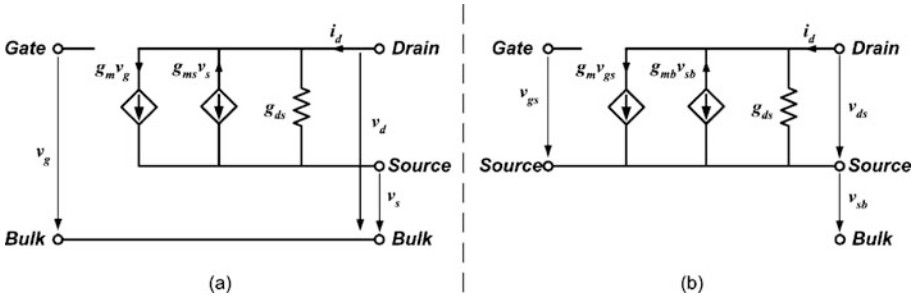

FIGURE 3.18. Small-signal MOS model in saturation: (a) bulk-referenced, (b) source-referenced

In saturation, the drain current in the bulk-referenced and source-referenced frameworks are respectively given in Equation 3.17.

$$i_d = g_m v_g - g_{ms} v_s + g_{ds} v_{ds} \qquad i_d = g_m' v_{gs} - g_{mb} v_{sb} + g_{ds} v_{ds} \quad \text{(EQ 3.17)}$$

The relationship between the source transconductance and the bulk transconductance is given in Equation 3.18.

$$g_{ms} = g_m' + g_{mb} \quad \text{(EQ 3.18)}$$

CHAPTER 4

System-Level Specifications

After this general introduction on short-distance fiber-optic communication systems, let us now focus on the receiver design. The receiver can be decomposed into several building blocks discussed below, which today are frequently integrated on separate dies (Figure 4.1). Commercial photodetectors are manufactured using III-V semiconductors like GaAs, InGaAs or InP. The transimpedance amplifiers, as well as the limiting amplifiers, use silicon or silicon–germanium (SiGe) bipolar processes to achieve a high bandwidth-power product. As clock recovery does not require the same gain-bandwidth product and is commonly built with logic gates, it is fabricated in standard CMOS. This technology-based breakup remains more or less valid for the limited amount of currently available multichannel receiver solutions [27, 2].

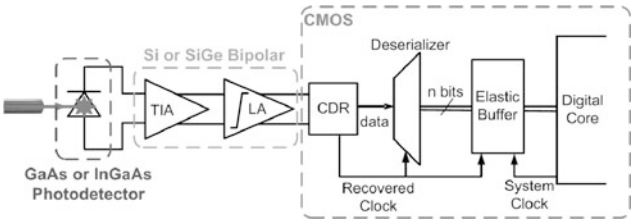

FIGURE 4.1. Components of a single-channel fiber-optic receiver

The rationale behind the goal of monolithic integration has already been discussed in the previous chapters. Following the introduction of the system-level specifications, this chapter describes a top-down methodology to obtain the block-level specifications from the system level constraints. Including all major parameters at the block level, this consistent approach allows the designer to operate optimum design choices with respect to device noise, data rate and bandwidth. It will appear that a pure CMOS

P. MULLER, Y. LEBLEBICI,
CMOS Multichannel Single-Chip Receivers for Multi-Gigabit Optical Data Communications,
DOI 10.1007/978-1-4020-5912-4_4, © 2007 Springer

implementation of the receiver is not a trivial task, illustrating why multi-chip solutions based on individual technology choices for each component today still dominate the market. Nevertheless, monolithic integration, driven by system cost and form factor, is a must for optical links to be adopted as I/O replacement.

4.1 Technology

Before addressing the design of such a high-speed CMOS receiver and compare it with state-of-the-art heterogeneous technology solutions, it is important to briefly discuss the advantages and limitations of ongoing CMOS technology scaling. Undeniably, RF CMOS, as well as wireline and fiber-optic CMOS transceivers, have become a reality through the scaling of MOS transistors. The continuous improvement of their unity current-gain frequency f_T (also called transit frequency, Equation 4.1 and Figure 4.2), a common performance indicator for a manufacturing process generation, demonstrates that available CMOS processes are indeed competitive with bipolar devices in terms of speed.

$$f_T = \frac{g_m}{2\pi \cdot C_g} = \frac{g_m}{2\pi \cdot (C_{gs} + C_{gd})} \qquad \text{(EQ 4.1)}$$

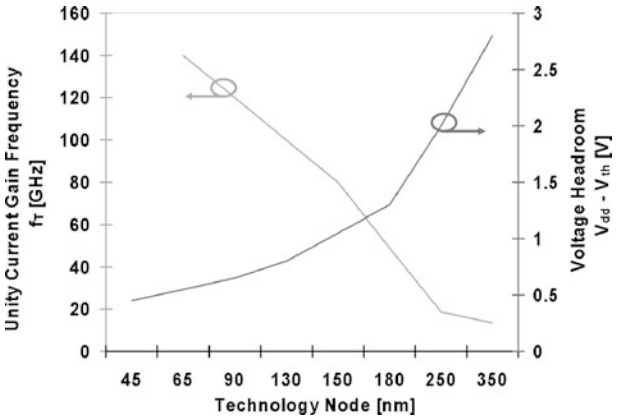

FIGURE 4.2. Unity gain frequency and supply voltage of different CMOS process generations

In this equation, C_g is the total input capacitance seen from the gate, while g_m is the transistor's transconductance. As the transit frequency does not take into account extrinsic device capacitance (e.g. drain junction capacitance C_d) which do not shrink as much as the gate from one technology node to the next, it represents an optimistic estimate of technology performance in RF and wireline applications. Designers frequently prefer to use f_{max} [28], defined as the frequency at which the capacitive component

of the input current equals the static component of a diode-connected MOS transistor (Equation 4.2). This parameter takes into account the parasitic drain capacitance that scales less efficiently than the gate dimensions. However, it is rarely published by the foundries, reducing its usage in a priori comparison of technology performance.

$$f_{max} = \frac{g_m}{2\pi(C_g + C_d)} \quad \text{(EQ 4.2)}$$

Device scaling however also requires thinning of the gate oxide, which determines the upper limit of the supply voltage in order to keep the electric field below breakdown (see Figure 4.2 for a comparison of supply voltages between CMOS and bipolar processes). Furthermore, threshold voltages of the devices do not scale at the same rate compared to the supply voltages. The decreasing voltage headroom in deep sub-micron CMOS, which is much lower than in bipolar technologies, limits the gain we can design for and the number of devices that can be stacked.

Beyond the fact that the transconductance of MOS transistors is intrinsically smaller than the one of bipolar devices, it has recently been shown that MOS transconductance g_m and output conductance g_{ds} do not scale properly over technology generations [29, 30]. Among other reasons, this is due to a reduction of carrier mobility in advanced processes, as well as to halo implants required to reduce short-channel effects and control punch-through [31]. The small-signal voltage gain per device (g_m/g_{ds}) and the transconductance to drain current i_d efficiency (g_m/i_d) for several deep-submicron processes are shown in Figure 4.3.

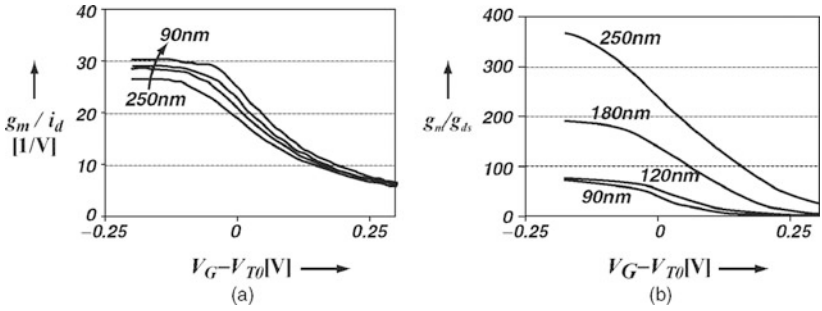

FIGURE 4.3. (a) Transconductance efficiency and (b) MOS transistor gain for four process generations in production [29]

It is obvious for the reader that almost constant transconductance efficiency over technology generations, combined with increasing output conductance due to short-channel effects, represents a real obstacle to high-gain amplifier design. Particularly in high-speed RF and wireline applications, the

speed requirements force the designers to operate the devices at high overdrive (V_G - V_{T0}) voltages, where the intrinsic device gain drops to lower two-digit numbers (Figure 4.3b). Although the high-frequency performance of each new CMOS process generation outpaces the previous ones, the decreasing gain per device increasingly becomes an issue when comparing the performance to bipolar implementations.

Due to the related constraints in terms of speed and supply voltage, the technology choice is directly influenced by the targeted data rate. Indeed, it is commonly accepted that the f_T of the technology must be ten times larger than the data rate f_B of the receiver [32]. Alternatively, Voinigescu [33] defines his criterion as $f_T > 4\,f_B$ and $f_{max} > 5\,f_B$ at $V_{DS} = V_{DD}/2$ over all temperature and process corners, which is close to the 10× requirement for peak f_T.

For illustration purposes, the building blocks of a multichannel receiver with a per-channel data rate of 2.5 Gb/s have been implemented in a 0.18 μm digital CMOS manufacturing technology.

4.2 System-Level Requirements

In order for optical I/Os to become a viable alternative to electrical interfaces, hundreds of data pads on a Pentium-equivalent chip must be replaced by optical receivers and transmitters operating under tight system constraints. High-performance microprocessor systems require robustness to device parameter variations, tolerance to supply voltage variations, operation at increased temperatures due to the system power dissipation and minimum sensitivity to supply and substrate noise, especially to switching noise. Further more, form factor, silicon area and per-link power consumption are key issues to be minimized.

Coupling the optical signal from the multimode fiber with a 50 μm core to the detector with a typical diameter of 20–70 μm for 10 Gb/s devices [34] may require microlenses to be integrated at the system level for the smaller diameter detectors. The standard fiber ribbon cables used in multichannel short-distance links have a manufacturing pitch of 250 μm. In addition to the photodiode itself, silicon area is occupied by the electronic part of the receiver, i.e. the amplifiers and the CDR. In order to be competitive, the die area occupied by an optical receiver shall be comparable or smaller than the one occupied by an equivalent electrical I/O interface, when scaled by the data rate, a metric one could call area-per-gigabit per second.

In comparison, electrical high-speed I/Os commonly use low-voltage differential signaling (LVDS), e.g. operating at 800 Mb/s. As a differential scheme, it occupies two bond pads with associated

electrostatic discharge (ESD) protection and the LVDS driver or receiver. In advanced technologies, the bond pad area can be as small as 50×50 um^2 with a pitch of 60 μm (and as low as 40 μm in staggered configuration). Combined with a commercially available intellectual property (IP) LVDS receiver operating at 800 Mb/s in a 0.18 μm CMOS technology, this solution occupies a total area of 28'000 μm^2, equivalent to 35 μm^2/(Mb/s). While the photodetector form factor appears to be competitive, it is today very hard to integrate a complete receiver with a comparable density. For sake of completeness, it should be mentioned that electrical serial links increase the data rate per I/O dramatically compared to LVDS and that they are indeed competitive in terms of silicon area, as shown by a complete 10 Gb/s transceiver occupying 1 mm^2 in a 0.13 μm CMOS process, equivalent to 100 μm^2/(Mb/s) [35].

It is commonly accepted that microprocessors use about 20% of their power consumption for I/O communications [36]. Considering high-density links with more than hundred I/O links and the need for hundreds of gigabit per second aggregate data rates in future designs, the power consumption-data rate ratio can be specified as [37]:

$$\frac{P}{f_b} = 20 - 30\frac{mW}{Gb/s} \qquad \text{(EQ 4.3)}$$

Splitting up this power budget between transmitter and receiver parts, and considering the fact that the laser and its driver require a considerable portion of this budget, we may state that the power budget for the receiver is between 10 and 15 mW/Gb/s. Despite the fact that higher input referred noise levels can be tolerated compared to long-haul receivers, the receiver power consumption is usually dominated by the amplification path. Indeed, due to the higher gain-bandwidth constraints in this section, a 2:1 ratio between amplification and CDR is a fairly good starting point for power budget allocation. The overall power consumption-data rate ratio improves when pushing the data rate closer to the limits of the considered technology generation. It goes beyond saying that these power specifications, determined from system constraints, are far from trivial to achieve at the transistor level.

4.3 Receiver System Specifications

Example specifications for one channel of 2.5 Gb/s per channel short-distance fiber optic receiver are presented in Table 4.1. They are based on well-known standards such as InfiniBand [38] and Gigabit Ethernet [39]. The stressed receiver specification defines the worst case horizontal eye closure with which the receiver performance is tested. While the gain specification of the receiver shall comply with the minimum *OMA* defined above, the sensitivity specification shall be derived from the stressed case. Compliance test points [38] for jitter measurement are located, as shown in Figure 4.4, at the input of

the receiver (TP3, in the optical domain, right before the photodetector) and at the input of the clock recovery circuit (TP4, after amplification).

TABLE 4.1. Example Receiver Specifications

Parameter	Minimum	Maximum	Units
Baud rate		2500	Mb/s
Rate tolerance		+/– 100	ppm
Average received power		–1.5	dBm
Optical modulation amplitude (OMA)	0.05		mW_{pp}
Receiver return loss	12		dB
Bit error ratio		10^{-12}	–
Stressed receiver sensitivity (OMA)	0.085		mW_{pp}
Stressed receiver ISI test	0.9		dB
Stressed receiver DCD component of DJ (at TX)	60		ps
Receiver electrical 3 dB upper cutoff frequency		2.8	GHz
Receiver electrical 10 dB upper cutoff frequency		6.0	GHz

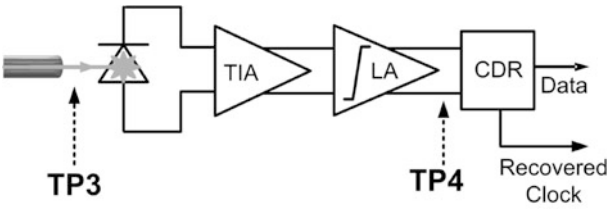

FIGURE 4.4. Jitter compliance test points

The jitter components, which specify the jitter tolerance of the clock and data recovery circuit at a given data rate, are given in Table 4.2. The values for total jitter do not include a sinusoidal jitter component. Taking as an example the InfiniBand specification, the CDR connected at TP4 "shall tolerate total jitter of 0.75 UI [...], which includes 0.10 UI of sinusoidal jitter (SJ) over a swept frequency from 1.5 MHZ to 1250 MHz", all values depending on the standard and data rate considered. The sinusoidal jitter specification, already introduced in Chapter 3, is shown in Figure 4.5.

TABLE 4.2. Receiver jitter tolerance specifications [38]

Compliance point	Deterministic jitter		Total jitter	
	UI	(e.g.) ps	UI	(e.g.) ps
TP3	0.30	120	0.53	212
TP4	0.40	160	0.70	280

Although the sum of TJ defined in Table 4.2 combined with the SJ component from Figure 4.5 would exceed 0.75 UI for jitter frequencies close to $f_B/2$, we understand that the absolute total jitter to be accommodated by the CDR is 0.75 UI. Indeed, as will be discussed in Chapter 8, the per-period sinusoidal jitter component remains below the specified SJ amplitude for low jitter frequencies.

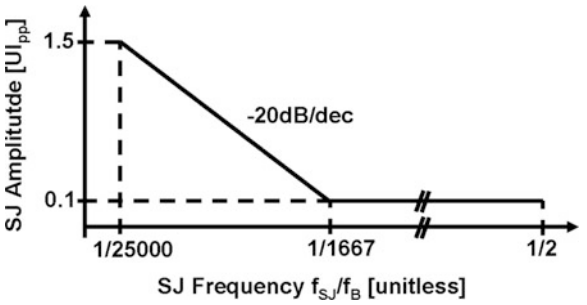

FIGURE 4.5. Jitter tolerance mask of a typical optical link standard

4.4 Subblock Parameters

In order to guarantee that the designed receiver will match the above specifications, these must be propagated down to the block level. Receiver sensitivity must be related to the top-level bit error ratio and requires most of the effort to be related to the contribution of each block.

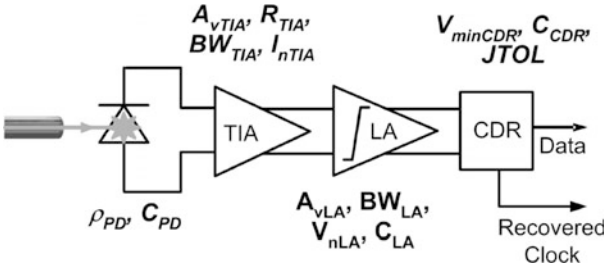

FIGURE 4.6. Receiver design parameters listed in Table 4.3

Before starting the mathematical development, we will introduce the major block-level specifications (Table 4.3), as shown in Figure 4.6. The photodetector (PD) is characterized by three parameters: the responsivity ρ_{PD} defines the amount of current delivered for a given received optical power, C_{PD} is the parasitic capacitance of the detector and I_{nPD} is the dark current, due to spontaneous carrier generation, contributing to the receiver noise budget. For many available photodetectors designed for

high-data rate applications, the dark current is negligible with respect to the other noise sources in the receiver.

TABLE 4.3. Block specifications

Parameter	Acronym	Unit
PD responsivity	ρ_{PD}	A/W
PD capacitance	C_{PD}	fF
PD dark current	I_{nPD}	nA_{RMS}
TIA voltage gain	A_{vTIA}	dB
TIA transimpedance gain	R_{TIA}	dBΩ
TIA bandwidth	BW_{TIA}	GHz
TIA input-referred noise current	I_{nTIA}	nA_{RMS}
LA voltage gain	A_{vLA}	dB
LA bandwidth	BW_{LA}	GHz
LA input-referred noise voltage	V_{nLA}	μV_{RMS}
LA input capacitance	C_{inLA}	fF
CDR input sensitivity	V_{minCDR}	mV_{PP}
CDR jitter tolerance	$JTOL$	UI
CDR input capacitance	C_{inCDR}	fF
CDR center frequency	f_0	GHz
CDR frequency tuning range	Δf_{TR}	GHz

The transimpedance amplifier (TIA) is defined by its voltage gain A_{vTIA}, the transimpedance gain R_{TIA}, the bandwidth BW_{TIA} and the input-referred noise current I_{nTIA}. The limiting amplifier (LA) is defined by its voltage gain A_{vLA}, the bandwidth BW_{LA} and the input-referred noise voltage V_{nLA}. Its input capacitance C_{inLA} is not an immediate design parameter at this point, but will be derived from the topology and sizing during the design phase. Finally, the clock recovery block specifications are input voltage sensitivity V_{minCDR}, jitter tolerance $JTOL$, the input capacitance C_{inCDR}, the center frequency f_0 and the frequency tuning range Δf_{TR}.

4.5 Transimpedance Amplifier Analysis

Before evaluation of system gain, bandwidth and sensitivity, it is preferable to spend some time on the analysis of the transimpedance amplifier. The analysis in Paragraphs 4.5.1 and 4.5.2 is based on [40], which results are then fed into the system-level analysis presented in Section 4.7. The reader should be aware that the following TIA analysis is not directly applicable to topologies which decouple the detector node from the transimpedance input node (see Paragraph 6.2.1 on “Common-Gate Input

Stage Amplifiers"). Adaptation of the presented theory to such topologies shall take into account the independence between TIA bandwidth and detector capacitance and is left to the interested reader.

4.5.1 Transfer Function

In comparison to voltage amplifiers, the transimpedance amplifier has a somewhat unusual configuration, due to the fact that it converts the input signal, a current, to a voltage at the output (Figure 4.7). Compared to a passive current-voltage conversion scheme, this transimpedance configuration interestingly divides the impedance as seen by the detector by the amplifier's open-loop voltage gain A_{vTIA}. In the initial calculations, the (parasitic) shunt capacitance C_f will be neglected for the moment.

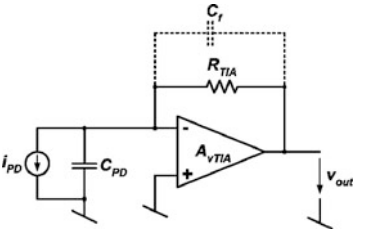

FIGURE 4.7. Transimpedance amplifier configuration

$$v_{out} = \frac{i_{PD} \cdot R_{TIA}}{1 + \dfrac{1}{A_{vTIA}(s)\beta(s)}} \qquad \text{(EQ 4.4)}$$

Feedback analysis of this circuit reveals the obvious result that the output voltage is proportional to the product of the input current and the feedback resistor (Equation 4.4). The loop gain $A_{vTIA}\beta$, defined by the open-loop gain A_{vTIA} and the feedback factor β, determines the frequency response of the TIA. Assuming that the nondominant amplifier pole is at frequencies comparable to the amplifier gain-bandwidth product GBW_{TIA}, the amplifier transfer function can be written as:

$$A_{vTIA}(s) = \frac{A^0_{vTIA}}{1 + s/\omega_{TIA}} \qquad \text{(EQ 4.5)}$$

where ω_{TIA} is the TIA open-loop bandwidth and A^0_{vTIA} is the DC open-loop gain. Also considering the external pole defined by the feedback resistor and the detector capacitance, the loop gain becomes:

$$A_{vTIA}(s)\beta(s) = \frac{A_{vTIA}(s)}{1 + sR_{TIA}C_{PD}} = \frac{A^0_{vTIA}}{(1 + sR_{TIA}C_{PD}) \cdot (1 + s/\omega_{TIA})} \qquad \text{(EQ 4.6)}$$

For $A_{vTIA}\beta >> 1$, the ideal current-voltage response is obtained. At higher frequencies, the dropping feedback factor β leads to increased requirements on open-loop gain, which however also decreases above the TIA open-loop bandwidth (Figure 4.8).

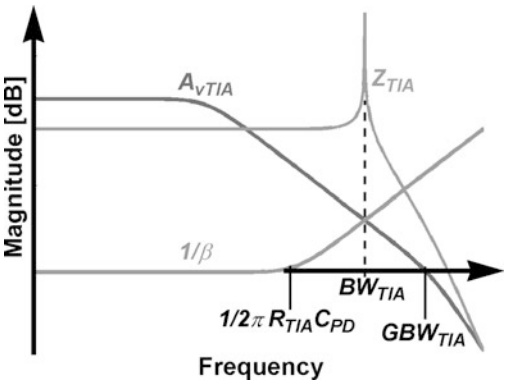

FIGURE 4.8. Typical transimpedance amplifier gain-bandwidth graph. (After Graeme [40].)

The intercept of open-loop amplifier gain and the inverse of the feedback factor corresponds to the geometric mean of the amplifier's gain-bandwidth product and the feedback pole. The 180° phase shift at this frequency creates peaking of the closed-loop transimpedance gain. Although the effective closed-loop bandwidth exceeds the intercept frequency by up to 50%, we will in the following consider it to be equal to the intercept frequency (Equation 4.7). Indeed, in physical implementations, the peaking is damped by the parasitic shunt capacitance in parallel with R_{TIA}, as well as the lower open-loop DC gain in very large bandwidth amplifiers. The low-frequency gain of the TIA shall however be large enough to obtain the $1/\beta$ intercept in the –20 dB/decade region. Otherwise, the optimum bandwidth-transimpedance gain trade-off cannot be achieved. If required, stability can be guaranteed by adding a feedback capacitor to achieve a total feedback capacitance of value $1/2\pi R_{TIA}BW_{TIA}$ in parallel with R_{TIA}.

$$BW_{TIA} = \sqrt{\frac{A_{vTIA}^{0} \cdot \omega_{TIA}}{2\pi \cdot R_{TIA} \cdot C_{PD}}} = \sqrt{\frac{GBW_{TIA}}{2\pi \cdot R_{TIA} \cdot C_{PD}}} \qquad \text{(EQ 4.7)}$$

Based on this equation for closed-loop bandwidth, the maximum feedback resistor value can be determined once the maximum achievable gain-bandwidth is known for given amplifier structure and manufacturing technology.

For the interested reader, the relationship between the closed-loop bandwidth and the open-loop amplifier gain can presented in two alternative ways. The residual error voltage at the amplifier input is given by the output voltage divided by the amplifier's open-loop gain. This error voltage is applied to the diode capacitance, absorbing part of the generated photocurrent to satisfy the resulting signal swing at the amplifier input. Another equivalent is the impedance-based interpretation, in which the dominant pole on the input node is determined by the detector capacitance and the equivalent input impedance seen by the detector, which is equal to the feedback impedance divided by the amplifier gain.

4.5.2 Noise Model

The model for the transimpedance amplifier noise analysis [40] is shown in Figure 4.9. Though alternative arrangements of the noise sources can be found in the literature, the presented figure provides the best understanding of the noise peaking theory. The amplifier is characterized by input-referred voltage noise of v_{nTIA}, associated to its power spectral density S_{nvTIA}, and current noise i_{nTIA} with its power spectral density S_{niTIA}. In CMOS implementations, the latter represents gate noise, which increases with frequency and may not be negligible at the considered data rates. While the photodetector dark current is neglected in the present model, the very large but finite (noiseless) equivalent resistance R_{PD} of the junction diode is of importance in the following calculations. v_{nR} represents the feedback resistor voltage noise.

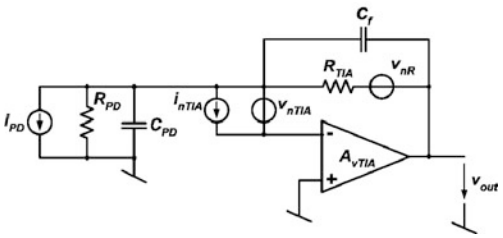

FIGURE 4.9. Noise model of the transimpedance amplifier. (After Graeme [40].)

The amplifier current noise is converted to a voltage $R_{TIA}\, i_{nTIA}$ through the feedback resistor and appears as such at the amplifier output in the same way the resistor voltage noise v_{nR} does. Due to the amplifier feedback, the amplifier voltage noise however is subject to a noise gain of:

$$A_{vn}(s) = \frac{1 + \dfrac{R_{TIA}}{R_{PD}} + sR_{TIA}(C_{PD} + C_f)}{1 + \dfrac{1}{A_{vTIA}} \cdot \left(1 + \dfrac{R_{TIA}}{R_{PD}}\right) + sR_{TIA} \cdot \left(C_f + \dfrac{C_{PD} + C_f}{A_v}\right)} \qquad \text{(EQ 4.8)}$$

At low frequencies, the large junction diode resistance keeps the noise gain close to unity. Replacing A_{vTIA} by its expression in Equation 4.5 and simplification of $1+1/A^0_{vTIA}=1$ and $R_{TIA}/R_{PD}=1$ lead to Equation 4.9.

$$A_{vn}(s) \cong \frac{1+sR_{TIA}(C_{PD}+C_f)}{1+s\left(\frac{1}{GBW_{TIA}}+R_{TIA}C_f+\frac{R_{TIA}(C_{PD}+C_f)}{A^0_{vTIA}}\right)+s^2\left(\frac{R_{TIA}(C_{PD}+C_f)}{GBW_{TIA}}\right)} \quad \text{(EQ 4.9)}$$

Considering further more that, for realistic component parameter values, $R_{TIA}C_f$ is the dominant term in the first order product, the Equation 4.10 represents a good approximation:

$$A_{vn}(s) \approx \frac{1+sR_{TIA}(C_{PD}+C_f)}{(1+sR_{TIA}C_f)\cdot\left(1+s\frac{(C_{PD}+C_f)}{C_f\cdot GBW_{TIA}}\right)} \quad \text{(EQ 4.10)}$$

Figure 4.10a illustrates the behavior of the noise gain in different frequency regions. From unity gain in Region 1, the noise gain starts to rise at frequencies around $1/2\pi R_{TIA}(C_f+C_{PD})$, where the diode capacitance begins to short-circuit the detector resistance (Region 2). At higher frequencies, around $1/2\pi R_{TIA}C_f$, the feedback capacitance shunts the feedback resistor and limits the available noise gain to $1 + C_{PD}/C_f$ (Region 3). Finally, the roll-off of the amplifier open-loop gain leads to a decrease in the noise gain at the upper end of the spectrum (Region 4). In multi-gigabit receivers, where $R_{TIA}C_{PD}$ is not very large, the poles delimiting Region 3 may overlap and result in a narrow peak, as shown in Figure 4.10b. In this case, Regions 2 and 4 abut and Region 3 is inexistent.

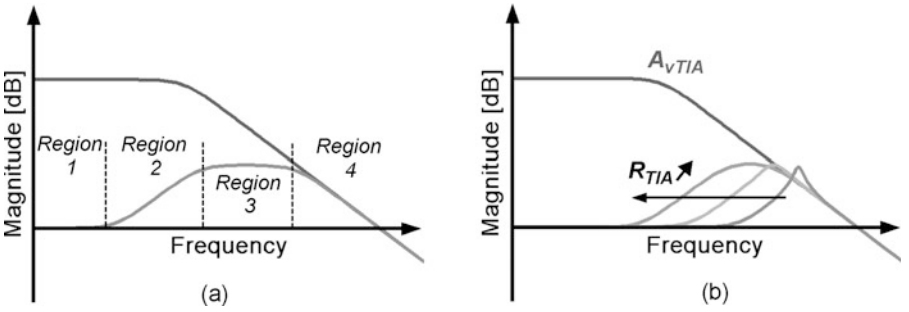

FIGURE 4.10. (a) Noise peaking behavior in TIAs and (b) noise peaking various values of R_{TIA}

An increase of R_{TIA} results in growing noise gain at lower frequencies, increasing at the same time the area under the noise peak. The resulting trade-off between increasing TIA gain and increasing noise peaking will be addressed in the following. Device flicker noise has been omitted from this derivation, as the amplifier is supposed to implement an internal offset compensation mechanism. In presence of noise peaking, it is also reasonable to neglect gate current noise. We can then obtain the RMS input referred noise current through quadratic summation and integration of the different components. The logarithmic scale of the frequency axis in Figure 4.10a may hide the fact that the noise in Regions 3 and 4 completely dominates the amplifier RMS noise. Setting C_f to the optimum value of $\frac{1}{2\pi R_{TIA} BW_{TIA}}$ defines a zero in the feedback factor at the closed-loop bandwidth of the TIA. This value maximizes the reduction of noise peaking without deterioration of the bandwidth. The integrated output noise of the amplifier and the feedback resistor can then be written as:

$$V_{noutTIA} \cong \sqrt{S_{nvTIA}^2 \cdot GBW_{TIA} \cdot \frac{C_f + C_{PD}}{C_f} + \frac{k_B T}{C_f}} \qquad \text{(EQ 4.11)}$$

In this equation, T is the absolute temperature in Kelvin and k_B is the Boltzmann constant, approximately $1.38 \cdot 10^{-23}$ J/K. The RMS noise decreases with increasing values of C_f, which justifies the value chosen above. After replacing C_f, some manipulation delivers the following equations for the total integrated output noise voltage:

$$V_{noutTIA} \cong \sqrt{\sqrt{2\pi \cdot R_{TIA} \cdot C_{PD} \cdot GBW_{TIA}} \cdot \left(S_{nvTIA}^2 \cdot GBW_{TIA} + \frac{k_B T}{C_{PD}} \right)} \qquad \text{(EQ 4.12)}$$

The input referred RMS noise current is simply obtained by division by the transimpedance R_{TIA}:

$$I_{nTIA} \cong \sqrt{R_{TIA}^{-3/2} \cdot \sqrt{2\pi \cdot C_{PD} \cdot GBW_{TIA}} \cdot \left(S_{nvTIA}^2 \cdot GBW_{TIA} + \frac{k_B T}{C_{PD}} \right)} \qquad \text{(EQ 4.13)}$$

4.5.3 Specification of Design Parameters

Based on the preceding analysis, we suggest the following flow for TIA parameter definition. First we determine the maximum achievable TIA gain-bandwidth product, where the parasitic amplifier poles shall not affect the stability (i.e. they shall be located around or beyond the GBW_{TIA}). In Figure 4.11, we notice that increasing R_{TIA} lowers or keeps constant the total input-referred noise.

At the same time, it reduces the impact of limiting amplifier noise, improving the overall sensitivity. For these reasons, the transimpedance gain is chosen, based on the closed-loop bandwidth specification, as large as allowed by the achievable TIA gain-bandwidth product. The closed-loop bandwidth also determines the value of C_f. From the total receiver noise budget defined at the system level, we obtain the noise budget for the TIA from Equation 4.13. Figure 4.11 shows the transimpedance gain, required GBW_{TIA} and input referred noise current as a function of R_{TIA} and S_{nvTIA}.

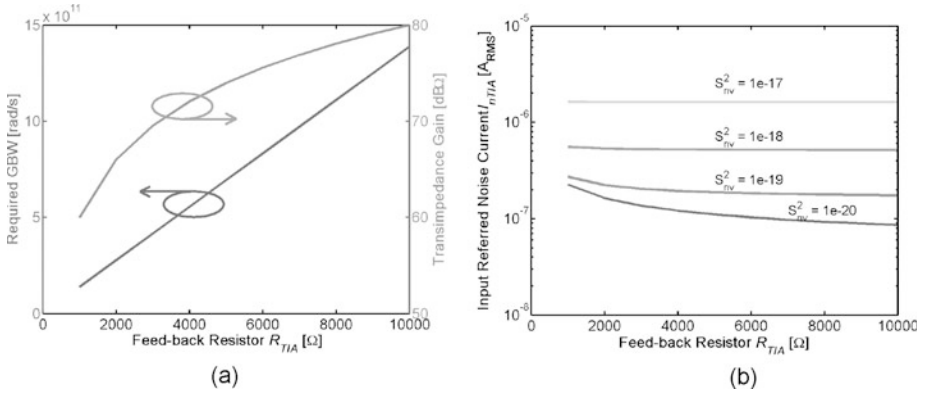

FIGURE 4.11. (a) Transimpedance gain and required GBW_{TIA}, (b) input referred noise current as a function of S_{nvTIA} and R_{TIA}

4.6 System Gain and Bandwidth Specifications

In the receiver chain, most block-level parameters influence more than one system-level specification. For this reason, we will first present the different relationships in the form of equations. Then, we will try to extract convenient values for the parameters to meet the various system constraints. The first requirement for the amplifiers is to deliver a sufficiently large signal to the CDR input for its sensitivity to be respected (Equation 4.14). Under this condition, we can safely neglect the CDR amplitude noise in the receiver noise calculations later.

$$OMA \cdot \rho_{PD} \cdot R_{TIA} \cdot A_{vLA} - V_{minCDR} \qquad \text{(EQ 4.14)}$$

We have already mentioned in the previous chapter that the bandwidth is constrained both by the noise budget and by the timing budget. Low bandwidth may add deterministic jitter in the form of intersymbol interference, making the CDR design tougher, while high bandwidth results in a larger noise bandwidth for the input noise. It has been shown earlier that the power spectral density concentrates most of the signal energy at lower frequencies (Figure 3.4). As a transimpedance amplifier bandwidth of 0.7–0.75 f_B hardly creates any ISI, this value is traditionally considered to be optimum with

respect to the noise-ISI trade-off [42]. We will discuss the validity of this assumption in the scope of short-distance CMOS receivers throughout the following section.

The bandwidth of a cascade of blocks (i.e. TIA + LA) being lower or equal to the lowest of the individual block bandwidths, the limiting amplifier bandwidth must be high enough to minimize the reduction in system bandwidth. Further more, because the output stages of the LA operate in large-signal regime, limiting of the output signal has an enhancement effect on the signal bandwidth [19]. A conservative value for the limiting amplifier bandwidth specification can be defined as 1.5 times the data rate.

4.7 Bit Error Ratio Evaluation

Bit error ratio calculations are based on the decomposition in horizontal and vertical eye closure. Vertical eye closure is due to (random) amplitude noise and intersymbol interference, while horizontal eye closure is caused by jitter components of different origins (Figure 4.12). While the clock recovery circuit is specified by the horizontal eye closure (i.e. jitter and timing specifications), the amplification section contributes to both horizontal and vertical eye closure. Indeed, bandwidth limitations in this section lead to deterministic jitter, whereas AM–PM conversion in the limiting amplifier translates amplitude noise into random jitter. We will see later that front-end bandwidth determines both random and deterministic errors in the system, while jitter tolerance and amplifier noise define how large errors the receiver can handle.

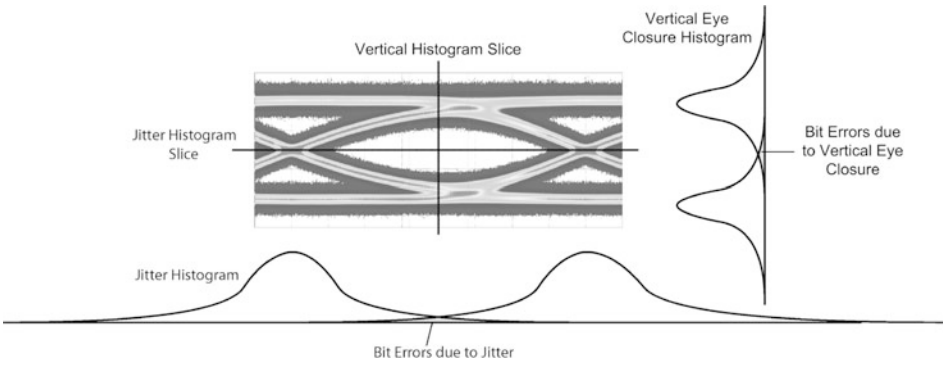

FIGURE 4.12. Illustration of bit errors due to horizontal and vertical eye closure

In the following BER calculations, we will rely on the definition of Q_{BER} to simplify the expressions. As previously defined, a bit error ratio below 10^{-12} requires $Q_{BER} > 7.035$. Maximum Integrated

products [43] provides the basic calculations for system-wide sensitivity analysis. We considerably extend this approach by introducing jointly (a) the noise-bandwidth trade-off in the amplification path and (b) the ISI versus bandwidth relationship. The result of this development provides valuable design insight to properly chose the bandwidth resulting in an optimum noise-ISI trade-off.

4.7.1 Vertical Eye Closure Contribution

Let us first consider the contribution of vertical eye closure to the bit error ratio. The total RMS input referred current noise of the receiver is:

$$I_{ntotRMS} = \sqrt{I_{nTIA}^{\,2} + \left(\frac{V_{nLA}}{R_{TIA}}\right)^2} \qquad \text{(EQ 4.15)}$$

The first term under the square root corresponds to the noise contribution of the TIA, whereas in the second term, the input-referred noise of the LA is divided by the TIA gain to be referred back to the receiver input. In order to fully take into account the influence of the feedback resistor value, the TIA input-referred current noise power spectral density (PSD) $S_{niTIA}^{\,2}$ is in the following computed from its output-referred voltage noise PSD $S_{nvTIA}^{\,2}$ divided by R_{TIA}. The reader should not be confused by the fact that $S_{nvTIA}^{\,2}$ is output-referred, while $S_{nvLA}^{\,2}$ is input-referred. Basically, both voltage noise PSDs are referred to the same signal node at the output of the TIA, which is at the same time the input of the LA.

The high transimpedance gain achievable in older technologies and at lower data rates allowed us to neglect the noise of limiting amplifiers in the past. This is assumption is no more valid in advanced CMOS processes because it becomes more and more difficult to achieve high gain at multi-gigabit data rates (see [37]). Both TIA and LA noise are considered to be Gaussian processes. Vertical eye closure is determined by both random and deterministic contributions, i.e. amplitude noise and intersymbol interference. As the latter is a deterministic process, it adds in amplitude to the input noise current. The input sensitivity can thus be written as:

$$I_{in_{min}} = 2Q_{BER} \cdot I_{ntotRMS} + 2 \cdot \frac{V_{ISI}}{R_{TIA}} \qquad \text{(EQ 4.16)}$$

It is related to the minimum optical modulation amplitude through the detector responsivity:

$$I_{in} = OMA \cdot \rho_{PD} \quad \text{(EQ 4.17)}$$

We have previously seen that the ISI component itself depends on the system bandwidth. In a general way, it can be considered to be composed of one term related to channel and transmitter contributions and a second term given by the receiver bandwidth. For noise optimization reasons, the receiver bandwidth is in most cases determined by the TIA. The worst-case intersymbol interference appears in a string of identical data followed by the exponential growth due to a single data edge.

$$V_{ISI} = ISI_{chan} + V_{PP} \cdot e^{-\frac{2\pi \cdot BW_{TIA}}{f_B}} \quad \text{(EQ 4.18)}$$

In standard transmitter compliance masks, acceptable unilateral vertical eye closure at the transmitter output is limited to 20% of the signal swing [38], which will be taken into account by replacing ISI_{chan} by ζV_{PP}, with $\zeta = 0.2$. Channel ISI being due to various phenomena from fiber dispersion to transmitter bandwidth, some effects may not cumulate with receiver bandwidth limitations. The reader should therefore be aware that the present approach may somewhat overestimate total ISI in the system. V_{PP}, the peak–peak voltage at the output of the transimpedance amplifier, can obviously be written as $V_{PP} = R_{TIA} I_{in}$. Feeding these results back into Equation 4.16, we obtain the receiver input sensitivity:

$$I_{in_{min}} = \frac{2Q_{BER} \cdot I_{ntotRMS}}{1 - 2\zeta - 2e^{-2\pi \cdot BW_{TIA}/f_B}} \quad \text{for } \zeta + e^{-2\pi \cdot BW_{TIA}/f_B} < 0.5 \quad \text{(EQ 4.19)}$$

Bandwidth values which do not respect the given boundary condition, though quite improbable to appear in real designs, would result in complete vertical eye closure, in which case the specified BER cannot be met. For a bandwidth of $0.75 f_B$, the denominator tends to one for small ζ, but decreases in the

general case. In any case, due to the bandwidth dependency of the numerator, a denominator close to one does not always optimize receiver sensitivity.

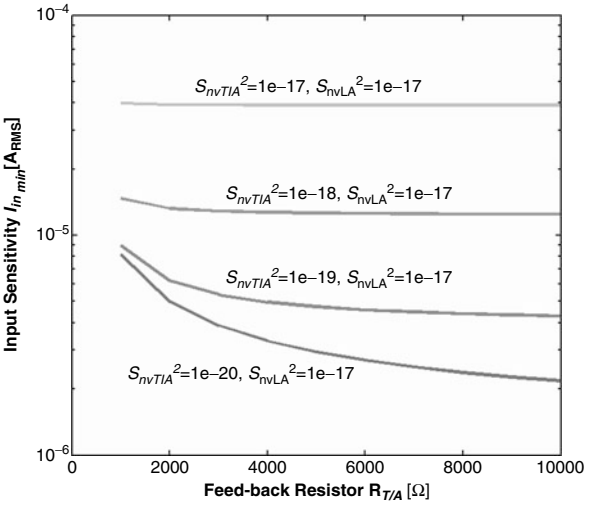

FIGURE 4.13. Input sensitivity due to vertical eye closure, plotted versus transimpedance gain and TIA and LA noise

Combining Equations 4.13, 4.15 and 4.19, we can plot the dependency of the input sensitivity on transimpedance gain and amplifier noise, where the closed-loop TIA bandwidth is kept constant and the TIA gain-bandwidth changes with R_{TIA} (Figure 4.13). This illustrates that the LA and feedback resistor noise contributions are negligible at large TIA noise densities. They only appear at low TIA noise densities and low transimpedance gain values. Plotting the receiver input sensitivity as a function of TIA bandwidth at constant TIA GBW gives additional design insight (Figure 4.14).

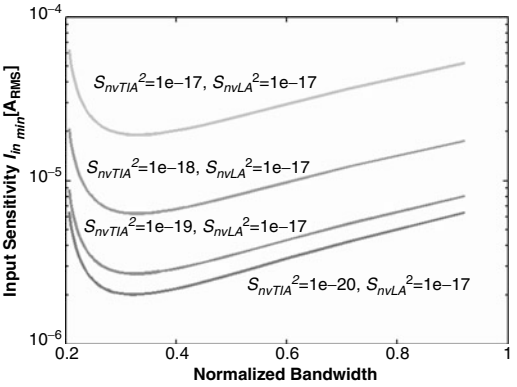

FIGURE 4.14. Input sensitivity due to vertical eye closure, plotted versus transimpedance bandwidth and TIA and LA noise

This result shows an optimum bandwidth which is relatively independent of the transimpedance amplifier noise PSD. Indeed, at high transimpedance gain, the amplifier noise is suppressed accordingly. Only strong ISI components counteract this relationship at very low bandwidths. Such a strong impact of device noise on the receiver sensitivity is due to the increasing Q_{BER} values at low BER.

4.7.2 Horizontal Eye Closure Contribution

The analysis of the jitter-BER relationship requires some additional mathematical derivation. Assuming first the presence of deterministic and random jitter sources and neglecting vertical eye closure, we can first write the Q-factor as [43]:

$$Q_{BER} = \frac{JTOL - DJ_{PP}}{2 \cdot RJ_{RMS}} \quad \text{(EQ 4.20)}$$

If we consider the fact that random jitter is due to additive white noise around the signal transitions, we can estimate the noise amplitude V_{nRMS} as a function of the 10–90% rise and fall times $t_{rt\text{-}ft}$ at the input of the LA and the signal swing V_{PP}:

$$RJ_{RMS} = \frac{t_{rt-ft}}{\frac{V_{PP}}{V_{nRMS}} \cdot 0.8} \quad \text{(EQ 4.21)}$$

The random noise sources considered here are the device noise sources in the amplifier chain, which are in fact the same sources previously considered in the vertical eye closure calculations. Indeed, the amplitude noise around the signal transitions is converted to jitter through AM–PM conversion in the limiting amplifier, as discussed in Section 7.1.

The transimpedance amplifier and photodetector can be expected to be linear for small signals like noise, allowing us to refer the noise voltage back to the receiver input:

$$\frac{I_{in}}{I_{ntotRMS}} = \frac{V_{PP}}{V_{nRMS}} \qquad \text{(EQ 4.22)}$$

Feeding this into Equation 4.20 and extracting the input sensitivity, we obtain:

$$I_{in_{min}} = \frac{2Q_{BER} \cdot t_{rt-ft} \cdot I_{ntotRMS}}{(JTOL - DJ_{PP}) \cdot 0.8} \qquad \text{(EQ 4.23)}$$

For a first order low-pass transfer function (as well as for higher order functions with reasonable damping), the rise-time bandwidth product can be estimated to 0.22 [43], leading to:

$$I_{in_{min}} = \frac{2Q_{BER} \cdot \frac{0.22}{BW_{TIA}} \cdot I_{ntotRMS}}{(JTOL - DJ_{PP}) \cdot 0.8} = \frac{0.55 \cdot Q_{BER} \cdot I_{ntotRMS}}{(JTOL - DJ_{PP}) \cdot BW_{TIA}} \qquad \text{(EQ 4.24)}$$

Comparable to the vertical eye closure, the horizontal eye closure is determined by both random noise sources, discussed above, and deterministic noise sources, included in the DJ_{PP} term in Equation 4.24. The latter term can be decomposed in a channel and transmitter part and a receiver part.

$$DJ_{PP} = DJ_{chan} + DJ_{RX} \qquad \text{(EQ 4.25)}$$

Indeed, the receiver may contribute to deterministic through the limited transimpedance amplifier bandwidth. A detailed analysis of deterministic jitter and its relationship to bandwidth limitations has been presented in [44]. According to this reference, the deterministic jitter contribution of a first order system with bandwidth BW operating at a data rate f_B is given by:

$$DJ_{RX} = \frac{f_B}{4\pi \cdot BW} \cdot \ln\frac{1+\alpha}{1-\alpha+\alpha^2} \quad \text{[UI]} \quad \text{where } \alpha = e^{-(2\pi BW)/f_B} \qquad \text{(EQ 4.26)}$$

The resulting optical input sensitivity in presence of jitter is shown in Equation 4.27, where the deterministic jitter in the receiver at a bandwidth of 0.75 f_B is negligible, as was ISI in the vertical eye closure analysis.

$$I_{in_{min}} = \frac{0.55 \cdot Q_{BER} \cdot I_{ntotRMS}}{\left(JTOL - DJ_{chan} - \frac{f_B}{4\pi \cdot BW_{TIA}} \cdot \ln \frac{1+\alpha}{1-\alpha+\alpha^2} \right) \cdot BW_{TIA}} \quad \text{(EQ 4.27)}$$

Considering a channel contribution to deterministic jitter of 0.4 UI and a jitter tolerance of 0.75 UI, the resulting input sensitivity in presence of timing jitter is shown in Figure 4.15.

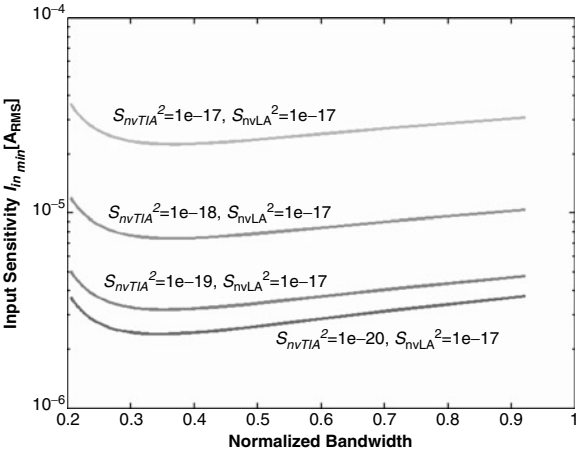

FIGURE 4.15. Input sensitivity in presence of jitter versus transimpedance gain and amplifier noise for *JTOL-DJ$_{PP}$* = 0.3 UI

Here, the optimum is weakly TIA noise dependent, but the slopes on either side of the optimum are moderate. It is also important to notice that the optimum input sensitivity values are approximately equal in horizontal and vertical eye closure analyses, which indicates that both closure mechanisms contribute equally to bit errors. However, when moving away from the optimum, vertical closure becomes dominant, requiring higher input sensitivity, at least for the considered values of *JTOL* and DJ_{PP}.

4.7.3 Combined Horizontal–Vertical Eye Closure Estimation

Horizontal and vertical eye closure are due to the same noise sources, i.e. they share the $I_{ntotRMS}$ factor and signal bandwidth. Due to bandwidth limitations in the channel and TIA, the random components converted into RJ at the data edge and the amplitude noise components are statistically correlated, as shown by the autocorrelation function of band-limited white noise of amplitude A_{RMS}:

$$R_{xx}\left(\frac{T}{2}\right) = A_{RMS}\sin c\left(2BW_{TIA}\frac{T}{2}\right) = A_{RMS}\sin c\left(\frac{BW_{TIA}}{f_B}\right) \quad \text{(EQ 4.28)}$$

Although the correlation term becomes small only for $BW_{TIA} > 0.8\,f_B$, we will neglect this effect in the following and leave the completion of the following development with correlation effects to the interested reader. We can now calculate the combined overall receiver sensitivity given in Equation 4.29, where it is important to notice that within the $(JTOL\text{-}DJ_{PP})BW_{TIA}$ product, the units must be chosen consistently (i.e. all factors in SI units – seconds and hertz – or in normalized units – UI and 1/UI).

$$I_{in_{min}} = Q_{BER}\cdot I_{ntotRMS}\sqrt{\left(1 - 2\zeta - 2e^{-2\pi\cdot BW_{TIA}/f_B}\right)^{-2} + \left(\frac{0.55}{(JTOL - DJ_{PP})\cdot BW_{TIA}}\right)^2} \quad \text{(EQ 4.29)}$$

$$\text{with } DJ_{PP} = DJ_{chan} + \frac{f_B}{4\pi\cdot BW_{TIA}}\cdot\ln\frac{1+\alpha}{1-\alpha+\alpha^2},\ \text{for } \zeta + e^{-2\pi\cdot BW_{TIA}/f_B} < 0.5$$

This equation illustrates the fact that the total amplifier noise I_{ntot} present in the circuit does not determine whether horizontal eye closure (jitter) or vertical eye closure dominate the bit error ratio. Indeed, the boundary between both regimes is defined by the sum under the square root, which is governed by the TIA bandwidth, the jitter tolerance and the channel properties ζ and DJ_{chan}. As the transimpedance amplifier bandwidth also acts on I_{ntot}, optimization of the above expression is not an easy task. In the following, we will provide two example approaches to this multivariable problem.

Again considering the TIA gain-bandwidth product as fixed, one can combine the previous horizontal and vertical eye closure results to plot total input sensitivity as a function of receiver bandwidth (Figure 4.16). This case appears when the TIA gain-bandwidth maximally exploits the available technology speed. Unsurprisingly, we observe again a sensitivity optimum which is more or less noise independent, but is a function of channel ISI and DJ. As already discussed, the increase of the

transimpedance gain, at the origin of the bandwidth reduction, suppresses the TIA core amplifier and limiting amplifier noise efficiently, thus pushing the optimum to low-bandwidth values.

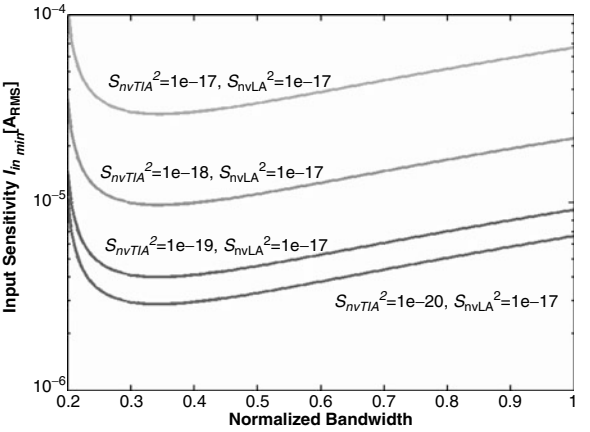

FIGURE 4.16. Input sensitivity versus TIA bandwidth at constant GBW_{TIA}

Let us consider now the more general case where the TIA gain-bandwidth is a free parameter (i.e. it can be adjusted as needed), TIA gain and bandwidth thus being chosen independently. In that case, Equation 4.29 can be decomposed in a total receiver noise factor $I_{ntotRMS}$ containing random sources of eye closure and the square root factor containing the deterministic contributions to eye closure. Noise power spectral density of each amplifier being a function of the amplifier topology and its power consumption, let us consider them to be constant. Under that assumption, $I_{ntotRMS}$ increases with increasing bandwidth, whereas the deterministic contributions decrease, as shown in Figure 4.17 for various noise PSDs.

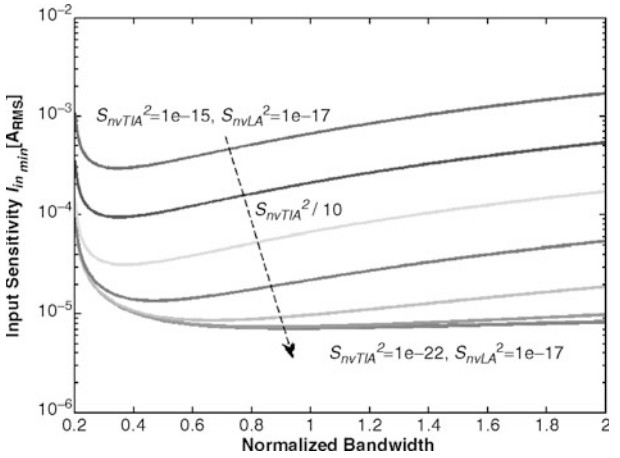

FIGURE 4.17. Input sensitivity versus TIA bandwidth at constant transimpedance gain

For given noise PSDs, the optimum TIA bandwidth can be determined with respect to sensitivity (Figure 4.18). Depending on the amount of channel ISI and DJ, the optimum at low-device noise levels is located between 0.75 f_B and 1.5 f_B, whereas this optimum moves to lower bandwidth values at increasing amplifier noise.

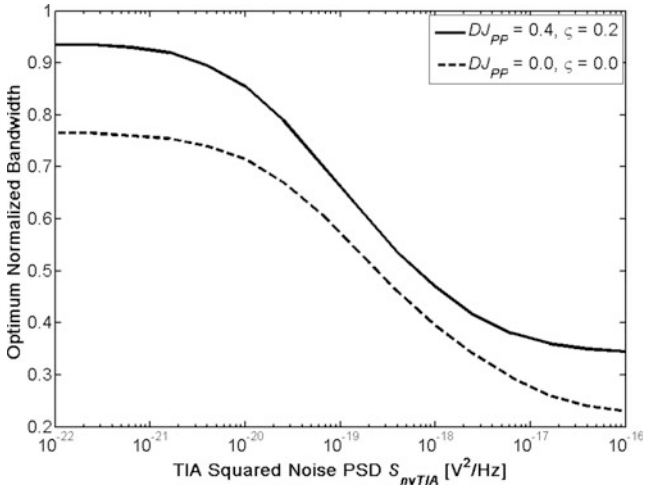

FIGURE 4.18. Optimum TIA bandwidth versus noise PSD

Interestingly, we find the the well-known bandwidth specification of 0.75 f_B, frequently presented as a rule of thumb for good receiver design, as an optimum for low-noise receivers and ideal channel conditions. With nonideal channel conditions, the optimum for low-noise receivers is located at higher bandwidths, but a look at Figure 4.17 shows that the minimal gain in sensitivity does not justify a corresponding increase in power consumption. This rule of thumb has its origins in the design of long-haul systems with very stringent jitter specifications using compound bipolar technologies with low-device noise levels. In such applications, any addition of deterministic jitter due to limited receiver bandwidth considerably harms the overall performance in that application, because of the already small horizontal eye opening of the received data.

The optimum receiver bandwidth for integrated short-distance receivers may however not match this rule of thumb. Today's advanced CMOS processes can achieve competitive jitter tolerance thanks to the high-transit frequency of deep submicron devices at the expense of intrinsically higher device noise. The presented results show that the well-known rule of thumb does not result in optimum receiver performance under these considerations and that the optimum receiver bandwidth must be carefully determined based on the noise-ISI trade-off.

4.8 Block Specification Flow

This section proposes a block-level specification derivation that can be used to define the silicon implementation of the receiver building blocks. Based on the fact that CMOS receivers tend to be TIA gain-bandwidth-limited, it summarizes the derivation of the block-level specifications from the system level with a numerical example. The calculations illustrate the flow shown in Figure 4.19, starting from the detector and following the noise-ISI trade-off previously introduced. At the receiver input, the detector is characterized by a minimum responsivity of 0.4 A/W at the desired wavelength and a capacitance of 600 fF, it's dark current being negligible.

The input capacitance of a reasonably sized limiting amplifier can be estimated to 200 fF. With this assumption, the maximum achievable TIA gain-bandwidth product is located around 50 GHz in the typical case. While the optimum bandwidth for GBW-limited designs is at 0.3 f_B, some margin should be taken for process variations against the onset of deterministic eye closure. A reasonable bandwidth specification for the typical case is given by BW_{TIA} = 0.5-0.6 f_B. The upper bandwidth specification was set to 2.8 GHz based on the already-mentioned communication standards. This bandwidth specifications can be met with a typical transimpedance value of 6.5 kΩ, leading to a transimpedance gain of 76 dBΩ. The minimum TIA open-loop gain is determined in order for the dominant pole to be located below the closed-loop bandwidth. In this case, we obtain a minimum gain of 40 (32 dB).

Based on the optical modulation amplitude, the detector responsivity and the input sensitivity of the current-mode logic gates in the CDR, characterized to 350 mV_{PP}, the limiting amplifier gain must exceed 12 (21 dB) over process corners and temperature variations. The minimum bandwidth has been previously defined to be 3.75 GHz when driving the estimated input capacitance of the CDR of about 40fF.

The jitter tolerance according to the standards mentioned in Paragraph 4.3 shall be 0.75 UI, with 0.4 UI of deterministic jitter. Although this choice makes the receiver operate in a vertical eye closure-dominated regime, it is sufficiently close to the intersection with the horizontal eye closure-dominated mode, where both regimes contribute equally to bit errors. During the CDR design phase, we use the remaining timing margin to relax the jitter tolerance specification, which allows us to accommodate a less tolerant CDR topology, presented in Chapter 8.

The sensitivity optimization also determines the maximum integrated noise current specification of 684 nA_{RMS} for an input current of 20 μA and $Q_{BER} = 7.1$. Allocating 90% of the noise voltage budget to the TIA, its input referred noise current shall be smaller than 615 nA_{RMS}, leaving the limiting amplifier with an input referred noise voltage budget of 103 μV_{RMS}.

The calculated specifications are summarized in Table 4.4, while Figure 4.19 graphically represents this top-down specification flow.

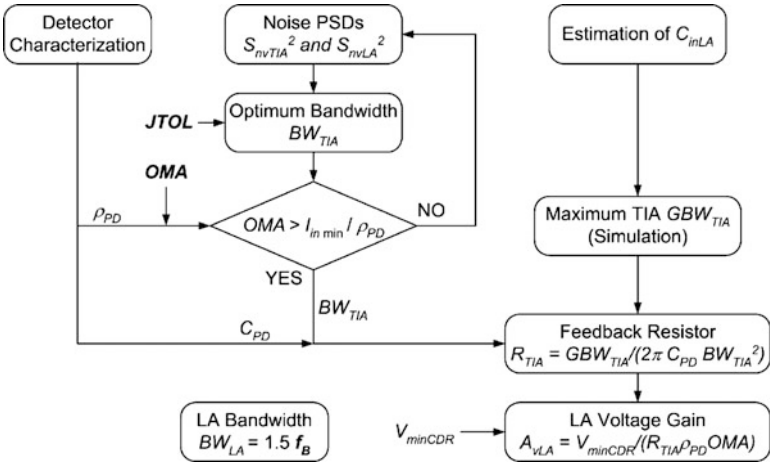

FIGURE 4.19. Specification flow for top-down propagation of receiver specifications (in bold the system level specifications)

TABLE 4.4. Calculated block specifications

Parameter	Acronym	Minimum	Maximum	Unit
PD responsivity	ρ_{PD}	0.4		A/W
PD capacitance	C_{PD}		600	fF
PD dark current	I_{nPD}		~0	nA
TIA voltage gain	A_{vTIA}	30		dB
TIA transimpedance gain	R_{TIA}	6.5k / 76		Ω / dBΩ
TIA bandwidth	BW_{TIA}	1.3	2.8	GHz
TIA output-referred voltage noise PSD	${S_{nvTIA}}^2$		1.0e–18	V^2/Hz
TIA input-referred integrated noise current	I_{nTIA}		615.0e–9	A/√Hz
LA voltage gain	A_{vLA}	21		dB
LA bandwidth	BW_{LA}	3.75		GHz
LA input-referred noise voltage	V_{nLA}		103	μV_{RMS}
LA input capacitance	C_{inLA}		200	fF
CDR input sensitivity	V_{minCDR}	350		mV_{PP}
CDR jitter tolerance	$JTOL$	0.75		UI
CDR input capacitance	C_{inCDR}		40	fF
CDR center frequency	f_0	2.5		GHz
CDR frequency tuning range	Δf_{TR}	–100	100	ppm f_0

CHAPTER 5

Silicon Photodetectors

To achieve complete monolithic integration of optical receivers on silicon substrates, we have to address the challenges of high-speed infrared photodetection in silicon. For obvious reasons, alternative substrates like GaAs or InP will not be discussed. First, the general principles of photodetection will be presented, then the dominant photodetector types for this application will be introduced.

5.1 Photodetection

Photons entering the silicon substrate eventually create electron-hole pairs, which then move, following drift and diffusion mechanisms, as any free carrier present in the surroundings. The average

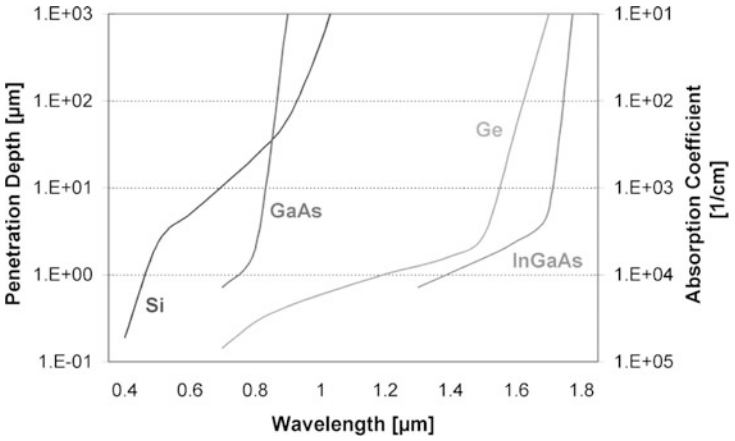

FIGURE 5.1. Optical penetration depths for various semiconductor materials. (After Davidson and Alexander [45,46].)

P. MULLER, Y. LEBLEBICI,
CMOS Multichannel Single-Chip Receivers for Multi-Gigabit Optical Data Communications,
DOI 10.1007/978-1-4020-5912-4_5, © 2007 Springer

penetration depth in silicon, i.e. the depth where carriers are generated, depends on the wavelength of the incoming light (Figure 5.1).

From the three major absorption mechanisms presented in Figure 5.2, only intrinsic band-to-band absorption is relevant in photoreceivers, which fortunately is the dominant mechanism in silicon. For an electron to transit from the valence band to the conduction band, the excitation energy must exceed the material bandgap energy E_g (E_g(Si)= 1.12 eV). This requirement determines an upper limit to the wavelength to be absorbed. In the case of silicon, the wavelength must be smaller than 1,110 nm to be possibly absorbed, defining the asymptotic behavior of the penetration depth at increasing wavelength shown in Figure 5.1. Today, lower fiber cost and the availability of VCSEL sources are the major drivers of the 850 nm short-distance window. In the future, the absorption capability of silicon at these wavelengths should even reinforce this situation.

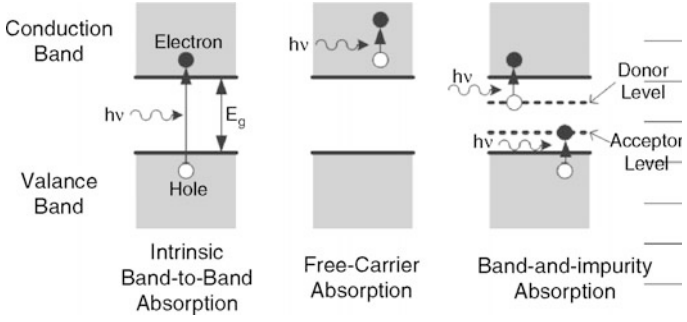

FIGURE 5.2. Light absorption mechanisms. (After Alexander [46].)

Electronic charge transport in semiconductors is due to either of two physical factors: an electric field causes drift currents of minority carriers, while a majority carrier concentration gradient leads to diffusion currents. As the movement of drift carriers is directed by the electric field, the effective charge displacement is considerably faster compared to the diffusion currents relying on stochastic thermal motion.

When applying an electric field around the absorption region, the generated electron-hole pairs are separated before recombination and dominantly collected at the contact regions of the device, a small amount being lost through recombination. Different devices have been built based on this principle and several topologies usable in optical communications, namely the PN and PIN diodes, the avalanche photodiode (APD) and the metal–semiconductor–metal (MSM) detector, will be further discussed.

5.2 PIN Photodiodes

The most basic structure required for creating a depletion region with an electric field is a p-n diode. The carriers absorbed in the depletion region will move relatively quickly to the cathode and anode and generate a photocurrent. As soon as the carriers move away from (or are generated away from) the center of the depletion region, the electric field drops and so does the drift velocity. This drawback can be circumvented by introducing an intrinsic region between the n-doped and p-doped regions (Figure 5.3). The resulting stack of p-doped material, intrinsic material and n-doped material is called a PIN diode.

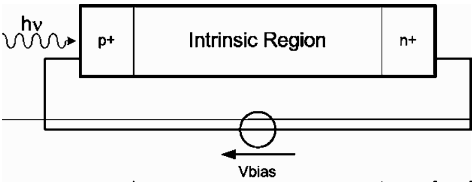

FIGURE 5.3. PIN photodiode structure

When a sufficiently high electric field is applied, the depletion region may span over the whole intrinsic part and the carriers will travel near their saturation velocity in the whole depletion region. In this situation, the diffusion currents are limited to the doped regions. The transit time defines the time an electron needs to travel from one contact to the other (Equation 5.1). Obviously, in order to obtain a good photodetector efficiency, this time shall be much smaller than the average carrier lifetime,

$$\tau_{transit} = \frac{l_a}{\mu_e \cdot E} \qquad \text{(EQ 5.1)}$$

with l_a the length of the absorption region, μ_e the mobility of electrons and E the electric field. The bandwidth of such a PIN photodiode is given by inverse of the geometric mean of transit time and RC time constant (Equation 5.2), where the latter is determined by the parasitic capacitance of the diode and the series combination of parasitic contact resistance R_s and equivalent load resistance R_L. Considering thin contact regions, such that the depletion region covers most of the absorption region, both terms depend on the length of the depletion region l_d (which is then equal to the length of the absorption region l_a). For a given device area A, typically determined by the diameter of the core of the fiber, there is a value for this parameter which optimizes the bandwidth (Figure 5.4).

$$BW[Hz] = \left[\left(\frac{2\pi \cdot l_d}{2.78 \cdot \mu_e \cdot E}\right)^2 + \left(2\pi \cdot (R_s + R_L) \cdot \varepsilon_0 \varepsilon_r \cdot \frac{A}{l_d}\right)^2\right]^{-\frac{1}{2}} \quad \text{(EQ 5.2)}$$

Furthermore, we must not forget that the length of the diffusion region also determines the quantum efficiency of the detector (neglecting reflection at the surface):

$$\eta = \left(1 - e^{-\alpha \cdot l_d}\right) \quad \text{(EQ 5.3)}$$

The detector responsivity can be written as a function of the quantum efficiency:

$$\rho_{PD} = \frac{\eta \cdot q}{h\nu} \quad \text{(EQ 5.4)}$$

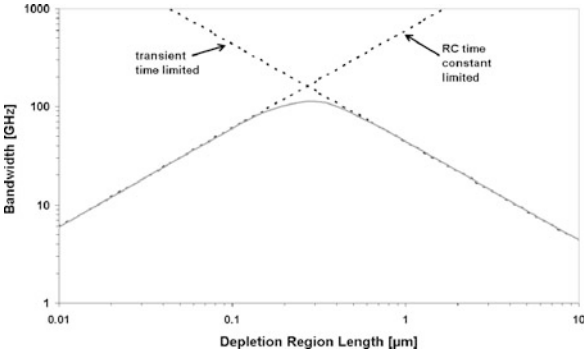

FIGURE 5.4. Illustration of the trade-off between the detector bandwidth and the length of depletion region for a silicon photodetector. (After Alexander [46].)

At 850 nm wavelength, the responsivity to efficiency ratio is equal to 0.685 A/W. The performances of today's compound semiconductor detectors with a responsivity around 0.6 A/W would require a quantum efficiency of 88%, resulting in a absorption length of 80 µm. As can be seen in Figure 5.4, this large number results in a high transit time constant and low bandwidth in a classical silicon PIN photodetector. Innovation is needed to improve the bandwidth-efficiency trade-off of silicon PIN diodes beyond their physical limitations.

5.3 Avalanche Photodiodes

Instead of generating one electron-hole pair per absorbed photon, the avalanche phenomenon in diodes at high reverse bias allows for the exploitation of carrier multiplication. At high reverse bias, velocity saturated carriers cause impact ionization, generating additional carriers, themselves subjected to the high electric field leading to more impacts (like an avalanche). A precise knowledge of the

achieved multiplication factor is obtained by thorough control of the reverse bias voltage. Several reasons however limit their usage in high-speed optical receivers.

Due to the randomness of the carrier multiplication and possible avalanche generation due to spontaneous electron-hole generation, APDs suffer of excess noise compared to PIN diodes. Furthermore, the avalanche needs some time to build up, intrinsically limiting the detection speed at high multiplication factors. Indeed, this time constant sums up with the RC and transient time constants previously introduced for PIN structures. This very particular "gain-bandwidth" trade-off is the reason why silicon APDs are not commonly used in multi-gigabit receivers.

5.4 Differential N-Well Detector

The most straightforward way to obtain a vertical photodetector in CMOS technologies uses an n+ diffusion in p-well or an n-well in p-substrate as shown in Figure 5.5. In both cases, a large amount of the incoming light at 850 nm wavelength is absorbed deep in the silicon substrate as given by the penetration depth in Figure 5.1. The n-well detector is usually preferred due to the deeper n-doped region. Nevertheless, the diffusion of these carriers in a region where the electric field is small limits the intrinsic bandwidth of such detectors to the lower megahertz range.

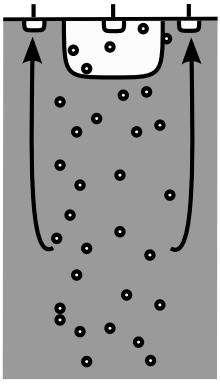

FIGURE 5.5. Vertical N-well photodetector. (After Hermans and Steyaert [47].)

Furthermore, these diffusion carriers generated deep in the substrate are shared by neighboring detectors in a multidetector array, potentially leading to interchannel crosstalk. In order to improve the frequency response, it has been proposed to subtract the response of a dark (i.e. not illuminated) detector

from the active detector [49–47]. For this purpose, active detection fingers are alternated with dark detection fingers capturing the deep carriers only (Figure 5.6).

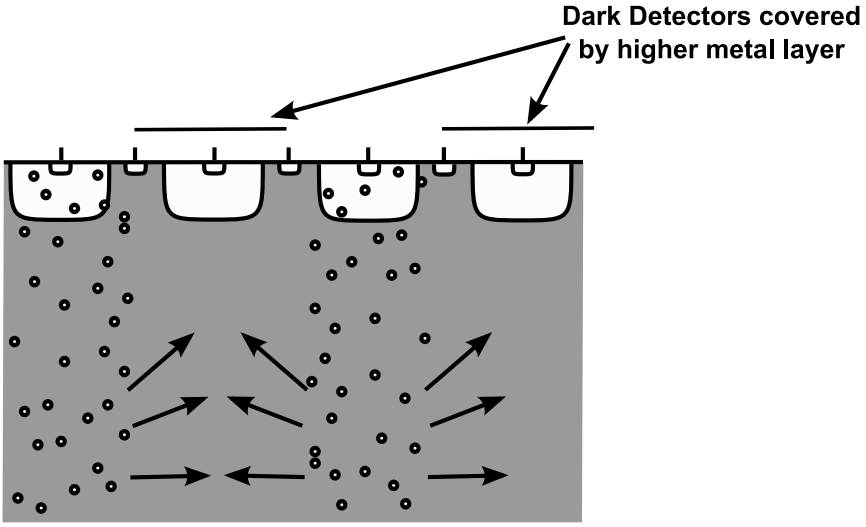

FIGURE 5.6. Differential photodetector. (After Hermans and Steyaert [47].)

Figure 5.7 shows the almost flat frequency response obtained by subtracting the dark detector response from the illuminated detector response. Today, the highest published data rate achieved with this cancellation technique remains below 1 Gb/s [48], while this technique also affects the detector sensitivity adversely.

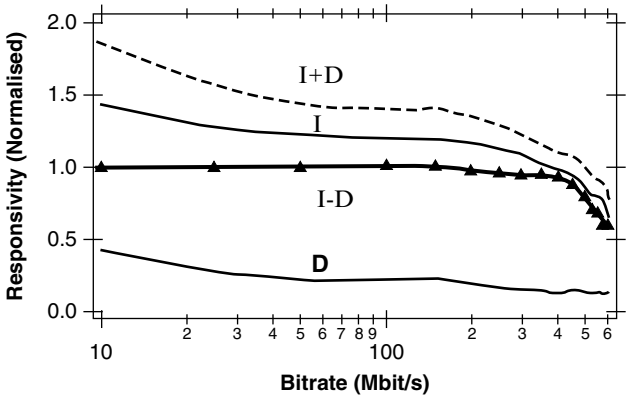

FIGURE 5.7. Frequency response of illuminated (I) and dark (D) detectors as well as their sum (I+Q) and difference (I–Q) [49]

5.5 Equalized PIN Detector

Instead of canceling the slow carriers, the overall frequency response of the detector can be compensated through equalization [50]. While this technique is more of a circuit technique than a detector technique, it deserves to be discussed in this chapter because of its direct impact on the detector performance. The block diagram of the concept is shown in Figure 5.8. Following the current-to-voltage conversion of the photocurrent, the received signal goes through an equalizer, compensating for the limited bandwidth of the detector.

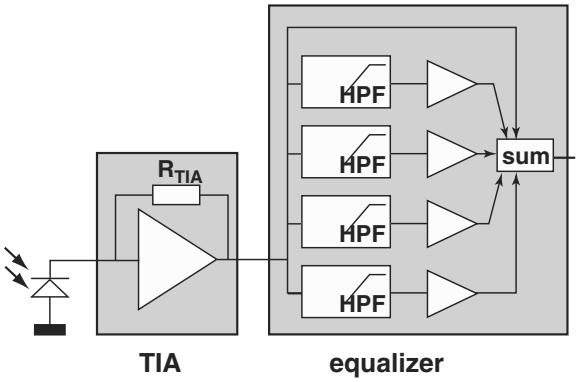

FIGURE 5.8. Equalized photodetector block diagram [50]

The analog equalizer presented in the referenced publication is based on programmable source degeneration of a common-source gain stage located between the transimpedance and the limiting

amplifiers. The simulated time response of this topology, considering process variations, is shown in Figure 5.9.

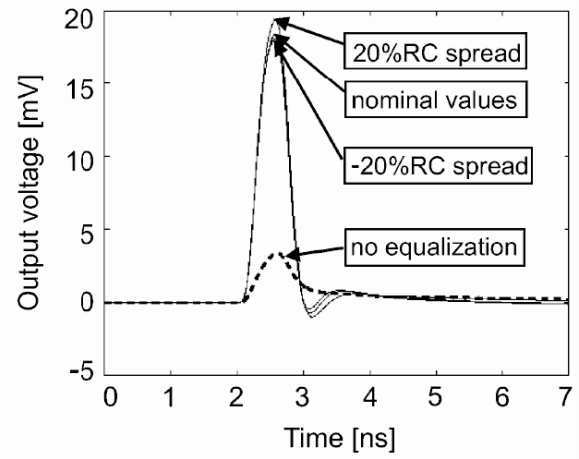

FIGURE 5.9. Simulated time response of the optical detector in the nominal case, the worst case process spread and without equalization [50]

As the frequency compensation is performed in the receiver circuit, no direct measurement of the detector responsivity can be made to compare with the previous results. However, the eye diagrams measured at the receiver output presented in the referenced publication with the equalizer respectively disabled and enabled easily highlight the performance of this technique (Figure 5.10). Indeed, eye opening and BER are considerably better when operating the receiver with equalization at 3 Gb/s compared to operation at 50 Mb/s (i.e. 60 × slower) without equalization.

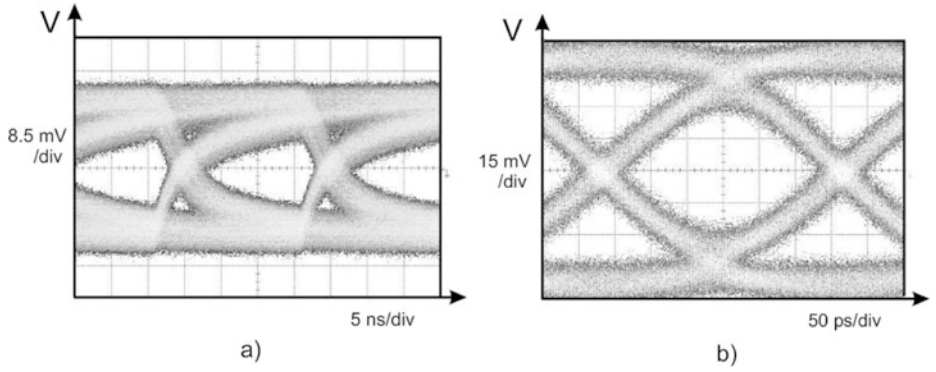

FIGURE 5.10. Measured eye diagram of the CMOS detectors at –19 dBm optical input power: (a) equalizer disabled at 50 Mb/s BER = 10^{-7}, (b) with equalizer enabled at 3 Gb/s BER = 10^{-11} [50]

5.6 Resonant Cavity-Enhanced Detector

Another solution to improve the intrinsic bandwidth-efficiency product of silicon detectors has been presented under the form of resonant cavity-enhanced (RCE) detectors, developed by Professor Ünlü's group at the Boston University Photonics Center [51].

5.6.1 The Resonant Cavity Structure

Including the absorption region of a PIN diode inside an ideal optical resonant cavity will keep the light circulating inside this cavity until it is fully absorbed. Such a Fabry–Pérot resonant cavity is delimited by two distributed Bragg reflectors (DBR, Figure 5.11) which, based on the principle of constructive and destructive interference, represent a reflecting surface at given wavelengths. While the thickness of the different layers of the DBR defines their reflectivity at a given wavelength, the resonant frequency can be tuned by accurate selection of the cavity length.

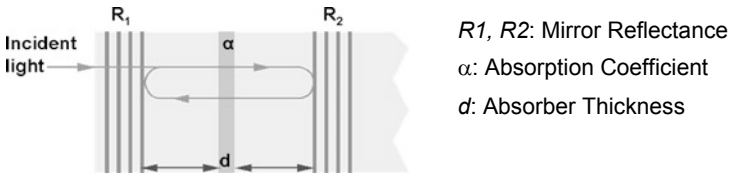

FIGURE 5.11. Fabry–Pérot resonant cavity structure [51]

The wavelength-dependency of the cavity's resonance peaks and the resulting increase in quantum efficiency are not a real problem in communication systems, because the laser sources' spectral distribution is narrow in comparison. Thermal drift of the laser wavelength however, must be limited to avoid excessive variations of the quantum efficiency. The simulation results in Figure 5.12 illustrate how the concept improves the detector responsivity without affecting the bandwidth. A significant improvement can be expected and operation at data rates of 10 Gb/s is achievable using the resonant cavity detector design discussed in the following.

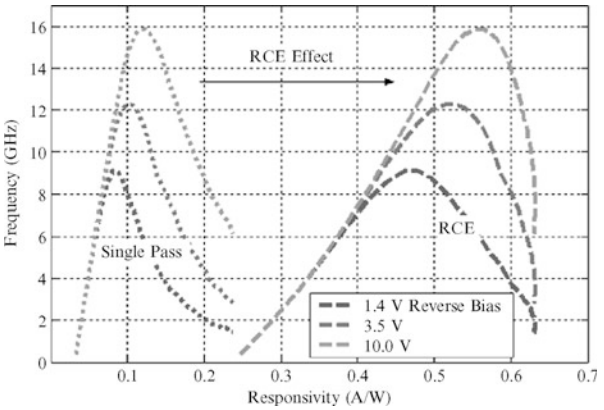

FIGURE 5.12. Simulated efficiency-bandwidth improvement of the RCE detector [51]

5.6.2 Distributed Bragg Reflectors

The performance of the resonant cavity depends on the reflectivity of the Bragg reflectors. These structures are made of alternating layers of quarter wavelength thick dielectrics with different refractive indexes. While high reflectivity can be obtained with more than ten periods of AlAs/GaAs, two periods of Si/SiO_2 are sufficient to obtain a reflectivity exceeding 90%, thanks to the higher refractive index contrast (60%) of these two materials (Figure 5.13).

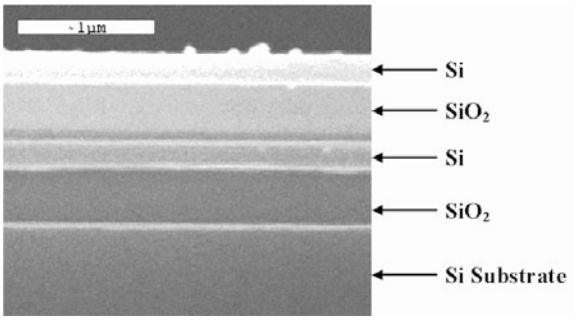

FIGURE 5.13. SEM of the buried distributed Bragg reflector [51]

The high-refractive index contrast also allows for the use of three-quarter wavelength thick layers instead of one-quarter wavelength thickness without excessive penalty in terms of optical selectivity

(i.e. stop-band width, Figure 5.14). This is in immense advantage because manufacturing constraints may not allow for the fabrication of layers as thin as one quarter wavelength.

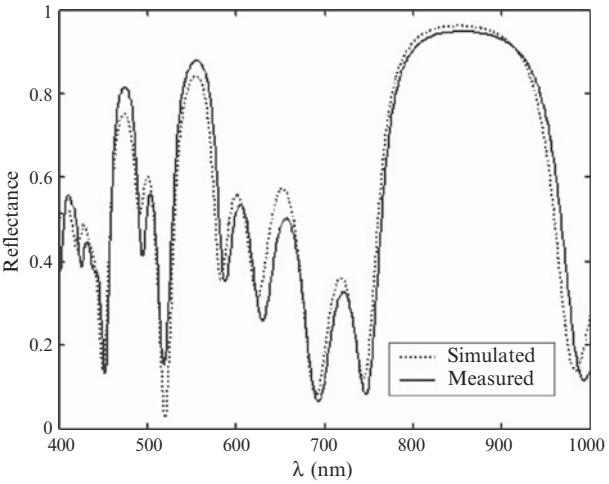

FIGURE 5.14. Reflectivity of the manufactured Bragg reflector (after epitaxial growth): Si thickness 174 nm, SiO_2 thickness 437 nm [51]

The commercial success of silicon-on-insulator (SOI) technologies, especially in the microprocessor market, has been used by the authors of this work to define a vertical detector structure using such Si/SiO_2 stacks in the substrate. The manufacturing process of their so-called double-SOI wafers, based on a conventional ion-cut process used to manufacture commercial SOI wafers [54], results in a monocristalline surface. This lattice allows for epitaxial growth and integration of standard semiconductor devices, while most previous work either resulted in a polycristalline surface or needed an exotic process technology.

5.6.3 Detector Fabrication

The vertical resonant cavity is delimited by the double-SOI-distributed Bragg reflector at the lower end and the high index contrast Si/air interface on the other, as shown in Figure 5.15. The detector thickness determines the spectral responsivity peaks of the resonant cavity and must be specified in

agreement with the wavelength of the laser used at the transmit side (Figure 5.16). It can be accurately controlled during epitaxial growth and fine-tuned by surface recessing.

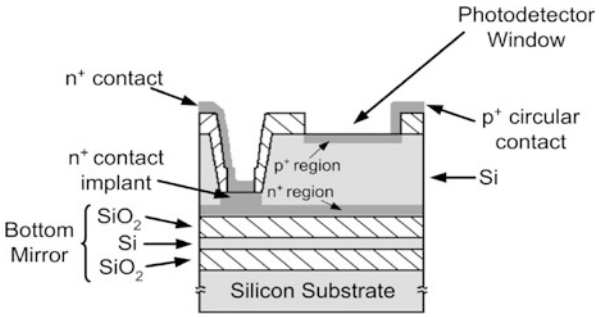

FIGURE 5.15. Illustrative vertical cut of the RCE photodetector [51]

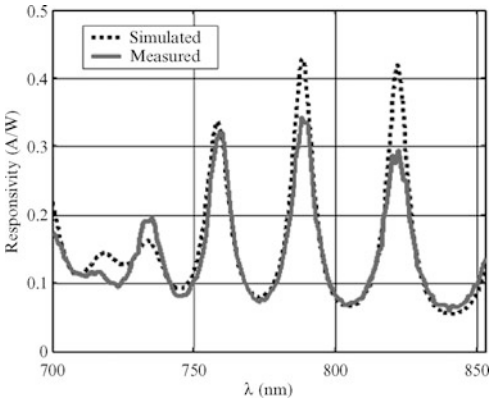

FIGURE 5.16. Measured spectral sensitivity of the manufactured detector [51]

Figure 5.17 illustrates the achieved bandwidth as a function of reverse bias. For bias voltages around 3 V, which can be generated on-chip in modern deep submicron processes with available I/O voltages of 3.3 V, bandwidth in excess of 7 GHz can be achieved, allowing for operation at data rates of 10 Gb/s.

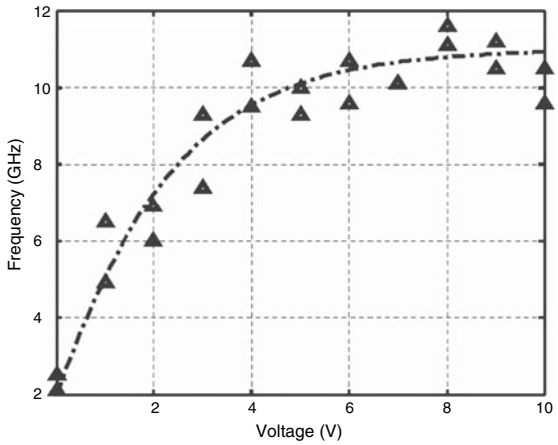

FIGURE 5.17. 3dB Bandwidth as function of bias voltage for 22 μm-diameter photodiodes [51]

A scanning electron microscope (SEM) picture of a 12×1 array of detectors is shown in Figure 5.18. In this design, the detector pitch of 250 mm was chosen to match the standard pitch for optical fiber ribbons.

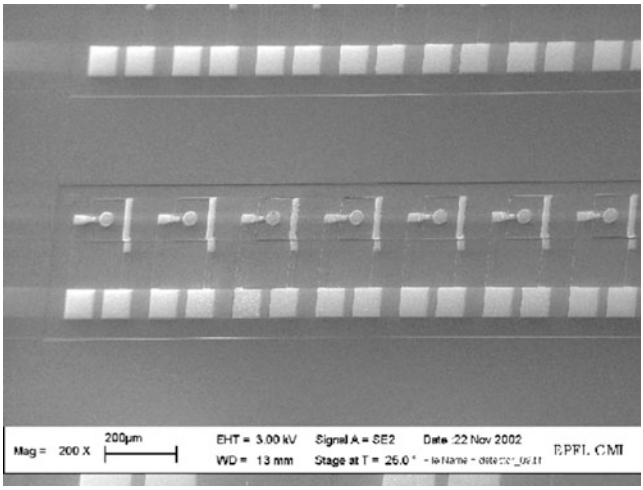

FIGURE 5.18. A 12 ×1 detector array (detector diameter is 40 μm and pitch is 250 μm) with bonding pads to be wire bonded to a corresponding receiver array [51]

5.7 Metal–Semiconductor–Metal Photodetectors

The metal–semiconductor–metal (MSM) detector is a planar detector compatible with current CMOS photolithography (Figure 5.19). The two metal contacts on silicon form back-to-back Schottky diodes. Modern process technologies allow for the fabrication of extremely narrow structures, minimizing the dark area of the detector due to contact regions.

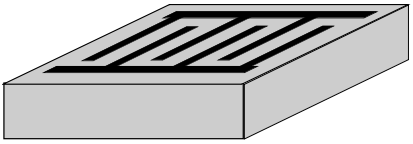

FIGURE 5.19. MSM photodetector

Due to its low intrinsic capacitance related to the finger dimensions, an MSM detector can potentially achieve much higher data rates compared to PIN detectors, as shown by GaAs detectors operated at frequencies of 60 GHz [52]. Increasing the electric field or reducing the finger spacing to further reduce the transit time results in a shallow high field region. As silicon has much larger penetration depth compared to GaAs, a considerable amount of carriers are generated deeper in the substrate where the electric field remains low, leading to slow tails in the detector's impulse response. For this reason, silicon MSM detector have not (or not yet) achieved a commercial breakthrough.

5.8 Lateral Photodetectors

For sake of completion, we would also like to mention the lateral detector shown in Figure 5.20, operating at 1,300 nm wavelength [53]. It is based on the implantation of silicon–germanium (SiGe) absorption islands at the end of a horizontal waveguide. In opposition to conventional detectors, the absorption is proportional to the length of these islands, thus to the capacitance of the detector, while the transit time is related to the thickness. However, the materials used in the referenced work are incompatible with operation at 850 nm wavelength.

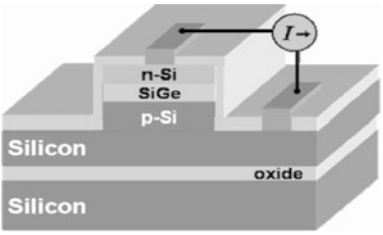

FIGURE 5.20. Horizontal SiGe waveguide-based photodetector [53]

5.9 Photodetector Characterization

High-speed detector characterization is far from being a trivial operation. Direct determination of frequency-domain characteristics may reveal quite difficult. While some of the previously presented results have been obtained with the detector connected to a transimpedance amplifier (and possibly to a limiting amplifier), pure detector characterization has to rely on time-domain pulse response measurements, as illustrated by these results obtained for the resonant cavity-enhanced detector.

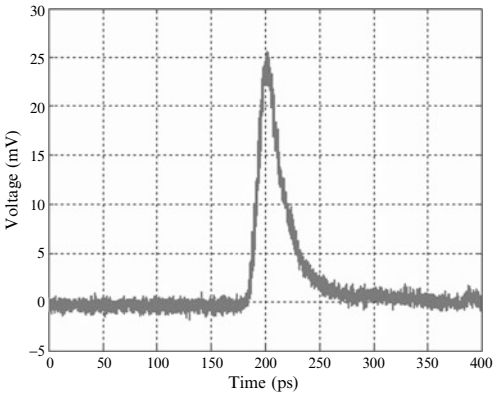

FIGURE 5.21. Measured time-domain response of a 22 μm RCE detector at 3V bias [51]

For this purpose, a Ti:Sapphire Picosecond Laser is used to generate a very short pulse with a full width at half maximum (FWHM) close to 1 ps and a repetition rate of several ten megahertz. The impulse response of the detector, probed on a microwave wafer probing station, is measured using a large bandwidth sampling oscilloscope. Figure 5.21 shows the obtained temporal response obtained at 3 V reverse bias from a 22 μm diameter RCE photodetector, achieving a FWHM value of 24.8 ps. The frequency-domain response can then be obtained through fast-Fourier transform calculation (FFT, Figure 5.22). For some detector types, time-domain measurements may exhibit a slow tail in the impulse response, due to slow carrier motion in some part of the detector. In the presented resonant cavity detector for example, the slow tail is due to carriers generated in the contact regions, where the carrier velocity is determined by diffusion time constants until they enter the depletion region.

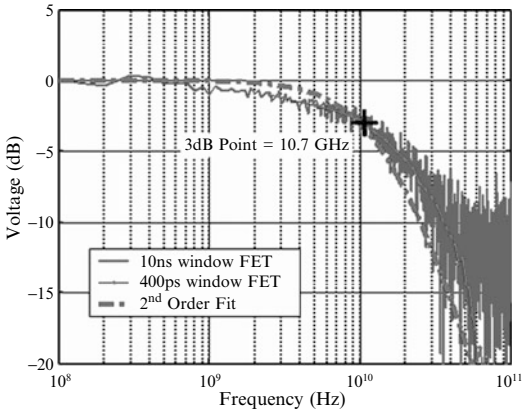

FIGURE 5.22. Calculated FFT transfer function of the same device at 4 V bias [51]

CHAPTER 6

Transimpedance Amplifier Design

The photocurrent generated by the photodetector through optoelectronic conversion of the incoming light must be conditioned to comply with the requirements of the clock and data recovery block. The CDR performs retiming and synchronization of the received data stream, but requires the input signal amplitude to be constant and larger than its input sensitivity. For this purpose, the photocurrent is converted to the voltage domain in the transimpedance amplifier (TIA). This current-voltage (I-V) conversion intrinsically provides signal amplification by the gain Z_{TIA}, commonly called *transimpedance gain*. Additional gain is then implemented in the limiting amplifier (LA) in the next step of the conditioning process.

6.1 Principles of I-V Conversion

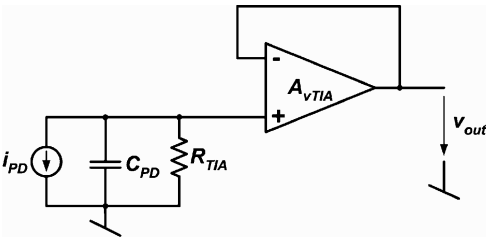

FIGURE 6.1. Basic I-V converter with voltage follower

The most basic current-to-voltage converter sends the photocurrent through a series resistor R_{TIA} and uses an amplifier with voltage gain A_{vTIA} in unity-gain configuration to buffer the signal (Figure 6.1). The I-V gain is given by the resistor value R_{TIA}, while the bandwidth, determined by the passive front-end, is given in Equation 6.1. In this expression, the photodetector capacitance C_{PD} is

P. MULLER, Y. LEBLEBICI,
CMOS Multichannel Single-Chip Receivers for Multi-Gigabit Optical Data Communications,
DOI 10.1007/978-1-4020-5912-4_6, © 2007 Springer

supposed to be the dominant term of the total input capacitance, compared to the amplifier input capacitance and other stray capacitances connected to this node.

$$BW = \frac{1}{2\pi R_{TIA} C_{PD}} \quad \text{(EQ 6.1)}$$

The trade-off between the I-V gain contributed by the resistor and the bandwidth represents the major limitation of this structure. Design for high gain in the passive stage leads to a large voltage signal at the amplifier input, minimizing the influence of the amplifier noise, but reducing the signal bandwidth. Low gain on the other hand reduces the signal-to-noise ratio (SNR) at the amplifier input, but increases the signal bandwidth considerably. Depending on the specifications for high gain or high bandwidth, this topology is also dubbed high-impedance amplifier or low-impedance amplifier, discussed in further details in [46].

Because of this impractical trade-off between noise and bandwidth, the feedback resistor topology has been developed (Figure 6.2), usually called *transimpedance amplifier* for its shunt feedback, which stabilizes the amplifier's transimpedance [56].

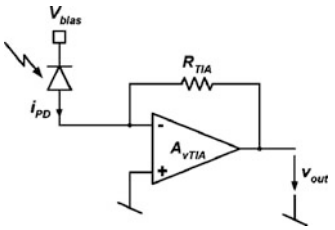

FIGURE 6.2. Simplified transimpedance amplifier topology

As the transfer function of this topology has already been presented in Section 4.5, only the important results for the block-level design will be recalled here. Figure 6.3 presents the overall single-ended amplifier configuration for frequency-domain analysis, where i_{PD} is the photocurrent generator from optoelectronic conversion of the incoming light and C_f a shunt feedback capacitor.

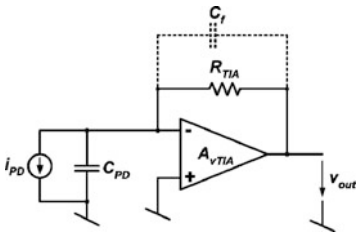

FIGURE 6.3. Transimpedance amplifier configuration

In presence of closed-loop gain peaking (Figure 6.4), C_f introduces a zero in the feedback function which saturates the feedback factor at higher frequencies. The phase shift at the intercept frequency is then below 180°, resulting in damped gain peaking in the closed-loop transfer function. Equation 6.2 presents the resulting I-V transfer as a function of the amplifier's frequency-dependent open-loop gain $A_{vTIA}(s)$ and the feedback factor $\beta(s) = \frac{1}{1 + sR_{TIA}C_{PD}}$.

$$v_{out} = \frac{i_{PD} \cdot R_{TIA}}{1 + \frac{1}{A_{vTIA}(s)\beta(s)}} \quad \text{(EQ 6.2)}$$

Assuming that the nondominant amplifier pole is at frequencies comparable to the amplifier gain-bandwidth product GBW_{TIA} and that the shunt capacitor is small ($C_f << C_{PD}$), the closed-loop bandwidth of the transimpedance amplifier is given by:

$$BW_{TIA} = \sqrt{\frac{GBW_{TIA}}{2\pi \cdot R_{TIA} \cdot C_{PD}}} \quad \text{(EQ 6.3)}$$

In fact, this value corresponds to the intercept of the amplifier open-loop transfer function and the feedback factor, illustrated in Figure 6.4. The effective bandwidth is about 50% larger as a benefit of peaking. As the peaking effect is mitigated by various issues in a real transimpedance amplifier design, this increase in bandwidth will not be taken into account in the following calculations. The gain peaking may among others be damped by the parasitic shunt capacitance in parallel with R_{TIA}, as well as the lower open-loop DC gain in very large bandwidth amplifiers. Indeed, in situations where the low-frequency gain of the TIA is small, the $1/\beta$ intercept may appear at frequencies below the –20 dB/decade region, in which case the optimum bandwidth-transimpedance gain trade-off may not be achieved.

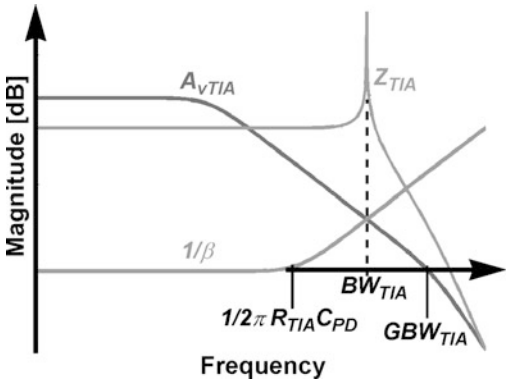

FIGURE 6.4. TIA gain-bandwidth graph without feedback capacitor. (After Graeme [40].)

Based on this equation for closed-loop bandwidth, the maximum feedback resistor value can be determined once the maximum achievable gain-bandwidth is known for a given amplifier structure and manufacturing technology.

6.2 Transimpedance Amplifier Topologies

For monolithic integration with digital processors, multichannel receivers must be particularly immune to power supply noise, ground bounce and substrate coupling, both from neighboring channels and from digital CMOS circuitry. Although single-ended structures (Figure 6.5) present lower noise levels and interesting power consumption per gigabit per second ratios, the use of a fully differential TIA design is preferred for optimal rejection of the above-mentioned noise sources.

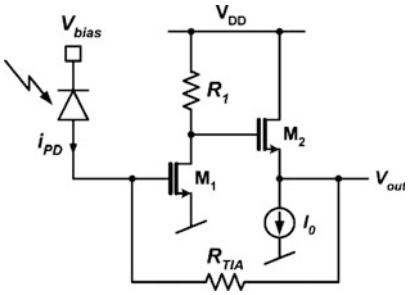

FIGURE 6.5. Basic single-ended transimpedance amplifier

Thanks to the improved device speed of deep submicron CMOS transistors, a number of fully differential high-speed CMOS transimpedance amplifiers has been published. However, with exception of the 4-channel single-ended receiver presented by Kromer et al. [57], the published circuits implement a single amplification channel only. Most TIA topologies can be classified in one of the two following categories: common-gate-input TIAs and voltage amplifier-based TIAs, separately discussed below.

6.2.1 Common-Gate Input Stage Amplifiers

In the previously discussed amplifier equations, we noticed that the TIA closed-loop bandwidth is limited by the dominant pole determined by the feedback resistor and the detector capacitance. The basic idea of common-gate input structures, and hence of regulated cascode structures which extend this concept, is the decoupling of the photodiode capacitance from the main amplifier and its feedback resistor. In advanced technologies, conventional common-gate (CG) input topologies (Figure 6.6a) only provide moderate isolation between its input and output nodes due to the limited transconductance and the high-output conductance of short-channel devices. While the low-current gain of the common-gate

stage affects the overall transimpedance gain adversely, the pole associated with the detector capacitance together with the TIA input impedance becomes nondominant. Furthermore, this pole does not affect the TIA stability as it does not appear in the feedback loop. At very high-data rates, these benefits largely balance the gain reduction appearing in the input stage.

The regulated cascode topology provides lower output conductance and thus improves the isolation of the TIA from the diode capacitance ([58], Figure 6.6b). The interested reader will notice similarities with the gain-boosting principle published by Bult and Geelen [59], both based on prior work by Hostika back in 1979 [60].

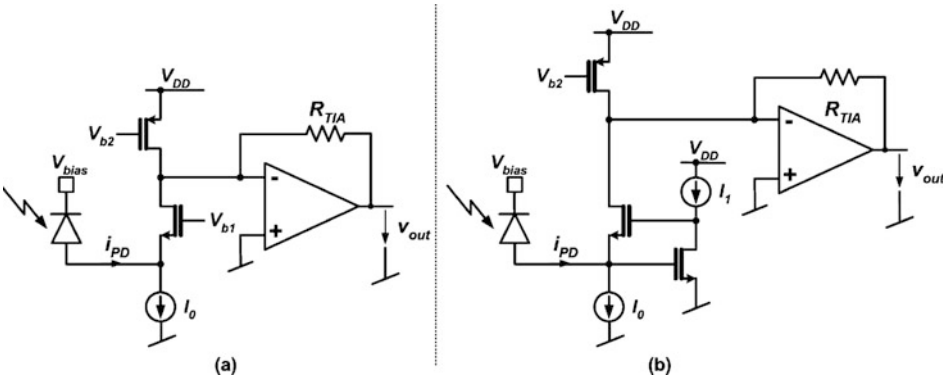

FIGURE 6.6. (a) Common-gate topology and (b) regulated cascode topology

Although common-gate amplifiers supposedly underperform regulated cascode structures, Kossel et al. [61] show a best-in-class 19 GHz bandwidth 45 dBΩ transimpedance amplifier consuming 6.5 mW. The tolerable input capacitance is limited to 400 fF, which would suffice for use with an integrated onchip photodetector. For further reading, it is also worth to mention the mathematical analysis of single-ended TIAs, including small-signal and noise analysis of common-gate, common-drain, common-source and regulated cascode topologies [62].

6.2.2 Voltage Amplifier Based TIAs

Several variants of voltage amplifier-based TIA topologies have been published. While the Cherry–Hooper topology is frequently used in limiting amplifiers, it appears sometimes in transimpedance amplifiers too [63]. Discussion of the advantages and drawbacks of the Cherry–Hooper amplifier is proposed in Chapter 7. Most voltage amplifier based TIAs implement the transimpedance conversion over a single amplifier stage, which is followed by one or more open-loop amplifier stages (Figure 6.7a).

The latter stages serve as signal buffer, some providing additional voltage gain and/or bandwidth enhancement.

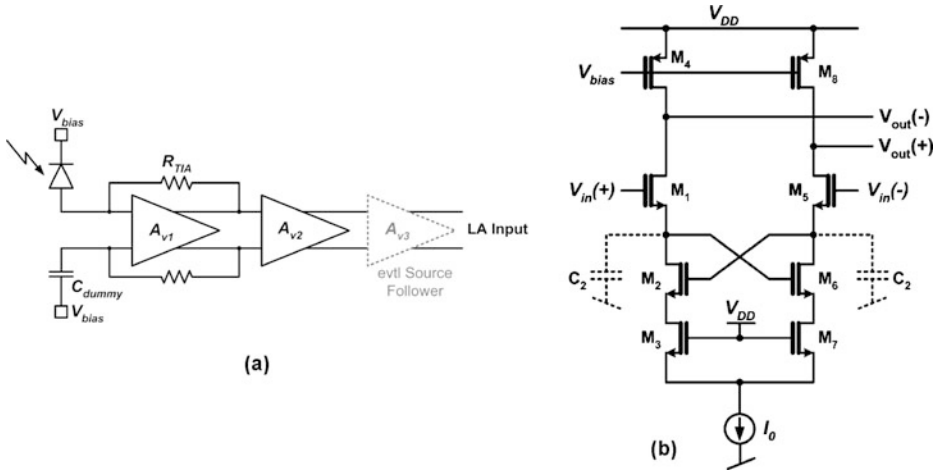

FIGURE 6.7. (a) Voltage amplifier-based TIA using one feedback stage, (b) negative resistively degenerated differential pair can be used as second stage for bandwidth enhancement

Tanabe et al. [64] for example propose a single-stage transimpedance amplifier followed by a bandwidth enhancement stage, which drives the source follower output driver. The bandwidth enhancement is obtained using a cross-coupled MOS-based negative resistive degeneration of the differential pair (Figure 6.7b). This stage contributes a zero and a pole located at frequencies higher than the zero to the overall transfer function. It also provides increased voltage gain compared to a conventional gain stage (Equation 6.4) at the expense of voltage headroom.

$$\frac{v_{out}}{v_{in}} = \frac{g_{m1}/g_{ds4}}{1-\dfrac{g_{ms1}}{g_{m2}}\cdot(1+g_{ms2}/g_{ds3})} \cdot \frac{1-s\cdot\dfrac{C_2}{g_{m2}}}{1+s\cdot\dfrac{C_2}{g_{ms1}\cdot(1+g_{ms2}/g_{ds3})-g_{m2}}} \qquad \text{(EQ 6.4)}$$

As shown in Equation 6.5, the zero location only depends on transconductance and gate capacitance. By proper design, it can be designed to compensate the output pole of the previous stage. The phase lag due to the right half plane zero is not a major issue, because this part of the amplifier is in open-loop configuration. However, its influence on the group delay might be worth a detailed analysis. The pole can be placed by careful selection of the different device sizes in the stage. The published example achieves a quite impressive bandwidth of 5.9 GHz bandwidth and a transimpedance gain of 59 dBΩ.

$$\omega_{z1} = -\frac{g_{m2}}{C_2} \qquad \text{(EQ 6.5)}$$

$$\omega_{p1} = \frac{g_{ms1} \cdot (1 + g_{ms2}/g_{ds3}) - g_{m2}}{C_2}$$

Addressing the problem from a different perspective, Analui and Hajimiri [65] propose a filter-based approach, where a cascade of passive and active elements is used to tune the frequency response as desired. This appealing technique allows for the design of a 59 dBΩ 9.2 GHz bandwidth TIA in a 0.18 μm CMOS technology. This publication is a proof of concept of the more general approach considering constant group delay bandwidth instead of magnitude bandwidth as the effectively important TIA specification. Indeed, a Bessel response with maximally flat group delay generates less ISI than a Butterworth response of equal magnitude bandwidth [42]. However, such an approach probably requires extensive tuning techniques to compensate for process variations in volume production.

6.2.3 Source Follower Output Stages

Before explaining the design choices operated in the TIA design, it is important to consider the performance of source follower structures in deep submicron technologies. The source follower topology, shown in Figure 6.8, is know for providing strong driving capabilities of capacitive and resistive loads at moderate input capacitance.

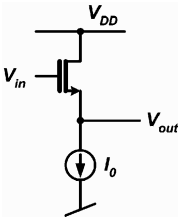

FIGURE 6.8. Source follower topology

High-speed operation requires the active device to be designed at minimum channel length. Equation 6.6 represents the voltage gain of such a configuration, which is smaller than one due to the combined action of body effect of NMOS transistors in the common substrate and the relatively large output conductance g_{ds} of the short-channel device. The influence of the output conductance could be reduced by increasing the active device's channel length. Maintaining gate and source transconductance values (respectively g_m and g_{ms} in a bulk-referenced transistor model) then requires an increase in device width, resulting in unacceptably large input capacitance. Furthermore, loss of headroom and the already mentioned signal attenuation associated with the source follower structure do not allow for the use of such a topology at the output of the TIA.

$$\frac{v_{out}}{v_{in}} = \frac{g_m}{g_{ms} + g_{ds}} \quad \text{(EQ 6.6)}$$

We should not forget during the design of the TIA that the amplifier output impedance in absence of a source follower output stage will somewhat affect the transimpedance gain. In fact, the amplifier being used in a current feedback configuration, the transimpedance gain becomes:

$$\frac{v_{out}}{i_{PD}} = \frac{1}{1 - \frac{1}{A_{vTIA}}} \cdot \left(R_F + \frac{R_2}{A_{vTIA}} \right) \quad \text{(EQ 6.7)}$$

In modern high-speed receivers, where R_F is in the kiloohm range and where the amplifier open-loop gain does not exceed 40 dB, this nonideality may be non-negligible.

6.2.4 Topology Discussion

While clock and data recovery circuits have less stringent requirements in that scope, today's gain-bandwidth demands cannot be met in the amplification front-end with a reasonable power consumption without relying on inductive peaking. Although inductors are believed to be area inefficient, peaking inductors do not require high quality factors and can thus be realized in an area-efficient fashion.

TABLE 6.1. Summary of TIA topologies

	Improvement technique	**Drawbacks/issues**
Common-gate input	Low-impedance current amplifier input stage	CG stage has low gain
Regulated cascode	Lower impedance current amplifier input stage	
Cherry–Hooper	Shunt feedback	Voltage headroom
Bandwidth-enhancement techniques		
Inductive peaking	Inductors resonate out capacitance	(Area)
Negative Miller capacitance	Cancels (partially) input capacitance	Capacitance matching
Cross-coupled degeneration	Additional pole-zero pair	Voltage headroom
Cascoded differential pair	Reduces Miller effect on differential pair	Voltage headroom
f_T - doubler	Two differential pairs	Power
Filter-based approach	Cascade of tuned filters	Inductor area

Among the topologies summarized in Table 6.1, the common-gate input amplifier is at first sight the most appealing one, because of the decoupling of the photodetector capacitance from the dominant closed-loop pole. This benefit is however limited by the short-channel devices in the first stage, which do not appear as an ideal current source. While the regulated cascode structure may improve this issue, it cannot completely eliminate the current loss in the common-gate stage, which has to be compensated for by additional transimpedance gain in the latter stages. Furthermore, the input capacitance of the

amplifier in absence of inductive peaking is not negligible in the deep submicron process in use, imposing an upper limit on the bandwidth improvements of this technique.

For illustration purposes, we will present the design procedure of a simple two-stage amplifier without source follower output stage (Figure 6.9), which could either be used for voltage-mode amplification or as a transimpedance amplifier following a common-gate input stage. We will consider the more complicated case of voltage-mode amplification, where the dominant pole is determined by the feedback resistor and the detector input capacitance. High closed-loop bandwidth requires the TIA to achieve a sufficient TIA gain-bandwidth product. While cascode structures enhance the DC gain by lowering the output conductance and hence lowering the dominant pole value, two-stage topologies increase the gain and the overall gain-bandwidth at the same time. As robustness and operation at low supply voltages over process variations and temperature is a major concern in SoC implementations, integration of cascode transistors in the first stage has not been considered. A differential pair in each stage uses only the faster NMOS devices without raising triode-region issues around the DC-operating point and further more allows for individual common-mode adjustment in each stage. The design procedure for this topology is presented in Paragraph 6.4.1.

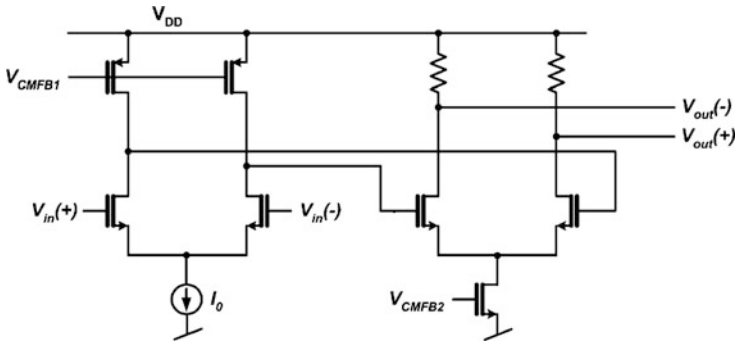

FIGURE 6.9. The implemented TIA core amplifier (feedback is omitted in this drawing)

6.3 Specifications

TABLE 6.2. Calculated block specifications

Parameter	Acronym	Minimum	Maximum	Unit
PD capacitance	C_{PD}		600	fF
TIA voltage gain	A_{vTIA}	30		dB
TIA transimpedance gain	R_{TIA}	6.5 k / 76		Ω / dBΩ
TIA bandwidth	BW_{TIA}	1.3	2.8	GHz
TIA input-referred integrated noise current	$S_{niTIA}{}^2$		615.0e–9	A/√Hz
LA target common-mode input voltage	V_{cminLA}	2/3 V_{DD}		V
LA input capacitance	C_{inLA}		200	fF

Before addressing the transistor-level design of the transimpedance amplifier, it is good to recall in Table 6.2 the specifications of this block as established in Chapter 4.

6.4 Transimpedance Amplifier Design

6.4.1 Description of the TIA Amplifier

For convenience, the two-stage amplifier topology is presented in Figure 6.10, including device naming conventions. It uses a maximum of three stacked devices to operate over all process and temperature corners and a supply voltage down to 1.62 V. It is based on two differential pairs (M_1–M_2 and M_5–M_6), which allow for the use of NMOS devices only in the signal path. The limiting amplifier input also being a differential-pair, the pole locations are dominantly determined by transistor widths and bias currents only, assuming one can neglect the wiring capacitance.

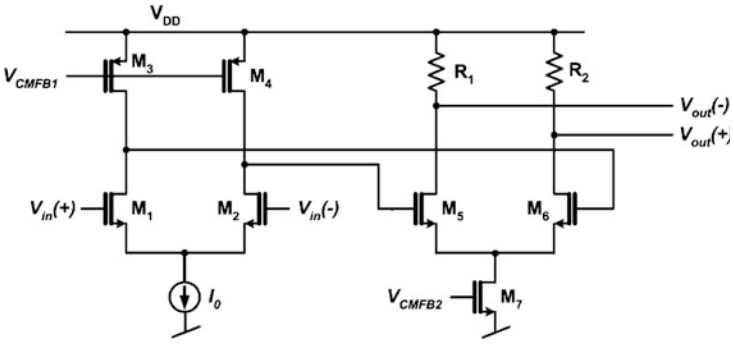

FIGURE 6.10. Transimpedance amplifier schematic

Amplifier stability constraints require the nondominant pole to be located at frequencies compared to the unity gain frequency, while the dominant pole should be located at much lower frequencies. For noise considerations, the dominant open-loop pole of the transimpedance amplifier should be located at the output of the first amplifier stage. When designing a large current-drive second stage with large differential pair devices (M_5 and M_6), this appears to be convenient, as these devices have a large input capacitance. The use of active loads M_3–M_4 instead of resistive loads in the first stage furthermore decreases the output conductance, increasing the voltage gain and bringing the dominant pole to lower frequencies. In this way, careful transistor sizing and biasing and joint design of the transimpedance and limiting amplifiers guarantees the amplifier stability without employing compensation capacitors.

This design requires common-mode feedback (CMFB) in both the first and second stages. In the first stage, the CMFB control voltage V_{CMFB1} adjusts the gate voltages of the active loads M_3–M_4, biasing the second differential pair around an optimum operating point. A separate CMFB loop adjusts the differential pair tail current in the second stage (M_7). In this way, the input common-mode of the limiting amplifier is controlled independently of process variations regarding bias currents and load resistor values. The common-mode feedback circuits use vertical metal–metal capacitors and standard polysilicon resistors with sufficiently high values not to affect the amplifier's DC gain (Figure 6.11). With capacitor values in both feedback circuits are 100 fF and resistor values of 27 kΩ, the influence of the CMFB components on the amplifier performance is not completely negligible, but remains minimal.

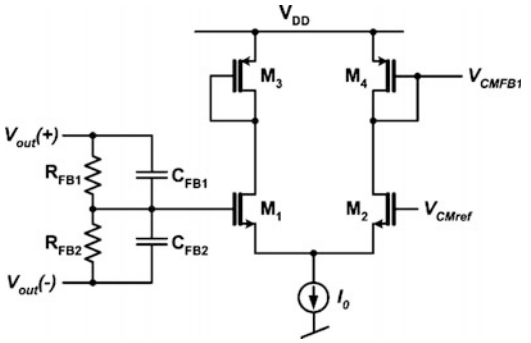

FIGURE 6.11. First stage common-mode feedback circuit

6.4.2 Design Procedure

In order to meet the transimpedance gain specification over process variations, the nominal feedback resistor value was chosen to be 2.095 kΩ. Indeed, the low-sheet resistance of available polysilicon resistors in this process leads to low-resistor width and consequently to weak process control. Variations

of 25% were considered and lead to this feedback resistor value, limited by the upper bound value of 2.7 kΩ. The reader should be aware that more advanced CMOS processes come with increasing sheet resistivity over technology generations, leading to the inverse situation, where small resistor values are hard to integrate with reasonable W/L ratios.

The device sizing was operated as following. The minimum closed-loop bandwidth of 1.8 GHz, corresponding to the maximum feedback resistor and a slightly augmented detector capacitance of 700 fF result in an amplifier unity gain frequency of:

$$GBW_{TIA} = BW_{TIA}^{2} \cdot 2\pi \cdot R_{TIA} \cdot C_{PD} = 38GHz \quad \text{(EQ 6.8)}$$

Of course this is based on the assumption of ultra-low capacitance electrostatic discharge (ESD) pad protections, as published, e.g. in [66]. The nondominant pole shall be located at frequencies at least four times larger than the closed-loop bandwidth:

$$f_{nd} = \frac{1}{2\pi \cdot R_2 \cdot C_{inLA}} = 4 \cdot BW_{TIA} \quad \text{(EQ 6.9)}$$

From this equation, we obtain the second-stage load R_2 = 100 Ω. With an output common-mode voltage target of 1.2 V (with respect to a 1.8 V supply), the tail current I_2 of the second stage becomes 12 mA. In the given technology, the current density at minimum channel length NMOS transistors for optimum transconductance is about I_D/W=50 μA/μm, resulting in a normalized transconductance of g_m/W = 340 μS/μm and an estimated normalized gate capacitance of C_g/W = 1.75 fF/μm. Based on these figures, the gain of the second stage is:

$$A_{v2} \cong g_{m2} \cdot R_2 = 0.122\text{S} \cdot 33\Omega = 4.0, \quad \text{(EQ 6.10)}$$

where g_{m2} is the transconductance of the second stage differential pair. The effective gain is slightly lower due to the differential pair output conductance neglected in the above calculation. To get an approximate value of the resulting second stage input capacitance, we neglect the Miller effect on the gate-drain overlap capacitance in presence of relatively low voltage gain in the second stage. Under this assumption, the estimated capacitance seen by the first stage is C_{g2}=210 fF. The transconductance of the first stage has then to be:

$$g_{m1} = \frac{GBW_{TIA}}{A_{v2}} \cdot 2\pi \cdot C_{g2} = 12.5\text{mS} \quad \text{(EQ 6.11)}$$

This transconductance can be designed for using an NMOS differential pair with a device width of 36 μm and a tail current of 3.6 mA. Again neglecting the Miller effect on the gate-drain overlap

capacitance, the amplifier input would contribute 63 fF to the total load on the input. Due to active loading, the gain of the first stage is difficult to estimate in absence of reliable short-channel models, but can be estimated from simulations to be in the order of 20. While the effective contribution to parasitic loading of the TIA input has not been calculated, it can be considered negligible compared to the detector capacitance. Table 6.3 gives an overview of the transistor-biasing conditions.

TABLE 6.3. Device dimensions and values

Device	Type	Width [μm]	Length [μm]	I_D [mA]	g_m [mA/V]	R [kΩ]
M_1/M_2	NMOS	36	0.18	1.8	12.2	–
M_5/M_6	NMOS	120	0.18	6	40.8	–
M_3/M_4	PMOS	72	0.18	1.8	–	–
R_0/R_1	RNPPO	3 (20//)	17	–	–	2.095

6.4.3 DC Compensation

The transimpedance amplifier block includes a DC current, dark current and offset compensation block. A single-ended integrator has been implemented to sink low-frequency current from the positive input node of the TIA (Figure 6.12) removing the DC components of the input current. The detailed DC components of the input current to be considered are:

- Photodetector *dark current*, due to the thermal generation of electron-hole pairs in the PIN structure. In the present detector, this component is negligible.
- Amplifier offset, due to device mismatch in the amplifier. This component has a Gaussian distribution with a zero mean value, i.e. the amplifier input offset voltage can have either positive or negative sign.
- The receive signal DC component. When receiving a logical "0", the detector generates little or almost no current, depending on the average received power. When receiving a logical "1", the detector generates a higher level of current. Data encoding guarantees a proper DC balance, i.e. a controlled density of binary data. The amplifier swing is maximized by removing the DC content of the signal before I-V conversion.

Complete amplifier offset compensation is guaranteed only if the equivalent amplifier input offset current remains smaller than the minimum signal DC input level. In the following, "DC compensation" and "offset compensation" are used as synonyms for the presented compensation mechanism. In some cases, depending on the particular amplifier topology, it may be desirable to use the DC component of the photocurrent to offset the TIA input common-mode voltage with respect to the regulated output common-mode. In this case, which does not apply to our design, the DC compensation should be applied to the opposite input. However, the obtained common-mode difference depends on the average photocurrent, thus on the average receive power, which may vary depending on temperature and

biasing conditions in the transmitter. This technique may particularly apply to common-gate input amplifiers.

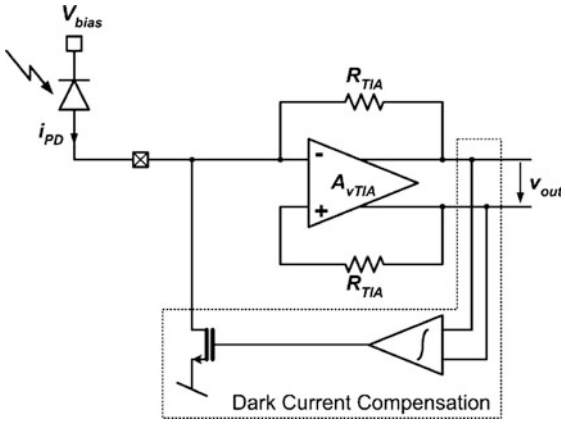

FIGURE 6.12. Dark current compensation structure

The integrating amplifier is implemented as a two-stage Miller-OTA (Figure 6.13). The Miller effect applied to the compensation capacitance C_c of 5.8 pF reduces the required capacitor area to obtain a low-frequency pole. Compared to a purely passive integration capacitor, the thermal noise of the second stage is however only filtered by the output capacitor C_{out} (260 fF) and contributes to the overall system noise at low frequencies. As shown in the simulation results in Paragraph 6.4.5, this noise contribution appears not to be excessive with respect to the required system performance. Finally, it should be mentioned that the lack of symmetry in the output stage adversely affects the power supply rejection ratio.

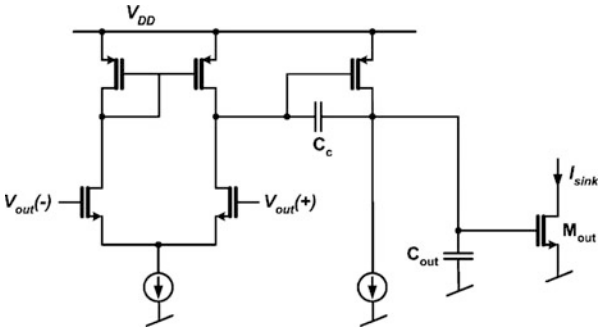

FIGURE 6.13. Miller-OTA as integrating amplifier with compensation and output capacitors and current sink M_{out}

6.4.4 Complete Block Diagram

The complete block diagram of the implemented transimpedance amplifier is shown in Figure 6.14. To achieve optimum symmetry, the fully differential design was propagated to the I/O pads, which contribute non-negligible parasitic capacitance and allow for connection of a dummy load equivalent to the detector capacitance. While access to both I/O pads for S-parameter testing remains of interest, a commercial implementation may benefit from lower bill of material by integrating the dummy capacitor on-chip.

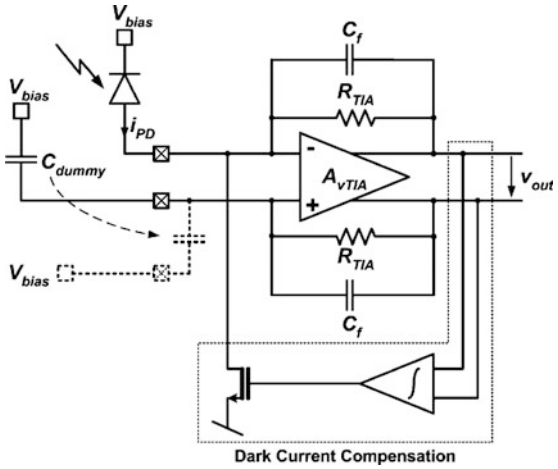

FIGURE 6.14. Complete TIA block diagram

It is crucial to understand that any capacitor mismatch between the two input branches does not only appear as a doubled additional capacitive load at the amplifier input, but also leads to common-mode to differential-mode conversion of high-frequency common-mode variations. This represents a major limitation of the fully differential TIA structure, because it is difficult to provide a good match between an on-chip metal–metal capacitor and an off-chip PIN detector capacitance, taking into account process, temperature and bias voltage variations.

With a DC gain close to 40 dB, the open-loop amplifier bandwidth happens to be hardly below the intersection with the feedback factor. As a result, the closed-loop phase shift does not reach 180° and no additional feedback capacitor is required to limit possible gain peaking. This result depends on both the photodetector capacitance and technology parameters like gate capacitance, transconductance and poly-silicon sheet resistivity and must thus be carefully simulated over the complete temperature range and all process corners. As the accuracy of transconductance and output conductance in the device model of the present digital CMOS process was not well known, dummy feedback capacitors (C_f in Figure 6.14)

should be placed in the layout to be connected through metal mask modification. In case the amplifier would show excessive gain peaking, a relatively inexpensive metal mask modification would connect a selectable amount of this capacitance in parallel with the feedback resistors in the following product integration.

Figure 6.15 shows the simulated open-loop amplifier gain, the feedback factor and the resulting closed-loop gain. The effect of the DC compensation is clearly visible in the closed-loop gain plot at lower frequencies. DC mismatch analysis shows a reduction of the amplifier offset from 58 mV without DC compensation to less than 1 mV with DC compensation enabled. In the present design, the intercept between the feedback factor and the amplifier open-loop gain curve is located around the 3 dB bandwidth of the amplifier, which results in an improved closed-loop stability at the expense of a little loss in the intercept frequency. On the other hand, we did not previously consider the bandwidth enhancement benefits of the slight gain peaking in transimpedance amplifiers, resulting in a closed-loop bandwidth of 1.85 GHz in the typical case. In most CMOS receivers, it is rather difficult to improve this figure once the design gets close to the limits of the technology, because any increase in gain-bandwidth in the first amplifier stage results in increased device dimensions and thus larger input capacitance, canceling the obtained transconductance improvement through reduction of the dominant pole.

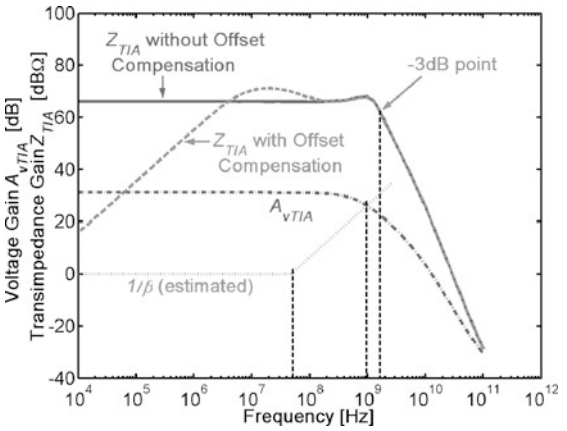

FIGURE 6.15. Simulated small-signal TIA transfer function

The transient simulation shows the correct current-to-voltage conversion at minimum and maximum input currents (Figure 6.16). Unlike in radio frequency (RF) applications, the amplifier linearity in the receiver input stage is not an important specification in fiber-optic receivers, due to the use of two-level signals. Even presence of large input signals, the output does not exhibit any slew-rate behavior

and is purely limited by settling-time constraints. This confirms the choice not to consider slew-rate issues in the amplifier design procedure.

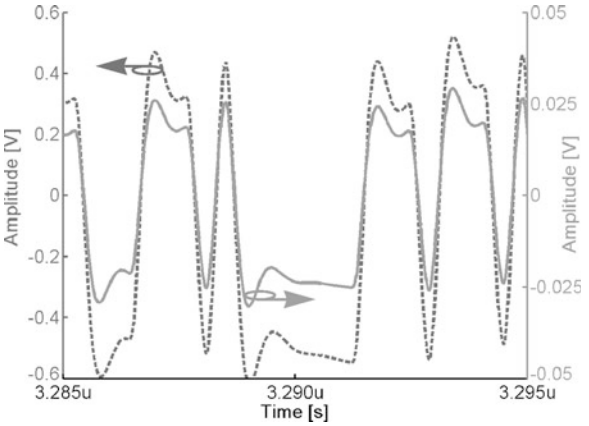

FIGURE 6.16. Transient simulation of the transimpedance amplifier with minimum (right axis) and maximum (left axis) output swing

6.4.5 Input Noise Estimation

As already discussed in the previous chapter, input referred noise is an important parameter in fiber-optic receiver front-ends. Although the flicker noise corner frequency in modern deep submicron processes may extend into the megahertz range, this contribution is usually negligible thanks to the use of offset compensation, though an example shows below that this may not be the case. Furthermore, it has been previously shown that, in the transimpedance amplifier designs analyzed in this work, the major noise contributions are located in the noise peaking frequency range. The output noise power spectral density simulated with the present TIA is shown in Figure 6.18.

Designers should have a critical look at noise simulations in deep submicron technologies. First, the quality of flicker noise parameters has to be assessed. Second, if critical devices are dimensioned at minimum channel length, increased thermal noise due to the presence of short-channel effects has to be expected, but is not always modeled consistently. Drain induced gate noise may or may not be considered negligible depending on the technology and the frequency of operation of the design.

Concerning the thermal noise excess factor γ, a moderate increase from the long-channel strong-inversion value of 2/3 is commonly accepted, although the physical reasons are still heavily discussed. Initial work pointed at the hot carrier generation and the velocity saturation effects in the pinched-off region of the channel close to the drain diffusion [67]. While hot carriers and velocity saturation effects in the pinched-off region have been eliminated as a possible cause by Chen and Deen [68], velocity

saturation inside the channel [69], the related carrier heating [70] and also channel-length modulation [68] are nowadays considered as possible sources of additional noise. Overall, a thermal noise excess factor value of 2 in short-channel designs in strong inversion regime appears consensual. Equation 6.12. gives the drain noise current power spectral density S_{ni}^2 as a function of the transistor transconductance, noise excess factor, Boltzmann constant k and absolute temperature T.

$$S_{ni}^2 = 4kTg_m\gamma \quad \text{(EQ 6.12)}$$

Although no noise measurements at the device level were available in the present technology for comparison, the simulated excess thermal noise factor for different values of the channel length L is shown below (Figure 6.17). It appears that the transistor model does take the increase in thermal noise at short channel length into account, particularly at higher current densities (i.e. in strong inversion regime). We also observe that in long-channel devices, the thermal noise factor tends to the well-known value of 2/3.

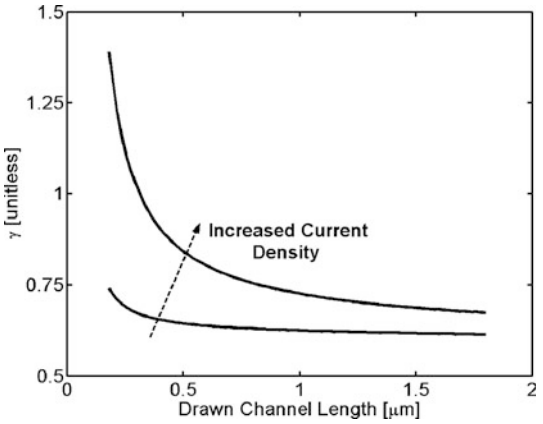

FIGURE 6.17. Excess thermal noise factor simulated as a function of channel length and current density

The output noise PSD plot of the transimpedance amplifier in Figure 6.18 shows the different noise regions already introduced in Chapter 4. It may be surprising that the flicker noise in Region 0 increases when the offset compensation is enabled. In fact, simulations with an ideal offset compensation amplifier show that this loop does indeed suppress low-frequency noise contributions. However, the implemented compensation amplifier has small bias currents (as long time constants are needed in this part) and small device sizes, resulting in considerable flicker noise. The second amplifier stage in particular does not see any integration effect from the dominant pole capacitor and directly adds to the input

signal. This additional low-frequency noise shows that the Miller-enhanced integration capacitor used in the compensation loop has to be used with caution. While the contribution of flicker noise to the total integrated noise remains small in presence of noise peaking, the additional noise at mid-range frequencies may contribute to up to 20% of additional integrated noise. The reader will understand that this may not be tolerable in applications requiring maximum receiver sensitivity, in which case the implementation of a conventional integrating amplifier is recommended.

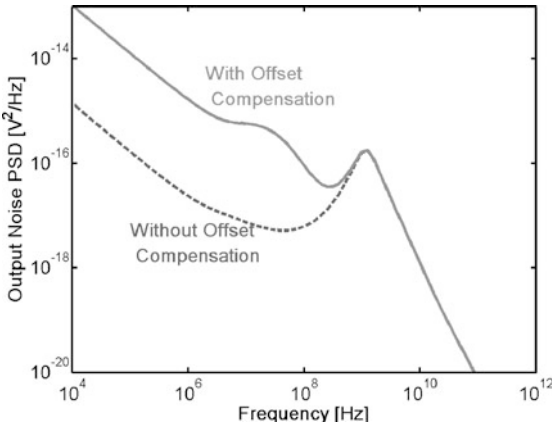

FIGURE 6.18. TIA output referred noise PSD simulation

The absence of a totally flat part of the noise peak (Region 3) is due to the relatively small $R_{TIA}C_{PD}$ time constant, as expected in such a multi-gigabit receiver, combined with the high flicker noise corner.

6.5 Block Layout

6.5.1 Channel Constraints

In order to allow for easy abutment, TIA, LA and output driver use the same row height with supply and ground routing on either side of the row, which should match the detector pitch of 250 μm, although chip core size may impose different form factors in some cases. A typical N-channel arrangement is shown in Figure 6.19. Though the respective silicon area depends on the topologies retained for each of the subblocks, the clock and data recovery part can be expected to occupy less than one third of the total area of a single channel in case a CDR topology with partial sharing is used. The gain-band-

width requirements, thus the power consumption and area indeed tend to be more stringent in the amplification front-end.

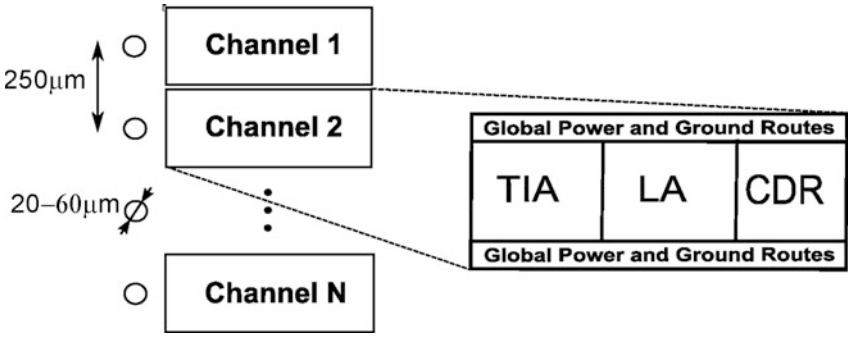

FIGURE 6.19. Typical floorplan of the optical receveiver array

Automatic dummy metal filling (also called "tiling") should be avoided on top of the matching-critical parts (i.e. differential pairs and active and passive loads), as increased mismatch due to asymmetrical metal filling in analog structures has been reported in the past [71].

6.6 Circuit Measurement

6.6.1 Frequency-Domain Measurements

The circuit performance of a transimpedance amplifier can be analyzed based on two different approaches. Time-domain measurements require the generation of a high-frequency-modulated current source, respectively the wire bonding of a photodetector to be stimulated with a modulated laser beam. Small-signal frequency-domain analysis on the other hand relies on the extraction of so-called S-parameters, corresponding to a two-port analysis of the design [72]. The transimpedance gain can be calculated from the measured S-parameters [73], where R_L is the 50 Ω load resistor present in the setup:

$$Z_{TIA} = R_L \cdot \frac{S_{21}}{1 - S_{11}} \qquad \text{(EQ 6.13)}$$

S-parameter analysis is frequently performed in a wafer probed setup, which guarantees optimum results provided the calibration of the setup is performed in an accurate manner. However, in case the dominant circuit pole of the retained transimpedance amplifier structure is defined by the detector capacitance, the physical absence of the detector in the setup would result in much wider signal bandwidth than the amplifier could offer in real-life configuration. This measurement therefore requires an integrated or bond-wire-connected detector.

6.6.2 Time-Domain Measurements

In order to mimic the signal current of the detector for time-domain measurements, one could consider adding an on-chip switch operating on a DC current source connected to I_{inDC} as shown in Figure 6.20a. However, beyond the transit frequency requirements for such a switch, the contribution of signal feedthrough and charge injection to the source current renders any control of the resulting signal current impossible. Instead, the pulsed current source shall be replaced by a Thévenin-equivalent voltage source supplied through a well-dimensioned passive network (Figure 6.20b). In that configuration, proper selection of C_c and R_{Th} is crucial to guarantee correct operation of this equivalent current source. A low value of R_{Th} for example may cancel the dominant input pole of the TIA, resulting in an optimistic bandwidth measurement.

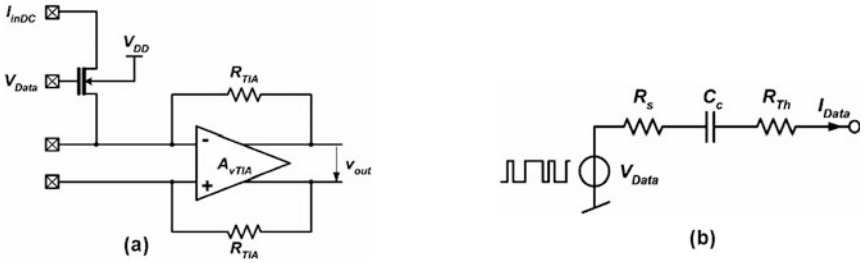

FIGURE 6.20. (a) Current switch for NRZ pattern generation and (b) Thévenin-equivalent passive network

CHAPTER 7

Limiting Amplifier Design

Following the photocurrent-to-voltage conversion performed by the transimpedance amplifier, the limiting amplifier provides additional voltage gain for the signal to satisfy the input sensitivity of the attached clock and data recovery circuit. The amplitude of the CDR input signal must not only exceed this value, the rise and fall times shall also allow for accurate detection of the zero crossings. Due to the signal amplification already operated in the TIA, noise may be less critical in the limiting amplifier, although this argument loses strength in advanced CMOS receivers operating at multi-gigabit data rates, where the achievable TIA gain tends to drop below the kiloohm barrier.

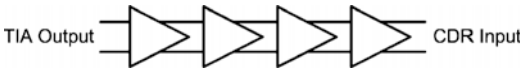

FIGURE 7.1. Typical cascade of gain stages used in limiting amplifiers

The limiting amplifier being implemented in an open-loop configuration, a cascade of broadband gain stages is perfectly suited for providing both high voltage gain and large bandwidth (Figure 7.1). After introduction of the signal-limiting concept and presentation of state-of-the-art topologies, two limiting amplifier structures are discussed to show the importance of inductive peaking. The interstage scaling amplifier design resulting of an optimization approach is compared to an inductive peaking amplifier, in which the potentiality of magnetic coupling is also analyzed.

7.1 Principles of Signal Limiting

The bandwidth of the transimpedance amplifier has been specified at the limit of ISI on-set for minimum integrated noise, defining at the same time the system bandwidth. In order for the overall bandwidth of the TIA–LA cascade not to be significantly affected by the LA bandwidth, the latter must

P. MULLER, Y. LEBLEBICI,
CMOS Multichannel Single-Chip Receivers for Multi-Gigabit Optical Data Communications,
DOI 10.1007/978-1-4020-5912-4_7, © 2007 Springer

be sufficiently larger than the transimpedance amplifier. A typical figure for the limiting amplifier bandwidth is 1.5 f_B. However, a larger signal bandwidth comes with a larger noise bandwidth, which has to be considered carefully to keep the LA noise down. This is implicitly taken into account in the system-level design approach in Chapter 4 by considering integrated LA noise.

The reader should be aware that small-signal bandwidth is a conservative measure of the limiting amplifier speed. Indeed, the latter stages in the cascade tend to operate in a large-signal regime, i.e. they experience complete switching, as illustrated in Figure 7.2. As explained in [19], the large-signal speed in a cascade of gain stages is limited by the speed of a single gate and not by the complete cascade. This fact is obviously exploited in combinatorial logic circuits, where the number of gates in the signal path determines the path delay, but not the path speed.

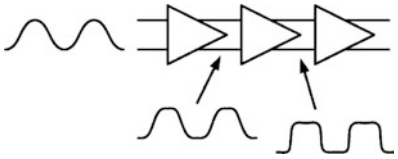

FIGURE 7.2. Large-signal operation in the LA gain stages

Differential amplifier stages are frequently considered to have high supply noise rejection. However, it is often neglected that differential circuits operating in large-signal regime are not supply and substrate noise insensitive. Indeed, they spend most of the time in a fully switched state, where one branch of the differential topology is off (i.e. conducts no current), while the other carries all the bias current. In this situation, differential analysis does not apply, as only the nodes in the active branch are subject to Ohm's and Kirchhoff's laws.

As a further issue, large-signal operation of an amplification stage corresponds to a nonlinear regime due to the saturation of the output levels. As such, the gain stages operating in large-signal regime may exhibit amplitude modulation to phase modulation (AM–PM) conversion. Without going into details, one can see that the harmonic components generated by the saturating differential pair experience different phase shift than the fundamental frequency at the dominant pole node, resulting in an amplitude-dependent signal phase. For a detailed discussion of the topic, the reader is again referred to [19]. This phenomenon is responsible of the conversion of additive amplitude noise into random jitter in the amplification path, which has been introduced in Paragraph 4.7.2.

Before addressing the various techniques and topologies developed to address the critical gain-bandwidth product requirement, the limitations and trade-offs of a simple cascade of gain stages will be presented.

7.2 Simple Limiting Amplifier Topologies

7.2.1 Cascade of Gain Stages

Let us consider again a cascade of M identical first order gain stages of gain A_{vi} and dominant pole ω_{pi}, shown in Figure 7.3. The transfer function of the cascade is given by the product of the transfer functions of the individual stages:

$$A_{vi}(s) = \left(\frac{A_{vDCi}}{1 + \frac{s}{\omega_{pi}}} \right)^M \qquad \text{(EQ 7.1)}$$

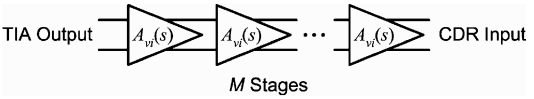

FIGURE 7.3. Cascade of *M* identical gain stages

In fact, identical bandwidth of all stages represents an intuitive optimum to the cascade problem. If one stage were slower than the others, it would determine the bandwidth of the complete cascade, while the other would burn additional power. By redistributing this additional power to the slow stage, the overall bandwidth can be improved, resulting in the presently discussed situation. We know that the bandwidth of such a cascade is given by Lee [72]:

$$BW = \frac{\omega_{pi}}{2\pi} \cdot \sqrt{\sqrt[M]{2} - 1} \qquad \text{[Hz]} \qquad \text{(EQ 7.2)}$$

Figure 7.4 illustrates that the bandwidth drops quickly when M becomes large, which is one reason why limiting amplifiers rarely employ more than 3–5 stages. The total gain of the cascade is given by:

$$Av_{DC} = \prod_{i=1}^{M} Av_{DCi} = Av_{DCi}{}^{M} \qquad \text{(EQ 7.3)}$$

It appears that the required gain could be distributed over a large number of stages as shown in Equation 7.3. However, when using low gain in the first stages to achieve large bandwidth, the designer must be aware that the noise of the later stages accumulates and results in low signal-to-noise ratio

(SNR). Some designs use higher order stages, which provide higher signal gain, for stages 1 and 2 to suppress the noise of the following amplification stages.

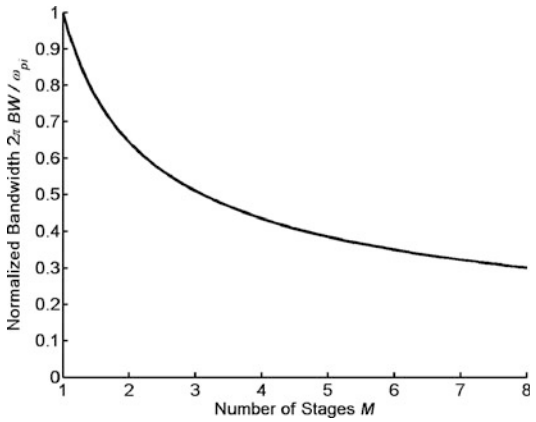

FIGURE 7.4. Bandwidth as a function of the number of cascaded stages *M*

7.2.2 Optimization Theory

Based on this simple structure, the optimum number of stages can be determined to achieve maximum bandwidth at a given gain A_{vDC}. Some mathematical development, presented in [72], leads to the following result: when considering identical gain stages, the optimum number of stages becomes $M_{opt} = 2 \ln Av_{DC}$, while the optimum gain per stage results in $Av_{DCi} = \sqrt{e}$.

While this approach leads to the optimum configuration for a cascade of amplifier stages, it does not in itself improve the gain-bandwidth trade-off imposed by a given technology. Furthermore, a cascade with such a low gain per stage suffers of the previously mentioned noise accumulation in the initial stages and is hardly practical for physical implementation. Application of this theory to a real-world design and a discussion of the trade-offs in terms of power consumption and noise accumulation have recently been published in [74].

7.2.3 Group Delay

Group delay defines the variation of the phase shift $\theta(\omega)$ experienced by the signal with respect to its angular frequency ω (Equation 7.4). In order for the data pattern not to be subject to intersymbol interference, all frequencies contained in its spectrum shall travel at approximately the same speed

through the cascade of amplifier. This is guaranteed if the group delay variation over the given frequency band is small.

$$GD = -\frac{\partial\theta(\omega)}{\partial\omega} \quad \text{(EQ 7.4)}$$

One could imagine designing the limiting amplifier considering group delay constraints instead of magnitude bandwidth. In such an approach, Bessel-type responses exhibiting maximally flat delay would outperform Butterworth-type responses preferred in the magnitude bandwidth approach [42]. We did not fathom this field and proof of the sufficiency of the group delay criterion as the only limiting amplifier bandwidth specification remains to be established. The interested reader can consider [75] and [76] for additional information.

7.2.4 Basic Gain Stage

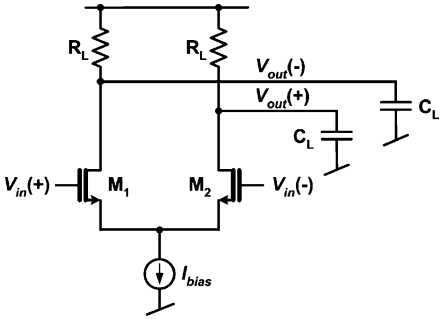

FIGURE 7.5. Resistively loaded differential pair

The simplest gain stage available in high-speed circuits is a resistively loaded differential pair (Figure 7.5), defined by its open-loop bandwidth (or dominant pole) $\omega_{pi} = (R_L C_L)^{-1}$ and the low-frequency gain $Av_{DCi} = g_{m12}R_L$, where g_{m12} is the transconductance of the differential pair transistors. In presence of non-negligible transistor output conductance g_{ds}, R_L shall be replaced by $R_L // g_{ds}^{-1}$ in both preceding expressions. The gain-bandwidth product (GBW) determines the maximum gain we can achieve for a given bandwidth:

$$GBW_{1\text{ stage}} = \frac{\omega_{pi}}{2\pi} \cdot Av_{DCi} = \frac{g_m}{2\pi C_L} \quad \text{(EQ 7.5)}$$

The GBW can be optimized by maximizing the amplifier transconductance and minimizing the load capacitance. As the input capacitance of each stage has to be driven by some preceding stage, the gain-bandwidth product in presence of identical gain stages is directly determined by the transconductance

to gate capacitance ratio. Design at minimum device channel length is thus required to fully exploit the high transit frequency of deep submicron transistors. Taking into account process and temperature variations, the available voltage swing V_{sw} on the resistor must remain lower than 0.5 V. Neglecting the output conductance of the differential pair, the maximum achievable gain with this topology is thus given by:

$$Av_{DCi}^{\max} = g_m R_L = \frac{g_m}{I_D} \cdot I_D R_L = \frac{g_m}{I_D} \cdot V_{sw}^{\max} \qquad \text{(EQ 7.6)}$$

With the available g_m/I_D ratios in advanced technologies (Figure 4.3b), the gain of a single stage hardly exceeds values between three and five. As the transistor's output conductance is not negligible with respect to the inverse of the load resistor, the effective DC gain is even somewhat smaller. While the architectural variations discussed in the following section somewhat extend the above-mentioned GBW limit, the calculated value gives a good information on the speed potential of a given technology. It is not surprising that it can be directly related to the technology benchmarks f_T and f_{max}.

7.3 Bandwidth Enhancement in Limiting Amplifiers

The performance of limiting amplifiers is characterized by the trade-off between bandwidth and open-loop gain. As the gain-bandwidth performance of such a design is related to the intrinsic device performance, several design solutions have been proposed to extend the achievable performance of the design. While most of these solutions are discussed in [19], an overview of the most appealing techniques is presented in the following.

7.3.1 Inductive Peaking

Inductive peaking is probably the most frequently used technique to enhance the operating frequency of broadband designs for a given technology. Basically, the inductor used in series with the load resistor resonates out the capacitive load represented by the following stage. While most conventional designs tend to connect the inductor in a T-coil configuration between the load resistors and the positive supply (Figure 7.6), alternative configurations may also be beneficial [75]. While nonidealities like parasitic capacitance and magnetically induced Eddy currents in the substrate affect the intrinsic resonant frequency of the inductor and the quality factor, they do not seriously affect the performance

of the circuit. Indeed, in absence of a resonant system, a high-quality factor is not required in gain peaking applications. Finally, the total inductor value is only constrained by the available silicon area.

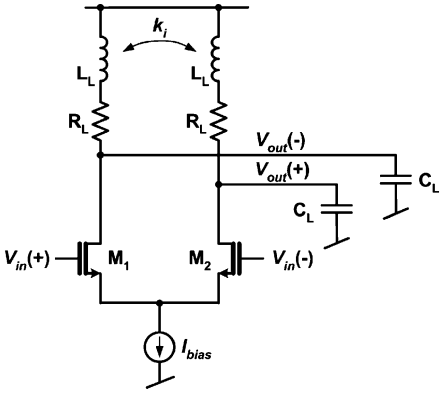

FIGURE 7.6. Inductively enhanced bandwidth stage

Unlike in oscillators, a perfect resonance between the inductor and the capacitance to be cancelled out is not required or even desired, as it would lead to peaking in the signal transfer function. Furthermore, the inductor series resistance can be used to partially replace the resistive load. As a result, a moderate quality factor Q is sufficient to improve the bandwidth of an amplifier stage. Potential magnetic coupling issues in multichannel receivers will be discussed later in this chapter.

7.3.2 Active Inductors

Active inductors replace the inductor by a MOS transistor in a particular configuration (Figure 7.7, [77]). The important gate-source voltage drop of the NMOS load can be circumvented by applying a boosted gate voltage to the transistor. This, however, requires generation and routing of an additional supply voltage that exercises excessive gate bias. Gate voltages exceeding the process' maximum ratings may lead to early breakdown of the gate oxide and failure of the circuit.

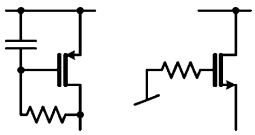

FIGURE 7.7. PMOS and NMOS in active inductor configuration

7.3.3 Capacitive Degeneration

Capacitive degeneration, as shown in Figure 7.8, improves the amplifier bandwidth in two ways. The amplifier's bandwidth is increased by a factor $(1+g_m R_S)/2$, while the DC gain is lowered by the same amount [19]. Second, the input capacitance seen by the preceding stage is also reduced by the same factor.

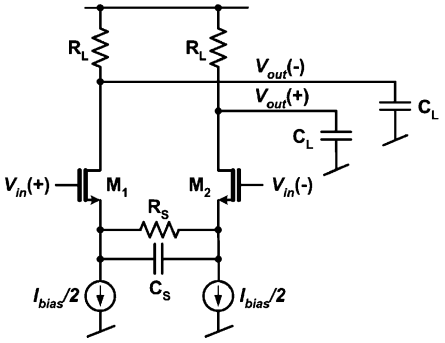

FIGURE 7.8. Capacitively degenerated amplifier stage

Due to the low voltage gain of such first-order stages, the benefits of the capacitive degeneration thus cannot always be exploited to their full extent. Dividing the DC gain by a factor of two would indeed lead to a number of LA stages with considerable noise accumulation in the initial stages.

A variant of this topology has been published in [78]. In addition to a split of C_S into two grounded capacitors, the authors use a cross-coupled transistor pair representing a negative impedance (Figure 7.9). While this topology eliminates the problematic gain reduction of the previous topology, it does not benefit of the reduction in input capacitance.

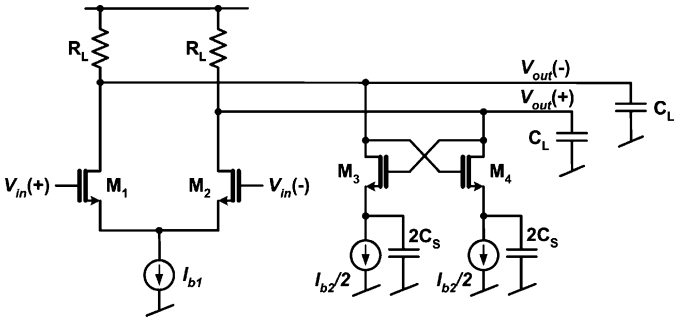

FIGURE 7.9. Alternative capacitively degenerated topology

7.3.4 Negative Miller Capacitance

The concept of negative Miller capacitance is illustrated in Figure 7.10. It uses capacitors C_M to reduce the capacitive load seen at the input [79, 80]. Intuitively speaking, as the capacitors are connected to the opposite output branch, they benefit of the 180° phase shift between the signals in both branches and add to the gate-drain overlap capacitance of transistors M_1 and M_2 with a negative sign. As a result, they compensate for the overall gate capacitance seen by the preceding stage.

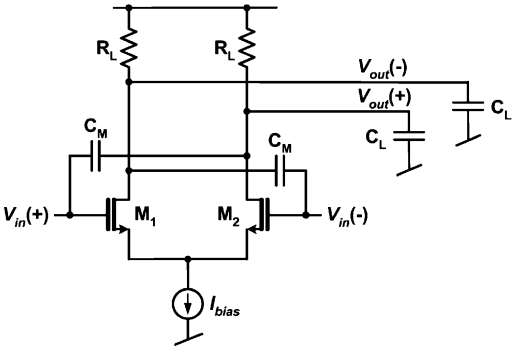

FIGURE 7.10. Gain stage with negative Miller capacitance

The efficiency of this configuration depends on the cutoff frequency of the capacitors, which is related to their series resistance. Especially when using gate-bulk capacitors, parameters like poly resistivity, poly contacts and bulk contacts are critical to guarantee that all the capacitor area is efficiently exploited at high frequencies of operation.

7.3.5 The Cherry–Hooper Amplifier

The so-called Cherry–Hooper topology, well known in bipolar circuits, uses local shunt feedback to improve the frequency response of a two-stage amplifier.

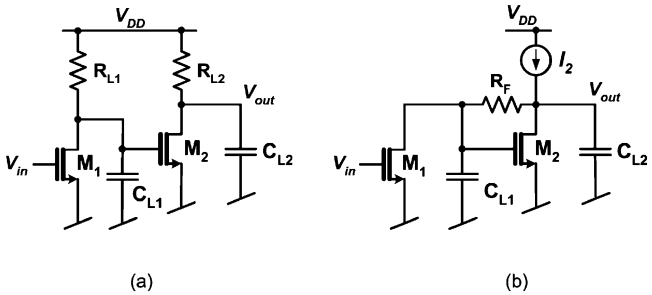

FIGURE 7.11. Illustration of the Cherry–Hooper topology: (a) amplifier without shunt feedback, (b) Cherry–Hooper amplifier with shunt feedback

Consider the single-ended two-stage amplifier in Figure 7.11a. Its frequency response is conditioned by the poles at the output of the first and second stages, determined by the load capacitance and the output conductance ($g_{ds} + 1/R_{L1}$) of the driving stage. High bandwidth can thus only be achieved at the cost of reduced gain per stage, which is of course proportional to the inverse of the output conductance. By introducing a feedback resistor between drain and gate of M_2 (Figure 7.11b), both poles tend to become dependent on the transconductance of M_2. Considering both poles to be at approximately identical frequencies (Equation 7.7) leads to a pessimistic estimation of the achieved bandwidth. In this equation, the term $g_{m2}R_F C_{GD2}$ represents the Miller capacitance of the second stage, which contributes to the loading of the first stage. For a detailed mathematical development of the transfer function, the reader is referred to [19].

$$\omega_{p1} = \omega_{p2} = \frac{2g_{m2}}{C_{L1} + C_{L1} + g_{m2}R_F C_{GD2}} \qquad \text{(EQ 7.7)}$$

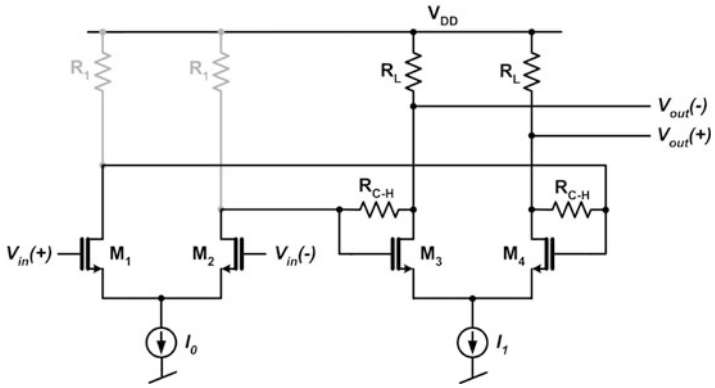

FIGURE 7.12. Differential Cherry–Hooper amplifier with resistive loads

The major drawback of the Cherry–Hooper topology, usually used in its differential form shown in Figure 7.12, is the voltage drop due to the bias currents flowing through the load and feedback resistors. Adding a current path through resistor R_1 improves this problem, at the expense of increased capacitance at the output of the first stage. Furthermore, the value of R_1 must be much larger than $1/g_{m2}$ not to alter the voltage gain, which again leads to non-negligible voltage drop in the resistor. PMOS active loads in place of R_1 would contribute prohibitive drain capacitance and are not recommended in high-speed designs.

Embedding source followers to drive the output load inside the topology provides additional benefits ([19], Figure 7.13a). Not only the frequency of the pole at the drain of M_3–M_4 increases due the reduced capacitive load, the feedback around M_5 and M_6 also contributes to lower output impedance of the source follower, which increases the frequency of the pole attached to the output node. A very similar topology has been proposed in [81], with the addition of resistor R_1. As shown through the mathematical developments in this reference, the additional resistor adds significant gain without deterioration of the bandwidth. However, the fact that the output is not picked at the outputs of the source followers, the major benefit of embedding these is lost. Indeed, additional source followers are required to cascade several stages using this topology. The reader should also be aware that the Cherry–Hooper amplifier is most beneficial when driven by small input signals, although interesting performances are guaranteed at any signal strength. As a result, some designers prefer to use it in the first stages of the limiting amplifier only.

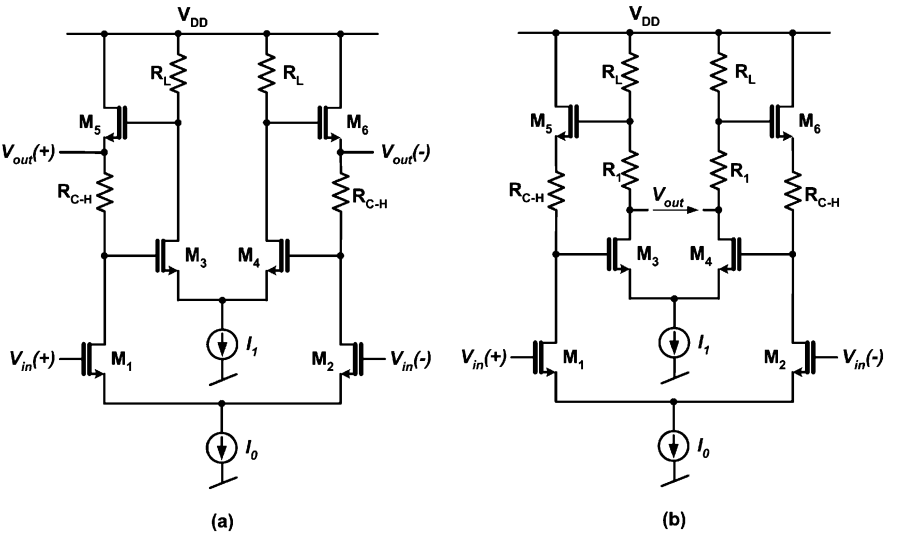

FIGURE 7.13. Two modified Cherry–Hooper topologies with embedded source followers

As a concluding remark, the voltage headroom requirements of the Cherry–Hooper structure render its implementation impossible at low supply voltages. However, the following paragraph introduces a derived topology with considerable potential in low-voltage designs.

7.3.6 Active Feedback Structure

An extension of the Cherry–Hooper concept has been presented in more recent years (in fact during the design phase of the presented circuits) by Galal and Razavi [80]. It uses a so-called *active-feedback architecture* applied to a five-stage limiting amplifier. The active-feedback topology, shown in Figure 7.14, can be compared to a Cherry–Hooper configuration with the advantage that it does not resistively load the second gain stage. Equation 7.8 (from Galal and Razavi [80]) shows the resulting gain-bandwidth product as a function of the transconductances of the first two stages and the capacitive loads C_{L1} and C_{L2} seen by either stage. Assuming the feedback is designed for a maximally flat response and the transconductance-load ratios of either stage is approximately equal, the resulting GBW is equal to the technology transit frequency f_T augmented by the ratio between f_T and the bandwidth of a single gain stage, f_{pi}. In the published circuit, this considerable bandwidth is furthermore extended with the already introduced negative Miller capacitor concept and inductive peaking techniques, to achieve a currently unequaled 10 Gb/s data rate in a 0.18 µm CMOS process.

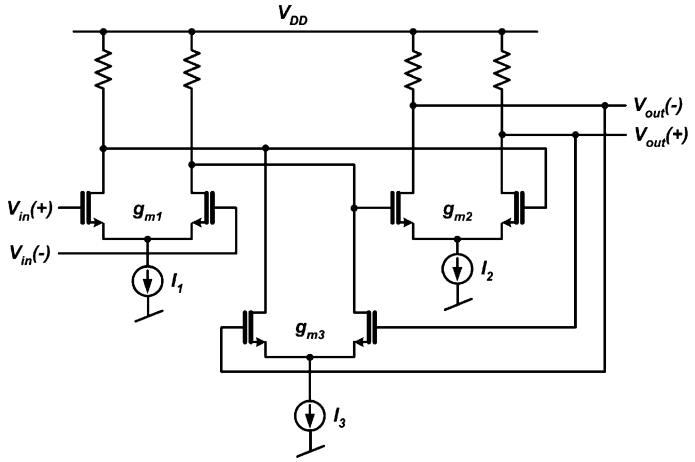

FIGURE 7.14. Active-feedback architecture

$$GBW = \frac{g_{m1}g_{m2}}{4\pi^2 C_{L1}C_{L2}} \cdot \frac{1}{f_{pi}} \approx f_T \cdot \frac{f_T}{f_{pi}} \qquad \text{(EQ 7.8)}$$

7.3.7 Interstage Scaling Technique

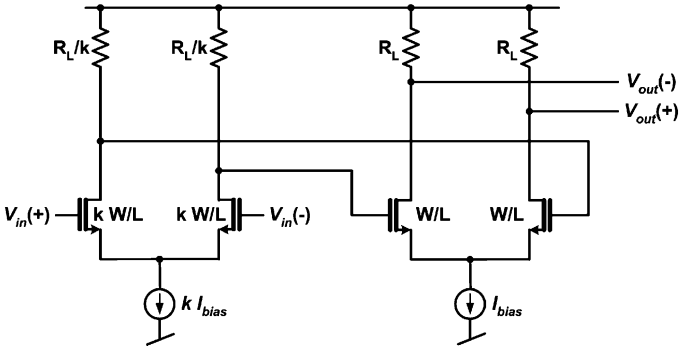

FIGURE 7.15. Differential gain stage and its load, defining the scaling factor k

We have previously seen that the gain-bandwidth performance of the limiting amplifier is determined by the transconductance-capacitance ratio of a single stage. In order to minimize the capacitance, it is obvious that all stages will be designed at minimum channel length, unless the speed requirements are sufficiently weak to allow longer channels. In the following, we will always consider the amplifier as being loaded by another differential stage, with all differential pairs operating at the same current density (Figure 7.15). By scaling each stage with respect to the succeeding stage by $k > 1$, the bandwidth of each stage is in first order increased by the same amount [77]. The scaling factor k the becomes the ratio of bias currents between driving stage and loading stage, as well as the ratio of transistor width between these stages.

7.3.8 Topology Discussion

The presented bandwidth-enhancement techniques are summarized in Table 7.1. Instead of going into the details of one of these techniques, we would like to highlight the impact of inductive peaking for the reader to understand why its use becomes almost mandatory at high-data rates. For this reason, we use simple resistively loaded stages and apply the interstage scaling technique on one hand, the inductive peaking technique on the other, to achieve sufficient gain-bandwidth to meet the specifications of our system. The second design, discussed in Section 7.6, is also used as a basis to analyze the potentiality of magnetic coupling in multichannel optical receivers with inductive peaking.

TABLE 7.1. Summary of LA bandwidth-enhancement techniques

	Improvement technique	Drawbacks/issues
Inductive peaking	Inductors resonate out capacitance	Area (?)
Active inductive peaking	MOS transistors replace inductors	Voltage headroom, gate oxide stress
Capacitive degeneration	Additional pole-zero pair	Gain reduction

TABLE 7.1. Summary of LA bandwidth-enhancement techniques

	Improvement technique	Drawbacks/issues
Negative Miller capacitance	Cancels (partially) input capacitance	Capacitance matching
Cherry–Hooper	Shunt feedback	Voltage headroom
Active feedback	Signal feedback without resistive loading	
Interstage scaling	Stage sizing	Power, area and input capacitance

7.4 Specifications

Sample limiting amplifier specifications, as established in Chapter 4, are presented in Table 7.2.

TABLE 7.2. Calculated block specifications

Parameter	Acronym	Minimum	Maximum	Unit
LA voltage gain	A_{vLA}	21		dB
LA bandwidth	BW_{LA}	3.75		GHz
LA input-referred noise voltage	V_{nLA}		103	μV_{RMS}
LA input capacitance	C_{inLA}		200	fF
CDR input sensitivity	V_{minCDR}	350		mV_{PP}
CDR input capacitance	C_{inCDR}		40	fF

7.5 Inductorless Limiting Amplifier Design

The limiting amplifier is composed of several gain stages, scaled according to the techniques discussed in Paragraph 7.3.7 (Figure 7.16). From the figures established in Paragraph 7.2.4 with a somewhat idealized view of parasitic capacitance, it can be seen that the bandwidth of a single stage at a gain of 3 does not exceed 10 GHz. Considering the bandwidth reduction when cascading multiple stages, the need for bandwidth enhancement in presence of parasitic capacitance becomes obvious. In the following, we will introduce an optimization procedure for proper selection of the two dominant parameters in this topology, the scaling factor k and the number of stages M.

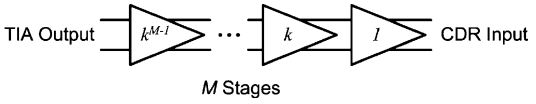

FIGURE 7.16. Inductorless LA topology

7.5.1 Interstage Scaling Optimization

The bandwidth of each stage can be increased using the interstage scaling technique previously discussed. If the stage (i + 1) loading the stage (i) is scaled down by a factor k, the bandwidth of stage (i)

achieved with an identical loading stage is multiplied by the same factor k. Interstage scaling is achieved by multiplying source current and transistor widths of the preceding stage by k and dividing the resistor values by k, resulting in identical gain and voltage swing. The drawback of this technique is the large input capacitance of the limiting amplifier, loading the transimpedance amplifier's output stage.

In the present design, the differential pairs operate close to the velocity saturation regime, where the transconductance cannot be expressed analytically. Nevertheless, we will develop an analytical expression of the DC gain, which then allows for bandwidth optimization at minimum input capacitance. Let us first write $C_u(i)$ the load capacitance per unit width of stage (i):

$$C_u(i) = \frac{C_G(i+1) + C_{DB}(i)}{W(i+1)} = C_W(1 + \alpha k) \qquad \text{(EQ 7.9)}$$

The load capacitance is given by the sum of the gate capacitance of the next stage $C_G(i + 1)$ and the drain junction capacitance of the actual stage divided by the width of the next stage. As both terms are proportional to the width of the device, we can rewrite C_u as a function of a technology-dependent parameter C_W, a parameter α, which is the ratio between junction capacitance and gate capacitance, and the scaling factor k. The interconnect capacitance, as well as the sidewall capacitance of the junction's narrow edges are neglected in Equation 7.10, assuming the presence of large devices. The overlap capacitance, included in the factor C_W, is multiplied by a typical gain value of 3 to estimate the Miller effect. The DC gain can then be written as a function of k and the number of stages M:

$$Av_{DC} = \left(\frac{g_{mu}}{2\pi BW} \cdot k \cdot \frac{1}{C_W \cdot (1 + \alpha \cdot k)} \cdot \sqrt{\sqrt[M]{2} - 1} \right)^M$$

$$\text{with } C_W = C_{ox} \cdot L + C_{oxside} \cdot 2 + C_{overlap} \cdot Av_{DCi} \qquad \text{(EQ 7.10)}$$

$$\text{and } \alpha = \frac{C_J \cdot W_J + 2 \cdot C_{JSW}}{C_W}$$

The achievable gain can be maximized through g_m/C_W, i.e. using minimum channel length devices and maximizing the current densities, as already discussed. Figure 7.17 shows the amplifier gain for different values of k and M and bandwidth specification of 3.75 GHz. The flat pane represents the mimimum gain specification, while the intersection determines the minimum values for k and M satisfying both the gain and bandwidth specifications.

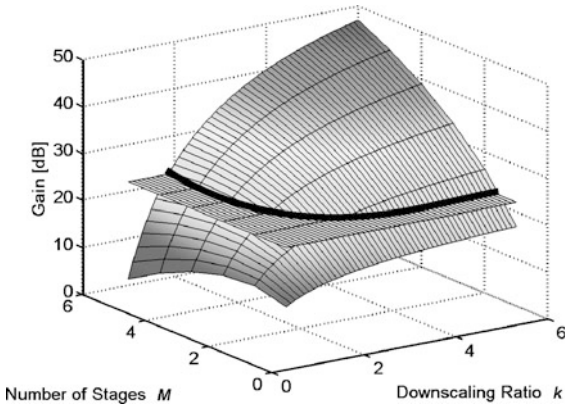

FIGURE 7.17. Achievable gain as a function of the number of stages *M* and the scaling *factor k* for a bandwidth of 3.75 GHz

$$C_{in} = k^M C_{load} \qquad \text{(EQ 7.11)}$$

Although scaling is good for the overall noise performance, the penalties in the LA input capacitance may render the transimpedance amplifier design excessively complex (Equation 7.11). It is preferable to achieve the required gain with a small number of amplifier stages and a larger scaling ratio, than with a large number of stages (Figure 7.18).

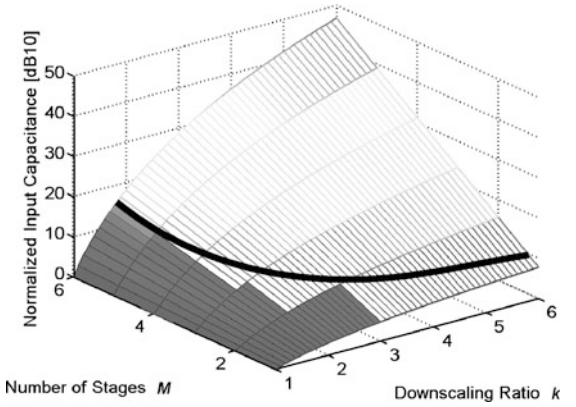

FIGURE 7.18. Resulting input capacitance of the Limiting Amplifier as a function of the number of stages *M* and the scaling factor *k*

In order to meet the gain requirements at minimum input capacitance, a scaling factor of 2 was chosen to be implemented in a 4-stage structure. These numbers result in a relatively simple transistor-level design.

The above equations have been further developed for the case when velocity saturation can be neglected. Equation 7.12 illustrates the fact that the effective device width does not influence the overall gain, which is completely defined by the scaling factor, the number of stages and technology parameters.

$$Av_{DC} = \left(\frac{1}{\pi BW} \cdot \sqrt{\frac{\mu C_{ox}}{2nL}} \cdot \sqrt{\frac{I_{Di}}{W_i}} \cdot k \cdot \frac{1}{C_W \cdot (1 + \alpha \cdot k)} \cdot \sqrt{M\sqrt{2} - 1} \right)^M \quad \text{(EQ 7.12)}$$

7.5.2 Transistor Dimensioning

Based on the above-mentioned figures, sample calculation of the power consumption of the individual stages of the limiting amplifier Figure 7.19 are presented in Table 7.3, starting with the final stage driving the CDR input capacitance.

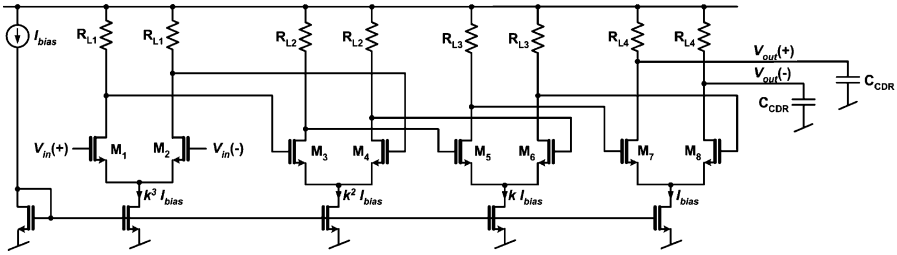

FIGURE 7.19. Core of the interstage scaling limiting amplifier

TABLE 7.3. Inductorless limiting amplifier parameters

	LA stage 1	LA stage 2	LA stage 3	LA stage 4	Load
I_D [mA]	16	8	4	2	–
C_{in} [units]	16	8	4	2	1

7.5.3 Discussion

The interstage scaling technique results in an impractically large input capacitance and power consumption. The following section discussed the inductive peaking amplifier design to allow for results comparison.

7.6 Design of an Inductive Peaking Limiting Amplifier

Before addressing the design of this amplifier, it is worth discussing potential inductive coupling effects when using spiral inductors in multichannel limiting amplifier structures. Indeed, one can imagine that the presence of multiple spiral inductors in close vicinity may lead to magnetic coupling between the different inductors.

7.6.1 The Inductor

The dominantly used inductor structure in differential limiting amplifiers is the T-coil based two-port inductor, which connects to one of the differential signals on each end and to the supply from the center tap (Figure 7.20). This structure maximizes the coupling coefficient k_i between the equivalent inductors in both branches, resulting in an increased effective inductor value.

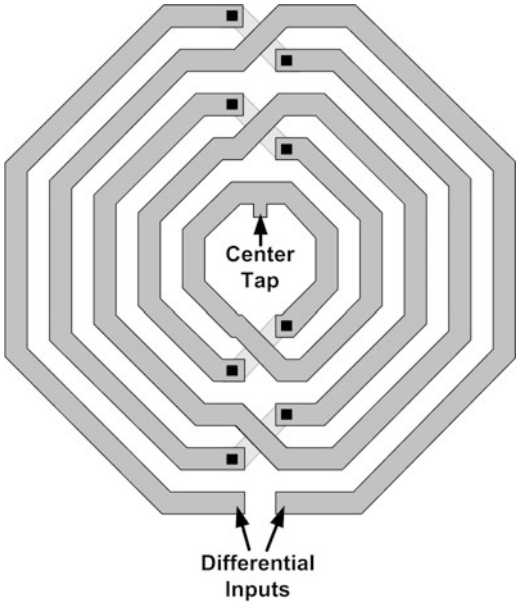

FIGURE 7.20. Illustration of a T-coil spiral inductor (wire spacing not to scale)

The use of multiple metal layers in series allows for implementation of larger inductor values at the expense of increased series resistance. Parallel metal layers on the other hand reduce the series resistance due to increased effective conductor diameter, but the improvement is limited by the relatively large via resistance encountered in modern processes. Furthermore, the lower metal layers have increased parasitic capacitance to the substrate, reducing the self-resonant frequency of the inductor

itself. As peaking inductors in amplifiers do not need high-quality factors, multiple metal layers could indeed be used to minimize the required silicon area. However, trading off the resulting series resistance with the load resistance in the amplification stage requires knowledge of the effective parasitic resistance. In absence of accurately characterized inductor model parameters in the present technology, we preferred to restrict our design to the standard top metal layer available in this digital CMOS process.

7.6.2 Magnetic Coupling Coefficient

Consider a circular spiral inductor with N_1 turns of radius a and center O and a point Q in the plane of the inductor, distant by l of the center of the inductor (Figure 7.21). The magnetic field at Q created by a current I_1 through the inductor, can be written as:

$$B_{21} = -\frac{\mu_0 \cdot I_1 \cdot N_1 \cdot a}{4\pi} \cdot \int_0^{2\pi} \frac{\sin(\phi + \delta)}{(l^2 + a^2 + 2al\cos\phi)} d\phi \qquad \text{(EQ 7.13)}$$

$$\text{where } \delta = \arcsin\left(\frac{a\cos\phi + l}{l^2 + a^2 + 2al\cos\phi}\right)$$

As the angle calculated by the inverse function of sin, arcsin(δ), is defined in [-π/2, π/2], the exact solution of this integral can be written in numerical analysis tools as: ArcTan[Cos(δ), Sin(δ)]. This result will be used later to calculate the value of the coupling coefficient.

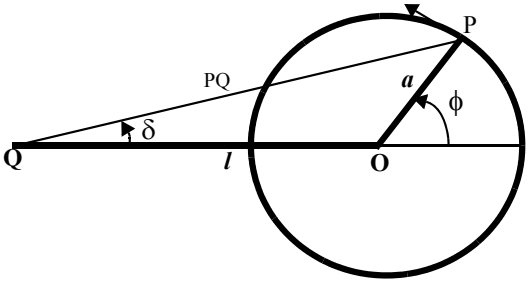

FIGURE 7.21. Calculation of the magnetic field in Q generated by current through a loop centered in O

Now consider a victim loop of radius b which center is at a distance d of point O and which is in the same plane than the aggressor loop. Assuming that the victim loop is far enough from the aggressor ($b << 2a$), the magnetic field is uniform over the surface of that inductor. The flux through that inductor can then be written as:

$$\Phi_{21} = B_{21} \cdot \pi \cdot b^2 = -\frac{\mu_0 \cdot I_1 \cdot N_1 \cdot a \cdot N_2 \cdot b^2}{4} \int_0^{2\pi} \frac{\sin(\phi + \delta)}{(l^2 + a^2 + 2al\cos\phi)} d\phi \quad \text{(EQ 7.14)}$$

The mutual inductance is defined by:

$$M_{21} = \frac{\Phi_{21}}{I_1} = -\frac{\mu_0 \cdot N_1 \cdot a \cdot N_2 \cdot b^2}{4} \int_0^{2\pi} \frac{\sin(\phi + \delta)}{(l^2 + a^2 + 2al\cos\phi)} d\phi \quad \text{(EQ 7.15)}$$

The coupling coefficient can finally be obtained from the following equation, which can be easily simplified in presence of identical inductors.

$$k_{ind} = \sqrt{\frac{M_{21}^2}{L_1 L_2}} = \frac{M_{21}}{L_1} = \frac{\Phi_{21}}{\Phi_1} = \frac{B_{21} \cdot \pi \cdot b^2}{B_1 \cdot \pi \cdot a^2} = \frac{B_{21}}{B_1} \quad \text{(EQ 7.16)}$$

In this formula, the magnetic field inside the source loop is given by $B_1 = -\frac{\mu_0 \cdot I_1 \cdot N_1}{2a}$. As a result, we can write the magnetic coupling coefficient as follows:

$$k_{ind} = \frac{a^2}{2\pi} \cdot \int_0^{2\pi} \frac{\sin(\phi + \delta)}{(l^2 + a^2 + 2al\cos\phi)} d\phi \quad \text{(EQ 7.17)}$$

Worst case coupling is expected for very close loops, the minimum distance being given by the sum of the inductor radius ($l = 2a$) plus some spacing to respect the layout design rules, which we will neglect in the following. The proximity of the loops may violate the uniform field assumption used to obtain the above results, but the resulting value of the coupling coefficient will hold as a first order approximation.

$$k_{ind} = \frac{a^2}{2\pi} \cdot \int_0^{2\pi} \frac{\sin(\phi + \delta)}{(4a^2 + a^2 + 4a^2 \cdot \cos\phi)} d\phi = \frac{1}{2\pi} \cdot \int_0^{2\pi} \frac{\sin(\phi + \delta)}{(5 + 4\cos\phi)} d\phi \quad \text{(EQ 7.18)}$$

where $\sin\delta = \frac{\cos\phi + 2}{a(5 + 4\cos\phi)}$

It can be seen from the above equation that the magnetic coupling coefficient between two inductors spaced at a constant factor of their radius hardly depends on this radius. As illustrated in Figure 7.22, the coefficient is constant with inductor radius and remains below 0.1.

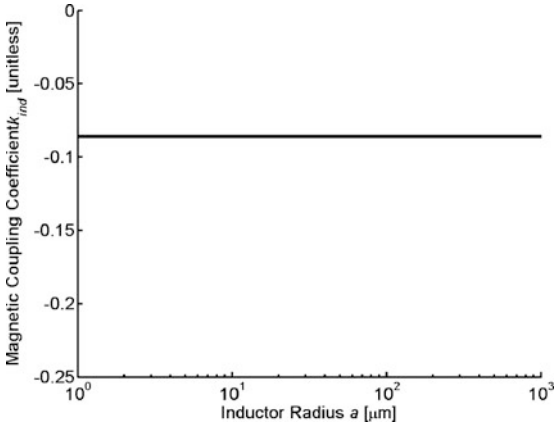

FIGURE 7.22. Mutual inductance coefficient as a function of inductor radius

When a T-coil inductor is excited by a differential signal, the current flow through the inductor is comparable to the one appearing in a conventional spiral inductor carrying a single excitation from the input to the output terminal. Indeed, the electric fields of both half-inductors in the T-coil add in amplitude and thus appear as a single inductor. As a result, it is acceptable to analyze the coupling between two T-coils by considering the coupling between conventional spiral inductors.

Although proximity effects may result in slightly larger effective coupling coefficients than the result shown in Figure 7.22, we can safely conclude that inductive coupling does not represent a major risk to this design. This assumption has been validated by simulation.

7.6.3 Design of the Amplifier Core

TABLE 7.4. Inductor design parameters

Outer diameter [μm]	Wire width [μm]	Wire spacing [μm]	Number of turns	Metal layer	Metal thickness [μm]
250	6.0	5.0	7	Met 6 (top layer)	0.86

To allow for a fair comparison of both designs, we use the same basic gain stage and replace the bandwidth enhancement through interstage scaling by inductive peaking. The resulting amplifier thus consists of three identical gain stages plus a swing-limited output stage (Figure 7.23). Initial inductor

values result from estimation of the capacitive load of each stage and can be fine-tuned by simulation of the small-signal bandwidth. The T-coil parameters shown in Table 7.4 can be determined using a 2.5D- or 3D-modeling tool (e.g. ASITIC [82]) to achieve the inductor specifications determined in the previous small-signal simulation.

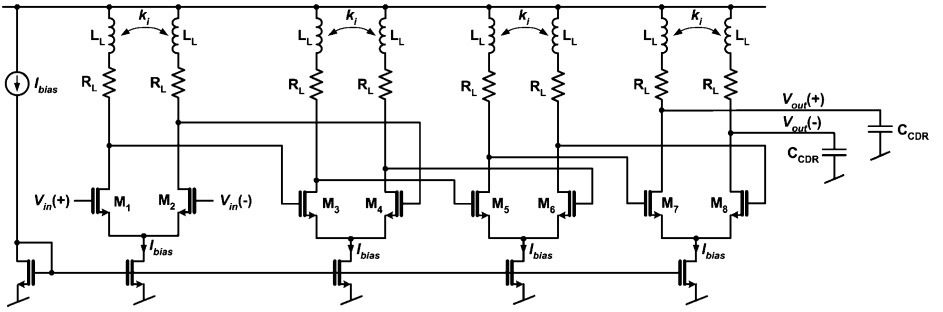

FIGURE 7.23. Limiting amplifier core

Typical performance parameters of the amplifier have already been introduced in Paragraph 7.5.2. Derivation of the transistor dimensions is straightforward; the results are presented in Table 7.5. Indicated bias currents are obviously nominal values. They are to be generated by an on-chip bias current generator (not presented here) to compensate for process and temperature variations.

TABLE 7.5. Inductive peaking limiting amplifier parameters

	LA stage 1	LA stage 2	LA stage 3	LA stage 4	Load
I_D [mA]	2	2	2	1	–
C_{in} [units]	1.2	1.1	1.1	1.1	1

7.6.4 Discussion

Inductive peaking results in a more convenient limiting amplifier design in terms of input capacitance and power consumption, while remaining competitive with its inductorless counterpart in terms of silicon area. Although magnetic coupling appears not to be an issue, pickup of noisy substrate currents by the inductors may become a source of interchannel cross talk. This source of coupling can however be mitigated by the use of properly connected patterned ground shields underneath the inductors, which however may reduce the self-resonant frequency of the inductor.

7.7 Complete Limiting Amplifier

7.7.1 Offset Compensation

The limiting amplifier core provides constant gain over a large range of frequencies down to DC. Eventual offsets in the differential stages, due to device mismatch, as well as low-frequency noise sources and potential drift due to thermal variations, can be of the same order of magnitude than the actual high-frequency signal. To avoid saturation of the amplifier output swing by these undesired low-frequency components, an offset compensation mechanism is implemented (Figure 7.24). A feedback loop reduces the low-frequency gain to minimize their contributions to the output signal and obtain a band-pass transfer characteristic of the overall amplifier. The passband requirements of the limiting amplifier have already been previously introduced (Section 3.2) and are verified by simulation.

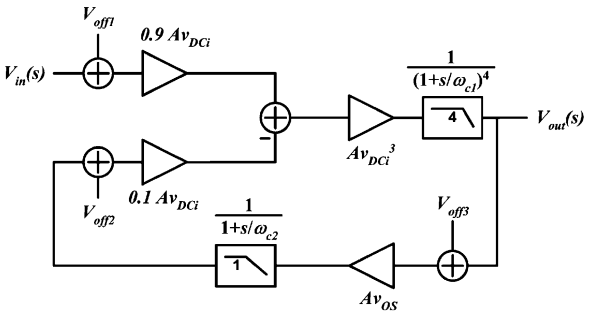

FIGURE 7.24. LA offset compensation mechanism

Equation 7.19 shows the result of this mechanism, commonly dubbed *offset compensation*, for different offset sources. We will use this term in the following although the suggested mechanism does not really compensate for the amplifier offset. The integrator pole also appears as a zero in the closed-loop form, therefore this system provides only finite attenuation at DC by reduction of the overall low-frequency amplifier gain. This fact is illustrated by the simulated small-signal transfer functions shown in Figure 7.25.

$$V_{out} = \frac{\frac{A_{vDCi}^4}{10}\left(1+\frac{s}{\omega_{c2}}\right)}{\frac{s^5}{\omega_{c1}^4\omega_{c2}}+\frac{s^4}{\omega_{c1}^4}+\frac{s}{\omega_{c2}}+\frac{A_{vDCi}^4 A_{vOS}}{10}+1}\cdot\left[9(V_{in}+V_{off1})-V_{off2}-\frac{A_{vOS}V_{off3}}{1+\frac{s}{\omega_{c2}}}\right] \quad \text{(EQ 7.19)}$$

The reader must be aware that this small-signal analysis does not take into account the limiting effect of the amplifier. Due to limiting, offsets are removed from the amplitude information of the

output signal and appear only in the form of duty-cycle errors. As a result, the feedback gain of the offset compensation loop is effectively lower than given by the small-signal analysis, especially at large input signals!

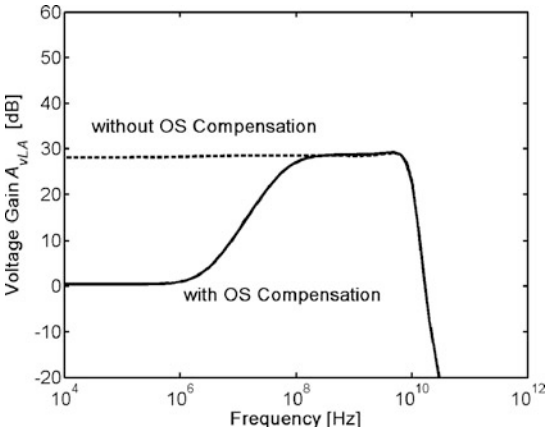

FIGURE 7.25. Simulated small-signal transfer function of the inductive peaking amplifier

In order for the amplifier to allow for low-frequency gain reduction, an additional input has to be provided. For this purpose, a small part of the bias current of the first stage is rerouted to an additional differential pair. This pair receives the output of the integrator implementing the pole ω_{c2}. Its output corresponds to the long-term mean output of the limiting amplifier. The additional differential pair then subtracts this signal from the main input signal (Figure 7.26).

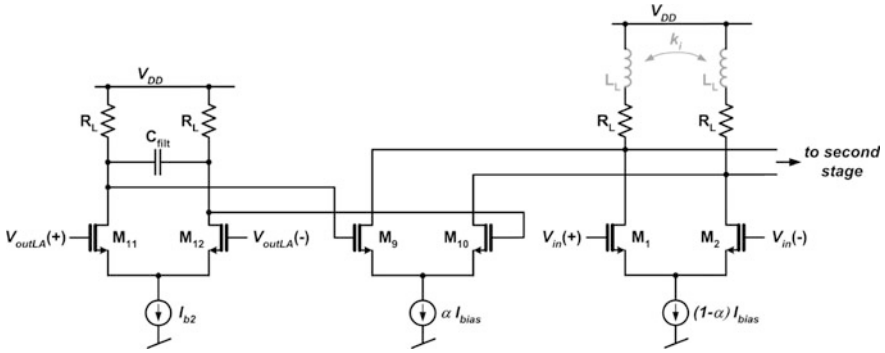

FIGURE 7.26. Modified LA input stage for offset compensation (α = 10%)

7.7.2 Complete Block Diagram

The complete limiting amplifier schematic is shown in Figure 7.27, which contains the limiting amplifier and the offset compensation blocks. This amplifier is a voltage amplifier, so no additional test features are required.

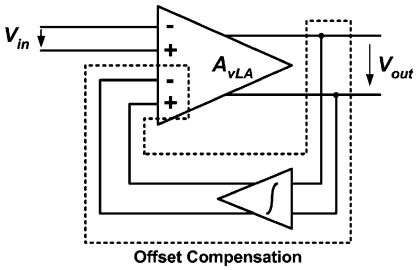

FIGURE 7.27. Complete limiting amplifier block diagram

7.7.3 Limiting Amplifier Noise Estimation

The input referred noise V_{nLA}^2 of a cascade of M identical gain stages can be expressed as a function of the noise of a single stage V_{ni}^2:

$$V_{nLA}^2 = V_{ni}^2 \sum_{j=1}^{M} \frac{1}{A_{vDCi}^{2(j-1)}} \quad \text{(EQ 7.20)}$$

Applied to a cascade of M stages scaled by a factor k, the noise of stage j depends on its load capacitance $C_L(j)$. The latter can be expressed as a function of the load capacitance C_{inCDR} of the LA and the stage index j. As a result, the total input referred amplifier noise becomes:

$$V_{nLA}^2 = \sum_{j=1}^{M} \frac{V_{ni}^2(j)}{A_{vDCi}^{2(j-1)}} = \frac{2k_B T}{A_{vDCi} C_L}\left(\gamma + \frac{1}{A_{vDCi}}\right) \sum_{j=1}^{M} \frac{1}{A_{vDCi}^{2(j-1)} k^{M-j}} \quad \text{(EQ 7.21)}$$

Based on the previously discussed transistor-level design, the inductive peaking amplifier input referred noise is about 347 μV_{RMS}. This value is based on an assumed excess noise factor of $\gamma = 1.0$ and a total load capacitance of 40 fF. The limiting amplifier output noise is shown in Figure 7.28. The effect

of offset compensation of the low-frequency noise components can be clearly seen, which are attenuated by two decades.

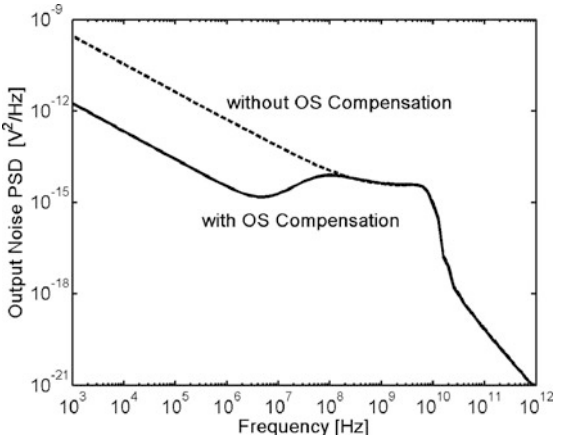

FIGURE 7.28. Inductive peaking LA output noise

The inductorless amplifier's input referred noise has been estimated and simulated at 1/4 of the inductive peaking design. This is obviously related to the much higher power consumption in the first stages. Nevertheless, the inductive peaking amplifier meets the noise specifications with a 22% margin, which is sufficient to cover other possible noise sources in the circuit (e.g. supply and substrate noise).

7.8 Block Layout

The layout constraints in terms of block sizing have already been discussed in Paragraph 6.5.1 with a target channel width of 250 μm in case of detector integration. Power supply and ground rails are placed on either side of this channel. The differential gain stages are systematically constructed from the current source in proximity of the ground rail to the load resistors and inductors connected to the positive supply. Although smaller inductors can perfectly meet the peaking specifications, larger inductors where implemented in Figure 7.29 to verify the importance of magnetic coupling. Indeed, two limiting amplifiers where integrated with adjacent spiral inductors for maximum potential inductive coupling. Additional power supply decoupling capacitors on either side of the amplifiers are easily

visible, as well as the output drivers. As already mentioned, the inductor area can be reduced by trading off load resistance, without impact on the circuit performance.

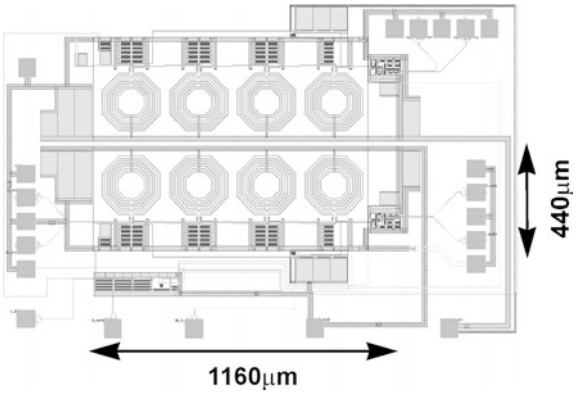

FIGURE 7.29. Two-channel inductive peaking limiting amplifier layout

7.9 Circuit Measurements

7.9.1 Measurement Setup for Time-Domain Measurements

All following measurements were done using an output buffer at the limiting amplifier output to drive the 50 Ω-load measurement equipment. This buffer only being used for characterization purposes, it is based on conventional differential stages and uses some of the above-mentioned bandwidth enhancement techniques (Figure 7.30). When designing such output buffers, it is important to use reliable models for I/O pads, ESD protection, bond wires, packaging and PCB traces if applicable (depending on the measurement setup and the application configuration).

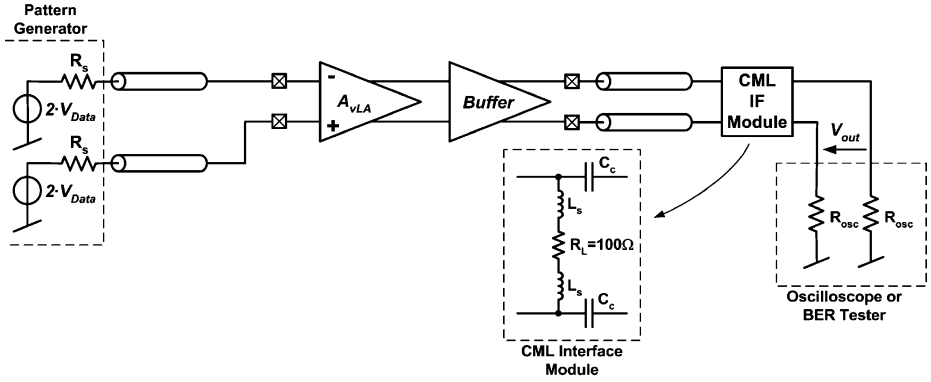

FIGURE 7.30. Limiting amplifier measurement setup

In case the limiting amplifier design to be characterized contains no termination loads at the input, it is tempting to include such terminations resistors right in front of the wafer probes. However, such a configuration would lead to impedance mismatches at both ends of the wiring segment between the probe connector and the circuit input, leading to a large amount of reflections, appearing as ISI in the measured eye diagram (Figure 7.31).

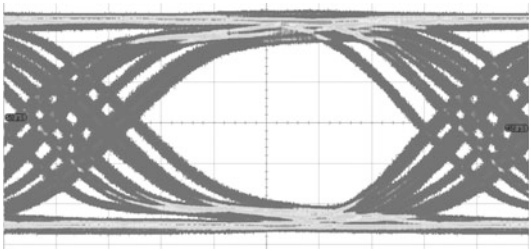

FIGURE 7.31. Measured eye diagram with termination resistor and 50 mV input signal

Omitting the termination at the probe connector somewhat improves signal reflections, as it results in a single impedance mismatch at the circuit input, while the source impedance remains matched. The reflections generated at the circuit input thus travel back to the source, where they are swallowed. The measurements are then almost free of inter-symbol interference, though the remaining impedance mismatch at the amplifier gate results in a different effective input voltage.

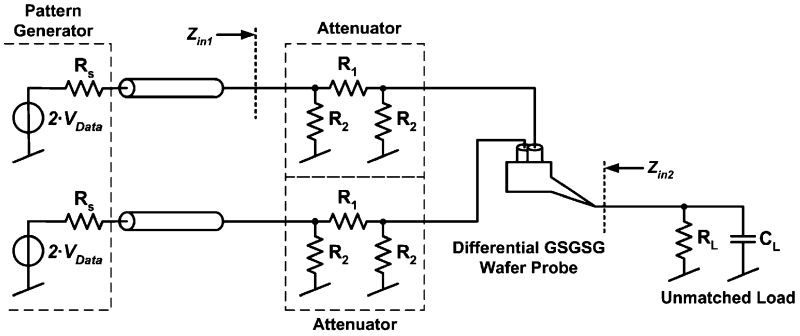

FIGURE 7.32. Input signal generation with attenuator serving as termination impedance

Instead of a termination resistor next to the probe, one can use a 20 dB attenuator (Figure 7.32). Equations 7.22 and 7.23 show the resulting impedance seen by the surrounding circuit from either side of the attenuator. The equivalent resistances in the Π-model of the microstrip attenuator correspond to $R_1 \approx 247.5\ \Omega$ and $R_2 \approx 61.1\ \Omega$. Considering the load impedance Z_L large compared to the characteristic impedance, we obtain the following impedance values: $Z_{in1} \approx 51.0\ \Omega$, while $Z_{in2} = 50.0\ \Omega$ as expected.

As a result, the signal source sees an almost terminated load, resulting in very few reflections, while the reflections issued by the unmatched load are swallowed by the 50 Ω matched impedance seen from the probe tip back into the attenuator.

$$Z_{in1} = \frac{R_1 + \dfrac{R_2 Z_L}{R_2 + Z_L}}{1 + \dfrac{R_1}{R_2} + \dfrac{Z_L}{R_2 + Z_L}} \cong \frac{R_1 + R_2}{2 + \dfrac{R_1}{R_2}} \qquad \text{(EQ 7.22)}$$

$$Z_{in2} = \frac{R_1 + \dfrac{R_2 R_s}{R_2 + R_s}}{1 + \dfrac{R_1}{R_2} + \dfrac{R_s}{R_2 + R_s}} \qquad \text{(EQ 7.23)}$$

Finally, the resulting attenuation of a 20 dB attenuator terminated by a much larger impedance can be calculated. This value allows for calculation of the voltage applied to the limiting amplifier input in the above described measurement setup based on the amplitude delivered by the pattern generator.

$$Att = \frac{\dfrac{R_2 Z_L}{R_2 + Z_L}}{R_1 + \dfrac{R_2 Z_L}{R_2 + Z_L}} \cong \frac{R_2}{R_1 + R_2} \cong 0.2 \qquad \text{(EQ 7.24)}$$

7.9.2 Eye Diagram Measurements

A typical eye diagram measured at 2.5 Gb/s and an input signal of 20 V_{PP} is shown in Figure 7.33. Slight onset of ISI combined with the relatively slow rise and fall times hint at slightly insufficient bandwidth, in the region of 0.5 f_B. Amplitude noise is clearly visible in the vertical eye opening, while jitter remains limited.

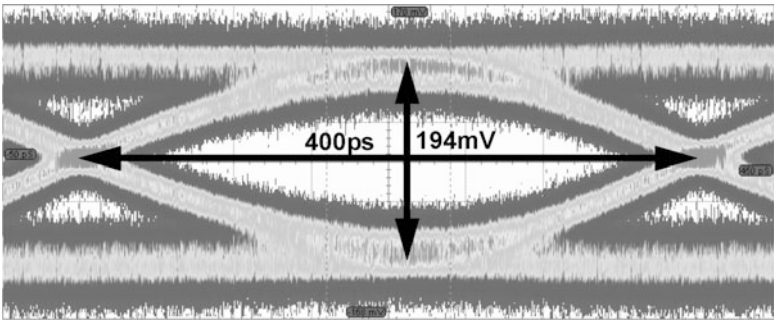

FIGURE 7.33. Measured eye diagram of the inductive peaking LA

The amplitude noise level at the output is determined based on the Q_{BER} value (Section 3.4) measured by the oscilloscope in eye diagram mode, relying on the assumption that intersymbol interference remains small and the noise amplitudes around either logic data level are approximately equal. As a result, the RMS noise amplitude at the amplifier output is given by:

$$V_{nout} = \sigma_{out} = \frac{V_{inPP}}{2Q_{BER}}, \qquad \text{(EQ 7.25)}$$

resulting in an output noise amplitude of 10.60 mV_{RMS}. The input referred noise voltage, calculated based on the voltage gain value of 20 dB, become 1.06 mV_{RMS}.

7.9.3 Magnetic Coupling Measurements

We concluded earlier that magnetic coupling between channels is negligible and should not perturbate the neighboring channels. To validate this result, two channels were implemented side-by-side, where one channel was flipped to achieve inductor proximity.

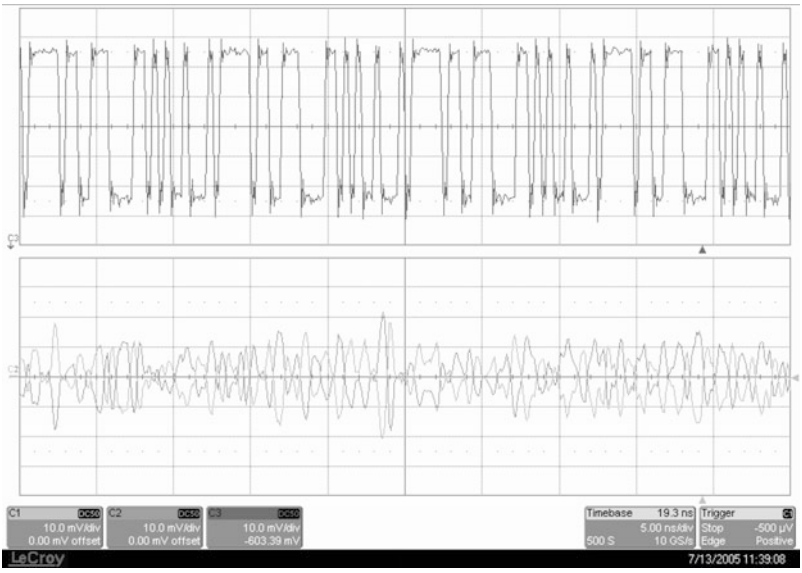

FIGURE 7.34. Magnetic coupling measurement showing the input stimulus (top) and both outputs of the victim channel (bottom)

While the aggressor channel is strongly overdriven with a 200 mV input signal, the victim channel has its input tied to a common-mode voltage of 2/3 V_{DD}, generated by a resistive divider with capacitive decoupling. As shown in Figure 7.34, the output of the victim channel hardly exceeds 40 mV_{PP}

differential for this relatively large aggressor input signal. This illustrates that the weak coupling effects between neighboring channels can be safely neglected.

7.9.4 Measurement Summary

Table 7.6 shows the performance of the inductive peaking limiting amplifier implemented in a digital 0.18 μm CMOS technology without thick top metal for inductors, while also including simulation results of the inductorless implementation for comparison.

TABLE 7.6. Summary of limiting amplifier measurement results

Specification	Simulation				Measurement		Unit
	w/o L		with L		with L		
	min	max	min	max	min	max	
Supply voltage	1.62	1.98	1.62	1.98	1.6	2.0	V
Bandwidth	3.7		4.5		1		GHz
Gain	25		23		20		dB
Input referred noise	51.5		238.7		1059.8		μV_{RMS}
Current consumption @25°C	60		7.5		6.4		mA
Area	0.40		0.51				mm^2

Finally, input referred noise can be calculated from the measurements of output noise by division by the amplifier gain, leading to the value of 1.06 mV_{RMS}. While a small part of the difference between simulated and measured input referred noise levels, diverging by a factor of 4.4, can be explained from the discrepancy between simulated and measured voltage gain, we believe that the dominant part of this error is due to ambient noise pickup in the wafer probed measurement setup, which is subject to several unavoidable ground loops and uses the ground contacts of the differential probes as a supply connection. Indeed, we were able to show the noise dependency on ambient switching noise due to power supplies, the presented noise measurements being done at reduced ambient switching noise.

7.10 Discussion

We have seen in this chapter that inductive peaking is a powerful technique to achieve high-bandwidth at low-power consumption. Technology scaling further favors the inductive peaking design because advanced processes come with thicker top metal layers further away from the substrate, resulting in reduced parasitics. Referring to the initially discussed topologies, one can say that optimum limiting amplifier performance is obtained by combining the inductive peaking technique with one or

several other bandwidth enhancement methods. In this direction, especially the active feedback topology enhanced with inductive peaking and negative Miller capacitance seems particularly promising for high-gain high-bandwidth-limiting amplifier design in very deep submicron processes.

CHAPTER 8 *Clock and Data Recovery Circuit*

At the output of the limiting amplifier, the amplified data signal with sharpened data edges is available for further processing, but unique interpretation of the received signal requires timing information. Serial communication links do not provide a synchronization signal on a separate channel and therefore the receiver must rely on the extraction of the timing information from the data stream. This *clock and data recovery* process can be performed in a similar manner in optical communication, electrical serial links, hard drive read-out channels, as well as in some memory interfaces. In the latter field, advocates and opponents of the clock forwarding scheme still debate about the advantages and drawbacks of per-channel clock recovery circuits.

Proper denomination of the whole synchronization process would be *clock recovery and data retiming*, better describing the behaviorally independent operations. Indeed the recovered clock is used to re-sample the incoming (or sometimes delayed) data to provide proper timing information, i.e. synchronicity, for the following blocks (Figure 8.1). However, in many implementations, the retiming operation is embedded in the clock recovery part to achieve improved circuit performance.

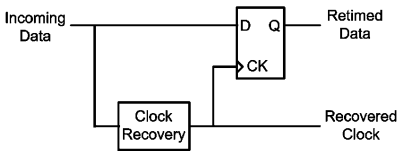

FIGURE 8.1. Clock recovery and data retiming principle

The major challenges of multichannel clock recovery circuit design have already been presented, namely, the achievement of low-power and low-area overhead with minimum impact on the circuit performance. First, an introduction to the important concepts in the design of clock recovery circuits is

P. MULLER, Y. LEBLEBICI,
CMOS Multichannel Single-Chip Receivers for Multi-Gigabit Optical Data Communications,
DOI 10.1007/978-1-4020-5912-4_8, © 2007 Springer

presented. Then, an overview of a variety of clock recovery topologies with their respective advantages and drawbacks allows the reader to fully understand the design choices operated later in this chapter. After discussion of appropriate topologies for multichannel clock recovery, a top-down design approach is presented which validates the jitter performance of the selected topology at the concept level, at the behavioral level and at the transistor level. Finally, the transistor-level design of the CDR is briefly discussed and characterization techniques are introduced.

8.1 Clock Recovery Principles

Historically, clock recovery was performed using resonant filter-based structures, either using discrete passive elements or resonant elements like surface acoustic wave (SAW) devices. In the era of systems-on-chip, requirements for small system form factors and minimum bill of material, the use of such off-chip components is strongly discouraged. In the long-haul market, where fiber-optic communications made their breakthrough first, phase-locked loop (PLL) topologies have become the mainstream solution, thanks to their capability of monolithic integration with a minimum number of external components, typically only one loop filter capacitor.

Beyond the fact that the PLL currently still dominates the field and is the most well-known clock recovery topology, we would like to use this CDR structure as a basis to the explanation of some concepts to the novice reader. A more detailed discussion of PLL-based clock recovery circuits can be found in [83]. For simplification, let us first consider a phase-locked loop with a periodic input signal (Figure 8.2). The three main building blocks are the phase detector (PD), the loop filter (LF) and the voltage-controlled oscillator (VCO). The phase detector compares the phase error between the input signal and the generated signal at the VCO output. The phase error is integrated in the low-pass loop filter and the resulting control signal tunes the VCO frequency, which in turn reduces the phase error.

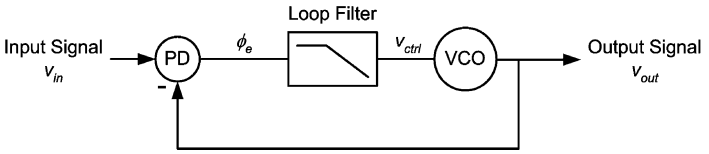

FIGURE 8.2. PLL with periodic input signal

The periodic input signal *f(t)* can be written as the sum of sine and cosine functions, namely, its Fourier series:

$$v_{in}(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} [a_n \cdot \cos(nt) + b_n \cdot \sin(nt)] \quad \text{(EQ 8.1)}$$

In most high data rate applications, we can neglect the higher order harmonics and focus on the fundamental signal, which can be written as:

$$v_{in}(t) = A \cdot \sin(\phi(t)) \quad \text{(EQ 8.2)}$$

The instantaneous value *f(t)* of a sine wave is defined by its peak amplitude *A* and its instantaneous phase $\phi(t)$, which can again be decomposed into a constant angular frequency $\omega_0 = 2\pi f_0$ and a time-varying phase angle $\theta(t)$. In phase and frequency modulation schemes used in wireless data transmission, this time-varying phase angle contains the information to be transmitted. In fiber optic links, it does not contain any useful information, but represents the phase noise due to channel and circuit imperfections like noise and limited bandwidth.

$$v_{in}(t) = A \cdot \sin(\omega_0 t + \theta(t)) = A \cdot \sin(2\pi f_0 t + \theta(t)) \quad \text{(EQ 8.3)}$$

In both types of applications, the phase-locked loop technique allows to extract the time-dependent phase information and track the input signal. The in-phase output signal v_{out}, locked to the input signal, guarantees a very small phase error $\phi_e(t)$ between the input phase $\phi_i(t)$ and the output phase $\phi_o(t)$. This situation is called *phase lock*. As frequency is the derivative of the instantaneous phase, constant phase error provides zero frequency error between input and output signals. The loop bandwidth being limited by the low-pass filter, these characteristics are not guaranteed for variations $d\theta(t)/dt$ at higher frequencies.

As already mentioned, the two-level amplitude modulation of the carrier in serial links can lead to data runs which do not contain any transitions. The low-pass behavior of the loop filter allows the PLL to memorize the signal frequency during these long runs without loss of phase lock.

The *capture range* defines the frequency span of the input signal for which phase lock can be achieved. For frequencies outside this span, the PLL will remain in a free-running situation. *Lock range* on the other hand defines the frequency span for which lock can be maintained once it is acquired. Without going into more details, these parameters depend on the characteristics of the PLL building blocks, among which the VCO's *tuning range*, defining the span of frequencies the oscillator can generate when driving its control voltage from one end to the other of the range.

In the following, the particularities of clock recovery from random data will be addressed. The input signal to be considered will be the NRZ data stream, while the output signal is the recovered clock signal.

8.1.1 NRZ Data Phase Detection

Combining the data retiming mechanism presented in Figure 8.1 with the PLL topology in Figure 8.2, we obtain the general PLL-based clock recovery structure presented in Figure 8.3.

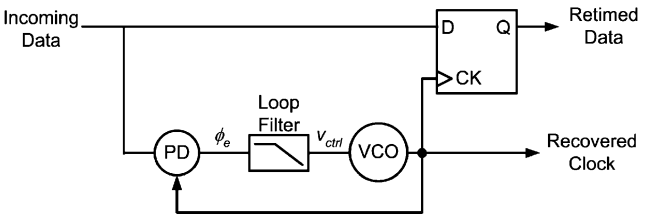

FIGURE 8.3. Overall PLL-based clock recovery structure

Several phase detector topologies with implicit edge detection on random data have been published. One well-known topology uses the incoming data D_{in} to sample the VCO clock signal CK_{VCO} (Figure 8.4, [83]). However, the flip-flop used in the phase detector and in the data sampler present unequal delays in the CK->Q and D->Q paths. This delay mismatch introduces a non-negligible static phase error at high data rates, resulting in reduced jitter tolerance at high frequencies.

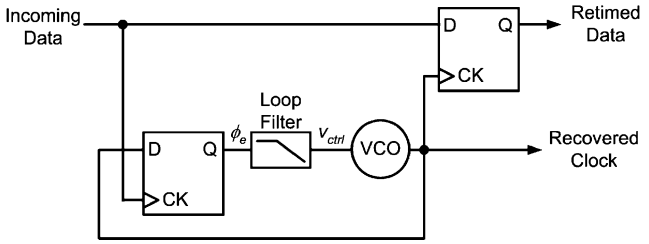

FIGURE 8.4. CDR based on a sampled clock topology

More convenient topologies for high data rate NRZ clock recovery have been published by Hogge [84] and Alexander [85]. Both phase detectors perform inherent retiming, canceling any static phase error resulting from possible path delays between clock and data.

8.1.2 Linear Phase Detector

The linear phase detector, introduced by Hogge, delivers output pulses with a duration that is proportional to the phase error between data and clock. The resulting PLL can be analyzed using linear system theory. The equivalent phase-domain model of the linear CDR is shown in Figure 8.5. The phase detector is represented by a gain block of value K_{PD}. The Laplace representation is used for the loop filter which also contains the charge pump, while the VCO is modeled using a gain factor combined with an integration. This integrative behavior is due to the fact that the signals are considered in the phase, whereas the characteristic tuned by the control voltage is the VCO frequency.

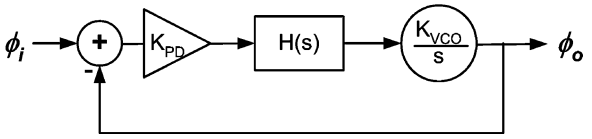

FIGURE 8.5. Equivalent linear model of the CDR

Charge pump topologies are commonly used in integrated PLLs to eliminate the large resistor required in the loop filter, changing at the same time the system behavior to discrete-time operation. Additionally, the use of a charge pump results in a second pole at the origin in the loop transfer function (the first being due to the VCO). A zero is thus required in the loop to guarantee the stability of the system. The resulting loop filter is shown in (Figure 8.6).

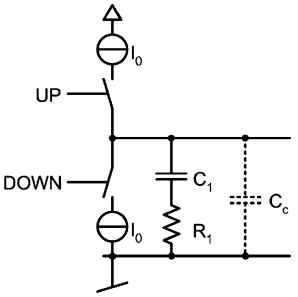

FIGURE 8.6. Second-order loop filter

A resistor R_1 placed in series with the main loop filter capacitor C_1 creates a zero at $1/(R_1C_1)$, but also converts the switched charge pump current into ripple on the control voltage at the output of the loop filter. This ripple can be attenuated by an additional shunt capacitor C_c, resulting in a third-order transfer function (Equation 8.4). This continuous-time representation of an inherently discrete-time circuit is of course based on the assumption that the loop bandwidth is much smaller than the frequency of operation [86]:

$$H(s) = \frac{2\zeta\omega_n s + \omega_n^2}{m\frac{2\zeta}{\omega_n}s^3 + (m+1)s^2 + 2\zeta\omega_n s + \omega_n^2} \quad \text{(EQ 8.4)}$$

In this equation, $m = C_c/C_1$, while the natural frequency ω_n and the damping factor ζ are defined as follows:

$$\omega_n = \frac{I_{PD}K_{VCO}}{2\pi C_1} \quad \text{(EQ 8.5)}$$

$$\zeta = \frac{RC_1}{2}\omega_n \quad \text{(EQ 8.6)}$$

Notice that in a 3rd order function, natural frequency and damping factor do not have the same meaning and do not characterize the loop behavior as they do for a 2nd order. Typical loop filters have $C_1 >> C_c$ ($m << 1$), which means that the third pole does not influence the peaking. With this hypothesis, the loop function reduces to a 2nd order function, where damping factor and natural frequency recover their well-known meaning (Equation 8.7).

$$H(s) = \frac{2\zeta\omega_n s + \omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \quad \text{(EQ 8.7)}$$

As a side note, such a transfer function must exhibit peaking because the closed-loop zero is located at a lower frequencies than the poles. The SONET standard specifies the maximum jitter peaking in the transfer function as 0.1 dB to minimize jitter transfer. In short-distance communications where jitter transfer is not critical, the jitter peaking requirements are less stringent. Loop stability however, is and determines the maximum acceptable peaking.

The relationship between closed-loop bandwidth and natural frequency is given by:

$$\frac{\omega_c}{\omega_n} = \sqrt{2\zeta^2 + \sqrt{4\zeta^4 + 1}} \quad \text{(EQ 8.8)}$$

Provided that low peaking is achieved, i.e. the loop is sufficiently damped ($\zeta > 4$), the transfer function tends to its asymptotic 2nd order function and the closed-loop bandwidth is equal to:

$$\omega_c = 2\zeta\omega_n \quad \text{(EQ 8.9)}$$

The loop bandwidth of conventional PLLs directly determines the system's capture range. As we will see later (paragraph 8.2.1), there are several reasons in clock recovery systems to add a frequency acquisition path, which then also determines the capture range.

In any case, the PLL loop bandwidth determines jitter tolerance performance. As jitter transfer and generation are of no importance to short-distance communications, it is "sufficient" to design the loop bandwidth larger than the jitter tolerance corner frequency. It is easily understood that the loop tracks jitter at frequencies below its bandwidth, exhibiting decreasing jitter tolerance at higher frequencies, as previously shown in the related jitter tolerance specification. Jitter tolerance can also be determined by the VCO tuning range. In this case, exhaustively discussed in [87], a sort of frequency slewing is exhibited by the loop. This issue however only appears at higher jitter frequencies when using a VCO with a reduced frequency-tuning range or in circuits with very large-loop bandwidth.

8.1.3 Binary Phase Detector

At high data rates, it becomes increasingly difficult to generate pulses which widths are proportional to the phase error, as required in linear phase detectors. Lack of proportionality between the phase error and the pulse width leads to a dead zone in the phase detector characteristic. The resulting phase error in the loop reduces the jitter tolerance performance. One attempt to circumvent this problem is the use of binary phase detectors, initiated by J. D. H. Alexander and nowadays extensively used in multi-gigabit receivers. The output of this type of phase detector, also called *bang–bang phase detector* (BB PD) for its hard switching characteristic (Figure 8.7), only indicates whether the recovered clock signal is early or late with respect to the received data edges.

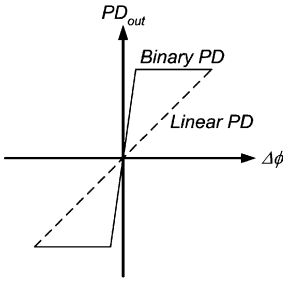

FIGURE 8.7. Bang–bang characteristic compared to a linear characteristic

Although bang–bang phase-locked loops (BB-PLLs) do not allow for linear analysis, some efforts have been undertaken recently to provide some mathematical relationships between the loop specifications and its parameters. While Razavi's work [88] considers both the linear in-lock regime and the binary regime at large phase error, we prefer to summarize here the analysis presented by Walker in [89], focusing on the binary regime of the bang–bang loop. By the way, most bang–bang detector implementation are in fact tri-state designs, which do not deliver any decision in absence of data transitions. This behavior optimally exploits the loop's memory effect during data runs. In the following, the term

bang–bang phase detector defines such tri-state detectors, as opposite to linear detectors, which themselves may also present tri-state implementations.

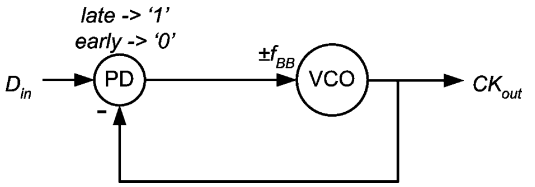

FIGURE 8.8. A first-order bang–bang PLL

In a first order bang–bang loop, shown in Figure 8.8, the VCO frequency toggles between f_0–f_{BB} and f_0+f_{BB}, where f_0 is the oscillator's free running frequency and f_{BB} the frequency step defined by the control voltage step and the VCO gain. The presence or absence of a variation of the control voltage during each cycle depends on the sign of the phase error. A first-order bang–bang loop thus never reaches the exact frequency of the incoming signal, but the average frequency minimizes the phase error. Phase lock is guaranteed as long as the input frequency does not exceed either selectable frequency of oscillation. The frequency switching between both ends of the tuning range results in additional *tracking jitter* (also called *hunting jitter*), which amplitude can be calculated as:

$$TrJ_{PP} = \frac{4\pi}{2\pi} \cdot \frac{f_{BB}}{f_0} = 2 \cdot \frac{f_{BB}}{f_0} \text{ [UI]} \qquad \textbf{(EQ 8.10)}$$

Low tracking jitter requires a relatively small tuning range of $2f_{BB}$. While a small tuning range affects capture and lock properties, it also limits the system's ability to track data jitter. In presence of sinusoidal data jitter of the form $\theta(t) = \frac{SJ_{pp}}{2} \cdot \sin(2\pi f_{SJ} t)$, the maximum jitter amplitude before appearance of a slewing phenomenon, called *slope overload*, is:

$$SJ_{PP} < \frac{2}{2\pi} \cdot \frac{f_{BB}}{f_{SJ}} \text{ [UI]} \qquad \textbf{(EQ 8.11)}$$

Interestingly, this result corresponds to the slope overload of a delta modulator used in early voice encoding schemes. It is valid for a periodic input signal, whereas the lower amount of data transitions in an NRZ data stream further reduces the tolerable amount of jitter.

Due to these limitations and the unsuitable trade-off between jitter tolerance and tracking jitter, a second order loop topology is preferred (Figure 8.9). The integral branch obviously improves jitter tolerance in combination with reduced frequency steps, resulting in less tracking jitter.

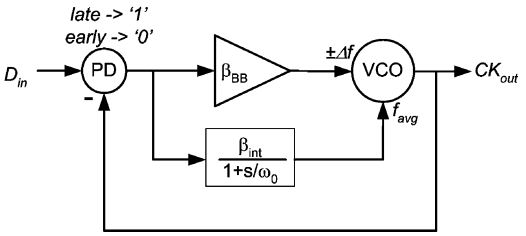

FIGURE 8.9. Second-order bang–bang PLL

The integral branch filters the PD output and adjusts the average VCO frequency with a gain β_{int}, performing low-frequency tracking. The bang–bang branch applies the unfiltered detector output with a gain of β_{BB} to the VCO and thus takes care of the high-frequency tracking. At very high jitter frequencies, jitter tolerance is still limited by the bang–bang step, but at lower frequencies, where more jitter must be tolerated, the integral loop takes over and tracks the low-frequency jitter. As a result, the bang–bang step can be reduced to keep tracking jitter low, while still achieving good jitter tolerance performance thanks to the integral path. The interested reader is referred to [89] for a detailed description.

8.2 CDR Topologies

Power consumption and silicon area are key factors to the success of a multichannel clock recovery circuit, while the influence of interchannel cross talk on the system performance must be considered. Sharing common parts of the CDR structure is an obvious step to reduce the system overhead. Relaxed frequency variations, run length and jitter specifications compared to long-haul systems allow for a simplification of the CDR topology, either by reduction of the loop filter size, or by implementation of alternative topologies, which would be unable to cope with SONET-type specifications. In order to compare their advantages and drawbacks, an overview of the topologies to be considered is presented. Over the last years, the growing number of electrical serial link applications has stimulated the research on alternatives to the conventional PLL/DLL topologies. As the terminology in this field is not totally harmonized, the following classification and especially the association of some published designs to given categories may be subject to discussion. In a general fashion, we will consider the following clock recovery categories, individually discussed below:

- PLL-based topologies
- Delay-locked loop (DLL)-based topologies
- Oversampling topologies
- Phase interpolating (PI) topologies
- Injection locking (IL) topologies
- Gated oscillator (GO) topologies

8.2.1 PLL-Based Topologies

PLL-based clock recovery systems, presented above, have several intrinsic advantages, like exact frequency acquisition, a loop filter memorizing the data frequency over long runs and last but not least, the exhaustive literature and design documentation available on this topology. On the other hand, each PLL is an independent system which cannot easily share components with others, limiting the amount of optimization in multichannel receivers. Although fully integrated DSP-based loop filter topologies have been presented at the transmit side [90], most loop filter topologies rely on passive components of relatively large values. Finally, the presence of multiple oscillators in a common substrate may lead to unwanted *injection locking*, which describes the situation where an oscillator locks to a parasitic signal superimposed to the supply, the substrate bias, a bias current or a reference voltage. This issue may however be alleviated by the usage of silicon-on-insulator technologies in the case of monolithic integration with the proposed photodetector.

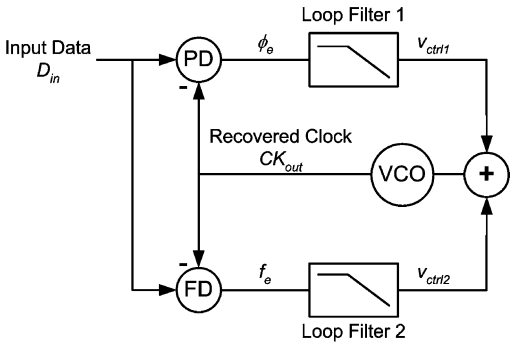

FIGURE 8.10. PLL-based CDR with phase and frequency acquisition loops

Another issue when using PLLs in serial link receivers is frequency acquisition. Due to the non-periodic input data, clock recovery circuits cannot use combined phase-frequency detectors and require an additional loop for frequency acquisition (Figure 8.10). Indeed, the phase detectors able to operate on NRZ data cannot provide frequency acquisition the way conventional phase-frequency detectors do.

The nonlinear operation applied to the input data for generation of spectral content at the data frequency also generates spectral power at different frequencies, depending on the received data pattern. A PLL without frequency acquisition may lock to some harmonic spectral content, leading to erroneous interpretation of the data stream called *harmonic lock* or *false lock* [91].

A frequency acquisition aid may also improve the acquisition speed when out-of-lock. Further more, due to the increasing manufacturing tolerances in modern processes, the free-running frequency of integrated VCOs exhibit considerable die-to-die variations. Again, a frequency acquisition path contributes to the solution of this problem. The design of this additional path however requires some care to guarantee the proper transition between frequency, acquisition and phase-acquisition regimes. Hard switching between the two paths, as suggested in some publications, is usually not recommended because the switching noise may suffice to bring the loop again back out of frequency lock. Frequency detectors with a dead zone are an elegant solution to this problem.

The reader should understand that frequency acquisition is an issue for all closed-loop clock recovery topologies discussed below. Stringent frequency stability requirements on the transmitter side and reduced data run length in 8b/10b encoded data streams however mitigate this issue and frequently allow for omission of this path, especially in receiver schemes based on a very precise reference clock (e.g. phase interpolation, injection locking).

8.2.2 DLL-Based Topologies

A way to circumvent the use of multiple oscillators on a single die is the centralized generation and distribution of a reference clock around the chip. In each receiver channel, this clock can be fed into a voltage/current controlled delay line (VCDL/CCDL) to be aligned with the received data stream (Figure 8.11). In most circuits, such a reference clock is readily available from the transmitter clock source, resulting in significant overhead reduction. It is interesting to graphically present the similarities of the ring oscillator-based PLL and the DLL clock recovery structures (Figure 8.12).

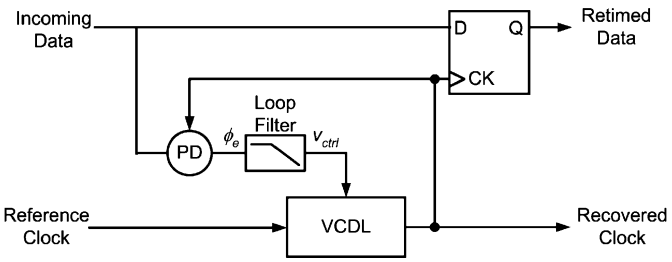

FIGURE 8.11. DLL-based clock recovery and data retiming circuit

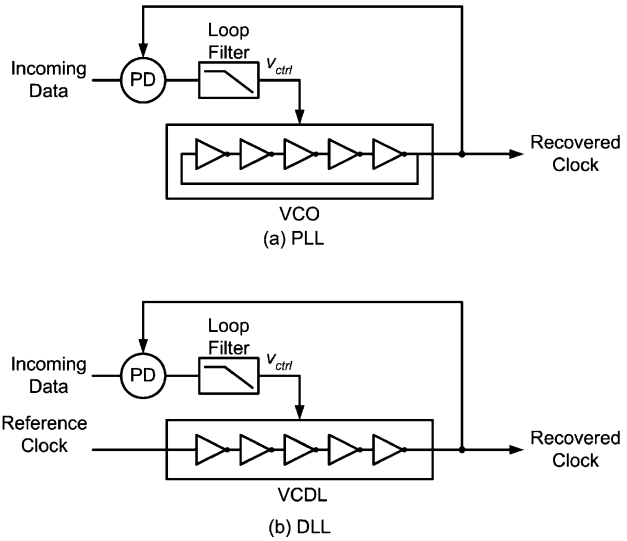

FIGURE 8.12. (a) PLL and (b) DLL clock recovery topologies. (After Yeung [92].)

The phase detector and the loop filter operate the same way than in a PLL structure. The control voltage delivered by the loop filter tunes the delay of the VCDL. While the VCO control voltage acts on the clock frequency, thus on the derivative of the phase, the VCDL control voltage directly acts on the clock phase. This means that the delay line does not introduce a pole in the loop transfer function as the VCO does. Delay-locked loops are thus of the same order than their loop filter, leading to more stable systems. Furthermore, they do not suffer of jitter accumulation due to the signal recirculation in the VCO. However, the delay line does not provide infinite delay tuning capability and for this reason cannot handle even small-frequency offsets between the receiver and the incoming data. Solutions to this problem have been presented. A control voltage correction scheme may compensate for the limited phase-capture range [93], while quadrature mixing is used in a topology somewhere between a DLL and a phase-interpolation scheme in [94]. Another alternative called “dual delay-locked loop” is in fact based on phase interpolation and will be discussed in the corresponding paragraph [95].

8.2.3 Oversampling Receivers

With the increasing speed performance of modern CMOS technologies, the use of oversampling strategies becomes appealing. As shown in Figure 8.13, a multiphase clock, generated either by a ring oscillator in a PLL or by a delay-locked loop, can be used to acquire several samples of the input data at

different instants within a clock cycle. The resulting data word, stored on a parallel array of retiming elements (i.e. flip-flops), is fed to a digital decision logic block, which selects the optimum sample out of the data word [96]. Due to the digital decision algorithm, this feedbackless topology contains a minimum amount of analog blocks. It is however very sensitive to mismatches in and between these blocks, like delay mismatch in the sampling clock paths. This sensitivity becomes critical when supplying multiphase clocks to several of these oversampling receivers using full-chip routing methods. Furthermore, the digital implementation of the so-called *phase picking* algorithm [97] occupies considerable chip area. It can however be expected to shrink conveniently with technology scaling and become appealing in terms of area and power consumption in very deep submicron processes.

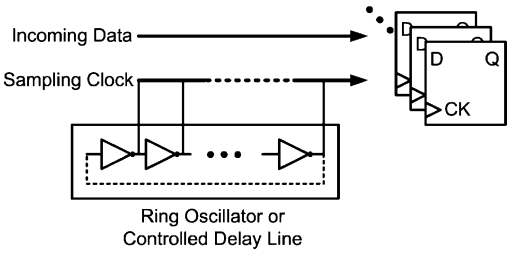

FIGURE 8.13. Multiphase clock generation and signal oversampling

8.2.4 Phase Interpolation

Another way of using a multiphase clock is presented in Figure 8.14, where the best-matching clock phase is selected to minimize the phase error. This *phase selection* strategy is quite close to the *phase alignment* presented in [98] and [99], where the data is sent through a delay line to obtain multiple data phases, followed by a selection multiplexer. As it is preferable to send a periodic signal through the delay line to avoid duty cycle distortion, the phase alignment topology shall be discarded to the benefit of phase selection.

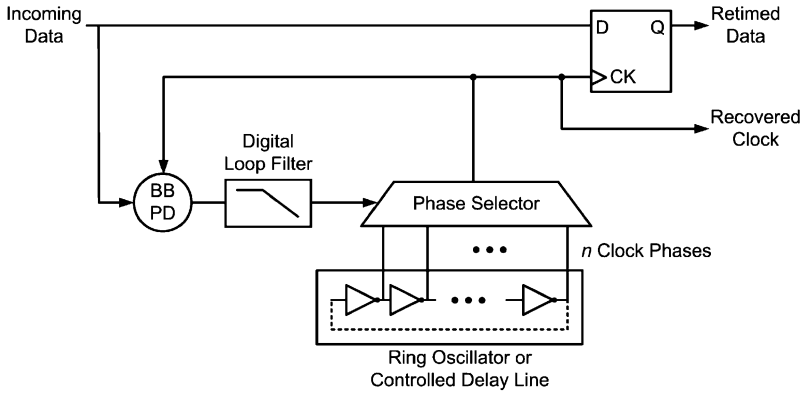

FIGURE 8.14. Phase selection

The granularity of the phase selection can be improved by weighted interpolation between two consecutive clock phases. As shown in Figure 8.15, the clock phase interval corresponding to the arriving data edges is selected first, based on the most significant bits (MSBs) of the loop filter. A finer tuning is achieved by mixing the bracketing phase edges using a weighting method based on the least significant bits (LSBs) of the loop filter [95, 100–102]. This topology can be implemented in a very compact fashion in modern processes using a bang–bang PD with a digital loop filter and a current-steering DAC [103].

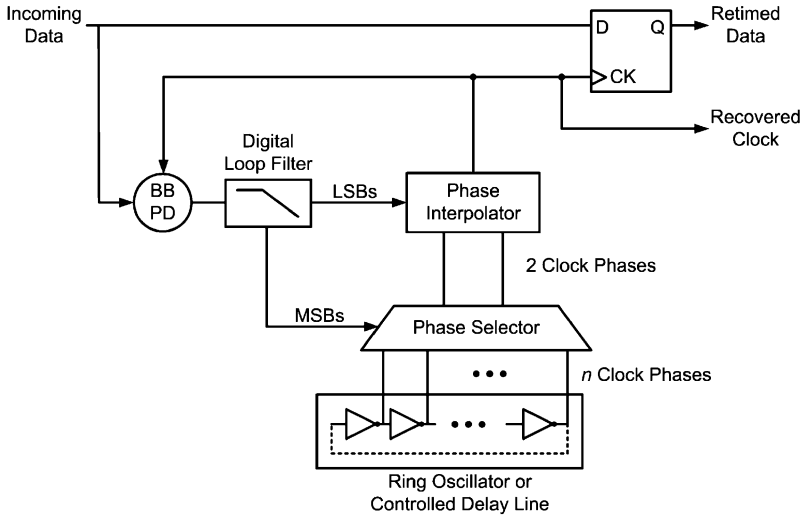

FIGURE 8.15. Phase interpolation structure

As the multiphase clock signal is delivered either by a ring oscillator or by a delay-locked loop and because of the similarities with phase alignment, phase interpolation designs frequently carry different names (e.g. “dual delay-locked loop”, “dual-loop PLL”, etc.). In some designs, quadrature clock signals are used as inputs to the phase interpolator, which in that case is called “phase rotator”. Quadrature phase interpolation requires only two differential clock signals to be routed, minimizing the overhead compared to oversampling techniques and nonetheless eliminating the need for one DLL per channel for phase generation.

8.2.5 Injection Locking

While injection locking has been mentioned earlier as a parasitic phenomenon which may harm the performance of an oscillator-based CDR, it can also be used as a means of synchronization of an oscillator with a stimulus [104–105]. Imagine that, in the phase interpolation structure presented above, we replace the phase interpolator with a ring oscillator (Figure 8.16). In this case, we can inject the weighted clock signals at the output of the phase selector into specific nodes of the ring oscillator. Obviously, the injection mechanism must be strong enough to override any parasitic injection from the supplies, the substrate or other potential noise sources.

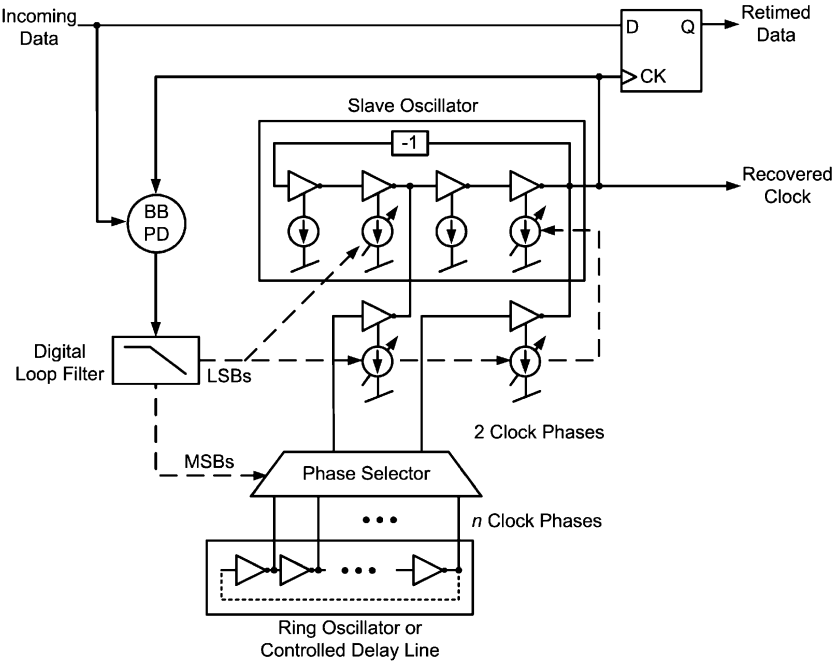

FIGURE 8.16. Injection-locking CDR topology

As the signal injected into the ring oscillator represents only a fraction of the circulating oscillation, the injection-locked oscillator acts as a low-pass filter and somewhat smooths duty-cycle distortion due to the injection mechanism [104]. The generated sampling clock phase thus exhibits much smoother variations compared to the steps of the PI clock. Again, this improvement in tracking jitter trades off with the slave oscillator lock range. Due to the additional ring oscillator, this topology may also result in a small area and power consumption overhead compared to the phase interpolation structure.

8.2.6 Gated Oscillator Topology

The gated oscillator (GO) topology [106] proposes a very different approach to the clock recovery problem. Its open-loop structure does not measure phase error between incoming data and sampling clock in any sense, but simply relies on the resynchronization of the sampling clock at each incoming data edge (Figure 8.17). This structure intrinsically offers a low device count, resulting in a very compact and low-power clock recovery solution. The free-running frequency of each oscillator is determined by the tuning signal delivered by a reference PLL. The possibility of sharing this reference PLL renders this topology even more appealing for multichannel communication links.

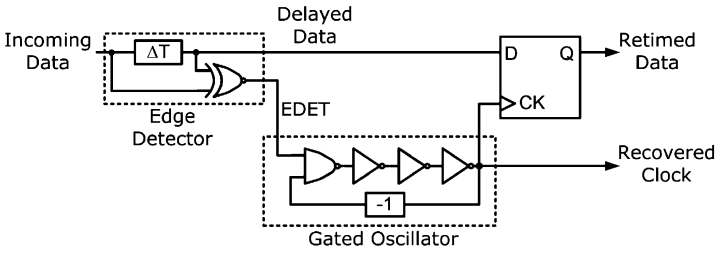

FIGURE 8.17. A single channel gated oscillator CDR

The gated oscillator receiver relies on the arrival time of only one transition, the immediately preceding one, for each bit to be synchronized. As such, it does not provide any jitter rejection mechanism, i.e. its jitter transfer characteristic is equal to one. Neither the immediately following data edge, nor the previous ones, are considered in the selection of the sampling instant. Beyond the oscillator phase information being reset by each arriving data transition, the oscillator runs at its free running frequency in absence of transitions (e.g. during data runs). In addition to its sensitivity to jitter, this structure suffers hence from increased timing errors in presence of frequency offset between the data source and the gated oscillator. However, in most transmission schemes, the incoming data frequency is only known

up to a given tolerance. Potential mismatch between the reference PLL and each local gated oscillator may cause additional frequency offsets to be taken into account in the circuit design.

8.3 Topology Discussion

Table 8.1 presents a summary of the advantages and drawbacks of the various topologies in the scope of clock recovery for short-distance multichannel data links.

TABLE 8.1. CDR Topology Comparison

Topology	Advantages	Drawbacks
PLL-based	Excellent jitter tolerance Tracks input frequency	One oscillator per channel (cross talk) Analog loop filter (area)
DLL-based	Reduced complexity Shared clock source	Complexity for plesiochronous operation Analog loop filter (area) Chip-level routing of clock signal
Oversampling	No feedback loops Little sensitivity to analog components	One DLL or PLL per channel or Chip-level routing of many clock signals Digital circuit complexity
Phase Interpolation (PI)	Good jitter tolerance Shared clock source Intrinsically digital loop filter	Chip-level routing of clock signal(s)
Injection Locking (IL)	Good jitter tolerance Shared clock source Smooth Phase Adjustment Intrinsically digital loop filter	One oscillator per channel (cross talk) Chip-level routing of clock signal(s)
Gated Oscillator (GO)	Low area Potentially low power No chip-level routing of clock signal	Lack of jitter tolerance One oscillator per channel (cross talk) No tracking of input frequency

Focusing on the drawbacks first and considering area, power consumption and straightforward place and route at the chip level as dominant constraints for multichannel receivers, the three latter topologies represent the most promising solutions to the problem. The absence of high-frequency signal routing promote the gated oscillator design to the first place of this selection, but the lack of jitter and frequency tolerance raise major questions regarding its performance. This topology will be used in the following to demonstrate how a systematic top-down design methodology can be applied to analyze the jitter performance of a given CDR topology, as well as guarantee that the final circuit design respects the initial high-level specifications.

8.4 Specifications

Right before going into the detailed analysis of the gated oscillator CDR topology, we will present typical CDR specifications as developed in Chapter 4 (Table 8.2). In oscillator-based multichannel receivers, the different channels may exhibit different free running frequencies due to device mismatch, temperature and supply voltage variations. For this reason, the specified span for tuning range presented below is considerably larger than its initial value determined at the system level. In very short distance processor-to-processor links, the jitter tolerance specifications can be relaxed compared to inter-server data links that are defined for distances up to 300 m.

TABLE 8.2. Clock recovery specifications

Parameter	Acronym	Minimum	Maximum	Unit
Input sensitivity	V_{minCDR}	350		mV_{PP}
Input capacitance	C_{inCDR}		50	fF
Frequency tuning Range	Δf_{TR}	–5%	5%	f_0
Jitter tolerance	$JTOL$	0.46		UI
Total jitter	TJ_{PP}	0.41		UI_{PP}
DJ contribution	DJ_{PP}	0.2		UI_{PP}

8.5 The Gated Oscillator Topology

In order to become a commercially viable alternative to conventional parallel I/O interfaces, multichannel serial receivers must achieve high data rates, low power consumption per channel and low silicon area. The gated oscillator topology is a good choice for such applications. It uses a shared PLL and one current-controlled oscillator (CCO) per channel (Figure 8.18). The advantages of this topology in multichannel applications have already been presented, but before addressing the low-level design, it must be shown that the achievable jitter tolerance and tolerable frequency offset meet the system specification. First, the circuit topology and operation will be described, then a detailed top-down timing analysis will be presented in Section 8.6.

8.5.1 The Multichannel Gated Oscillator CDR

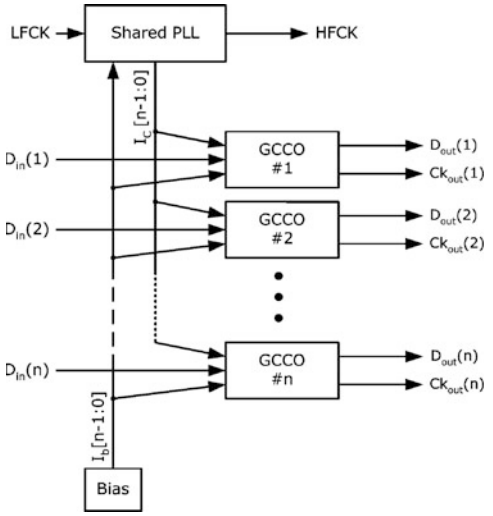

FIGURE 8.18. Multichannel GCCO CDR

As shown in Figure 8.18, the shared PLL generates a high-frequency clock *HFCK* at 2.5 GHz by frequency multiplication by 16 of a very precise low-frequency 156.25 MHz reference clock *LFCK*. It is based on a current-controlled oscillator and the control current of this oscillator is mirrored. One sample of the control current $I_c[n\text{-}1:0]$ is delivered to the CCO in each receiver channel, tuning them to a free-running frequency close to 2.5 GHz. The precision of the frequency tuning is only limited by the oscillator matching and the matching of the control currents. Current-controlled oscillators are preferred to their voltage-controlled counterparts due to better matching. Additional bias currents $I_B[n\text{-}1:0]$ are distributed by a central bias generator for proper biasing of logic gates and buffers.

8.5.2 The Clock Recovery Core

The clock recovery circuit for a single channel is composed of three majors parts (Figure 8.19): the edge detection, the gated oscillator core and a sampling element. For simplicity, the schematic is drawn in a single-ended fashion, whereas the hardware implementation uses fully differential current-mode

logic gates. The "-1" sign in the feedback loop represents the inversion operated on the fully differential signal through wire crossing.

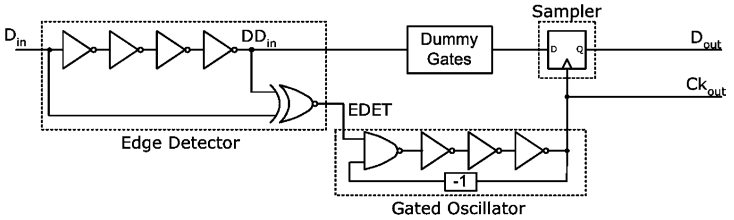

FIGURE 8.19. Gated oscillator schematic with edge detection

The edge detection is performed by an XOR gate comparing the incoming data D_{in} with a delayed copy of the same signal called DD_{in}. At each incoming data edge, a falling pulse is generated on the active low *EDET* signal. This pulse triggers the synchronization operation in the gated oscillator. The duration of the *EDET* pulse is determined by the delay line in the edge detector as well as potential delay mismatch between both inputs in the XOR gate. As we will see later, the *EDET* pulses shall not be shorter than half a clock period, in which case the XOR delay mismatch should be negligible. To allow for delay matching with the oscillator delay stages, the edge detector delay line is built using current-mode logic NAND gates.

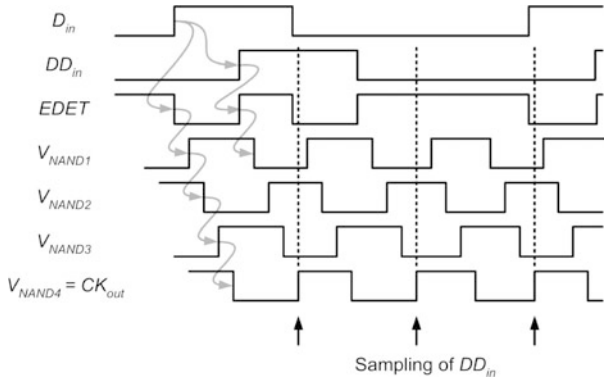

FIGURE 8.20. Gated oscillator timing diagram

The ring oscillator in the GO core is built using four identical NAND gates. The three latter stages have the unused input nodes connected to appropriate bias voltages. The use of identical gates results in better control of the delays compared to a mixed NAND-inverter combination, as well as good matching with the NAND gates used in the edge detector delay line. When the gated oscillator receives the

falling *EDET* signal, the output of the first NAND gate goes high (Figure 8.20). As a property of the ring oscillator, the propagation delay of each gate is equal to $T_{CK}/2n$, where n is the number of stages in the ring oscillator. As long as EDET stays low, this NAND gate inhibits propagation of any data transitions received by the feedback loop from the oscillator output. Once D_{in} propagated through the edge detector delay line to DD_{in}, the rising edge of *EDET* releases the first oscillator NAND gate, which goes back to its inverting operation with respect to the feedback signal.

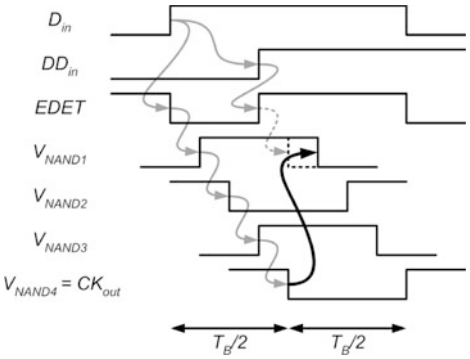

FIGURE 8.21. Race condition due to short *EDET* signal

If the *EDET* signal rises before the feedback signal has risen to "1" under the influence of the same *EDET* edge, an edge will not be generated at the output of the first NAND gate in all situations. If such an edge is not generated, an output clock edge will not occur half a clock cycle later and the data will not be sampled correctly. As this situation can be considered to correspond to a late arrival of the feedback signal, we say the circuit is subject to a *race condition* (Figure 8.21).As the propagation delay from the *EDET* input to the clock output and thus to the feedback input is half a clock cycle, the *EDET* pulse duration must in any case exceed this value. The reader should however also be aware that longer pulse durations result in degraded jitter tolerance. Indeed, in case a second data edge arrives early, a very short EDET rising pulse may be generated (Figure 8.22). The resulting clock pulse may be that

small that it vanishes in the oscillator or results in hold-time violations in the sampling flip-flop. Either case would result in a missing sampling instant, i.e. a dropped bit.

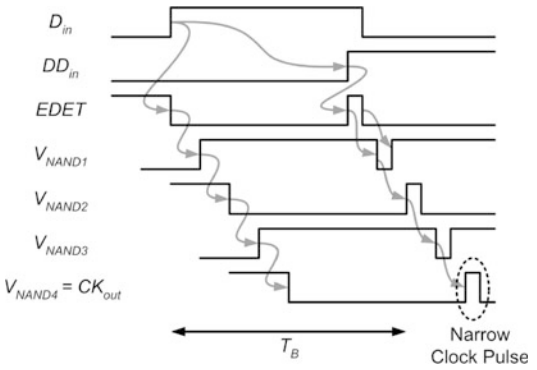

FIGURE 8.22. Race condition due to short *EDET* signal

When released, the oscillator resumes operation at its free-running frequency set by the control current delivered by the shared PLL. As the rising *EDET* edge takes half a clock cycle to propagate to the oscillator output, the delayed data is sampled in the sampling flip-flop (FF) with this same delay with respect to the data edge. Parasitic delays in the *EDET* path are compensated for using dummy gates in the data path. It is important to understand that, in presence of jitter, sampling half a clock period after the data edge does not always correspond to sampling in the middle of the eye, as performed by most conventional CDR topologies (Figure 8.23).

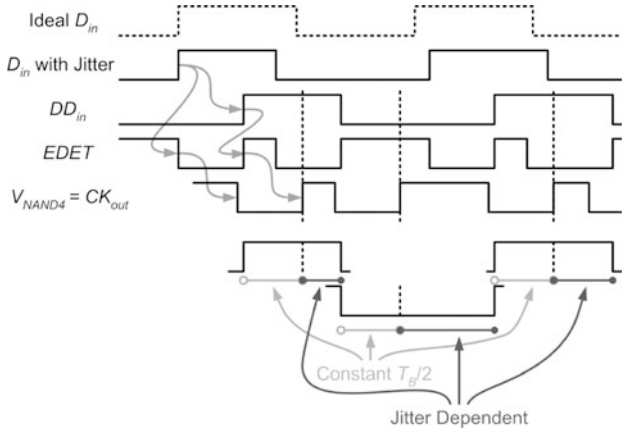

FIGURE 8.23. Sampling after $T_{CK}/2$

The alignment of the sampling instant on each data edge results in unity jitter transfer, which does not only affect the periodicity of the recovered clock signal, but also enhances the jitter seen by the sampling FF on the following data edge. Imagine an incoming edge being late due to jitter, shifting the generated clock edge by the same amount (bottom part of Figure 8.23). If the next data edge is early, it will not only be shifted by its own jitter to the center of the eye (defined by the clock edge), but by the sum of both jitter contributions. In consequence, all jitter contributions on both data edges are appearing in the eye diagram as being applied to the second data edge.

Finally, the sampled data D_{out} and the recovered clock CK_{out} are delivered to the output. As this topology does not have any long-term memory with respect to data rate and low-frequency jitter, it is fairly intuitive that in presence of frequency offsets, frequent data edges must be received to avoid excessive drift of the sampling instant from the desired location (Figure 8.24). This issue is accentuated by the fact that device mismatch and supply voltage variations between the CDR oscillator and the shared PLL oscillator are enhanced by physical distance on the silicon die.

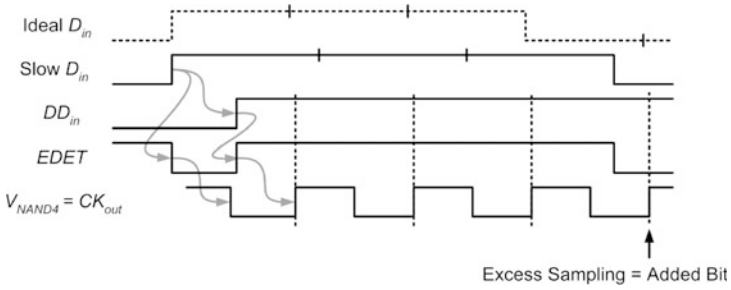

FIGURE 8.24. Drift of sampling point in presence of frequency offsets (exaggerated)

8.5.3 CDR Top-Down Design Methodology

While gated oscillator structures have been used in the past in burst-mode communications, jitter and frequency tolerance requirements in the present application are more stringent. For the above-mentioned reasons, a thorough analysis of jitter tolerance (JTOL) and frequency tolerance (FTOL) performance of the gated oscillator topology is mandatory before addressing the transistor-level design. In order to propagate and refine the bit error ratio estimations at the different abstraction levels, from the system level down to the device level, we introduce a specification driven top-down design methodology for such gated oscillator clock recovery circuits (Figure 8.25). After the selection of the clock recovery topology, statistical Matlab-based simulation is used to get an estimate of jitter and frequency tolerance. A time-domain topology-based behavioral model is developed using a hardware description language (HDL) that allows in-depth sensitivity analysis of various topology parameters, like gate

delays, thermal noise and voltage levels. If the simulation results of the time-domain analysis are in agreement with the statistical results, we can step over to the transistor-level design. The HDL model includes most block-level parameters of the device-level design. Particularly the acceptable oscillator jitter is first determined at the statistical level, then verified at the behavioral level. The bias conditions of the ring oscillator are obtained from the obtained value using Hajimiri's theory [108], as discussed in Section 8.8. Finally, the classical analog flow through layout, parasitic extraction and post-layout simulation phases, guarantees the verification of the design before submission for manufacturing. The presented top-down design methodology may of course be adapted to other CDR structures as well.

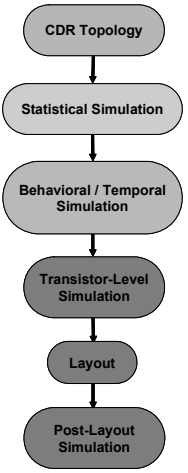

FIGURE 8.25. Specification driven top-down design methodology

8.6 Statistical Modeling of the Gated Oscillator

The highest level of abstraction in this flow models jitter components based on their statistical distributions. Jitter tolerance and tolerance of frequency offsets are estimated based on the resulting bit error ratio through purely statistical analysis, using Matlab. This model does not take into account implementation details, but delivers good insight on the jitter and frequency tolerances in limited simulation time.

8.6.1 Jitter Statistics

Before discussion of the statistical model, we describe the jitter statistics of the different jitter components applied to the receiver (Table 8.3).

TABLE 8.3. Jitter specifications for statistical modeling

Jitter Type	Acronym	Units	Value	PDF
Random jitter	*RJ*	UI_{RMS}	0.015(1)	Gaussian
Deterministic jitter	*DJ*	UI_{PP}	0.2	Dual-Dirac
Sinusoidal jitter	*SJ*	UI_{PP}	swept(2)	Sinusoidal
GCCO Clock jitter	*CKJ*	UI_{RMS}	TBD(3)	Gaussian

(1) The RMS value is obtained from a total jitter (TJ) specification of 0.41 UI_{RMS}, resulting in $RJ_{PP} = TJ_{PP} - DJ_{PP} = 0.21$ UI_{PP}, then divided by the PP to RMS conversion factor for a BER of 10^{-12}.

(2) Sinusoidal jitter amplitude and frequencies are swept to obtain the jitter tolerance graph to be compared with Figure 4b.

(3) To be defined. An acceptable upper bound for the clock jitter is to be obtained from this analysis.

The basic concepts of statistical jitter analysis, as discussed in [107], will be briefly introduced. In this reference, the bit error ratio is defined by the integral of the error probability over the range of possible sampling instants t_s.

$$BER = \int_{-\infty}^{\infty} P[\varepsilon, t_s] \qquad \text{(EQ 8.12)}$$

The bit error probability over the full range of t_s itself depends on the probability of error at each possible sampling instant $P[\varepsilon|t_s]$ and the probability of sampling at this instant $P[t_s]$.

$$P[\varepsilon, t_s] = P[\varepsilon|t_s] \cdot P[t_s] \qquad \text{(EQ 8.13)}$$

In this approach, channel jitter contributions are assigned to the data edges, while jitter sources in the CDR oscillator are associated with the sampling instant. This calculation is suitable for conventional clock recovery circuits, which use the loop filter memory to average out jitter and track the average data edge location in time. It is not suitable, however, for the analysis of the gated oscillator CDR, where the oscillator phase is realigned at each incoming data edge. As a result of this retiming strategy, channel and receiver jitter, as well as frequency offsets, accumulate differently on both edges of the eye diagram. In the following, we will propose an alternative model, which considering an ideal sampling clock and assigns all jitter contributions to the data edges. This approach results in more accurate modeling of the gated oscillator CDR performance.

First, the jitter sources and offsets corresponding to both data edges in the eye diagram will be calculated. The total jitter probability density function, calculated separately for the first (i.e. left) and second (i.e. right) data edge, is obtained by convolution of the different jitter distributions. The reader should understand that in this development, the term *jitter* is used in its largest sense, i.e. it also includes gate delays and frequency offsets. The sources of jitter can be easily found by analyzing the data and clock paths in the schematic, starting from the *EDET* generation at the output of the edge detector and closing the loop at the sampler.

Possible static delays between data and clock can be due to mismatch between active gates and dummy gates (e.g. data-clock delay `d2ck_del`), as well as intentionally introduced delays obtained by tapping a different output of the ring oscillator (`delta_samp_point`). Frequency offset of the oscillator is accumulated as long as no new data edge arrives. This duration is determined by the number of consecutive identical bits `CID` (Figure 8.26).

```
edge1_off = delta_samp_point + d2ck_del + delta_f_rel*0.5;
edge2_off = delta_samp_point + d2ck_del + delta_f_rel*(CID-0.5);
```

FIGURE 8.26. Modeling of timing offset due to static delays and frequency offset

8.6.2 Random Jitter Components

Oscillator jitter accumulates over a time period ΔT according to Equation 8.14. Unlike the free-running oscillator modeled in [108], the gated oscillator is regularly reset. We can consider the case of short-term accumulation of jitter only and neglect correlated noise sources like up-converted 1/f noise.

$$\sigma_{CKJ} = \kappa\sqrt{\Delta T} \qquad \text{(EQ 8.14)}$$

In this relationship, κ is a technology and design dependent proportionality constant used in the calculation of the oscillator bias currents. Knowing the maximum number of CID defining ΔT, we can calculate κ to achieve a given phase noise, respectively a given oscillator jitter performance. The present bit error ratio estimation determines the maximum tolerable oscillator jitter to meet the system specifications. A shown in Section 8.8, this approach allows us to obtain design parameter values for minimum system power, which is one of the main constraints of the application. Random data jitter, due to amplitude noise to jitter conversion in the receiver's amplification stages, is not subject to any recirculating mechanism the way oscillator is and does thus not exhibit comparable accumulation properties. Random jitter sources in the edge detection delay line of the CDR are supposed negligible.

```
edge1_sigma = gvco_k * sqrt(0.5);
edge2_sigma = gvco_k * sqrt(CID-0.5) + sqrt(2) * RJ;
```

FIGURE 8.27. Oscillator jitter accumulation model

As the jitter affecting the first data edge is transferred to the clock edge, the effective random jitter of the second data edge with respect to the clock edge is weighted by a factor √2 (Figure 8.27). This model does not take into account the fact that random jitter on either data edge is correlated, as shown by the nonzero correlation coefficient of Gaussian white noise of limited bandwidth BW (Equation 8.15).

$$\rho_x(\Delta t) = \mathrm{sinc}(2BW\Delta t) \quad \textbf{(EQ 8.15)}$$

Indeed, for a single-period time interval between two data edges and a bandwidth of $0.75 f_B$, the correlation is 66.5%. Inclusion of this reduction in effective random jitter seen by the CDR in the model is left to the interested reader.

8.6.3 Deterministic Jitter Components

As for random jitter, the deterministic jitter components on an incoming data edge are transferred to the clock. The left data edge is thus exempt of deterministic jitter when being sampled. However, deterministic jitter on the second data edge is doubled, because DJ components on consecutive data edges are considered uncorrelated. The deterministic jitter probability density function (PDF) is modeled with a so-called dual-Dirac distribution, as suggested in [21]. A detailed discussion of the dual-Dirac model and how it is used to estimate the BER from statistical measurement results is presented in [109].

The convolution of the dual-Dirac distribution with a conventional random jitter PDF leads to the results Figure 8.28. The reader can easily understand the split of the RJ Gaussian PDF at each data edge into a sum of two Gaussian PDFs offset by the DJ Dirac functions (Equation 8.16).

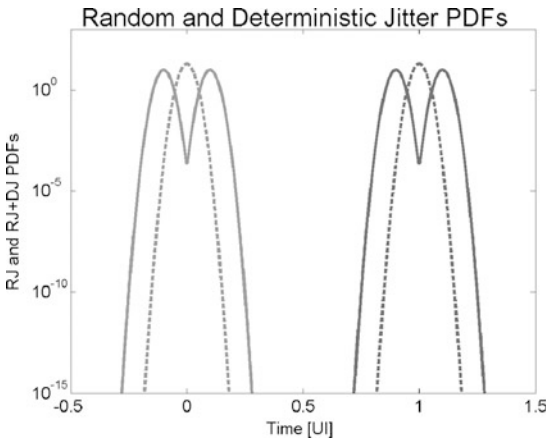

FIGURE 8.28. RJ and combined RJ–DJ probability density functions

$$p_{RJDJ}(t) = 0.5 \cdot \frac{1}{RJ_{RMS}\sqrt{2\pi}} \cdot e^{-\frac{\left(t-\frac{DJ}{2}\right)^2}{2RJ_{RMS}^2}} + 0.5 \cdot \frac{1}{RJ_{RMS}\sqrt{2\pi}} \cdot e^{-\frac{\left(t+\frac{DJ}{2}\right)^2}{2RJ_{RMS}^2}} \quad \text{(EQ 8.16)}$$

This equation, written for the PDF of a single data edge, also holds for the particular case of the gated oscillator, where different RJ and DJ values are applied. The implementation code is shown in Figure 8.29. As for random jitter, the second edge jitter is affected by deterministic jitter doubling due to the jitter transfer of the first edge jitter to the clock edge.

```
pdf_rjdj_edge1=normpdf(x, edge1_off, edge1_sigma);
pdf_rjdj_edge2=0.5*normpdf(x, 1.0-2*DJ/2+edge2_off, edge2_sigma)
              +0.5*normpdf(x, 1.0+2*DJ/2+edge2_off, edge2_sigma);
```

FIGURE 8.29. Combined RJ–DJ PDF Model

The resulting RJ–DJ PDF affecting the second data edge in the case of a gated oscillator CDR is illustrated in Figure 8.30. One can observe the broadening of the probability density function compared to the Figure 8.28.

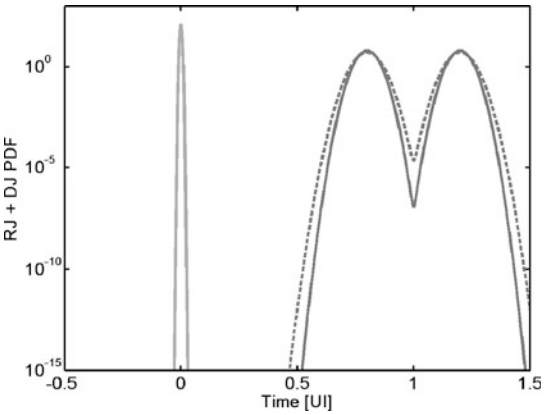

FIGURE 8.30. Combined RJ–DJ PDF in a GO CDR for CID = 1 (solid) and CID = 5 (dashed)

8.6.4 Sinusoidal Jitter Model

For sinusoidal jitter, the fundamental goal is the consideration of the variation of the sinusoidal modulation between two data edges, which will be called *differential sinusoidal jitter* (*dSJ*).

Let us consider an input signal *ck(t)* subject to the sinusoidal phase modulation $\varphi_{SJ}(t)$. For simplicity of the developments, we consider in the following a modulated periodic (sinusoidal) signal instead

of a random data signal. The results in terms of SJ contributions however remain valid for any data sequence with known maximum run length. Equation 8.17 shows the time-domain description of a sinusoidal clock signal of peak amplitude A_{ck} and frequency f_{ck} modulated by a sinusoidal jitter component $\varphi_{SJ}(t)$ of peak amplitude A_{SJ} and frequency f_{SJ}. As the clock recovery circuit operates on the data transitions only, the assumption of the modulated signal being sinusoidal does not limit the generality of the development.

$$\begin{aligned} ck(t) &= A_{CK} \cdot \sin[2\pi \cdot f_{ck} \cdot t + \varphi_{SJ}(t)] \\ \varphi_{SJ}(t) &= A_{SJ} \sin(2\pi \cdot f_{SJ} \cdot t) \end{aligned} \qquad \text{(EQ 8.17)}$$

The probability density function of a sinusoidal waveform of peak amplitude A_{SJ} is given by:

$$p_{SJ}(x) = \begin{cases} \dfrac{1}{\pi \cdot \sqrt{A_{SJ}^2 - x^2}} & \text{if} \quad (|x| < A_{SJ}) \\ 0 & \text{if} \quad (|x| > A_{SJ}) \end{cases} \qquad \text{(EQ 8.18)}$$

Based on the fact that the relationship between the sampling instant and the jitter waveform affecting consecutive samples is known, sinusoidal jitter cannot be considered to be a stochastic process. In this paragraph, the probability density function of the differential sinusoidal jitter from a time instant t_1 to a time instant $t_2 = t_1 + \Delta t$ will be developed. Figure 8.31 shows how the amount of differential jitter depends not only on the time interval Δt, but also on the initial time instant t_1.

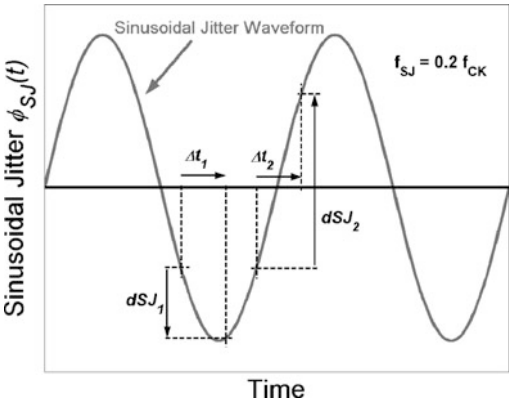

FIGURE 8.31. Influence of initial time and time interval on the differential sinusoidal jitter

The general approach to the problem will first develop the independent sinusoidal jitter PDF at a randomly chosen time t_1, then the conditional probability of the differential sinusoidal jitter, which depends on the initial position on the jitter sine wave at instant t_1. Finally, the time-independent differential SJ PDF will be obtained from these two distributions.

Let us first consider an arbitrarily chosen rising clock edge at time t_1, the PDF of the phase modulation due to sinusoidal jitter being given by Equation 8.18. The difference in SJ phase modulation between instant t_1 and arbitrarily chosen time instant $t_2 = t_1 + \Delta t$ is defined by:

$$\Delta\varphi_{SJ} = \varphi_{SJ}(t_2) - \varphi_{SJ}(t_1) = A_{SJ} \cdot [\cos(2\pi \cdot f_{ck} \cdot t_1) \cdot \sin(2\pi \cdot f_{SJ} \cdot \Delta t) - \sin(2\pi \cdot f_{ck} \cdot t_1) \cdot (1 - \cos(2\pi \cdot f_{SJ} \cdot \Delta t))] \quad \text{(EQ 8.19)}$$

This phase difference $\Delta\varphi$ represents the difference in phase modulation of the clock signal. If we take t_2 corresponding to another rising clock edge, Δt can be written as a function of the clock period T_{ck} and the number of clock periods n_{ck} between the two edges (Equation 8.20),

$$\Delta t = n_{ck} \cdot T_{ck} - (\varphi_{jit}(t_2) - \varphi_{jit}(t_1)) \quad \text{(EQ 8.20)}$$

where φ_{jit} represents the phase modulation due to all jitter sources. In order to solve this recursive relationship, combined with the dependency on other a priori unknown jitter contributions, we only consider the average value of Δt given by Equation 8.21.

$$\Delta t \approx CID \cdot T_{ck} = \frac{CID}{f_{ck}} \quad \text{(EQ 8.21)}$$

Knowing the PDF of the SJ phase at instant t_1 (Equation 8.18), as well as the probability $P[\Delta\varphi_{SJ}|\varphi_{SJ}(t_1)]$ of an SJ phase shift of $\Delta\varphi_{SJ}$ is known for each $\varphi_{SJ}(t)$, we can calculate the PDF associated with *dSJ*. The corresponding Matlab code, which calculates the probability density function of the differential sinusoidal jitter as a function of the SJ amplitude, normalized frequency and the number of

```
if abs(sin(TWOPI_fmod_t(i))) < 1
  prob_phi_t1_ft(i) = 0.5 /
    (pi*sqrt((sj_amp_pp/2)^2-(sj_amp_pp/2*sin(TWOPI_fmod_t(i)))^2));
  % 0.5 because the total probability is related to two points on the
  % single-period sine wave
else if abs(sin(TWOPI_fmod_t(i))) == 1
  % provides correct calculation of the boundary probabilities
  prob_phi_t1_ft(i) = 1 / (pi*sqrt((sj_amp_pp/2)^2
    -(sj_amp_pp/2*sin(TWOPI_fmod_t(i)-2*pi*fmod_t_step))^2));;
  else
    prob_phi_t1_ft(i) = 0;
  end
end
```

FIGURE 8.32. Calculation of the basic sinusoidal jitter PDF

CID, is shown below and explained step-by-step (Figures 8.32–8.34). The index `i` is the vector index used in the calculations to define the number of bins (or samples) used to represent the probability density functions.

According to Equation 8.18, the first part calculates the sinusoidal jitter PDF of $P[\varphi_{SJ}(t_1)]$ (represented by `prob_phi_t1_ft`) as a function of the instantaneous SJ phase (represented by `TWOPI_fmod_t`, corresponding to $2\pi f_{mod}\, t$), thus as a function of time. In this calculation, the boundary conditions require some particular attention and a factor 0.5 is introduced to take into account that each probability level in Equation 8.18 corresponds to two points on the sine wave in the time domain, which is considered here.

The next part calculates the phase $\varphi(t)$ for each instant t, as well as the associated-phase increment $\Delta\varphi$ (corresponding to `phi_t2_phi_t1`) as a function of the data rate `Tck`, the number of CID and the instantaneous SJ phase. This value is further scaled and rounded (`phi_t2_phi_t1_roundscale`), because it is later on used as a vector index of the final probability density function.

```
phi_t1(i) = sj_amp_pp/2 * sin(TWOPI_fmod_t(i));
phi_t2_phi_t1(i) = delta_phi_offs +
   sj_amp_pp/2 * (cos(TWOPI_fmod_t(i))*sin(TWOPI_fmod_Tck_CID)
   - sin(TWOPI_fmod_t(i))*(1-cos(TWOPI_fmod_Tck_CID)));
phi_t2_phi_t1_scale(i) = phi_t2_phi_t1(i) / prob_phi_t2_step;
phi_t2_phi_t1_roundscale(i) = round(phi_t2_phi_t1_scale(i));
```

FIGURE 8.33. Calculation of the initial SJ phase and the SJ phase increment

Up to this point, the probability of obtaining $\Delta\varphi_{SJ}$ is represented as a function of time. In order to calculate the conditional probability $P[\Delta\varphi_{SJ}|\varphi_{SJ}(t_1)]$, we need the probability to be expressed as a function of the phase. For this purpose, we use the scaled and rounded value of the differential phase `phi_t2_phi_t1_roundscale` as an index for the probability of the differential phase, called `prob_delta_phi_temp`. The '`temp`' extension is due to the fact that the result will be subject to some additional conditioning before revealing the true differential sinusoidal jitter PDF. Incremental addition is needed as several initial phase values $\varphi_{SJ}(t)$ combined with their associated $\Delta\varphi_{SJ}$ may be mapped to a single bin of the differential phase histogram.

```
x_delta_phi_temp(phi_t2_phi_t1_roundscale(i)) =
   phi_t2_phi_t1(i) - delta_phi_offs;
prob_delta_phi_temp(phi_t2_phi_t1_roundscale(i)) =
prob_delta_phi_temp(phi_t2_phi_t1_roundscale(i)) + prob_phi_t1_ft(i);
```

FIGURE 8.34. Calculation of the differential SJ PDF

Zero values in some histogram bins are due to the limited number of samples used for the calculations. In order to obtain a smooth plot, these bins are removed, which does not alter the integral of the histogram. Then, the histogram is smoothed by averaging with the two neighboring bins. These operations are not required from a mathematical point of view, but results in a more explicit probability density plot.

Figure 8.35 shows the resulting probability density functions for a given set of jitter amplitudes and frequencies. While jitter amplitude SJ_{PP} determines the maximum possible extent of the differential sinusoidal jitter distribution, the differential SJ amplitude and the location of its peaks depend on the normalized jitter frequency and the number of consecutive identical bits.

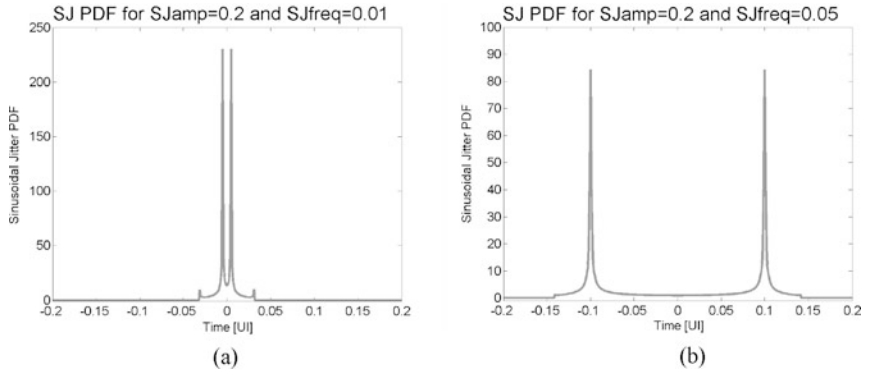

FIGURE 8.35. SJ probability density functions for CID = 5 and (a) SJ_{amp} = 0.2 UI_{PP} and SJ_{freq} = 0.01 f_B, (b) SJ_{amp} = 0.2 UI_{PP} and SJ_{freq} = 0.05 f_B

8.6.5 Bathtub Curves and BER Estimation

Based on the probability density functions for the different jitter contributions, we can calculate the the total jitter PDF for both data edges by convolution of the RJ–DJ and the SJ PDFs. The cumulative distribution functions (CDF) for both data edges, resulting from integration of the corresponding PDFs (Equation 8.23), are shown in Figure 8.36. In the case of gated oscillators, the first data edge

exclusively suffers from oscillator jitter, while the second jitter distribution is obtained by convolution of the RJ–DJ PDF with the sinusoidal jitter previously described.

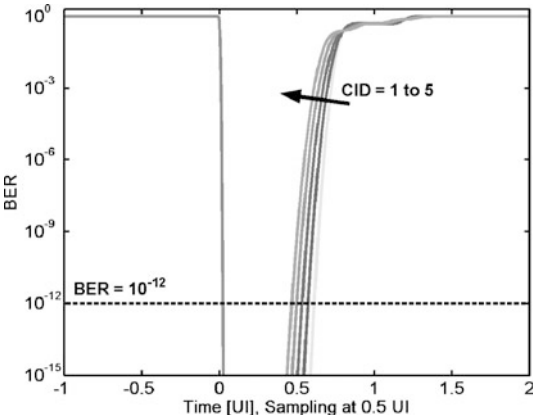

FIGURE 8.36. Bathtub curves for SJ_{amp} = 0.2 UI_{PP}, SJ_{freq} = 0.05 f_B, RJ = 0.015 UI_{RMS}, DJ = 0.2 UI_{PP}, $FTOL$ = 0 and CID = 1...5 (not weighted)

The amount of random jitter to be applied is derived from the total peak–peak jitter specification, then converted to an RMS value according to Equation 8.22.

$$RJ_{RMS} = \frac{TJ_{PP} - DJ_{PP}}{k_\sigma} \quad \text{(EQ 8.22)}$$

$$CDF_{edge1} = \int_t^\infty PDF_{edge1}\,dt$$
$$CDF_{edge2} = \int_{-\infty}^t PDF_{edge2}\,dt \quad \text{(EQ 8.23)}$$

The bit error ratio can be calculated (Equation 8.24) from the sum of both CDFs at the sampling point, multiplied by the transition probability (generally considered to be 0.5). Summation of the CDFs is equivalent to integration of the area under the PDF curves. In order to consider data edges beyond the previous or next sampling instants (–0.5 UI and 1.5 UI) as full errors, i.e. without weighting with the transition density, the bit error ratio is augmented by the sum of integrals of these intervals.

$$BER = p_{trans} \cdot (CDF_{edge1} + CDF_{edge2}) \quad \text{(EQ 8.24)}$$

Estimating the bit error ratio for the maximum number of CID overestimates the number of errors, as it does not take into consideration the fact that only few data transitions in an encoded data pattern

are subject to these conditions. In the presented model, the total BER is calculated as a weighted sum of the bit error ratios corresponding to n CID, where n is swept from 1 to CID_{max} = 5. The weights are determined by the probability of occurrence of n CID in an encoded random data stream. Considering a transition probability of 0.5, the probability of n CID is given by $p(\text{CID} = n) = 2^{-n}$. As 8 b/10 b encoding does not allow for n exceeding 5, each probability is the divided by $\sum_{n=1}^{5} p(\text{CID} = n)$ to guarantee that the sum of all probabilities equals 1. The calculated bit error ratio for given sinusoidal jitter amplitude-frequency pairs is shown in Figure 8.37. According to the jitter tolerance mask plotted at SJ_{PP} = 0.1 UI, the specified JTOL can tightly be achieved at a BER=10^{-12}, keeping in mind the relaxation of the deterministic jitter specifications with respect to the InfiniBand standard initially considered.

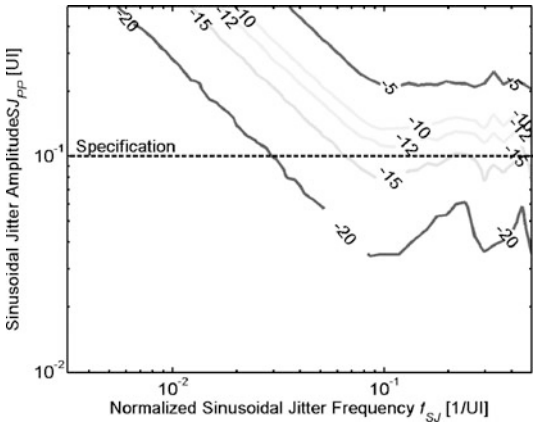

FIGURE 8.37. BER estimation for *RJ* = 0.015 UI_{RMS}, *DJ* = 0.2 UI_{PP}, FTOL = 0

Figure 8.38 shows the BER estimation in presence of a 1% frequency offset. The result highlights the sensitivity to frequency offsets between the transmitter and the receiver, as the specified bit error ratio cannot be met within all cases of the jitter tolerance mask. The bumps in the BER curves are due

to the dependency of differential sinusoidal jitter on the normalized jitter frequency (i.e. the ratio between the SJ frequency and the data rate), as shown in the SJ analysis.

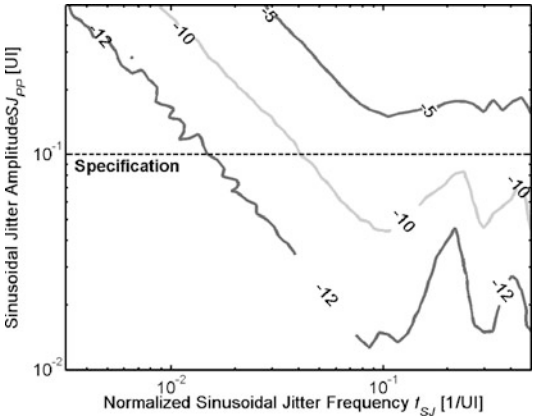

FIGURE 8.38. BER Estimation for *RJ* = 0.015 UI_{RMS}, *DJ* = 0.2 UI_{PP}, *FTOL* = 1%

In conclusion, the statistical modeling reveals that the simplicity and compactness of the gated oscillator topology result in limited tolerance to frequency mismatch and jitter. This topology cannot achieve sufficient jitter tolerance to apply for long-haul clock recovery or inter-server interconnects of 10–300 m length. However, in the box, where jitter is limited due to the reduced communications distances, this clock recovery topology can fully exploit its benefits: small silicon, low power and optimal scalability due to the use of digital building blocks only. Based on the presented analysis, the acceptable RMS clock jitter *CKJ* is specified as $0.01UI_{RMS}$ at very low-frequency offsets, allowing for a good power consumption/bit rate efficiency.

8.7 Time-Domain Modeling

The pattern generator code presented in this section has been previously released in [110].

8.7.1 CDR Model

The statistical model previously discussed gives a good estimate of the achievable performance, but it cannot verify the time-domain behavior of the selected topology. This verification is however necessary to analyze the influence of gate delays, various jitter sources, mismatch effects and also amplitude restoring mechanisms on jitter and frequency tolerance. Behavioral modeling using a hardware description language (HDL) is a convenient approach to this problem.

As the clock recovery circuit has an event-driven behavior and operates on two-level (binary) signals, the CDR time-domain model is based on a pure VHDL description. VHDL-AMS components are used in the test bench to generate the input stimuli and allow for proper representation of the output waveforms in the form of eye diagrams. The behavioral model of the clock recovery circuit is very close to the actual implementation shown in Figure 8.19 and contains all important gate delays and jitter contributions to be considered.

```
entity cdr_gcco is
  generic (
  cdr_gcco_k: real;-- CCO gain [Hz/A]
  cdr_gcco_fc: real;-- Free-running frequency [Hz]
  cdr_gcco_cc0: voltage;-- Control current mid-point [C]
  cdr_gcco_jit_sigma: real); -- defined as a ratio, e.g 1%->0.01
  port (
    […]);
end entity cdr_gcco;

architecture bhv of cdr_gcco is
  […]
begin
  calc_delay0: process
  begin
    awgn(seed1, seed2, mean, sigma, jitter);-- gaussian random gen.
    delay0 <= 1 ps * 1.0e12/ (8.0*(cdr_gcco_fc+cdr_gcco_k*(cctrl-
cdr_gcco_cc0))) * (1.0+jitter);
    wait for delay0;
  end process calc_delay0;

  […]    -- calculation of the three remaining delays

  (0) <= transport (vinv4(0) and cdr_gcco_trig(0)) and (cdr_gvco_enable
and cdr_gcco_nreset) after delay0;
  vinv1(1) <= transport (vinv4(0) nand cdr_gcco_trig(0)) and
(cdr_gvco_enable and cdr_gcco_nreset) after delay0;
  vinv2 <= transport not(vinv1) after delay1;
  vinv3 <= transport not(vinv2) after delay2;
  vinv4 <= transport not(vinv3) after delay3;
  cdr_gcco_ckout <= not(vinv4);
end bhv;
```

FIGURE 8.39. VHDL code of gated CCO

As shown in Figure 8.39, the delay of each gate in the ring oscillator is adjusted with respect to the control current `cctrl` coming from the shared PLL (the nominal value being `cdr_gcco_cc0`) and a jitter amplitude which is computed as a random white noise source following a Gaussian distribution [111, 112]. The output of each stage (`vinv1-vinv4`) is affected with the output of the preceding stage after the delay calculated above. In order to avoid suppression of events in the event stack under variable delay conditions, **`transport`** statements are required. In the current implementation, the

pseudo-random number generators used by the different delay cells use different seed values, but their sufficiency to achieve complete statistical decorrelation has not been further analyzed.

8.7.2 Input Data Source with Extended Jitter Model

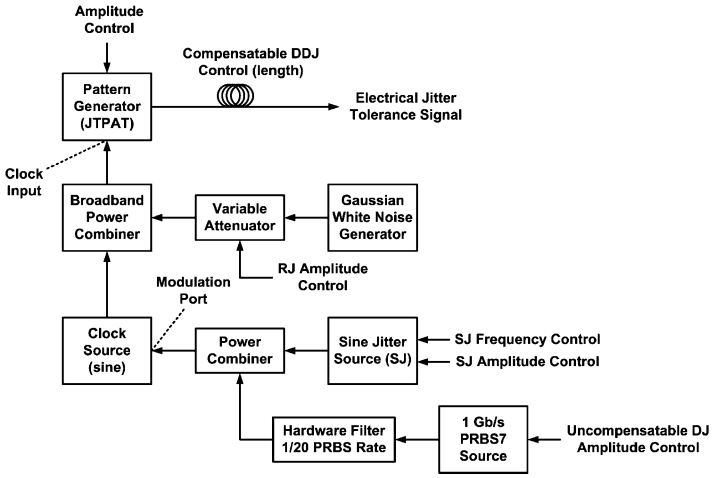

FIGURE 8.40. Block diagram of the electrical signal tolerance source. (After INCITS Technical Commitee [113].)

Time-domain generation of a data pattern exhibiting jitter characteristics with the previously introduced statistical properties is difficult. For this reason, we built a behavioral model of the data source recommended for hardware jitter tolerance compliance testing by most communication standards (Figure 8.40). The different components of this data source are discussed in the following. Most jitter contributions are applied to the locally generated reference clock. At each rising edge of this reference clock, the pattern generator emits a new bit of the data pattern.

Some of the contributions to deterministic jitter mentioned in Paragraph 3.7.1, especially those related to bandwidth limitations, can be compensated for by signal processing manipulations like predistortion and equalization. In order to obtain a close to real-life data pattern where part of the deterministic jitter can be compensated for, DJ is generated based on two independent mechanisms. The *compensatable deterministic jitter*, designated by the acronym DDJ (which stands for *data-dependent jitter*), is introduced by a bandwidth limitation mechanism applied at the output of the pattern generator. The *uncompensatable deterministic jitter* source, designated by UDJ, is built of a 1.0 Gb/s PRBS7 pattern generator, which data is sent through a 50 MHz bandwidth 1st-order low-pass filter.

$$\frac{v_{UDJ}}{v_{PRBS7}} = \tanh\left(\frac{7 \cdot T_{ck}}{2 \cdot \tau}\right) \cong \tanh(1.1) \cong 0.8 \qquad \text{(EQ 8.25)}$$

Due to low-pass filtering, the peak amplitude of the UDJ signal is lower than the input PRBS7 pattern as given by Equation 8.25 as a function of the longest run length (i.e. $7T_{ck}$ in PRBS7) and the filter time constant $\tau = \frac{20 \cdot T_{ck}}{2\pi}$. To compensate for this undesired attenuation, a gain block must be added after the filter. In the final model, this gain is numerically adjusted by simulation to improve the compensation according to the particular PRBS7 pattern. It is interesting to notice that the DDJ mechanism also results in partial vertical eye closure, while the UDJ mechanism does not.

The UDJ component is then combined with the sinusoidal jitter component to modulate the phase of the reference clock. When setting the parameters of the SJ component, care should be taken to avoid the sinusoidal jitter frequency to be derived from the clock frequency by an integer or rational division. Insufficient coverage of the stochastic properties of sinusoidal jitter may otherwise result in excessive worst or best case simulations. As a side comment, an interesting note on the validity of sinusoidal jitter testing at large jitter amplitudes has been presented in [114]. Indeed, as shown by the probability density function of a sine wave, the data edge density of the incoming data will be located close to the center of the eye for high SJ amplitudes (>0.5 UI), which may lead loop-based clock recovery systems to false lock.

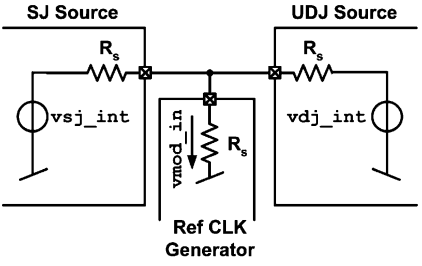

FIGURE 8.41. Equivalent schematic of the power combiners

After combination of the sinusoidal jitter component and the UDJ component, this signal is applied to the modulation input of the reference clock generator. The power combiners are modeled as Thévenin-equivalent voltage sources with a source resistance of 50 Ω driving a shared load resistance of 50 Ω (Figure 8.41), where the ideal voltage sources deliver twice the output voltage, as used in conventional measurement equipment. For a number *n* of sources driving the load according to this principle,

the attenuation of the individual signal is given by Equation 8.26, the equivalent code is shown in Figure 8.42.

$$v_{mod\_in} = \sum_{i=1}^{n} \frac{2 \cdot v_n}{1+n} \quad \text{(EQ 8.26)}$$

The values of the internal voltages `vsj_int` and `vdj_int` are twice the desired output voltages, in order for the output voltages `vsj_out` and `vdj_out` to correspond to the expected value when terminated with a matched load.

```
vsj_out == vsj_int + isj_out * 50.0;
vdj_out == vdj_int + idj_out * 50.0;
imod_in == isin_clk + idj_out
vmod_in == imod_in * 50.0;
```

FIGURE 8.42. Equivalent VHDL-AMS code of the power combiner

A sinusoidal clock source is used to allow proper phase modulation by the previously mentioned components (Equation 8.27). All voltages are centered around v_{mid} (for simplicity equal to ground, could also be set to $V_{DD}/2$), A_{clk} and A_{mod} are the respective amplitudes set for the reference clock and the voltage at the modulation input.

$$vsin\_clk = v_{mid} + A_{clk} \cdot \sin(2\pi f_{clk} t + d\phi)$$
$$\text{with} \quad d\phi = \frac{vmod_{in} - v_{mid}}{A_{mod}} \cdot k_{VCO} \cdot 2\pi \quad \text{(EQ 8.27)}$$

Figure 8.43 shows the corresponding code, where again the generated internal value is equal to twice the desired value. The `NOW` statement is used to recover the time information from the simulator and `dph` is the phase modulation represented by $d\phi$ in Equation 8.27.

```
vsin_clk_int == 2.0 * (0.0 + 0.9 * sin(MATH_2_PI*freq*NOW+dph));
```

FIGURE 8.43. Sinusoidal reference clock code

Random jitter is added in power to the modulated clock signal, which is then applied to the clock input of the pattern generator. The conversion from additive noise to jitter being illustrated in Figure 8.44, the conversion factor is given by the slope at the mid-point crossing of the clock signal (Equation 8.28).

$$\sigma_{AN} = RJ_{RMS} \cdot \frac{1\text{UI}}{\pi \cdot A_{clk}} \quad \text{(EQ 8.28)}$$

The principle of adding random noise is arguable due to the resulting invalid additional clock edges, as discussed in [115]. In our implementation, the probability of immediate successive zero crossings is somewhat mitigated by the hysteresis implemented at the clock input of the pattern generator. When applying larger amounts (e.g. 0.02 UI_{RMS}) of random jitter in simulation, the large number of invalid additional clock edges rendered the analysis of the bit error ratio and eye diagram impossible. For this reason, it is frequently recommended to add the RJ component to the combined SJ and UDJ components at the modulation input of the clock generator.

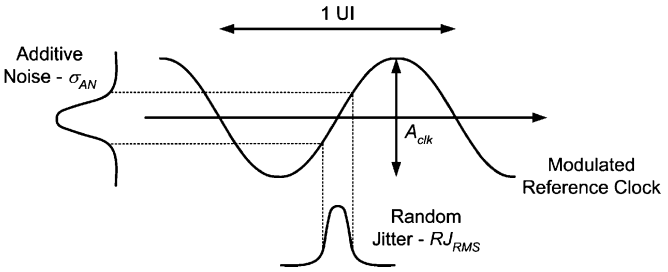

FIGURE 8.44. Additive noise to random jitter conversion. (After Baumer and Engelen [115].)

The pattern generator of the standardized jitter tolerance source for system compliancy verification generates a so-called *compliant jitter tolerance pattern* (CJTPAT), composed of the data words representing the worst-case condition for jitter tolerance measurement of the receiver. The *compliant* term indicates that the effective stress pattern is embedded into data frames including delimiters, fill words and CRC, which are required to test physical links. For verification of a CDR circuit, use of the *jitter tolerance pattern* (JTPAT), which corresponds to the payload of the CJTPAT, is recommended and implemented in the present model. In the current version of the model, the JTPAT is hardcoded as a data vector repeated iteratively (Figure 8.45).

At the output of the pattern generator, the resulting pattern is bandwidth limited by a first order filter of time constant τ, resulting in compensatable deterministic jitter. Considering a periodic pattern centered around its DC value, the filter bandwidth required to obtain a given amount of deterministic jitter is given by Equation 8.29, where $T_B = 1/f_B$ is the bit period and T_{MIN} and T_{MAX} respectively the shortest and longest run length in the pattern. Convenient cutoff frequency settings for some common values of deterministic jitter are proposed in Table 8.4. However, applying the JTPAT or a PRBS signal results in more severe jitter due to the transitions between long and short CID sequences, which are difficult to estimate mathematically. Some simulated correspondence points between cutoff frequency and DDJ for the JTPAT are also given in Table 8.4.

```
architecture dig of jtpat_core is
  constant d30p3_pat: std_logic_vector(19 downto 0) :=
  "10000111000111100011";
  constant d21p5_pat: std_logic_vector(19 downto 0) :=
  "10101010101010101010";
  signal counter : natural := 100;
begin
  gen_proc : process (clk, rst_b)
  begin
    if rst_b = '0' then
      data_out <= '0';
    elsif clk'EVENT and clk='1' then
      if counter < 100 then
        data_out <= d30p3_pat(counter mod 20);
      else
        data_out <= d21p5_pat(counter mod 20);
      end if;
      if counter = 130 then
        counter <= 0;
      else
        counter <= counter + 1;
      end if;
    end if;
  end process gen_proc;
end dig;
```

FIGURE 8.45. JTPAT pattern generator code

$$DDJ[UI] = \frac{\tau}{T_B} \cdot \ln\left(\frac{1 + e^{\frac{-T_{MIN}}{\tau}}}{1 + e^{\frac{-T_{MAX}}{\tau}}}\right) \qquad \text{(EQ 8.29)}$$

TABLE 8.4. Filter cutoff frequencies and associate DDJ values for $T_{min}=T_B$ and $T_{min}=5T_B$

f_c [MHz]	DDJ – periodicity assumption [UI_{PP}]	DDJ – JTPAT [UI_{PP}] (simulated)
292.0	0.50	
340.0	0.40	
400.0	0.30	
411.0		0.41
500.0	0.20	0.27
672.0	0.10	
1000.0	0.031	0.033
2000.0	0.001	0.001

Figure 8.46 shows the eye diagram of the data output by the JTPAT data source with different jitter components applied. Of course the simulator time step influences the accuracy of the obtained eye opening. Simulation of a 1μs pattern at 2.5 Gb/s (2′500 bits) with a maximum time step of 2 fs on a SPARC Ultra 10 (400 MHz CPU) takes approximately 7 min.

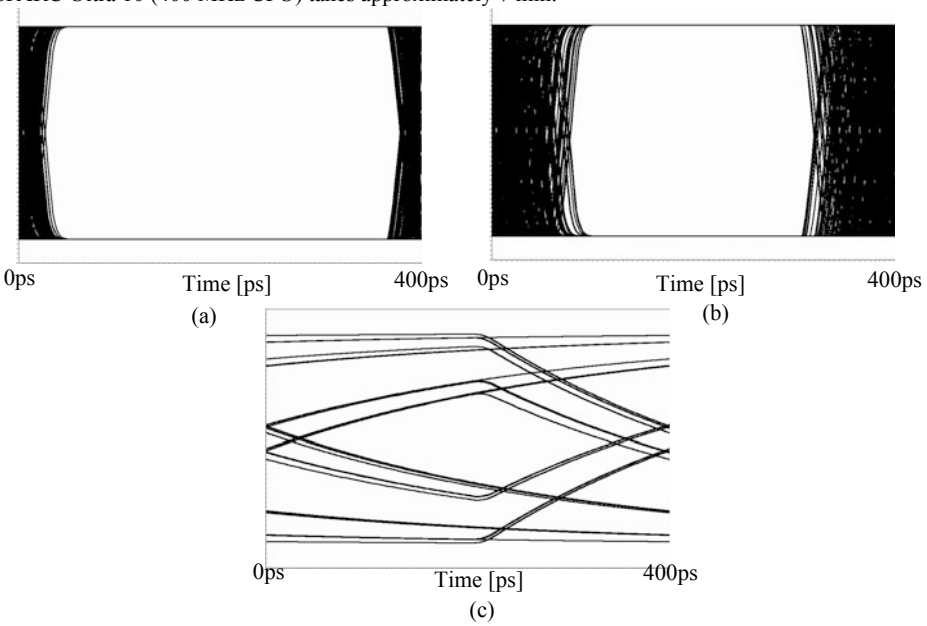

FIGURE 8.46. VHDL-AMS JTPAT data source with normalized amplitude for (a) SJ_{amp} = 0.1 UI_{PP} SJ_{freq} = 0.0211 f_B, (b) UDJ = 0.4 UI_{PP} and (c) DDJ = 0.4 UI_{PP}

The generated data pattern is sent to the gated oscillator-based clock recovery circuit. When aligning the data to be sampled (input signal of the sampling element) with respect to the sampling edges of the recovered clock, we obtain the eye diagram shown in Figure 8.47. Unlike an eye diagram obtained from the output data of the sampling element, this figure allows us to observe the remaining timing

margins with respect to the sampling instant located at 0.5 UI. The difference in jitter accumulation between both data edges is easily distinguishable.

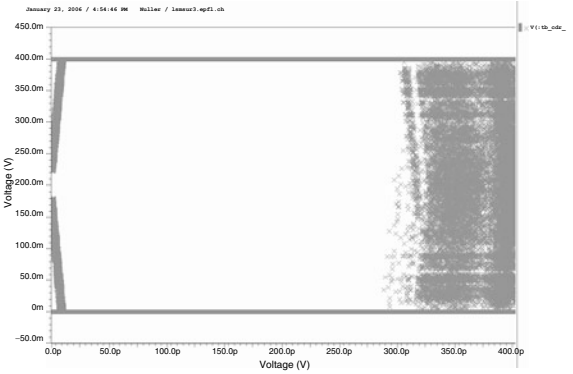

FIGURE 8.47. Eye diagram at the CDR output when applying data from the JTPAT source with SJ_{amp} = 0.1 UI_{PP}, SJ_{freq} = 0.0211 f_B, UDJ = 0.1 UI_{PP}, $DDJ \approx$ 0.1 UI_{PP} (f_C = 672 MHz), RJ = 0.001 UI_{RMS} and CKJ = 0.01 UI_{RMS} (no frequency offset)

Based on the presented models, the accurate behavior of the CDR circuit have been verified. Among other issues, the importance of the accuracy of the edge detection pulses has been analyzed and will be presented below. In combination with a bit error ratio test module (which can be included in the form of a VHDL or VHDL-AMS model), BER calculations can also be performed. Due to the large number of bits required to achieve statistical significance [117], the determination of the bit error ratio based on a given number of errors in the data stream however requires several weeks of simulation time on a state-of-the-art computing system.

8.7.3 Duration of Edge Detection Pulse

In the initial design, the edge detector delay was specified to be exactly half a nominal clock cycle, so that the reset signal would not disturb the oscillator duty cycle. Behavioral time-domain simulation showed however that this choice leads in presence of delay variations to a race condition between the oscillator output feedback and the EDET signal, resulting in the absence of synchronization. As shown in Figure 8.48, for synchronization to be effective, the clock edge generated in the oscillator by the falling EDET signal must propagate to the oscillator output and be fed back to the input gate before the

EDET signal may rise. Otherwise the information of this edge be used a single time for sampling the data, then vanish.

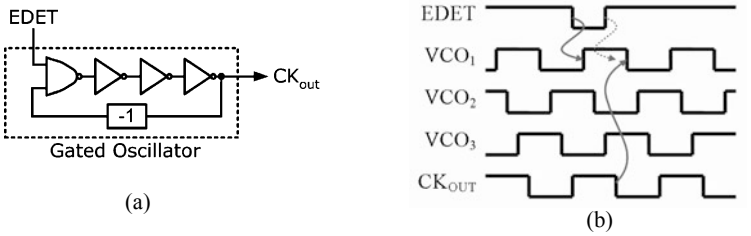

FIGURE 8.48. (a) Gated oscillator core, (b) race condition in the oscillator reset

This issue is solved by increasing the duration of the EDET pulses, achieved by increasing the delay of the delay line. However, this results in duty-cycle distortion of the recovered clock and an optimum trade-off between both issues is to be found. This optimum depends on the variability of the delay line delay, which can also be simulated by adding delay mismatch in the model. This example indicates the strength of combining statistical simulation with time-domain simulation, where both levels of abstraction have an important role to play in the complete design flow.

8.8 Transistor-Level Design

From statistical simulations we have determined the estimate of acceptable jitter performance of the oscillator, which can now be used for power-aware transistor-level design of the clock recovery circuit. As transistor-level clock recovery design is widely discussed in the literature, only the important issues of the design optimization of the gated oscillator topology will be presented here, completing the top-down design flow introduced earlier.

8.8.1 Current-Mode Logic Cells

The gated oscillator clock and data recovery circuit has been implemented using current-mode logic (CML) cells. The conventional CMOS logic design style has several drawbacks when operating at high speed:

- The load is the equivalent of two gate capacitances
- The signal swing is $V_{DD} - V_{SS}$
- The power consumption scales with f_{CLK}^2
- Supply and substrate noise is generated at the frequency of operation

According to Kwan and Shams [119], the differential signaling current-mode logic cells (Figure 8.49) can operate about two times faster than conventional CMOS cells at comparable power dissipation. This result probably has to be mitigated by the increased wire load in a larger design due to the increased number of routing constraints. Nevertheless, this logic family certainly has a speed advantage and comes with additional benefits:

- Reduced input capacitance
- Lower power supply and substrate noise
- Reduced sensitivity to environmental perturbations (capacitive or magnetic cross talk) due to routing symmetry.

As a result of the reduced noise sensitivity, this architecture can afford lower noise margins and thus operate at reduced signal swing. Beyond the intrinsic speed improvement resulting of reduced signal swing, this in turn also results in reduced capacitive coupling effects and closes the noise versus noise-margin loop. Additionally, a differential implementation of the CDR matches perfectly the differential limiting amplifier output interface.

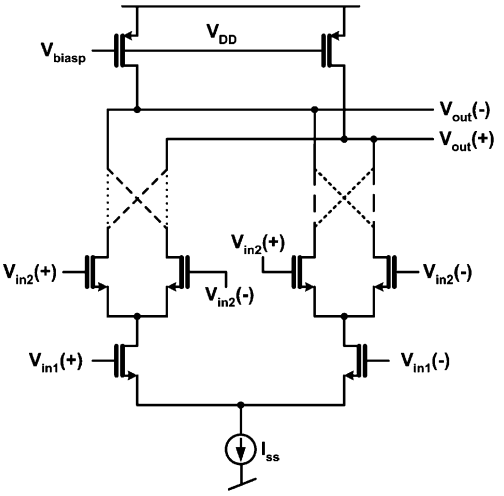

FIGURE 8.49. A simple current-mode logic cell (multifunction two-input gate)

Finally, a two-stage CML gate can perform multiple logical operations, only depending on the routing of its input nodes and output nodes of each highest stage.

8.8.2 Low-Power CDR Design

The four-stage current-controlled ring oscillator in the CDR is built using current-mode logic (CML) gates [116]. It was previously mentioned that completely switching differential stages do not offer the improved supply noise immunity over single-ended stages. However, in the case of oscillators, we are mostly concerned about phase noise and not about amplitude noise. As shown by Hajimiri et al. [108], the amplitude restoring mechanism of an oscillator renders it most sensitive to noise injection around the zero crossing instant. In this situation, the gain stage is in equilibrium and thus optimally rejects common-mode perturbations like supply noise. As a result, differential ring oscillator implementations provide optimal first-order rejection of supply and substrate noise.

In order to obtain perfect symmetry between all four stages, the two-stage CML gate performing the gating operation at the oscillator input is used for all stages (Figure 8.50). The tuning current delivered by the shared PLL determines the biasing of each stage, while the values of the load resistors operating in triode region are adjusted by a so-called *replica-bias circuit* to guarantee constant signal swing [120].

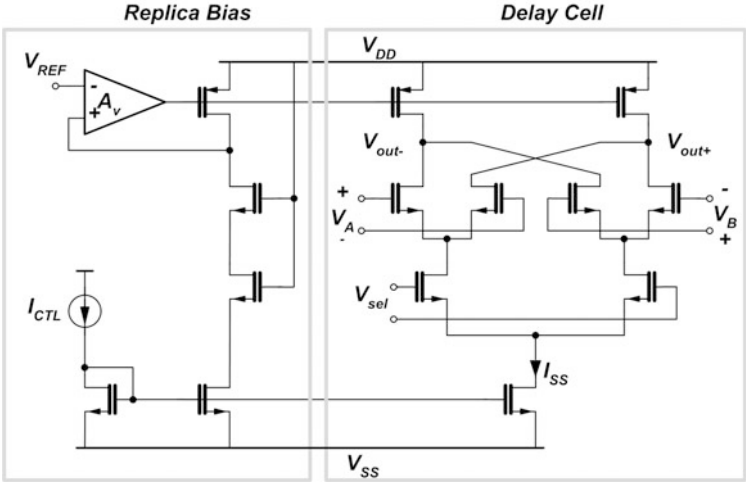

FIGURE 8.50. The two-stage CML gate used in the gated oscillator

From the oscillator jitter parameter κ_{min} determined in Paragraph 8.6.5, the bias current of one oscillator cell can be determined. Indeed, Hajimiri et al. [108] introduces the relationship between the κ_{min} and the oscillator bias current of one stage I_{ss}, as well as other parameters defined below:

$$\kappa_{min} = \sqrt{\frac{8}{3\eta} \cdot \frac{k_B T}{I_{ss}} \cdot \left(\frac{1}{V_{char}} + \frac{1}{V_{swing}}\right)} \qquad \text{(EQ 8.30)}$$

η is a proportionality constant between the rise/fall-time and the gate delay and is typically close to 1. k_B is the Boltzmann constant, while T is the absolute temperature expressed in Kelvin. The characteristic voltage V_{char} of the logic gate is given, according to the above-mentioned reference, by $(V_{GS} - V_T)/\gamma$ under the long-channel assumption, while it becomes $(E_c L)/\gamma$ in the short-channel regime, where γ is the excess thermal noise factor, E_c the critical electric field and L the channel length. The increase of the excess thermal noise factor in short-channel devices has been discussed in Paragraph 6.4.5 on page 91 and the use of $\gamma = 2$ has been suggested. V_{swing} is the maximum amplitude at the output of an oscillator stage determined by the setting of the reference voltage V_{ref}. V_{swing} has to be somewhat larger than the effective amplitude of oscillation, because the effective oscillation amplitude V_{eff} of an M-stage ring oscillator is given by:

$$V_{eff} = V_{swing} \cdot \frac{4}{\pi} \cdot \cos\left(\frac{\pi}{M}\right) \qquad \text{(EQ 8.31)}$$

The detailed development of the quasi-linear analysis leading to this result is presented in [121]. Based on the expression of κ_{min}, we can design the circuit to operate at the minimum power level required by the jitter specifications. The specified bias current and signal swing determine the value of the load resistor. Sufficient gain per stage is obtained by correct dimensioning of the differential pair transconductance, which in return also determines the load capacitance seen by the preceding stage. In order to obtain a good wire load model, post-layout simulation is essential. The wire parasitics indeed contribute to around 30% additional load capacitance compared to the schematic design. The simulated eye diagram in Figure 8.51 shows the output data after retiming with the recovered clock signal, which triggers the sampling flip-flop.

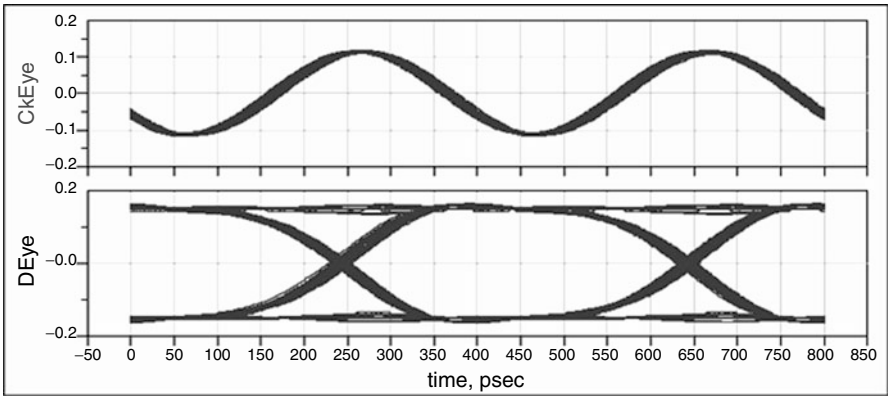

FIGURE 8.51. Eye diagram of the output data (bottom) with the recovered clock (top) [118]

A seven-channel CDR has been fabricated in a 0.18 μm CMOS technology, occupying a silicon area of 45′000 μm^2 per channel (Figure 8.52). For test purposes, it uses three separate replica bias circuits, controlling respectively the delay line, the oscillator and the fixed-delay gates like the sampler. Merging the delay line bias circuit with the fixed delay bias circuit would result in a 10% silicon area reduction.

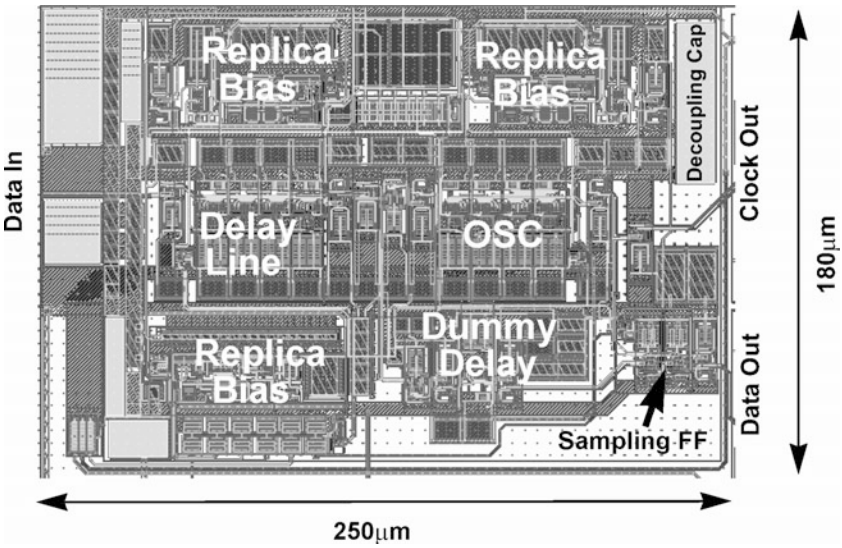

FIGURE 8.52. Gated oscillator CDR layout [118]

While the shared PLL occupies 0.1 mm^2, one CDR channel consumes 7.2 mW and the PLL 10.2 mW from a 1.8 V power supply.

8.8.3 Scaling Theory

As the presented topology focuses on area and power awareness, it is interesting to estimate its potential when porting to smaller feature size technologies. This development is based on first-order models neglecting some device parameters that do not scale perfectly. In deep submicron technologies, we must consider a scaling approach where dimensions and voltages scale differently [122]. Let us consider the scaling factors *a* for dimension scaling and *b* for voltage scaling, where both parameters are larger than one in a smaller feature size technology. In high-speed wireline circuits, the large majority

of the devices operating not in, but close to the velocity saturated regime. Using the assumption of velocity saturation to simplify the analysis, we obtain the scaling rules presented in Table 8.5.

TABLE 8.5. General scaling rules for short-channel devices

Parameter	Scales with
W, L, t_{ox}	$1/a$
C_{ox}	a
V_{DD}, V_{T0}	$1/b$
Area	$1/a^2$

Under the same assumption, the transconductance can be approximated by Laker and Sansen [26],

$$g_{msat} = WC_{ox}\mu\varepsilon_c \quad \text{(EQ 8.32)}$$

where ε_c is the critical longitudinal electric field and μ the mobility, which are considered constant with scaling. The transit frequency f_T, an indicator of the intrinsic device speed has been defined in Equation 4.1, reproduced in Equation 8.33. As the gate-source capacitance in saturation is proportional to the total gate capacitance, constant transconductance and scaled gate capacitance result in f_T increasing proportionally with K.

$$f_T = \frac{g_m}{2\pi \cdot C_{gs}} \sim \frac{WC_{ox}\mu\varepsilon_c}{2\pi \cdot C_{ox}WL} = \frac{\mu\varepsilon_c}{2\pi \cdot L} \sim a \quad \text{(EQ 8.33)}$$

As the velocity saturation operation is characterized by a linear relationship between the drain current and the gate voltage, the corresponding drain current I_{Dsat} is given by $g_{msat}(V_G\text{-}V_{T0}\text{-}nV_S)$, scaling as $1/b$. In conclusion, the area of the CDR scales as $1/a^2$, the power consumption as $1/b^2$, while the data rate is increased by a.

For illustration purposes, let us estimate the performance improvements when porting the current 0.18 μm design into a 90 nm CMOS technology. This step corresponds to two major technology generations, represented by $a = 180/90 = 2$ and $b = 1.8/1.2 = 1.5$. The new design is expected to operate at 5 Gb/s, occupying an approximate area of $11'000\mu m^2$ and consuming 4.8 mW, resulting in a normalized area of $2'200\mu m^2$ /Gb/s and a normalized power consumption below 1 mW/Gb/s. As speed/power and speed/area figures improve when data rates get closer to the technology limits, it can be safely assumed that a 90 nm GO-CDR operated at 10 Gb/s would achieve even better performance with respect to these metrics.

8.9 CDR Measurements

Now that we have seen how the performance of a particular CDR topology can be evaluated beforehand and how the jitter specifications can be propagated down to the transistor-level design, let us discuss the measurement techniques which allow for the characterization of such a CDR circuit.

8.9.1 Eye Diagram Measurements

Eye diagram measurements show the output data in an eye diagram, triggered by the recovered clock (Figure 8.53). In the measurement shown below, the clock recovery circuit receives a PRBS7 input data stream at 2.5 Gb/s and delivers both the recovered clock and retimed data signals. Proper measurement of the receiver jitter tolerance requires jitter to be fed into the pattern generator for phase modulation. Alternatively a complete link using a nonideal transmitter and a nonideal communication medium can be measured, varying the transmit power and/or the transmit jitter characteristics.

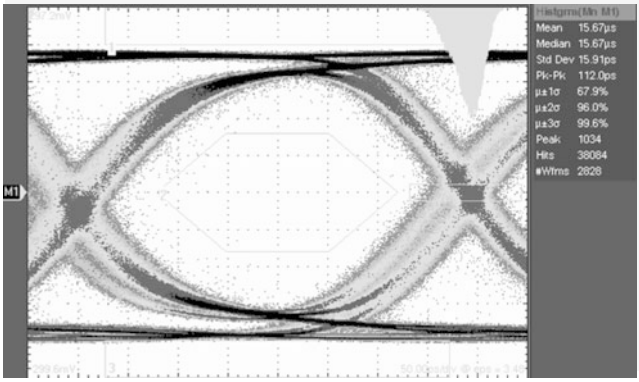

FIGURE 8.53. Measured eye diagram of a gated oscillator CDR output data operated at 2.5 Gb/s

The receiver output is typically taken after the retiming flip-flop. While this is the correct procedure for analyzing the bit error ratio of the complete receiver, it does not allow for direct determination of the receiver's jitter tolerance based on the eye diagram. Determination of JTOL from an eye diagram requires the input of the retiming element to be fed to the oscilloscope, where it is sampled with the recovered clock signal.

However, even in this configuration, exact eye diagram measurement is not always possible due to the way the oscilloscope uses the trigger signal. In the case of a *sampling* oscilloscope, the delay between the trigger event and the first sample is in the nanosecond range. This does not allow the scope to take into account short-term variations of the clock frequency as they appear due to the very short-term

synchronization of the gated oscillator CDR. *Real-time* oscilloscopes provide various sampling modes. In *single-shot* configuration, a user-defined number of samples is taken based on a single trigger event, which again does not consider individual sampling events. In *sequence* mode, acquisition of a single bit interval based on a trigger event is possible, but the dead-time between samples results in the loss of a large number of bits in the sequence.

8.9.2 Bit Error Ratio Measurements

Accurate estimation of the CDR performance and in particular the bit error ratio at the CDR output, as well as related jitter tolerance, can only be obtained from a bit error ratio tester (BERT), which compares the input and output data streams of the device under test (DUT, Figure 8.54).

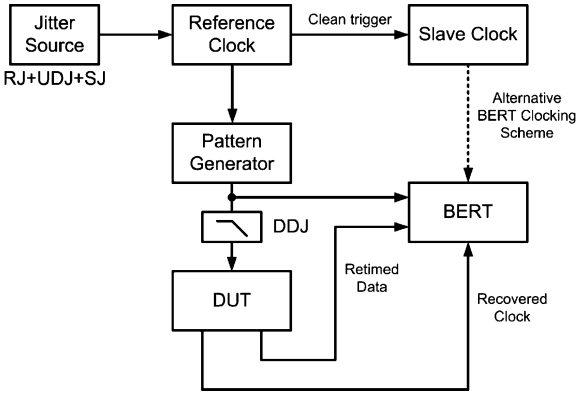

FIGURE 8.54. BER test measurement setup

In this scheme, data jitter is applied to the reference clock source, which triggers the pattern generator. The data pattern, known by the BERT, can be fed through a low-pass filter to add deterministic jitter. For reasons of setup complexity, the explicit addition of a DDJ component may in some cases be omitted, while the limited bandwidth of the board-level routing contributes some DDJ to the input data. The data pattern is applied to the DUT (i.e. the CDR), which recovers the clock signal from the received data and samples the data with this recovered clock signal. The same recovered clock signal is then used to determine the sampling instants in the BERT. By adjusting the BERT sampling delay between recovered clock and retimed data, the minimum BER value can then be determined. In the case of the GO-CDR, the possibility of using the recovered clock as a trigger for the BER tester depends on the shape of the recovered clock, which is subject to duty-cycle distortion inherent to this topology. In this situation, a stable slave clock can be used to trigger the BERT under certain conditions on jitter and frequency offsets ("Alternative BERT Clocking Scheme" in Figure 8.54).

An exhaustive measurement of the sampled eye opening requires the possibility to tune the sampling instant and the threshold voltage inside the retiming part of the CDR, using, e.g. a tunable delay line and a variable decision threshold. In such a configuration, the BERT can sweep the complete eye and determine the error ratio at each point of the eye, resulting in excellent knowledge of the design margins.

The gated oscillator CDR measurements revealed an interesting phenomenon. While short-term measurements (e.g. 5 min) result in zero detected bit error, longer-term measurements reveal a higher bit error ratio. Indeed, the bit errors appear by bursts of several million errors a time, the repetition rate of the bursts being related to the center frequency of the reference PLL and thus to the control current of the CDR (Figure 8.55, measured over intervals of 5 min).

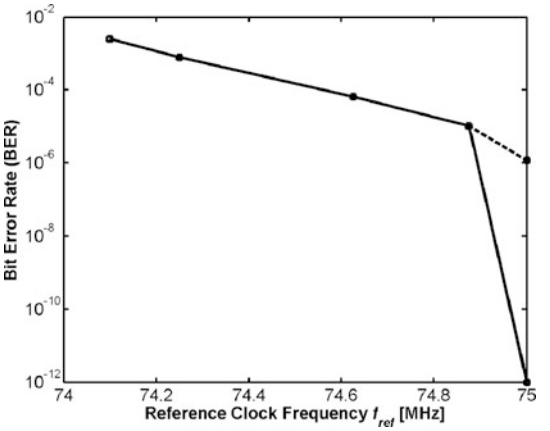

FIGURE 8.55. BER sensitivity to CDR center frequency measured over 5 min (dashed result is measured over 30 min)

Tuning of the center frequency resulted in a single error interval during measurements done over a duration of 30 min, equivalent to 99,9444% of error free intervals, however also equivalent to a BER of $1.147\ 10^{-6}$. We believe that residual frequency offset combined with ambient deterministic and random jitter result in eventual drop of a sampling event due to a very short clock pulse. Unlike the case of regular bit errors, this "lost" bit results in loss of synchronization in the PRBS sequence and thus appears as a large amount of bit errors in the BER result. As expected, the channel that was imple-

mented with a three-stage delay line instead of a four stage delay line experienced dramatically higher BER due to lacking synchronization.

TABLE 8.6. Jitter components and BER measured over 30 min

Jitter component	Jitter amplitude	BER	# Error intervals
no jitter		1.1473e-6	1
RJ_{RMS}	0.0077 mUI		
UDJ_{PP}	0.185 UI		
DDJ_{PP}	–		
SJ_{PP}	0.74 UI		
f_{SJ}	50 kHz	3.7574e-6	2
	2 MHz	4.0455e-6	3

This fact renders accurate jitter tolerance measurements very difficult, as the measurement duration required to achieve statistical significance in presence of such sporadic events is very long. Nevertheless, we tested the jitter tolerance up to the maximum jitter amplitude accepted by the pattern generator and the maximum sinusoidal jitter frequency achievable with the arbitrary waveform generator used. The jitter parameters applied and the corresponding BER results obtained in measurements over 30 min are shown in Table 8.6. Again, bursts of bit errors strongly determine the bit error ratio and limit the statistical significance of these measurements. It has not been verified in how far the combination of the CDR with an appropriate elastic buffer would suffice to suppress these error bursts. Table 8.7 summarizes measurement results of the gated oscillator CDR, implemented in a 0.18 μm CMOS technology.

TABLE 8.7. CDR simulation and measurement results overview

Specification	Simulated	Measured	Unit
Supply voltage	1.6–2.0	1.8	V
Data rate	2.5	1.0–2.5	Gb/s
Clock Jitter @ 2.5 GHz			
w/o trimming	4	12	ps
with trimming		4.1	ps
Bit error ratio/error free intervals	@ 2.5 Gb/s	@ 1.0 Gb/s	
no jitter	$< 10^{-12}$ / –	1.1473e-6 / 99.9444	– / %
jitter according to Table 8.3, FTOL = 0	$< 10^{-11}$ / –		–
jitter according to Table 8.6		4.0455e-6 / 99.8333	– / %

TABLE 8.7. CDR simulation and measurement results overview

Specification	Simulated	Measured	Unit
Current Consumption @ 25°C, 1.8 V			
PLL	10.5 @ 2.5 Gb/s		mW
CDR - 1 channel	8.5	~ 60mW @ 2.5Gb/s	mW
Output Drivers - 1 channel	~2 x 23.5		mW
Bias	~1.5		mW
Area			
PLL	0.10		mm^2
CDR (w/o Drivers)	0.06		mm^2

8.10 Discussion

In this chapter, we presented different CDR topologies with their advantages and drawbacks when used in the context of short-distance multichannel data transmission. A jitter-estimation-based top-down design flow was introduced and applied to the design of a gated oscillator based clock recovery circuit. The flow highlights the strength and limitations of the studied topology at an early stage. While the time-domain simulation allows for functional verification of the design, the specification propagation to the transistor-level leads to a hardware implementation that meets the high-level specifications. Furthermore, the consistent oscillator jitter specifications available to the designer result in a power-jitter optimized schematic, essential to the achievement of the power constraints. In a general fashion, such a specification-driven design approach can be applied to the analysis and design of most CDR topologies and delivers consistent early performance estimates for such a design.

CHAPTER 9 *Conclusions*

Short-distance fiber-optic chip-to-chip links provide next-generation processors with the I/O bandwidth required to fully exploit their immense data processing throughput. Economic viability however asks for a low-cost and low-overhead implementation, which can be optimally satisfied by fully integrated silicon receivers. Throughout this book, the reader has been shown how to address the upcoming challenges and trade-offs in the design of such CMOS-integrated receivers.

Knowing the substantial bandwidth available in the optical domain and remembering that the electrical properties of the transmitter and receiver ultimately limit the data rates of optical communication links, it is tempting to push the electro-optical interfaces as far as possible inside the computing parts of the processors. While optical computing seems destined to remain a far-out technology for some time to come, optical networking has become an immovable cornerstone of the connected world. As a result, it makes perfect sense to think about integration of optical links on silicon processors to replace global connections in the large multicore arrangements currently planned by the major manufacturers. Of course, miniaturization and CMOS process compatibility of optical add-drop multiplexers, waveguides and silicon transmitters and receivers represent major challenges to the engineering community, but the computing power leveraged by such reconfigurable on-chip optical networks combined with large-scale multiprocessor arrays would represent a quantum leap in manufacturable processing power.

P. MULLER, Y. LEBLEBICI,
CMOS Multichannel Single-Chip Receivers for Multi-Gigabit Optical Data Communications,
DOI 10.1007/978-1-4020-5912-4, © 2007 Springer

References

[1] G. E. Moore, "Cramming more components onto integrated circuits", *Electronics*, vol. 38, April 1965

[2] B. Madhavan, "Multi-Gigabit/s/channel Parallel Optical Data Link Design in CMOS Technology", Ph.D. Thesis, University of Southern California, Los Angeles, August 2000

[3] D. Erni, "Optical Backplane Communication on FR-4-Boards", *The Annual CTI Micro- and Nano-Technologies Event*, Neuchâtel, Switzerland, November 2004

[4] K. H. Gulden et al., "VCSEL Arrays for High Speed Optical Links", *Gallium Arsenide Integr. Circ. (GaAs IC) Symp. Annu. Tech. Dig.*, pp. 53–56, October 2001

[5] B. Casper et al., "A 20Gb/s Forwarded Clock Transceiver in 90nm CMOS", *IEEE Int. Solid-State Circ. Conf. Dig. Tech. Papers*, February 2006

[6] P. Dainesi et al., "Fast and Efficient Light Intensity Modulation in SOI with Gate-All-Around Transistor Phase Modulator", *Conference on Lasers and Electro-Optics*, May 2005

[7] M. Siegert et al., "Efficient Optical Coupling Between a Polymeric Waveguide and an Ultrafast Silicon MSM Photodiode", *IEEE J. Sel. Top. Quantum Electron.*, vol. 4, no. 6, November/ December 1998

[8] S. Radovanovic, A.-J. Annema, B. Nauta, "A 3-Gb/s Optical Detector in Standard CMOS for 850-nm Optical Communication", *IEEE J. Solid-State Circ*., vol. 40, no. 8, pp. 1706–1717, August 2005

[9] H. Rong et al., "An all-silicon Raman laser", *Nature*, vol. 433, no. 7023, pp. 292–294, January 20, 2005

[10] S. Coffa, "Light from Silicon", *IEEE Spectrum*, October 2005

[11] N. Savage, "Linking with Light", *IEEE Spectrum*, August 2002

[12] P. Mahoney et al., "Clock Distribution on a Dual-Core, Multi-Threaded Itanium®-Family Processor", *IEEE Int. Solid-State Circ. Conf. Dig. Tech. Papers*, vol. 48, pp. 292–293, February 2005

[13] C. T. Gray, W. Liu and R. K. Cavin III, "Timing constraints for wave-pipelined systems", *IEEE Trans. Comput.-Aided Des. Integr. Circ. Syst*., vol. 13, iss. 8, pp. 987–1004, August 1994

[14] N. Holonyak Jr. and M. Feng, "The Transistor Laser", *IEEE Spectrum*, February 2006

[15] D. R. Goff, "Fiber Optic Reference Guide", 3rd edn., Focal Press, Woburn, MA, 2002, http://www.fiber-optics.info/fiber-history.htm

[16] Maxim Integrated Products, "Extinction Ratio and Power Penalty", *Application Note HFAN-2.2.0*, rev.0, May 2001

P. MULLER, Y. LEBLEBICI,
CMOS Multichannel Single-Chip Receivers for Multi-Gigabit Optical Data Communications,
DOI 10.1007/978-1-4020-5912-4, © 2007 Springer

[17] P.-G. Fontolliet, “Systèmes de Télécommunications”, Traité d’électricité vol. XVIII, *Presses Polytechniques et Universitaires Romandes*, Lausanne, Switzerland, 1996

[18] Maxim Integrated Products, “NRZ Bandwidth – LF Cutoff and Baseline Wander”, *Application Note HFAN-09.0.4*, rev.0, September 2002

[19] B. Razavi, “Design of Integrated Circuits for Optical Communications”, McGraw-Hill, New York, 2003

[20] Maxim Integrated Products, “Optical Signal-to-Noise Ratio and the Q-Factor in Fiber-Optic Communication Systems”, *Application Note HFAN-09.0.2*, rev.0, February 2002

[21] M. A. Kossel and M. L. Schmatz, “Jitter Measurements of High-Speed Serial Links”, *IEEE Des. Test Comput*., vol. 21, no. 6, pp. 536–543, November–December 2004

[22] N. Ou et al., “Jitter models for the design and test of Gbps-speed serial interconnects”, *IEEE Des. Test Comput*., vol. 21, iss. 4, pp. 302–313, July–August 2004

[23] American National Standard/ATIS, “Telecom Glossary 2000”, *http://www.atis.org/tg2k*, ANS T1.523–2001

[24] B. Minch, “ECE 315: Introduction to Microelectronics”, Chalmers University of Technology, Göteborg, Sweden, September 2002: http://www.s2.chalmers.se/undergraduate/courses0405/ess065/doc/MOS-models-Minch.pdf

[25] Y. Tsividis, “Operation and Modeling of The MOS Transistor”, 2 nd Edition, McGraw-Hill, New York, 1999

[26] K. R. Laker and W. M. C. Sansen, “Design of Analog Integrated Circuits and Systems”, McGraw-Hill, New York, 1994

[27] R. Annen, M. Bossard and S. Werffeli, “Circuits for parallel optical interconnects”, *Workshop on Parallel Optical Interconnects Inside Electronic Systems*, Design, Automation and Test in Europe, Paris, France, 2004

[28] M. Steyaert et al., “RF Integrated Circuits in Standard CMOS Technologies”, *Proc. Eur. Solid-State Circ. Conf*., September 1996

[29] A.-J. Annema et al., “Analog Circuits in Ultra-Deep-Submicron CMOS”, *IEEE J. Solid-State Circ.*, vol. 40, no. 1, pp. 132–143, January 2005

[30] J.-O. Plouchart et al., “Application of an SOI 0.12-µm CMOS technology to SoCs with low-power and high-frequency circuits”, *IBM J. Res. Dev.*, vol. 47, no. 5/6, 2003

[31] J. S. Dunn et al., “Foundation of RF CMOS and SiGe BiCMOS technologies”, *IBM J. Res. Dev.*, vol. 47, no. 2/3, 2003

[32] Z.-G. Wang et al., “2.5-Gb/s 0.35-µm CMOS ICs for Optic-Fiber Transceiver”, *IEEE Proc. Int. Conf. on Electron., Circ. Syst. 2001*, vol. 2, pp. 689–692, September 2001

[33] S. Voinigescu et al., “Circuits and Technologies for Highly Integrated Optical Network IC’s at 10 Gb/s to 40 Gb/s”, *Proc. Custom Integr. Circ. Conf.* 2001, pp. 331–338, 2001

[34] Albis Optoelectronics AG, “PDCS65T - Short/Long Wavelength 10 Gb/s Photodiode Chip”, *Product Description*, rev. 0.33, April 2005

[35] J. L. Zerbe et al., “Equalization and Clock Recovery for a 2.5–10-Gb/s 2-PAM/4-PAM Backplane Transceiver Cell”, *IEEE J. Solid-State Circ.*, vol. 38, no. 12, pp. 2121–2130, December 2003

[36] P. Hofstee, “Architecting Interconnect”, *12th Topical Meeting on Electrical Performance of Electronic Packaging*, http://www.ewh.ieee.org/soc/cpmt/newsletter/200312/epep.html, Princeton, NJ, October 2003

[37] C. Kromer et al., “A Low-Power 20-GHz 52-dBΩ Transimpedance Amplifier in 80-nm CMOS”, *IEEE J. Solid-State Circ.*, vol. 39, no. 6, June 2004

[38] Infiniband Trade Association IBTA, “InfiniBand Architecture Specification”, *Volume 2 - Physical Specifications*, rel. 1.0.a, June 2001

[39] IEEE Computer Society, “IEEE Std 802.3 - 2002, Part 3: Carrier sense multiple access with collision detection (CSMA/CD) access method and physical layer specifications”, March 2002

[40] J. Graeme, “Photodiode Amplifiers: OPAMP Solutions”, McGraw-Hill, New York, 1996
[41] G. Fontana and A. Tonello, “Exact design solutions for photodiode transimpedance amplifiers based on FET input OP-AMPs.”, *Technical Report DIT-02-037*, Informatica e Telecomunicazioni, University of Trento, 2002
[42] H. Ransijn, “Wideband Amplifiers for high speed fiber optic receivers and transmitters”, *ISSCC 2002 Short Course on Wideband Communications*, San Francisco, February 2002
[43] Maxim Integrated Products, “Optical Receiver Performance Evaluation”, *Application Note HFAN-03.0.2*, rev.0, March 2003
[44] J. F. Buckwalter, “Deterministic Jitter in Broadband Communication”, Ph.D. Thesis, California Institute of Technology, Pasadena, January 2006
[45] Michael W. Davidson, “Molecular Expressions Microscopy Primer: Photomicrography – Interactions of Photons with Silicon”, http://micro.magnet.fsu.edu/primer/java/digitalimaging/ccd/quantum, National High Magnetic Field Laboratory, FL
[46] S. B. Alexander, “Optical Communication Receiver Design”, SPIE Press and IEE, San Diego, CA, 1997, p. 73
[47] C. Hermans and M. S. J. Steyaert, “A High-Speed 850-nm Optical Receiver Front-End in 0.18-μm CMOS”, *IEEE J. Solid-State Circ.*, vol. 41, no. 7, pp. 1606–1614, July 2006
[48] C. Rooman *et al.*, “Inter-chip optical interconnects using imaging fiber bundles and integrated CMOS detectors”, *Proc. 27th Eur. Conf. on Optical Commun. (ECOC)*, vol. 3, pp. 296–297, September–October 2001
[49] C. Rooman, D. Coppée and M. Kuijk, “Asynchronous 250-Mb/s Optical Receivers with Integrated Detector in Standard CMOS Technology for Optocoupler Applications”, *IEEE J. Solid-State Circ.*, vol. 35, no. 7, pp. 953–958, July 2000
[50] S. Radovanovic, A.-J. Annema and B. Nauta, “A 3-Gb/s Optical Detector in Standard CMOS for 850-nm Optical Communication”, *IEEE J. Solid-State Circ.*, vol. 40, no. 8, pp. 1706–1717, August 2005
[51] M. K. Emsley, “Silicon Resonant Cavity Enhanced Photodetector Arrays For Optical Interconnects”, Ph.D. Thesis, Boston University, Boston, 2003
[52] K. S. Mobarhan and J. Valdmanis, “Faster photodiodes meet variety of applications”, *Laser Focus World*, August 1999
[53] M. Salib, L. Liao, R. Jones, M. Morse, A. Liu, D. Samara-Rubio, D. Alduino, M. Paniccia, “Silicon Photonics”, *Intel Tech. J.*, vol. 8, iss. 2, May 2004
[54] M. Bruel, “The History, Physics, and Applications of the Smart-Cut Process”, *MRS Bull.* vol. 23, no. 12, pp. 35–39, December 1998
[55] M. Gökkavas et al., “Resonant Cavity Enhanced Photodiodes with a Broadened Spectral Peak”, *Proc. IEEE LEOS 2001 Annu. Meet.*, vol. 2, pp. 768–769, November 2001
[56] C. A. Holt, “Feedback Amplifiers in Electronic Circuits: Digital and Analog”, *Wiley,* New York, 1978
[57] C. Kromer et al., “A 100-mW 4 x 10 Gb/s Transceiver in 80-nm CMOS for High-Density Optical Interconnects”, *IEEE J. Solid-State Circ.*, vol. 40, no. 12, pp. 2667–2679
[58] E. Säckinger and W. Guggenbühl, “A high-swing, high-impedance MOS cascode circuit,” *IEEE J. Solid-State Circ.*, vol. 25, no. 1, pp. 289–298, February 1990
[59] K. Bult and G. Geelen, “A Fast-Settling CMOS Op Amp with 90dB DC-gain and 116MHz Unity-Gain Frequency”, *IEEE Int. Solid-State Circ. Conf. Dig. Tech. Papers*, pp. 108–109, February 1990
[60] B. J. Hosticka, “Improvement of the Gain of MOS Amplifiers”, *IEEE J. Solid-State Circ.*, vol. SC-14, no. 6, pp. 1111–1114, December 1979
[61] M. Kossel et al., “Wideband CMOS Transimpedance Amplifier”, *IEE Electronics Letters*, vol. 39, no. 7, pp. 587–588, 3rd April 2003
[62] S. M. Park and H.-J. Yoo, “1.25-Gb/s Regulated Cascode CMOS Transimpedance Amplifier for Gigabit Ethernet Applications”, *IEEE J. Solid-State Circ.*, vol. 39, no. 1, pp. 112–121, 2004

[63] R. Tao and M. Berroth, "A 10 Gb/s Fully Differential CMOS Transimpedance Preamplifier", *Proc. Eur. Solid-State Circ. Conf.*, pp. 549–552, September 2003
[64] A. Tanabe et al., "A Single-Chip 2.4-Gb/s CMOS Optical Receiver IC with Low Substrate Cross-Talk Preamplifier", *IEEE J. Solid-State Circ.*, vol. 33, no. 12, December 1998
[65] B. Analui and A. Hajimiri, "Bandwidth Enhancement for Transimpedance Amplifiers", *IEEE J. Solid-State Circ.*, vol. 39, no. 8, pp. 1263–1270, August 2004
[66] S. Galal and B. Razavi, "40-Gb/s amplifier and ESD protection circuit in 0.18-/spl mu/m CMOS technology", *IEEE J. Solid-State Circ.*, vol. 39, iss. 12, pp. 2389–2396, December 2004
[67] D. P. Triantis, A. N. Birbas and D. Kondis, "Thermal Noise Modeling for Short-Channel MOSFET's", *IEEE Trans. Electron Devices*, vol. 43, no. 11, pp. 1950–1955, November 1996
[68] C.-H. Chen and M. J. Deen, "Channel Noise Modeling of Deep Submicron MOSFETs", *IEEE Trans. Electron Dev.*, vol. 49, no. 8, pp. 1484–1488, August 2002
[69] K. Han et al., "Complete High-Frequency Thermal Noise Modeling of Short-Channel MOSFETs and Design of 5.2-GHz Low Noise Amplifier", *IEEE J. Solid-State Circ.*, vol. 40, no. 3, pp. 726–735, March 2005
[70] A. S. Roy and C. C. Enz, "Compact Modeling of Thermal Noise in the MOS Transistor", *IEEE Trans. Electron Devices*, vol. 52, no. 4, pp. 611–614, April 2005
[71] H. P. Tuinhout and M. Vertregt, "Characterization of Systematic MOSFET Current Factor Mismatch Caused by Metal CMP Dummy Structures", *IEEE Trans. Semiconductor Manufacturing*, vol. 14, no. 4, pp. 302–310, November 2001
[72] T. H. Lee, "The Design of CMOS Radio-Frequency Integrated Circuits", 2nd edn., Cambridge University Press, New York, 2004
[73] F. Giannini et al., "A High Gain-Bandwidth Product Distributed Transimpedance Amplifier IC for High-Speed Optical Transmission Using Low-Cost GaAs Technology", *Gallium Arsenide Applications Symposium (GAAS)*, Milano, September 2002
[74] E. A. Crain and M. H. Perrott, "A 3.125 Gb/s Limit Amplifier in CMOS With 42dB Gain and 1μs Offset Compensation", *IEEE J. Solid-State Circ.*, vol. 41, no. 2, pp. 443–451, February 2006
[75] J. Weiss et al., "A DC to 44-GHz, 19-dB Gain Amplifier in 90-nm CMOS Technology using Capacitive Bandwidth Enhancement", *IEEE Int. Solid-State Circ. Conf. Dig. Tech. Papers*, pp. 514–515, February 2006
[76] R. Schaumann and M. E. van Valkenburg, "Design of Analog Filters", Oxford University Press, New York, 2001
[77] E. Säckinger and W. C. Fischer, "A 3GHz 32dB CMOS Limiting Amplifier for SONET OC-48 Receivers", *IEEE Int. Solid-State Circ. Conf. Dig. Tech. Papers*, pp.158–159, February 2000
[78] S. Badel and Y. Leblebici, "An Inductorless Peaking Technique Applied to MOS Current-Mode Logic Gates", *Proc. 22nd Norchip Conference*, November 2004
[79] J. A. Mataya, G. W. Haines and S. B. Marshall, "IF Amplifier Using C_c Compensated Transistors", *IEEE J. Solid-State Circ.*, vol. 3, iss. 4, pp. 401– 407, December 1968
[80] S. Galal and B. Razavi, "10-Gb/s Limiting Amplifier and Laser/Modulator Driver in 0.18-μm CMOS Technology", *IEEE J. Solid-State Circ.*, vol. 38, no. 12, pp. 2138–2146, December 2003
[81] C. D. Holdenried, M. W. Lynch and J. W. Haslett, "Modified CMOS Cherry-Hooper amplifiers with source follower feedback in 0.35μm technology", *Proc. Eur. Solid-State Circ. Conf.,* pp. 553–556, September 2003
[82] A. M. Niknejad, "ASITIC – Analysis and Simulation of Spiral Inductors and Transformers for ICs", University of California, Berkeley, CA, http://rfic.eecs.berkeley.edu/~niknejad/asitic.html

[83] B. Razavi, "Design of Monolithic Phase-Locked Loops and Clock Recovery Circuits – A Tutorial", in *Monolithic Phase-Locked Loops and Clock Recovery Circuits – Theory and Design*, Ed. B. Razavi, IEEE Press, New York, 1996

[84] C. R. Hogge, "A Self-Correcting Clock Recovery Circuit", *IEEE J. Lightwave Technol*, vol. LT-3, pp. 1312–1314, December 1985

[85] J. D. H. Alexander, "Clock Recovery From Random Binary Signals", *Electronics Lett*, vol. 11, pp. 541–542, October 1975

[86] F. M. Gardner, "Charge-Pump Phase-Lock Loops", *IEEE Trans. Commun.*, November 1980

[87] L. M. DeVito, "Clock Recovery and Data Retiming", *MEAD Education/EPFL Advanced Engineering Course on PLLs and Clock & Data Recovery*, Lausanne, Switzerland, June–July 2003

[88] J. Lee, K. S. Kundert and B. Razavi, "Analysis and Modeling of Bang-Bang Clock and Data Recovery Circuits", *IEEE J. Solid-State Circ*, vol. 39, no. 9, pp. 1571–1580, September 2004

[89] R. C. Walker, "Designing bang-bang PLL's for clock and data recovery in serial data transmission systems", in *Phase-Locking in High-Performance Systems*, Ed. B. Razavi, IEEE Press, New York, 2003

[90] Silicon Laboratories, "SiPHY® OC-48/STM-16 SONET/SDH Transceiver", *Datasheet*, rev. 1.3, June 2005

[91] L. M. DeVito, "A Versatile Clock Recovery Architecture and Monolithic Implementation", in *Monolithic Phase-Locked Loops and Clock Recovery Circuits – Theory and Design*, Ed. B. Razavi, IEEE Press, New York, 1996

[92] E. F.-Y. Yeung, "Design of High-performance And Low-cost Parallel Links", Ph.D. Thesis, Stanford University, California, January 2002

[93] X. Maillard, F. Devisch and M. Kuijk, "A 900-Mb/s CMOS Data Recovery DLL Using Half-Frequency Clock", *IEEE J. Solid-State Circ.*, vol. 37, no. 6, pp. 711–715, June 2002

[94] T. H. Lee et al., "A 2.5V CMOS Delay-Locked Loop for an 18 Mbit, 500Megabyte/s DRAM", *IEEE J. Solid-State Circ.*, vol. 29, no. 12, pp. 1491–1496, December 1994

[95] S. Sidiropoulos and M. A. Horowitz, "A Semidigital Dual Delay-Locked Loop", *IEEE J. Solid-State Circ.*, vol. 32, no. 11, pp. 1683–1692, November 1997

[96] J. Kim and D.-K. Jeong, "Multi-Gigabit-Rate Clock and Data Recovery Based on Blind Oversampling", *IEEE Commun. Mag*., pp. 68–74, December 2003

[97] C.-K. K. Yang et al., "A 0.5-μm CMOS 4.0-Gbit/s Serial Link Transceiver with Data Recovery Using Oversampling", *IEEE J. Solid-State Circ.*, vol. 33, no. 5, pp. 713–722, May 1998

[98] R. Hofmann et al., "SiGe BiCMOS Broadband Phase-Aligner Circuit from 1Gb/s to 11Gb/s", *IEEE Int. Solid-State Circ. Conf. Dig. Tech. Papers*, pp. 420–421, February 2001

[99] D.-L. Chen, "A Power and Area Efficient CMOS Clock/Data Recovery Circuit for High-speed Serial Interfaces", *IEEE J. Solid-State Circ.*, vol. 31, no. 8, pp. 1170–1176, August 1996

[100] K. Y. K. Chang et al., "A 0.4-4-Gb/s CMOS Quad Transceiver Cell Using On-Chip Regulated Dual-Loop PLLs", *IEEE J. Solid-State Circ.*, vol. 38, no. 5, pp. 747–754, May 2003

[101] H. Partovi et al., "A 62.5Gb/s Multi-Standard SerDes IC", *Proc. IEEE Custom Integr. Circ. Conf.*, pp. 585–588, September 2003

[102] R. Kreienkamp et al., "A 10-Gb/s CMOS Clock and Data Recovery Circuit with an Analog Phase Interpolator", *Proc. IEEE Custom Integr. Circ. Conf.*, pp. 73–76, September 2003

[103] M. Y. He and J. Poulton, "A CMOS Mixed-Signal Clock and Data Recovery Circuit for OIF CEI-6G+ Backplane Transceivers", *IEEE J. Solid-State Circ.*, vol. 41, no. 3, pp. 597–606, March 2006

[104] H.-T. Ng et al., "A Second-Order Semidigital Clock Recovery Circuits Based on Injection Locking", *IEEE J. Solid-State Circ.*, vol. 38, no. 12, pp. 2101–2110, December 2003

[105] H. Lee et al., "Burst Mode Packet Receiver using a Second Order DLL", *IEEE Symp. VLSI Circ. Dig. Tech. Papers*, pp. 264–267, June 2004
[106] M. Banu and A. Dunlop, "A 660MWs CMOS Clock Recovery Circuit with Instantaneous Locking for NRZ Data and Burst-Mode Transmission", *IEEE Int. Solid-State Circ. Conf. Dig. Tech. Papers*, pp. 102–103, February 1993
[107] Maxim Integrated Products, "Jitter in Digital Communication Systems, Part 2", *Application Note* HFAN-4.0.4, 2002
[108] A. Hajimiri, S. Limotyrakis and T. H. Lee, "Jitter and Phase Noise in Ring Oscillators", *IEEE J. of Solid-State Circ*., vol. 34, no. 6, pp. 790–804, June 1999
[109] R. Stephens, "Jitter analysis: The dual-Dirac model, RJ/DJ, and Q-scale", *White Paper*, Agilent Technologies, ver. 1.0a, December 2004
[110] P. Muller and Y. Leblebici, "Pattern Generator Model for Jitter-Tolerance Simulation", *The Designer's Guide Community*, http://www.designers-guide.org/Modeling/JTOL_rev1.0.pdf
[111] Xilinx, "Additive White Gaussian Noise Core v1.0"
[112] G.Monnerie et al., "Modelling of transient noise sources with VHDL-AMS and normative spectral interpretation", *Forum on Specification & Design Languages,* 2003
[113] INCITS Technical Committee T11, "Fibre Channel – Methodologies for Jitter and Signal Quality Specification - MJSQ", *Draft Technical Report*, rev.14, June 2004
[114] A. Gupta, B. Tailor and J. Francis, "The Deficiencies in Measuring Input Jitter Tolerance (IJT) using Sine wave modulated Jitter", *IEEE 802.3 Higher Speed Study Group*, March 2001
[115] H. Baumer and J. van Engelen, "XAUI TX/RX Jitter Specifications", *IEEE P802.3ae 10Gb/s Ethernet Task Force*, July 2001
[116] M. I. Elmasry, "Nanosecond NMOS VLSI Current Mode Logic", *IEEE J. of Solid-State Circ*., vol. 17, issue 2, pp. 411–414, April 1982
[117] Maxim Integrated Products, "HFTA-05.0: Statistical Confidence Levels for Estimating BER Probability", *Application Note 703*, October 2000
[118] A. Tajalli, "Clock and Data Recovery Circuits for Multi-Link Optical Receivers", Ph.D. Thesis, Sharif University of Technology, Teheran, Iran, February 2006
[119] T. W. Kwan and M. Shams, "Design of multi-GHz asynchronous pipelined circuits in MOS current-mode logic", *Proc. of the 18th Int. Conf. on VLSI Des.*, January 2005
[120] T. H. Toifl et al., "Analysis of Parameter-Independent PLLs with Bang-Bang Phase Detectors", *Proc. IEEE Int. Conf. Electron., Circ. Syst.*, vol. 2, pp. 299–302, September 1998
[121] B. Schraner, "Design of an Integrated Voltage-Controlled Oscillator in CMOS", M.S. Thesis, EPFL, Lausanne, Switzerland
[122] J. M. Rabaey, A. Chandrakasan and B. Nikolic, "Digital Integrated Circuits, A Design Perspective", 2nd edn., Prentice Hall Electronics and VLSI Series, NJ, 2003
[123] T. Iwata et al., "A 5Gbps CMOS Frequency Tolerant Multi Phase Clock Recovery Circuit", *Symp. VLSI Circ. Dig. Tech. Papers*, pp. 82–83, July 2002
[124] M. Nogawa et al., "A 10Gb/s Burst-Mode CDR IC in 0.13μm CMOS", *IEEE Int. Solid-State Circ. Conf. Dig. Tech. Papers*, pp. 228–229, February 2005

Index

A

Absorption
  depth, 58, 61
AC-coupling, 17
Amplifier
  bandwidth, 38, 75
  limiting, 14, 95–97, 99, 100, 108, 111, 112, 116, 117, 119–121, 123, 125, 126
  transimpedance, 36–39, 42, 45, 47, 48, 50, 71, 73–78, 81–83, 85, 89, 90, 92, 95, 96, 110

B

Bandwidth, 1, 3–5, 8, 10, 11, 16, 19
Baseline wander, 17, 18
Bathtub curve, 22, 158, 159
BER. *See* Bit error ratio
BERT. *See* Bit error ratio tester
Bit error ratio, 19
Bit error ratio tester, 169, 177
Bragg reflector, 65–67

C

Capacitance
  gate, 78, 84, 87, 103, 109, 170, 175
  gate-source, 175
  input, 30, 36, 53, 74, 79, 80, 88, 108, 110, 144
  interconnect, 109
  junction, 30, 109
  load, 99, 104, 109, 119, 173
  matching, 80, 108
  Miller, 74, 104, 106, 126
  parasitic, 35, 59, 87, 100, 108
Cascode, 76, 77, 80, 81
Cherry-Hooper, 77, 80, 103–106, 108
CID. *See* Consecutive identical digits
Clock recovery
  delay-locked loop, 137, 141
  gated oscillator, 142–145
  injection locking, 141
  multi-channel, 4
  oversampling, 136, 138
  phase interpolation, 138–141
  phase-locked loop, 128
  topologies, 137, 138, 161
CML. *See* Current mode logic
Common-Gate, 36, 76, 77, 80, 81, 86
Conductance
  output, 28, 31, 76, 77, 79, 81, 84, 87, 99, 100, 104
  trans, 28, 30, 31, 78, 79, 84, 87, 88, 90, 99, 106, 109, 173, 175
Consecutive identical digits, 17
Correlation, 50, 153
Current mode logic, 54, 146, 170–172

D

Dark current
  compensation, 85–86
Data
  encoded, 137, 159
  scrambled, 14, 17
DC-Balance, 17
Deep sub-micron, 31
Degeneration, 78, 80, 102
Delay-locked loop, 136, 138, 141
Design methodology, 143, 149, 150
DLL. *See* Delay-locked loop
Duty-cycle distortion, 23, 142, 170, 177

P. MULLER, Y. LEBLEBICI,
CMOS Multichannel Single-Chip Receivers for Multi-Gigabit Optical Data Communications,
DOI 10.1007/978-1-4020-5912-4, © 2007 Springer

E

Edge
  data, 22, 45, 49, 127, 140, 142, 147–149, 151–153, 158
  detection, 18, 19, 130, 145, 146, 152, 169
  detector, 146, 147, 152, 169
  missing, 148
Encoding, 4, 14, 17, 18, 85, 134, 160
Equalization, 63, 64, 163
Estimation, 49, 89, 119, 158
Eye
  closure, 21, 33, 43–51, 54
  diagram, 20–22, 64, 122, 123, 151, 166, 169, 173, 176

F

Fiber
  core, 26, 59
  diameter, 26, 59
  monomode, 6, 26
  multimode, 7, 14, 26
Frequency tolerance, 4, 143, 149, 150, 161
fT-Doubler, 80
FTOL. *See* Frequency tolerance

G

Gain, 30, 36, 38, 40, 42, 44, 97, 99
Gain-Bandwidth, 29, 33, 38, 41, 42, 46, 50, 51, 53, 61, 76, 81, 88, 98, 100, 106, 107
Gaussian distribution, 22–24, 85, 162
Group delay, 78, 79, 98

I

Inductive
  coupling, 112, 115, 120
  peaking, 80, 95, 107, 112, 116, 120, 123
Inductor
  active, 101
  spiral, 112, 113, 115, 120
Injection locking, 136, 137, 141, 143
Interstage scaling, 95, 107–109, 111
Intersymbol interference, 20, 21, 42–45, 98, 122, 124
Intrinsic region, 59

J

Jitter
  bounded uncorrelated, 23
  clock, 151, 161, 179
  data dependent, 23, 163
  deterministic, 23, 34, 48, 49, 53, 153, 163, 177
  model, 154, 163
  periodic, 23
  random, 22, 23
  sinusoidal, 25, 34, 151, 154–161, 164
Jitter generation, 24, 25
Jitter tolerance, 24, 25, 34–36, 43, 49, 50, 53, 54, 130, 133
Jitter transfer, 24, 25, 132, 133, 142, 154
JTOL. *See* Jitter tolerance

L

Laser, 1, 7, 10, 14, 33, 65, 68, 71
Limiting amplifier
  bandwidth, 43, 95, 96, 99, 112
  gain, 54
  topologies, 43, 54, 97
Low-voltage differential signaling, 3, 32, 33

M

Measurement, 92, 93, 121–125, 176
Metal-semiconductor-metal, 58, 70
Methodology. *See* Design methodology
Miller
  amplifier, 86
  compensation, 86
  negative capacitance, 103, 108, 126
Model
  behavioral, 149, 161–163
  jitter, 154, 163
  noise, 39
  statistical, 150–161
Modulation, 4, 9, 13–16, 26, 34

N

Noise
  amplitude, 47, 124
  model, 39
Non-return-to-zero, 4, 15
NRZ. *See* Non-return-to-zero

O

Offset
  compensation, 90, 117–118
OMA. *See* Optical modulation amplitude
On-Chip Interconnects, 10
Optical modulation amplitude, 15, 34, 45, 54
Optimization, 4, 45, 50, 54, 95, 98, 108, 109, 136

Oscillator
current-controlled, 144, 145
gated, 142, 143, 145, 146, 149–152, 170, 174
ring, 137, 138, 141, 146, 150
voltage-controlled, 128
Oversampling, 138, 139, 141

P

Phase detector
Bang-bang, 133, 134
binary, 133–135
Hogge, 131
linear, 131–133
Phase interpolation, 139–142
Phase noise, 129, 152, 172
Phase rotator, 141
Phase-locked loop, 4, 19, 128, 129, 133
Photodetector
avalanche, 60
differential, 61, 62
MSM, 58, 70
PIN, 59–60
resonant cavity enhanced, 65–69
vertical, 61
Photodiode. *See* Photodetector
Plesiochronous, 26, 143
PLL. *See* Phase-locked loop
PRBS. *See* Pseudo-random bit sequence
Process, 3, 13, 17, 30–33, 44, 53, 54
Pseudo-random bit sequence, xix, 163, 164, 166, 176, 178
Pseudo-random, 163

Q

Q-factor, 47

R

Reflection, 60
Regulated cascode, 76, 77, 80, 81
Resonant cavity, 65–67, 71
Return-to-Zero, 4, 15, 16
Run length, 17, 135, 137, 155, 164, 166
RZ. *See* Return-to-Zero

S

Scaling
technology, 107, 108, 111
theory, 174–175
Scrambling, 14, 17
Sensitivity
receiver, 34–36, 50–53
spectral, 68
Signal limiting, 95
Silicon-on-insulator, 4, 67, 136
Spectrum, 1, 16, 18, 40, 98
Synchronous, 6

T

Time constant, 3, 59–61, 79, 89–91, 164, 166
Transconductance
bulk, 28
source, 28, 79
Transimpedance amplifier
bandwidth, 42, 48, 50
gain, 38
topologies, 74, 76–81
Transit frequency, 30, 53, 93, 100, 106, 173

W

Wander. *See* Baseline wander
Wavelength, 1, 6, 7, 13, 24, 53, 58, 60, 61, 65–68, 70