Springer Series in Advanced Microelectronics 54

Mohamed Atef
Horst Zimmermann

Optoelectronic Circuits in Nanometer CMOS Technology

Springer

Springer Series in Advanced Microelectronics

Volume 54

The Springer Series in Advanced Microelectronics provides systematic information on all the topics relevant for the design, processing, and manufacturing of microelectronic devices. The books, each prepared by leading researchers or engineers in their fields, cover the basic and advanced aspects of topics such as wafer processing, materials, device design, device technologies, circuit design, VLSI implementation, and subsystem technology. The series forms a bridge between physics and engineering and the volumes will appeal to practicing engineers as well as research scientists.

More information about this series at http://www.springer.com/series/4076

Mohamed Atef · Horst Zimmermann

Optoelectronic Circuits in Nanometer CMOS Technology

Springer

Mohamed Atef
Faculty of Engineering
Assiut University
Assiut
Egypt

Horst Zimmermann
Institute of Electrodynamics, Microwave and Circuit Engineering
TU Wien
Vienna
Austria

ISSN 1437-0387 ISSN 2197-6643 (electronic)
Springer Series in Advanced Microelectronics
ISBN 978-3-319-27336-5 ISBN 978-3-319-27338-9 (eBook)
DOI 10.1007/978-3-319-27338-9

Library of Congress Control Number: 2015960396

Printed on acid-free paper

This Springer imprint is published by SpringerNature
The registered company is Springer International Publishing AG Switzerland

Preface

Highly integrated communication systems are required to fulfil the growing demand for higher data rates in telecommunication networks. The optical fiber links are the best candidates to deal with large volumes of data since they provide superior performance compared to conventional electrical links in terms of bandwidth, channel loss, electromagnetic interference, reflection, and crosstalk. Optical receivers and transmitters are known to be the most important building blocks in optical communication systems. But since the scaling of CMOS structure sizes, clock frequencies of digital logic grew tremendously and the chip area necessary for logic functions decreased dramatically. Digital signal processing, equalization, error correction and the physical layers causing a lot of overhead became much more important. Therefore there is a general trend to integrate optical receivers and transmitters with a lot of digital circuitry together reducing their price in large volume production considerably. Low-cost and high performance optical receivers are required for high data rate telecommunication networks.

For plastic optical fiber receivers, fully integrated optical receivers with integrated silicon photodiodes provide advantages over hybrid implementations, including low-cost, reduced parasitic capacitance and no bond-wire inductance. Nanometer CMOS technologies have been rapidly advanced, enabling the implementation of integrated optical receivers for data rates of several Giga-bits per second. In particular, low-cost silicon CMOS optoelectronic integrated circuits become very attractive because they can be extensively applied to short-distance optical communications, such as local area network, chip-to-chip and board-to-board interconnects.

The different chapters in this book give a brief overview of the optoelectronics applications for long-haul optical communication systems, short distance optical communication like fiber to the home, in-home network and optical interconnects. CMOS optoelectronics for short distance optical communications and optical interconnect will be also discussed.

Optical sensor technology is another growing field of application for nanometer optoelectronic CMOS circuits. One very important field is image sensors, where the

pixel count of camera chips grew up to more than 10 megapixels thanks to the shrinking of CMOS structure sizes. New applications are time-of-flight based distance sensors also needing small-sized transistors for processing the distance information. Optically based medical investigation methods like positron emission tomography (PET) and magnetic resonance imaging (MRI) require a high spatial resolution, i.e. many pixels, and complex signal processing fostering nanometer CMOS circuits.

In the beginning of this book, Chap. 1 introduces the motivation for using optoelectronic integrated circuits in different applications like long and short distance optical communications, optical interconnects, image sensors, and medical applications.

Chapter 2 provides the description for optical communications fundamentals including optical communication building blocks, optical transmitter, and optical receiver. Also optical, data formats, binary data formats, multilevel signaling, DC balance code, eye diagram, bit error rate (BER), sensitivity, noise models, bandwidth and rise/fall times, intersymbol interference (ISI), jitter, nonlinearity, power penalty, dynamic range will be discussed.

Chapter 3 includes the mathematical models and physical prosperities for photodiodes, optical absorption, photocurrent generation, carrier diffusion, carrier drift, photodiode capacitance, photodiode bandwidth, quantum efficiency, internal quantum efficiency, optical quantum efficiency, photodiode responsivity, photodiode dark and noise currents, as well as photodiode small-signal and noise equivalent circuit models.

Chapter 4 introduces discrete photodiodes for visible light and infrared light. Also the photodetectors connected via bond wires or via flip-chip technique are provided.

Chapter 5 provides different types of integrated photodiodes in nanometer CMOS technologies like classical PN junctions, double-junction photodiodes, PW/DNW/P-substrate double photodiode P+/NW/P-sub avalanche double photodiode, P+/NW/P-substrate photodiode with guard, finger photodiodes, PIN photodiode, spatially modulated light detector, triple junction photodetector, and avalanche photodiodes. The end of this chapter dedicates itself to a comparison of the performance of the different photodiodes.

Chapter 6 discusses transimpedance amplifiers (TIAs) and their gain, bandwidth, and noise. The effect of manufacturing technology and different TIAs topologies are provided including simplest preamplifier, open loop TIAs, common gate input stage, regulated-cascode TIA, inverter based common-drain feedback TIA, and shunt-shunt feedback TIA. Frequency response, noise analysis of shunt feedback TIA, noise of ideal TIA, TIA with common-source input stage, multistage inverter based CMOS TIA, noise canceling TIA, inverter based cascode TIA, and differential TIA are covered. Advanced techniques for gain control gain, compression, bandwidth enhancement techniques for TIAs, super-gm technique, inductive peaking, active inductive peaking, and negative capacitance will be discussed.

Chapter 7 discusses equalizer types including passive equalizer, active equalizer, source degeneration, continuous time linear equalizer (CTLE) with multi-shunt-shunt feedbacks, inductive load equalizer, adaptive equalization, and continuous time adaptive equalizer. Discrete time adaptive equalizer, continuous time FIR filter implementation, discrete time FIR filter implementation, nonlinear equalization, decision feedback equalizer (DFE), maximum likelihood sequence estimator (MLSE) are described.

Chapter 8 presents circuit descriptions, results for post amplifiers, cascaded gain stages, differential post amplifier, amplifier with automatic gain control, limiting amplifier, offset compensation, broad band amplifier techniques, cherry-hooper amplifiers, interleaved active feedback, and transit-frequency doubler.

Chapter 9 introduces laser and modulator drivers, LEDs, specifications, rise and fall times, modulation current, extinction ratio, turn-on delay (TOD), output voltage (compliance voltage), laser driver circuit design, pre-driver, output driver, high voltage laser driver, laser automatic power control, and modulator drivers.

The last Chap. 10 gives circuit descriptions and experimental results of opto-electronic circuits in nanometer CMOS technology, fully integrated optical receivers in 180 nm CMOS, in 65 nm CMOS, and in 40 nm CMOS, as well as infrared optical receivers with external photodiode in 90 nm CMOS and in 40 nm CMOS. Optical sensors, 2D image sensors, 3D image sensors, and medical sensors are also introduced.

The authors thank Dr. Ascheron from Springer for initiating this book and his team for technical support with the text processor in good cooperation.

Assiut, Egypt — Mohamed Atef
Vienna, Austria — Horst Zimmermann
2016

Contents

About the Authors

Mohamed Atef received the B.Sc. and M.Sc. degrees in electrical engineering, electronics and communications from Assiut University, Egypt, in 2000 and 2005 respectively. From 2006 to 2007 he was a researcher in Czech Technical University in Prague, Department of Microelectronics, improving the optical properties of quantum dots. He received his Ph.D. in 2010 from Vienna University of Technology, Institute of Electrodynamics, Microwave and Circuit Engineering, and then worked at this institute as a post-doctoral researcher to the end of 2012. Since September 2015, M. Atef is a visiting Professor to BiCASL Lab, Shanghai Jiao Tong University, China. He is currently an Associate Professor in Assiut University, Egypt. His research interests are in the area of integrated circuits, opto-electronic integrated circuit, and communications over plastic optical fiber. He is an author of the Springer book 'Optical Communication over Plastic Optical Fibers: Integrated Optical Receiver Technology'. Furthermore, he is author and co-author of more than 45 scientific publications. Since 2012 he is senior member IEEE.

Horst Zimmermann received the Dr.-Ing. degree in the Fraunhofer Institute for Integrated Circuits (IIS-B), Erlangen, Germany in 1991. Then, Dr. Zimmermann was an Alexander-von-Humboldt Research-Fellow at Duke University, Durham, N.C., where he worked on diffusion in Si, GaAs, and InP. In 1993, he joined the Chair for Semiconductor Electronics at Kiel University, where he lectured optoelectronics and worked on optoelectronic integration. Since 2000 he is full professor for Electronic Circuit Engineering at Vienna University of Technology, Austria. His main interests are in design

and characterization of analog deep-sub-micron, nanometer CMOS and optoelectronic integrated (Bi)CMOS circuits as well as electronic-photonic integration. He is author of the two Springer books 'Integrated Silicon Optoelectronics' and 'Silicon Optoelectronic Integrated Circuits' as well as co-author of 'Highly Sensitive Optical Receivers', 'Optical Communication over Plastic Optical Fibers: Integrated Optical Receiver Technology', 'Analog Filters in Nanometer CMOS' and Comparators in Nanometer CMOS Technology. Furthermore, he is author and co-author of more than 450 scientific publications. Since 2002 he is senior member IEEE.

Chapter 1
Why Optoelectronic Circuits in Nanometer CMOS?

Highly integrated communication systems are required to fulfill the growing demand for higher data rates in telecommunication networks. The optical fiber links are the best candidates to deal with large volumes of data since they provide superior performance compared to conventional electrical links in terms of bandwidth, channel loss, electromagnetic interference, reflection, and crosstalk. Optical receivers and transmitters are known to be the most important building blocks in optical communication systems. But since the down-scaling of CMOS structure sizes, clock frequencies of digital logic grew tremendously and the chip area necessary for logic functions decreased dramatically. Digital signal processing, equalization, error correction and the physical layers causing a lot of overhead became much more important. Therefore there is a general trend to integrate optical receivers and transmitters with a lot of digital circuitry together reducing their price in large volume production considerably. Low cost and high performance optical receivers are required for high data rate telecommunication networks. Fully integrated optical receivers with integrated silicon photodiodes provide advantages over hybrid implementations, including low-cost, reduced parasitic capacitance, no bond-wire inductance. Nanometer CMOS technologies have been rapidly advanced, enabling the implementation of integrated optical receivers for data rates of several Giga-bits per second. In particular, low-cost silicon CMOS optoelectronic integrated circuits become very attractive because they can be extensively applied to short-distance optical communications, such as local area network, chip-to-chip and board-to-board interconnects. The next sections in this chapter give a brief overview of the current long-haul optical communication systems, short distance optical communication like fiber to the home, in-home network and optical interconnect. CMOS optoelectronics for short-distance optical communications and optical interconnect will be also discussed. Optical sensors is another growing field of application for nanometer optoelectronic CMOS circuits. One very important field are image sensors, where the pixel count of camera chips grew up to more than 10 Megapixels thanks to the shrinking of CMOS structure

M. Atef and H. Zimmermann, *Optoelectronic Circuits in Nanometer CMOS Technology*, Springer Series in Advanced Microelectronics 54,
DOI 10.1007/978-3-319-27338-9_1

sizes. New applications are time-of-flight based distance sensors with high pixel counts also needing small-sized transistors for processing the distance information. Optically based medical investigation methods like PET and MRI require a high spatial resolution, i.e. many pixels, and complex signal processing fostering nanometer CMOS circuits.

1.1 Long-Haul Communication

The silica glass optical fiber (GOF) has three transmission windows, around the wavelengths 850 nm, 1.31 and 1.55 µm, Fig. 1.1 [1]. The single-mode fiber (SMF) has a loss of about 0.3 dB/km at the 1.31µm wavelength. Because the lower loss of 0.2 dB/km at 1.55 µm this wavelength is preferred for long-haul communication. The first transmission window around 850 nm is used for short reach optical communication due to its higher loss compared to 1.55 and 1.31 µm. For traditional fiber, centered at approximately 1310 nm is a window of 200 nm and the total bandwidth in this region is about 25 THz. Centered at 1550 nm is a window of similar size ($\Delta\lambda = 200$ nm), which consists of three bands, S-band (1460–1530 nm), C-band (1530–1560 nm), and L-band (1560–1630 nm). Combined, these two transmission windows provide a theoretical bandwidth of 50 THz [2]:

$$BW = \frac{c}{\lambda^2}\Delta\lambda \tag{1.1}$$

where $c = 3 \times 10^8$ m/s.

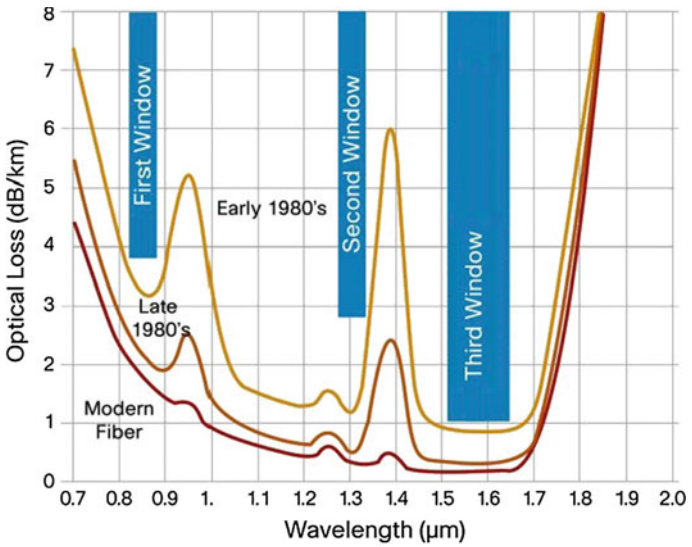

Fig. 1.1 Attenuation versus wavelength and transmission windows [1]

By 1980 the long distance optical fiber communication cables under the sea were tested. The optical fiber has a very low loss allowing long distances between amplifiers/repeaters. Optical communication links are capable of sending 1.28 Tbit/s data over a distance of up to 4000 km, without any electrical regeneration and a 2.4 Tb/s WDM signal (120 channels × 20 Gb/s) was transmitted over 6200 km [3]. An ultra-low loss high frequency coaxial copper cable operating at 40 GHz has an attenuation of about 2100 dB/km and a capacitance of 80 nF/km [4]. Due to the copper cables' high loss, communication systems using copper cables longer than one kilometer require signal repeaters for satisfactory performance. Many copper cables are required to replace one fiber cable to achieve the same data rate distance product. The high data rate distance product of the optical fiber is due to the absence of high loss, capacitive loading, inductive effects, and of cross talk common to long parallel electric conductor lines.

1.2 Fiber to the Home (FTTH)

Often 1.55 μm(0.2 dB/km) and 1.3 μm(0.3 dB/km) wavelengths are used for fiber to the home (FTTH). The 850 nm wavelength, however, also can be used for short-range data communication like FTTH, although its higher loss compared to 1.55 and 1.3 μm wavelengths disqualifies 850 nm for long-haul applications. For 850 nm wavelength, the commercial low-cost vertical-cavity surface-emitting lasers (VCSEL) can be used as transmitters. The loss of a multi-mode GOF is about 2.5 dB/km for a wavelength of 850 nm. These loss values are very low compared to the copper wire losses. For example, a coaxial cable has a typical loss of 500 dB/km for a signal at 10 GHz. A typical multi-mode fiber has a bandwidth distance product of 2 GHz· km at 850 nm whereas a Cat-6 unshielded twisted pair (UTP) has a bandwidth distance product of only 20 MHz· km [5] (Fig. 1.1). The most important driving forces for FTTH are the broadband applications such as IPTV, videophone, and online gaming which need more bandwidth than can be achieved by electrical links. The time taken to transfer large files (10 Gbit) using various access technologies like ADSL and VDSL is around 80 h and 50 min respectively. So, there is a limitations of ADSL and VDSL access network technologies to move such large data files over a few meters. Only FTTH provides realistic transfer times of 6 min compared to the other technologies [6]. In the near future, due to the large required data rate VDSL will be replaced by FTTH [6]. The biggest barrier for massive FTTH deployment is that the costs of FTTH systems are still high compared with VDSL. VDSL equipment costs are very low and VDSL equipment does not need installation of new copper or fiber cables, and the bandwidth provided by VDSL is enough for current applications. The transition from copper wire to optical fiber systems to migrate to FTTH cannot be achieved until new forms of technology capable of delivering high data rates with substantial cost reductions are developed. FTTH needs hardware and non-hardware improvements, such as transmission system components, network design, construction, operation, and maintenance [6]. The development of high performance low cost optical receivers is one of the important needs to enable low cost FTTH.

1.3 In-Home Network

In-building networks are using a wide range of transmission media like coaxial copper cables, twisted copper pair cables, free-space optical links, wireless links, etc. Each of these networks is optimized for a particular set of services; this complicates the introduction of new services and the creation of links between services (such as between video and data services). A single broadband multi-services network could provide an efficient solution to host and connect all existing and upcoming services together. The target for the data rate delivered to homes could be up to 1.25 Gbit/s in case of fiber to the home (FTTH) or up to 125 Mbit/s in case of very high bit rate digital subscriber line (VDSL2) technology [7]. The in-building network must not be a bottleneck for the future broadband services such as high definition internet protocol television (IPTV), multi-room and multi-vision configuration and high definition video transfer via the TV set. The home network can be used to share multimedia contents stored in a storage medium inside the house for offline future use [7]. Figure 1.2 shows a view for the next generation of home networks [8]. Copper links like twisted pair and coaxial cables are used to deliver telecom services within the building. For the future required broadband services the copper transmission medium suffers from serious shortcomings. Twisted pair has a limited bandwidth and is susceptible to electromagnetic interference (EMI). Coaxial cables have a larger bandwidth, but large thickness and the large effort required to establish a reliable connection introduces practical problems during installation. Moreover, the coaxial cable is not perfectly immune to EMI and has a certain attenuation. Optical fibers are used for long-haul communication and they represent an alternative for short-reach transmission inside building because of their higher bandwidth and they offer complete immunity to EMI [9, 10].

Glass optical fibers (GOF), however, are not suitable for use in building network because of the high cost involved for precise handling. Also the high bandwidth of the GOF is never necessary in this short distance application. On the other hand, it is important to have very simple and low-cost solutions. The cheap Poly-Methyl-Methacrylate Plastic Optical Fiber (PMMA-POF) is an excellent candidate for implementing such a short distance network. POF systems provide benefits compared to GOF and copper wire, which include simpler and less expensive components, operation in the visible range (the transmission windows are 530, 570 and 650 nm), greater flexibility and resilience to bending, shock and vibration, ease in handling and connecting (standard step-index POF core diameters are 1 mm compared to 8–100 μm for glass fibers), use of simple and inexpensive test equipment. Finally, POF transceivers require less power than copper transceivers. These advantages make POF very attractive for use within in-building networks [10]. To overcome the problem of the POF's high transmission loss very sensitive receivers must be used to increase the transmitted length over PMMA POF. There are two methods to solve the SI-POF's problem of limited bandwidth: the first method is to use multilevel signaling like Multilevel Pulse Amplitude Modulation (M-PAM) and Multilevel Quadrature Amplitude Modulation (M-QAM) [11]. Also the data rate limited by the POF's bandwidth can

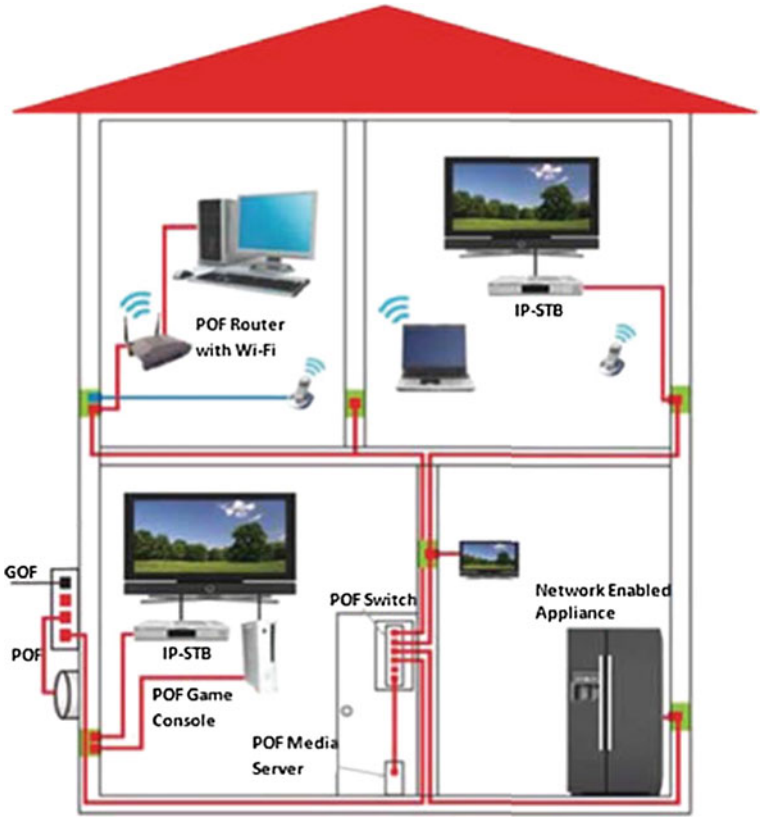

Fig. 1.2 The next generation of home network [8]

be increased by using spectral efficient modulation techniques like Discrete Multi-Tone (DMT) [12–15]. In the second method, equalization techniques can be used to compensate for the SI-POF limited bandwidth. Fixed or adaptive equalizers can be used for pre/post-equalization with different digital or analog equalization methods to increase the POF's bandwidth [11].

1.4 Optical Interconnects

Until now, electrical transmission is preferred for low data rates ($\leq$10 Gb/s) and short distance communication systems ($\leq$tens of meters) because of its lower material, electrical transmitters and receivers costs. The data rate distance product of electrical communication links is limited by physical loss coming from dielectric and skin effect losses of electrical interconnection, inductive and capacitive effects

involved with the wire line, and cross talk between adjacent lines. For the required future higher bandwidths, the wire-line interconnects are approaching the theoretical channel capacity limit defined by Shannon theory [16]. For example, the widely used FR-4 copper trace material has a loss around 2.0–3.0 dB/inch at 12.5 GHz. The insertion loss and return loss increases with respect to the data rate. At 25 Gbps the insertion loss is far beyond the equalization dynamic range. The current electrical transmission technology cannot scale to 25 Gbps unless additional components, such as a repeater, which adds more power, cost, and complexity, are used [17]. The data rate of electrical lines is limited to:

$$B \leq B_o \frac{A}{L^2} \tag{1.2}$$

where A is the wire cross-sectional area, L is the wire length, and B_o is a constant = 10^{16} to 10^{18} b/s for off chip lines [18]. For high data rate and long wire length the required wire cross-section area should be increased. This will lower the line density and limit the maximum number of lines for parallel transmission. According to Moore's law the speed is doubled every 18 months. As data rates and wiring density is increasing, the difficulties of electrical interconnection through wires are increasing. The optical fiber has a negligible loss compared to the electrical wire. So, optical links are attractive for relatively long lines at high data rates and limited cross-sections [18]. As the data rate increases the microprocessor I/O pin numbers increase. The optical interconnect can provide the solution to the large number of high data rate I/O pins and minimize the power consumption per pin. The electrical signal charges the wire line to the signaling voltage. The total energy is:

$$E_s \geq C_l {V_r}^2 \tag{1.3}$$

where C_l is the line capacitance and V_r is the signaling voltage. The longer the wire is the larger the capacitance and the higher the energy [18]. For an optical link, the optical energy required to charge the total capacitance C_d of the photodiode and the TIA input capacitance is [18]:

$$E_P \geq C_d V_r \frac{h\omega}{e} \tag{1.4}$$

where the voltage $\frac{h\omega}{e}$ is numerically equal to the photon energy in electron-volts. There is potentially an energy benefit for optical link when [18]:

$$C_d V_r \frac{h\omega}{e} < C_l V_r^2 \tag{1.5}$$

The optical connections will have a lower dissipated energy than the electrical connection for long distance and low photodiode capacitance. The required future higher data rates and lower power consumption make the optical interconnect competitive with electrical connect, leading to board-to-board, chip-to-chip and on chip

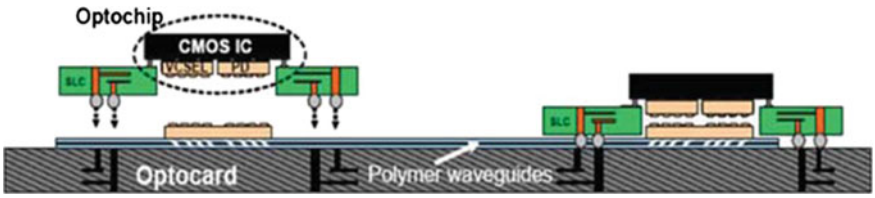

Fig. 1.3 Low cost optical printed circuit board with polymer waveguides [20]

optical communications. Optical links provide higher bandwidth for board-to-board and chip-to-chip on PCB connections, this enables higher connection density, lower latency, reduced power consumption and electromagnetic noise [19]. An example for low cost optical printed circuit boards (Polymer Waveguides) is shown in Fig. 1.3 [20]. On chip optical interconnects offer a potential solution for addressing global wiring issues. Optical Input/Output (I/O) could be a potential first application. On chip interconnects are not expected to replace Cu wires on low-k dielectrics in the lower layers of the on-chip interconnect stack because of delay and power considerations [21].

1.5 Optical Receivers

Long-haul optical communications works at 1.55 and 1.31 μm wavelengths which cannot be detected by silicon photodiodes and InGaAs photodetectors are needed. For in-building optical communication 650 nm laser diodes/VCSELs are well-suited for plastic optical fiber (POF) links. Large photodiodes with responsivities of about 0.5 A/W at 650 nm can be fabricated in cheap silicon technology. Short reach optical communication using cheap 850 nm VCSELs can use silicon photodetectors with data rates less than several Gb/s and responsivity smaller than hundreds of mA/W. For higher data rates and responsivity GaAs photodetectors are needed. The photonic layer needed for the III/V optoelectronic devices can be integrated following two different ways. In the first approach, wire-bonding or flip-chip bonding, the photonic devices are fabricated separately and bonded to the electronics chip. In the second approach, the photonic devices are integrated heterogenously on the circuits by wafer bonding or die to wafer bonding.

However, optical components' costs are still too high for these high volume markets of short-reach interconnects. Low cost, high volume CMOS compatible fabrication processes are required to enable that photonic connections replace the electrical connections. Implementation of optical interconnects for signaling, clock distribution and I/O requires development of a number of optical components like photodetectors and light sources [19]. The optical receiver costs can reduce by using a low-cost semiconductor process. Photodetectors and transistors in GaAs technology achieve superior performance, but GaAs processes are quite costly and they attain

Table 1.1 Time of optical communications commercial deployment (copper displacement) [20, 22]

	WAN	LAN	Rack-Rack	Card-Card	On PCB	Chip-Chip	On Chip
Length	Multi-km	<1 Km	<100 m	>1 m	Several cm	10 s of mm	10 s of μm
Links	1	1–10	10 s	100 s	1000 s	10 Ks	100 Ks
Date	80s	90s	2000	2010	>2012	>2015	>2020

only limited integration scales. Silicon processes are cheap and allow a high density of integration. An integrated receiver with a monolithically integrated photodetector has economical advantage regarding to the solution with an external photodiode. In single-chip solutions the photodiode bond pad and the receiver input bond pad are not present and since parts of the receiver circuit can be placed around the photodiode, the single chip solution usually requires considerable less chip area than the sum of chip areas of the two-chip solution specially for smaller-size photodiodes. The mass production is one of the drivers to reduce the integrated system cost. Table 1.1 shows the time of optical communication's commercial deployment (copper displacement) [20, 22]. It is clear from Table 1.1 that the large volumes of optical interconnects (PCB, chip to chip, on chip) will be the main drivers to reduce the costs of OEICs.

Figure 1.4 is shown to demonstrate the technical advantages of an optoelectronic integrated circuit (OEIC) [23]. The solution of using a packaged photodiode and a packaged receiver IC soldered to a printed circuit board (PCB) is shown in Fig. 1.4a. CPAD1 is the capacitance of the bond pad of the photodiode. LB1 is the inductance of the bond wire inside the photodiode package. Additionally there are the capacitance of the pin of the photodiode package CPIN1, the capacitance of the copper line on the printed circuit board CLINE and the capacitance of the pin of the receiver package CPIN2. LB2 is the inductance of the bond wire inside the receiver package.

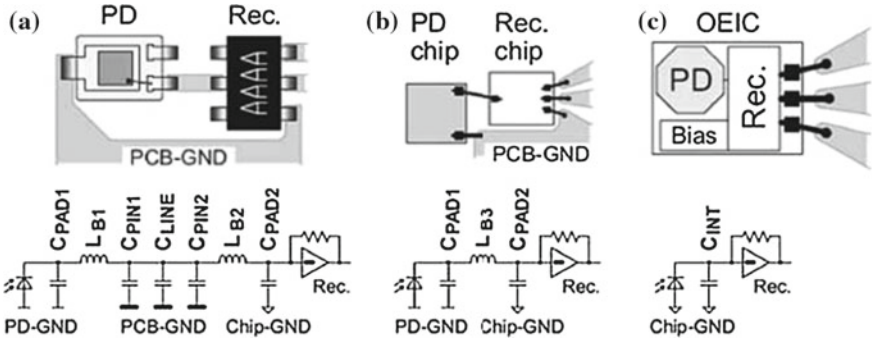

Fig. 1.4 Different ways of connecting the PD and the receiver: **a** packaged photodiode connected to packaged receiver, **b** photodiode die connected to receiver die via bond wires, **c** fully integrated optical receiver [23]

The capacitance of the bond pad at the receiver input is CPAD2. The mentioned parasitic capacitances will add to the junction capacitance of the photodiode and will reduce the bandwidth and worsen the noise behavior of the optical receiver. The parasitic inductances of the bond wires may lead to unwanted gain-peaking in the frequency response of the receiver input stage. Since the photodiode and the parasitic capacitances are not referenced to the same ground as the receiver, it is possible, that noise signals (e.g. switching noise) might be picked up from the ground lines.

The number of parasitic components is much smaller, if both the photodiode package and the receiver package are omitted (Fig. 1.4b). The photodiode chip and the receiver chip are glued to the printed circuit board. The photodiode is connected to the receiver with a bond wire (LB3) directly. The grounds of the photodiode chip and the receiver chip can be connected to the PCB ground with bond wires (as shown in the figure) or the chips can be glued to the PCB with a conductive glue if the chips provide a suitable backside contact. Photodiode and amplifier can be connected analoguously when they both are mounted on a lead frame within one package instead on a PCB.

If the photodiode is integrated to the receiver chip, almost all parasitic components at the input node of the receiver are avoided (Fig. 1.4c). The only remaining component is the parasitic capacitance of the interconnection line between the integrated photodiode and the input stage (C_{INT}). Usually, this capacitance is negligibly small. Therefore neither the bandwidth nor the sensitivity of the optical receiver is degraded by parasitic components. Additionally, the avoidance of bond wires minimizes the risk, that external electromagnetic fields or currents in neighboring bond wires couple noise to the input node of the transimpedance amplifier. Figure 1.5 shows the die photo of a fully integrated optical receiver with integrated photodiode for application together with POF.

Photodetectors can either be integrated on-die, bonded to a CMOS die, or on-package. Ge-based Metal-Semiconductor-Metal (MSM) and PIN diode photodetectors have received significant attention since they have the potential to be CMOS compatible. The key parameters of photodetectors are: high responsivity, high bandwidth at reverse bias less than or equal to 1 V and, stable operation up to 100 °C [19, 21]. Light sources can be off-die (package or board), bonded to a photonics die, or integrated on-die (typically adding III–V capabilities to a CMOS platform). Quantum dots nanophotonic material and high index contrast structures may play a strong role in future optical interconnections. To overcome the indirect silicon band-gap to achieve silicon lasers, silicon nanocrystals are employed. The requirements are high switching speeds. Total power requirements depend on the application, but are typically below 1W. Currently, the key concerns include wavelength stability, reliability at operating conditions and cost of laser arrays [19, 21].

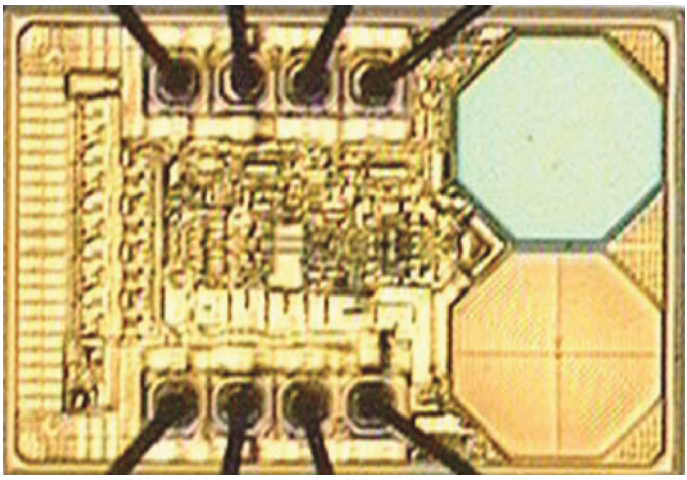

Fig. 1.5 Die photo for fully integrated optical receiver with integrated photodiode [11]

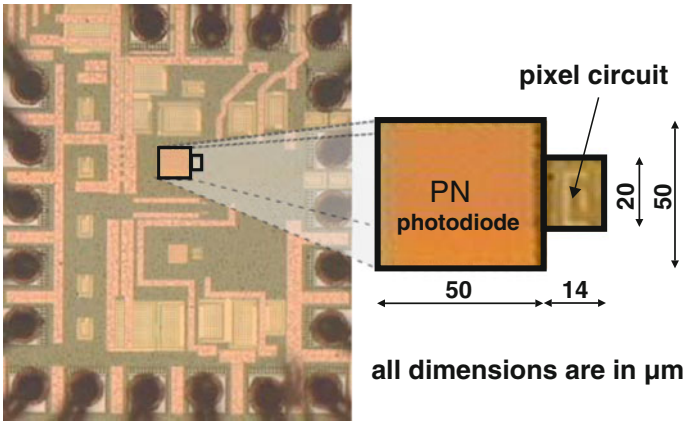

Fig. 1.6 Die photo of a distance sensor pixel in 90 nm CMOS [26]

1.6 Optical Sensors

Optical sensors span a wide range of applications. The ones being most relevant to nanometer CMOS OEICs are image sensors, time-of-flight 3D cameras, and optical sensors for medical applications. In these three applications a high spatial resolution is required. Advanced CMOS APS image sensors contain more than 10 million pixels with a pixel of about 2 µm. Such a small pitch requires very small transistors within the APS pixel, which aims into the nanometer scale CMOS processes. Time-of-flight 3D cameras record not only a two-dimensional image like the CMOS APS imagers.

These 3D cameras also determine the distance to the object point corresponding to each pixel simultaneously. Therefore, they need more transistors in each pixel than the APS pixels. The photodiode area within the 3D pixel has to be larger than that in a APS pixel in order to collect more light, which is necessary to achieve a good distance measurement accuracy being limited by the signal-to-noise ratio. Therefore, small transistors are needed aiming also towards nanometer scale CMOS. Figure 1.6 shows one pixel of a distance sensor in 90 nm CMOS. Up to 40,000 pixels have been realised in one 3D-sensor chip [24]. The newest 3D vision chip contains 2.1 million pixels [25].

Medical diagnostics requires highly sensitive photodetectors and a high spatail resolution. One application is positron electron tomography, where scintillation crystals are used to detect positrons. Such a crystal emits photons in the visible spectral range, which can be detected by silicon photodetectors. Furthermore, a high time resolution is needed implying a complex high-speed signal processing, which aims towards nanometer CMOS OEICs [27].

References

1. White Paper, *Fiber Types in Gigabit Optical Communications* (Cisco Systems, 2006), pp. C11–463661-00
2. B. Mukherjee, *Optical WDM Networks, Chapter 2* (Springer, New York, 2006)
3. G.P. Agrawal, *Fiber-Optic Communication Systems*, 3rd edn. (Wiley, New York, 2002)
4. http://www.gigalink-mce.net
5. M.S. Filip Tavernier, *High-Speed Optical Receivers with Integrated Photodiode in NanoscaleC-MOS*. (Springer, NewYork, 2011)
6. C. Lin (ed.), *Broadband Optical Access Networks and Fiber-to-the-Home Systems Technologies and Deployment Strategies, Chapter 8* (Wiley, Chichester, 2008)
7. ETSI TS 105 175–1 V1.1.1(2010–01), Access, Terminals, Transmission and Multiplexing (ATTM); Plastic Optical Fibre System Specifications for 100 Mbit/s and 1 Gbit/s (2010). http://www.etsi.org/WebSite/homepage.aspx
8. http://fiberopticpof.com
9. H.P.A. van den Boom, W. Li, P.K. van Bennekom, I.T. Monroy, G.D. Khoe, High-capacity transmission over polymer optical fiber. IEEE J. Sel. Top. Quantum Electron. **7**(3), 461–469 (2001)
10. P. Polishuk, Plastic optical fibers branch out. IEEE Commun. Mag. **44**(9), 140–148 (2006)
11. M. Atef, H. Zimmermann, *Optical Communication over Plastic Optical Fibers: Integrated Optical Receiver Technology* (Springer, Berlin, 2013)
12. R. Gaudino, E. Capello, G. Perrone, G. Perrone, M. Chiaberge, P. Francia, G. Botto, advanced modulation format for high speed transmission over standard SI-POF using DSP/FPGA platforms. in *POF Conference 2004*, (Nuerberg, 2004), pp. 98–105
13. F. Breyer, S. Lee, S. Randel, N. Hanik, PAM-4 signalling for gigabit transmission over standard step-index plastic optical fibre using light emitting diodes, in *34th European Conference and Exhibition on Optical Communication (ECOC 2008)*, vol. 3 (Brussels, Belgium, 2008) pp. 81–82
14. S.C.J. Lee, F. Breyer, D. Cardenas, S. Randel, A.M.J. Koonen, Real-time gigabit DMT transmission over plastic optical fibre. Electron. Lett. **45**(25), 1342–1343 (2009)
15. S.C.J. Lee, F. Breyer, S. Randel, R. Gaudino, G. Bosco, A. Bluschke, M. Matthews, P. Rietzsch, H.P.A. van den Boom, A.M.J. Koonen, Discrete multitone modulation for maximizing

transmission rate in step-index plastic optical fibers. J. Lightwave Technol. **27**(11), 1503–1513 (2009)
16. J.G. Proakis, M. Salehi, *Fundamentals of Communication Systems*. (Pearson Prentice Hall, 2005)
17. White Paper, *Overcome Copper Limits with Optical Interfaces*. (Altera Corporation, 2011) pp. WP-01161-1.1
18. D.A.B. Miller, Device requirements for optical interconnects to silicon chips. Proc. IEEE **97**(7), 1166–1185 (2009)
19. A European Roadmap for Photonics and Nanotechnologies. Published by the MONA consortium (2008)
20. F. Doany, Power-efficient, high-bandwidth optical interconnects for high performance computing. in *Hot Interconnects conference*, (Santa Clara, 2012)
21. International Technology Roadmap for Semiconductors (ITRS): Interconnect (2011)
22. ZRL I/O Link Technology Group, ZRL Photonics Group, Optical Interconnects: Intra-system Data Transfer with Light. Foil-set for Internet-download and General Media Usage, (IBM, Zurich, 2005)
23. M. Fortsch, Monolithically Integrated Optical Receivers for Low-Cost Data Communication and Optical Storage Systems. Ph.D. Dissertation, Vienna University of Technology, 2007
24. P. Tech, Pmd tech 41k-s datasheet. PMD Tech. http://de.pluspedia.org/wiki/PMDTechnologies. Accessed Jan 2016
25. S. Koyama, K. Onozawa, K. Tanaka, Y. Kato, A 3D 2.1 Mpixel image sensor for single-lens camera systems, in *IEEE International Solid-State Circuits Conference(ISSCC 2013)*, (San Francisco, USA, 2013) pp. 492–493
26. M. Davidovic, G. Zach, K. Schneider-Hornstein, H. Zimmermann, Range finding sensor in 90 nm CMOS with bridge correlator based background light suppression. in *ESSCIRC*, (Seville, 2010), pp. 298–301
27. L. Braga, L. Gasparini, L. Grant, R. Henderson, N. Massari, M. Perenzoni, D. Stoppa, R. Walker, An 8 × 16-pixel 92kSPAD time-resolved sensor with on-pixel 64ps 12b TDC and 100 MS/s real-time energy histogramming in 0.13 μm CIS technology for PET/MRI applications. *IEEE International Solid-State Circuits Conference(ISSCC 2013)*, (San Francisco, USA, 2013), pp. 486–487

Chapter 2
Optical Communications Fundamentals

The necessary fundamentals for the analysis and design of optical communication links will be introduced in this chapter. The transmitter, receiver (transceiver) and optical fiber channel for optical communication systems will be discussed. The features of binary and multilevel data formats will be shown. Measuring the quality of the random signal using the eye diagram will be explained. The different noise sources in an optical receiver will be introduced. Binary and multilevel formulas for the bit error rate (BER) will be calculated and receiver sensitivity will be the defined. The effect of bandwidth limited systems on random data, intersymbol interference (ISI), and jitter will be discussed.

2.1 Optical Communication Building Blocks

2.1.1 Optical Transmitter

Figure 2.1 shows a block diagram for an optical communication system. On the transmitter side a time-division multiplexer (MUX) combines N parallel low data rate streams into a fast serial data stream with N times higher bit rate. A laser driver drives the laser diode or a modulator driver drives a modulator. The laser driver modulates the current of a laser diode (LD). The modulator driver modulates the voltage across a modulator, which in turn modulates the light intensity from a continuous wave laser. Some laser drivers/modulator drivers require a clock at the data rate (or half rate) to perform data retiming [1]. The laser diode coverts the electrical signal to an optical one. The optical signal emitted from the laser is coupled to an optical fiber and travels over the fiber to reach the optical receiver.

M. Atef and H. Zimmermann, *Optoelectronic Circuits in Nanometer CMOS Technology*, Springer Series in Advanced Microelectronics 54,
DOI 10.1007/978-3-319-27338-9_2

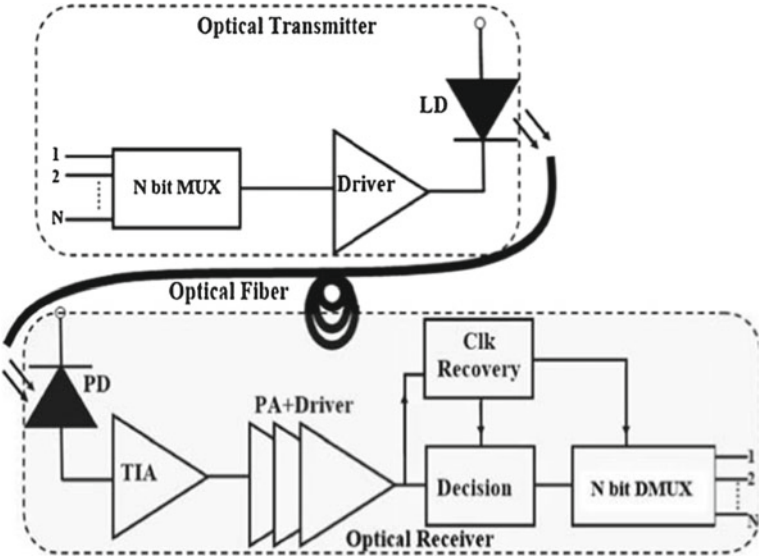

Fig. 2.1 Block diagram for an optical communication system

2.1.2 Optical Receiver

On the receiver side, the photodetector receives the optical signal coupled from the optical fiber and converts it to an electrical current signal. The current signal generated by the photodetector is usually small due to the fiber attenuation after long distance and the low responsivity for the photodetector in nanometer CMOS technology. The photodetector is connected to a transimpedance amplifier (TIA) to convert and amplify the photocurrent into a voltage signal. The TIA should give high gain, high bandwidth and low-noise. A TIA with automatic gain control (AGC) can be used to enable the TIA to work at high input photocurrents; increasing the dynamic range. The post amplifier (PA) amplifies the signal coming from the TIA to give enough output voltage swing for the decision circuit as well as the clock and data recovery. The TIA and PA require high bandwidth from near DC to the maximum data rate with low jittering. The output driver gives enough current to drive the following circuits and to make impedance matching. The PA can be a limiting amplifier (LA) where the output voltage swing of the amplifier is kept constant independent of the input signal level which is working with binary data formats. The other type of the PA is a linear amplifier where the output voltage signal is proportional to the input signal; the linear amplifier is required for analog signals. The clock recovery circuit retrieves the clock from the data which is needed for the decision circuit and DMUX. The decision circuit regenerates the high quality data to reduce the effects of limited bandwidth and timing jitter produced by the cascaded amplifiers. The DMUX splits the serial high data rate stream back into the original N parallel low bit rate streams. The MUX

and DMUX are digital integrated circuits which must work at high data rate with low power dissipation. The clock recovery circuit does not require broadband operation like TIA and PA [2]. The next chapters will discuss in detail the design of optical receivers and transmitters and their performance.

2.1.3 Optical Fiber

Optical fibers have a core from silica with refraction index n1 surrounded by a cladding material with a lower index of refraction (n2) as shown in Fig. 2.2. Light is kept in the core by total internal reflection. Coating is a material that is put over the cladding of an optical fiber to protect it from the environment. The buffer material (generally plastic) is used as a jacket to protect optical fiber from physical damage. Unlike the cladding, the jacket is physically distinct from the fiber core. In the next subsection the fiber dispersion and loss will be discussed.

2.1.3.1 Modal Dispersion

Step-Index Multimode Fiber: Due to its large core, some of the data light rays may travel a direct route, whereas others travel in a zigzag way. Different paths cause the different groups of light rays (modal dispersion) to arrive separately at the receiving point. The transmitted pulse begins to spread out, losing its well-defined shape Fig. 2.3a. There is a need to leave a spacing between data pulses to prevent ISI which

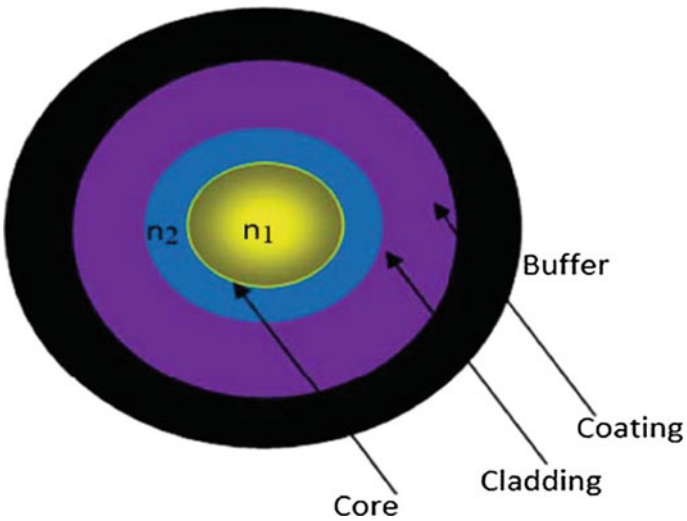

Fig. 2.2 Optical fiber structure

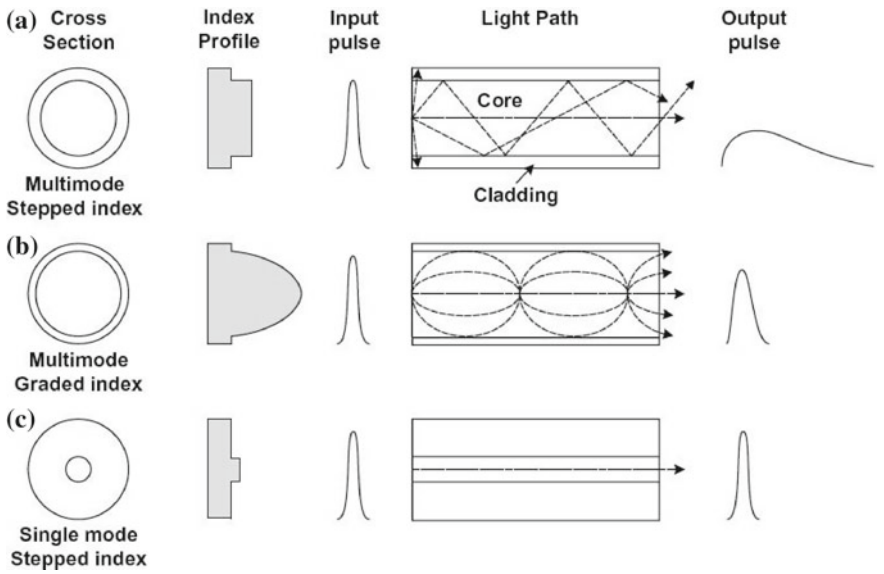

Fig. 2.3 Characteristics of three basic types of optical fibers [3]

limits the amount of information that can be sent per second. The transmission bandwidth is normally given as bandwidth-length-product. The limiting bandwidth-distance product for the step-index MMF fiber can be approximated by [4]:

$$BW.L \leq \frac{n_2 C}{{n_1}^2 \Delta} \tag{2.1}$$

where $\Delta = (n1 - n2)/n1$ is the fractional index change at the core-cladding interface. Clearly, Δ should be made as large as possible in order to couple maximum light into the fiber. However, a very small Δ (less than 0.001) is required for the purpose of optical communications to reduce the modal dispersion. A transmission bit rate—distance product of about 100 Mbps.km is typical for step index multimode fibers. This type of fiber is best suited for short reach communications. The PMMA SI-POF is other type of MMF which has a core from PMMA larger in diameter (1 mm) than the silica MMF core. Due to the large difference in the fractional index ($\Delta = 0.08$) the typical bandwidth for 100 m of PMMA SI-POF is 60 MHz.

Graded-Index Multimode Fiber: This type of fiber contains a core in which the refractive index decreases gradually from the center towards the cladding. The higher refractive index at the center makes the light rays near the axis slower than those near the cladding. Due to the graded index, light in the core curves smoothly off the cladding rather than the sharp zigzag in case of the step-index fiber, reducing the travel distance, see Fig. 2.3b. The shortened path and the higher speed allow light at the periphery to arrive at about the same time as the slow but straight rays in the core axis. The output pulse suffers from less dispersion than the step-index fiber.

The limiting bit rate—distance product obtained for the graded-index MMF fiber can be approximated by [4]:

$$BW.L \leq \frac{8C}{n_1 \Delta^2} \tag{2.2}$$

The graded index multimode fibers show bit rate—distance products less than 10 Gbps.km. This type of fiber is best suited for last mile connections like in FTTH.

Single mode Fiber: The single-mode fiber has a small core diameter of about 4–10 μm. The change of the refraction index between core and cladding is less than for multimode fibers. The effects of pulse distortion due to different travel times of modes do not take place for single-mode fibers (no modal dispersion), see Fig. 2.3c. The single-mode fibers show transmission bit rate—distance products around 100 Gbps.km. Thus single mode fibers are well suited for long-haul optical communications.

2.1.3.2 Chromatic Dispersion

The refractive index of GOF is a function of the light wavelength, so the spectral components of a transmitted pulse are traveling with different velocities along the fiber. Hence chromatic (material) dispersion broadens optical pulses beyond their time slot, leading to intersymbol interference (ISI). The group velocity associated with the fundamental mode is frequency dependent because of chromatic dispersion. As a result, different spectral components of the pulse travel at slightly different group velocities, a phenomenon referred to as group-velocity dispersion (GVD), intramodal dispersion, or simply fiber dispersion [4]. The effect of dispersion on the bit rate—distance product can be estimated by:

$$BW.L \leq \frac{1}{D \Delta \lambda} \tag{2.3}$$

where D is called the dispersion parameter and is expressed in units of ps/(km·nm). For standard single-mode GOF, D is relatively small in the wavelength region near 1.3 μm [D ≤ 1 ps/(km·nm)]. For a semiconductor laser, the spectral width $\Delta\lambda$ is around 2–4 nm. The bit rate—distance product of such optical communication systems can reach 100 Gbps. km [4].

2.1.3.3 Loss

As an optical signal propagates over a fiber, it is attenuated because of material absorption, Rayleigh scattering, and other effects. Material absorption can be divided into two types. Intrinsic absorption losses correspond to absorption by material used to make fibers, and the extrinsic absorption which relates to losses caused

by impurities within the fiber material. The Rayleigh scattering is a fundamental loss mechanism arising from local microscopic fluctuations in density. Fiber loss is a limiting factor which reduces the signal power reaching the optical receiver. The maximum transmission distance is limited by fiber losses. The optical receivers need a certain minimum amount of received optical power for generating a high quality electrical signal [4]. Single-mode (SM) GOF has a 4–9 μm core diameter and three transmission windows, around the wavelengths 850 nm, 1.31 and 1.55 μm, (see Fig. 1.1). SMF has a loss of about 2.1 dB/Km@850 nm, 0.3 dB/km @ 1.31 μm wavelength, and the lowest loss is 0.2 dB/km@1.55 μm. Multimode (MM) GOF has a 50 or 62.5 μm core diameter and two transmission windows, around the wavelengths 850 nm and 1.31 μm at 850 nm the MMF has a loss of about 2.5 and 3.5 dB/km for the 50 and 62.5 μm core diameter, respectively. For 1.31 μm wavelength the attenuation of MMF is 0.8 and 1.4 dB/km for the 50 and 62.5 μm core diameter, respectively. In general, we can say the smaller the core diameter the lower the attenuation. Due to its low transmission loss the SMF is used for long-haul communication whereas the higher loss MMF is used for short distance communication [4, 5]. The 1 mm core PMMA POF has transmission windows at 520 nm (0.07 dB/m), 570 nm (0.06 dB/m) and 650 nm (0.14 dB/m). POF systems provide benefits compared to GOF, which include simpler and cheaper components, operation in the visible range, greater flexibility and resilience to bending, shock and vibration, ease in handling and connecting (standard step-index POF core diameters are 1 mm compared to 9–62.5 μm for glass fibers), use of simple and inexpensive test equipment [6–8].

2.2 Data Formats

2.2.1 Binary Data Formats

The non-return-to-zero (NRZ) and return-to-zero (RZ) formats are the most commonly used modulation formats in optical communication systems, see Fig. 2.4. For NRZ the signal is high to transmit ONE bit and is low to transmit a ZERO bit. When the signal is high, it stays high for the entire bit period T. The inverse of the bit period is the data rate. The transmitted data is random in nature and the ONE and the ZERO occur typically with an equal probability. For RZ the format, shown in Fig. 2.4b, the bits ONE, occupy only half of the bit period. The RZ format requires less signal to noise ratio compared with the NRZ. Its is also more immune to fiber imperfections; as RZ has narrower pulse width; so pulse spreading has a negligible effect on the adjacent bits. On the other hand, RZ has a higher bandwidth because of its shorter pulses; so a higher bandwidth circuitry is needed for RZ.

A PRBS sequence is always used to simulate and measure optical communication circuits because it is difficult to generate a really random binary signal. The PRBS is used as a fictitious data waveform. A typical generated PRBS sequence has a lengths vary between $2^5 - 1$ and $2^{31} - 1$. Also the PRBS sequence is used to scramble the real

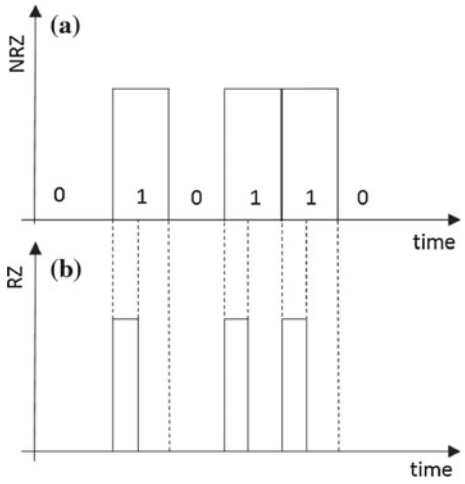

Fig. 2.4 Binary data formates **a** NRZ and **b** RZ

data before transmission. Data scrambling eliminates long sequences of the same bit value. A long sequence of the same bit value causes baseline wander, also the CDR circuitry cannot extract the correct clock signal [9].

2.2.2 Multilevel Signaling

The use of multilevel signaling has applications in increasing the information bandwidth over optical fiber links, thus obviating the necessity to replace fibers whose capacity is fully utilized at a given binary bit rate. The communication end-to-end delay can be minimized and the reduction in signal to noise ratio (SNR) caused by multilevel signaling is not significant over short distances of optical fiber. By using multilevel modulation the cost and complexity are reduced by using one channel compared to binary signal over multiple channels. By using multilevel signaling over a single channel instead of multiple channels, there is no attenuation due to wavelength multiplexing at the transmitter and demultiplexing at the receiver, no multiple optics precision alignment variances, no reliability issues associated with individual channel failures and no skew management and associated delay. Wavelength Division Multiplexing (WDM) requires N lasers whereas multilevel signaling reduces the system cost by using just one similar laser. The principle of multilevel signaling is to use a larger alphabet of M symbols to represent data, so that each symbol can represent more than one bit of data. As a result, the number of symbols that need to be transmitted is less than the number of bits (the symbol rate is less than the bit rate, hence the bandwidth is compressed. Figure 2.5 shows an example for a four-level scheme. If M is the number of distinct signal levels, then each symbol

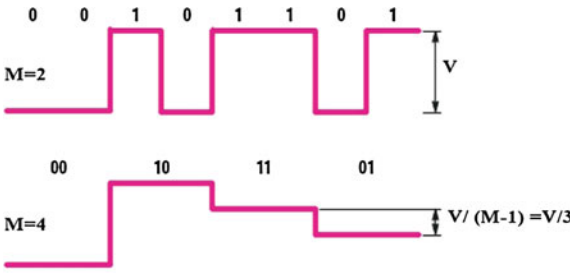

Fig. 2.5 Binary and multilevel signaling

now carries $N = log_2(M)$ bits of information, and the overall data rate (DR) rises to [10]:

$$DR = R_s log_2(M) \tag{2.4}$$

where R_s is the symbol rate. No additional bandwidth is required for this increase. The increased information rate comes either at the expense of added transmitter power or an increased error rate at the receiver. The multilevel signaling scheme is more sensitive to non-linearities and noise than the binary schemes. There is an extra power penalty by using M-levels compared to a binary signal [10]:

$$Power\ Penalty = 10log_{10}(M-1)dB \tag{2.5}$$

Such amplitude margin loss leads to the need for increased receiver sensitivity in addition to receiver complexity to decode the incoming multilevel signal. This increases the receiver physical area and costs.

2.3 DC Balance Code

Most of the optical transmission systems do not pass the DC-component because of AC-coupling. In a standard binary NRZ transmission system, AC-coupling generates the unwanted baseline wander. This wander in the resulting signal is due to the removal of the signal DC components below the low cut-off frequency. Baseline wander is particularly evident for long sequences of ones or zeros. One of the most commonly used DC balance code is the 8B/10B coding, described in details in [11]. The 8B/10B code maps each possible 8 bit sequence into new 10 bit sequence providing a good DC-balance. The DC balance coding, besides solving the AC-coupling problem, also eases clock recovery; as it avoids long sequence of bits without transitions. The negative effect of coding is to increase the transmitted data rate by 1.25 times the actual data rate, for example 1 Gbps will be coded to 1.25 Gbps.

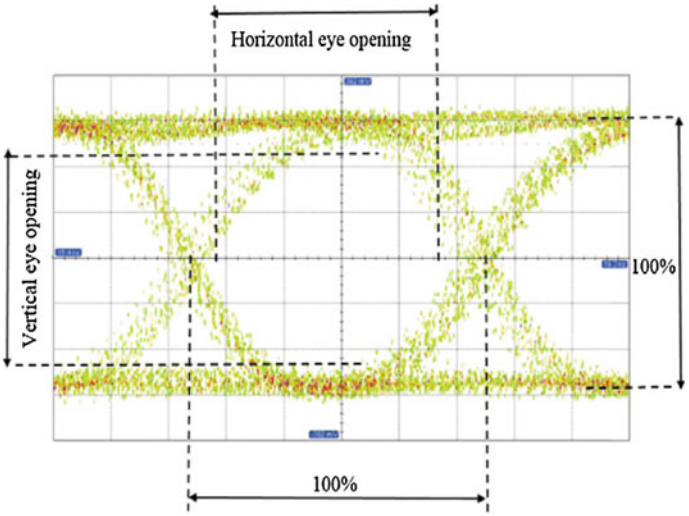

Fig. 2.6 Eye diagram

2.4 Eye Diagram

In an eye diagram, different bit segments with all possible bit transitions are superimposed as shown in Fig. 2.6. For 3 bits of data the PRBS length is $2^3 - 1$. The superimposed segment possibilities are 00X, 01X 10X, 11X. The advantage of the eye diagram over the linear signal representation is that all possible bit transitions can be displayed in a compact representation. Signal quality can be judged from the eye shape. The vertical and the horizontal eye opening are important characteristics of the eye diagram to aid in measuring the signal quality. The vertical eye opening is measured at the sampling instant (center of the cycle) and is expressed as a percentage of the full eye height. The horizontal eye opening is measured at the slice level and is expressed as a percentage of the bit interval. The vertical eye closure is caused by noise and ISI, and the horizontal eye closure is caused by deterministic jitter, pulse width distortion, and noise. The eye closure depends also on the PRBS length because the receiver has a low cutoff frequency. The eye closure will become worse with increasing the PRBS length [1].

2.5 Bit Error Rate (BER)

A detailed analysis for the symbol error rate (SER) can be found in [12]. Consider M amplitude levels centered on zero, with $M = 2^l$. If the M-PAM signals are represented geometrically as M one dimensional signal point values:

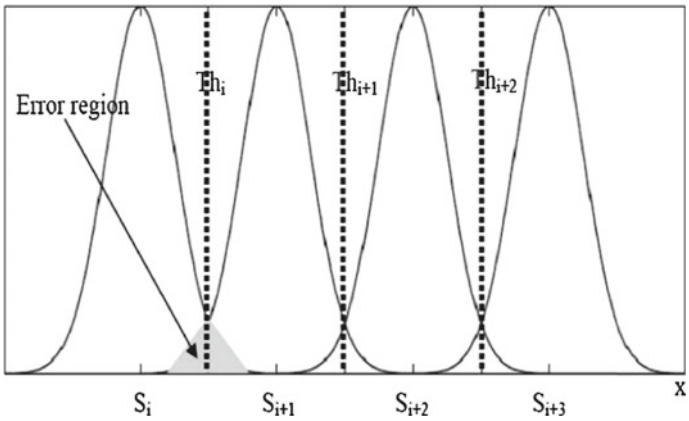

Fig. 2.7 M-PAM signals geometrical representation as M one dimensional signal point values

$$s_m = \sqrt{0.5 E_g A_m}, m = 1, 2, ...M \tag{2.6}$$

where E_g is the energy of the basic signal pulse.

Placing the threshold levels as shown in Fig. 2.7 helps in evaluating the probability of errors. We note that if the $M^t h$ amplitude level is transmitted, the demodulation output will be:

$$r = s_m + n \tag{2.7}$$

The noise variable n has zero mean and variance $\sigma^2 = 2N_o$ and all amplitude levels are equally likely a priori. The average probability of a symbol error is simply the probability that the noise variable n exceeds in magnitude one-half of the distance between levels. The average symbol energy is given by:

$$E_{av} = \frac{(M^2 - 1)}{6} d^2 E_g \tag{2.8}$$

where the Euclidean distance between adjacent signal points is $d\sqrt{2E_g}$. The SER can be written in terms of average symbol energy (2.8):

$$SER = \frac{2(M-1)}{M} Q\sqrt{\left(\frac{6E_{av}}{(M^2-1)N_o}\right)} \tag{2.9}$$

It is customary to plot the probability of the symbol error (SER) to use SNR per bit as the basic parameter, $E_{av} = Log_2(M) E_{bav}$ where E_{bav} is the average bit energy:

$$SER = \frac{2(M-1)}{M} Q\sqrt{\left(\frac{6Log_2(M)E_{bav}}{(M^2-1)N_o}\right)} \tag{2.10}$$

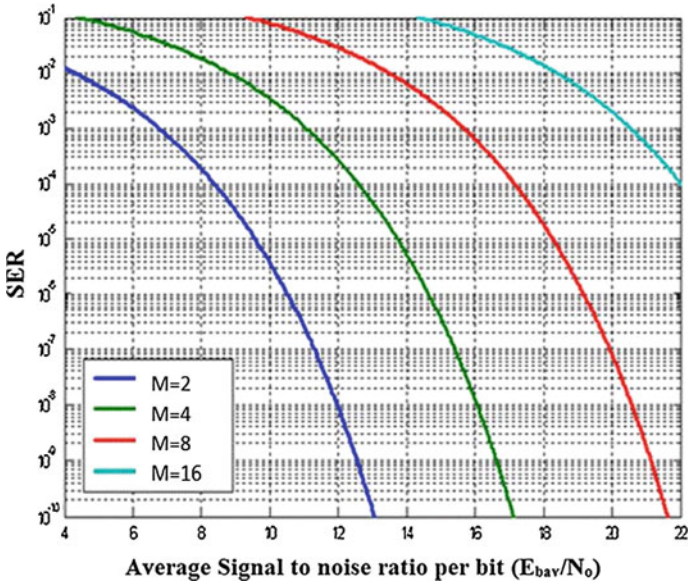

Fig. 2.8 Symbol error rate for M-PAM signal

BER can be obtained by dividing SER by the number of bits per symbol $Log_2(M)$. The SER presented in (2.10) is illustrated in Fig. 2.8 as function of average signal-to-noise ratio per bit (E_{bav}/N_o).

For binary signals the number of levels M is equal to 2. The quality factor Q which measures the quality of the signal is defined by:

$$Q = (V_H - V_L)/(\sigma_H + \sigma_L) \tag{2.11}$$

where V_H and V_L are the signal amplitude for high and low value, respectively, and σ_H and σ_H are their standard deviations. By using M = 2 and substituting (2.11) into (2.10) plus considering the function $Q(x) = 0.5erfc(x/\sqrt{2})$, the BER for a binary signal can be calculated by:

$$BER = 0.5erfc(\frac{Q}{\sqrt{2}}) \approx \frac{1}{\sqrt{2\pi}} exp\left(\frac{-Q^2}{2}\right)\left(1 - \frac{1}{Q^2}\right) \tag{2.12}$$

The BER as function of the quality factor Q (noise distance) is plotted in Fig. 2.9.

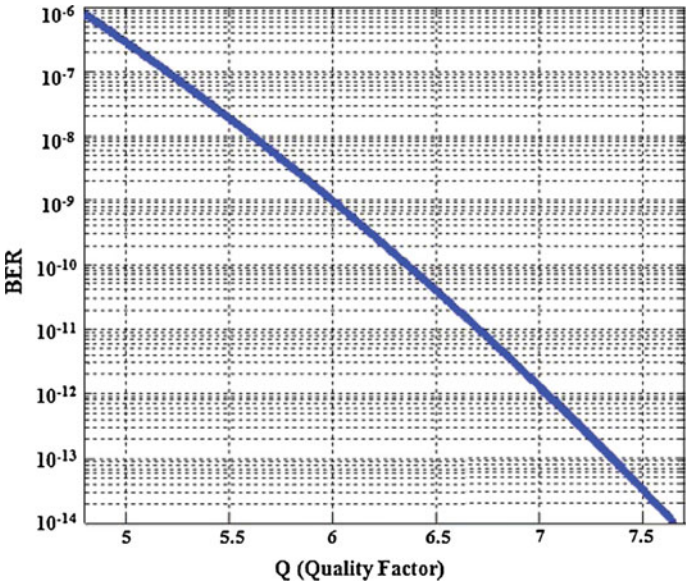

Fig. 2.9 Bit error rate as a function of the quality factor

2.6 Sensitivity

The sensitivity of an optical receiver is defined as the minimum average optical power (P_{sens}= $P_{av,min}$@BER) necessary to achieve a certain BER. The photodetector converts the optical power to photocurrent (I_{ph}):

$$I_{ph} = R \cdot P_{in,opt} \tag{2.13}$$

where R is the responsivity of the photodiode and expressed in A/W. The photocurrent is then amplified by the TIA gain and the PA gain to get the output voltage.

$$V_{out} = I_{ph} \cdot TIA_{gain} \cdot PA_{gain} = R \cdot P_{in,opt} \cdot TIA_{gain} \cdot PA_{gain} \tag{2.14}$$

The quality factor is calculated as the signal-to-noise ratio at the output:

$$Q = \frac{V_{out}}{V_{noise,out}} = \frac{R \cdot P_{sens} \cdot TIA_{gain} \cdot PA_{gain}}{I_{noise,rms} \cdot TIA_{gain} \cdot PA_{gain}} = \frac{R \cdot P_{sens}}{I_{noise,rms}} \tag{2.15}$$

$$P_{sens} = \frac{Q I_{noise,rms}}{R} \tag{2.16}$$

The quality factor calculated by (2.11) can be rewritten to be:

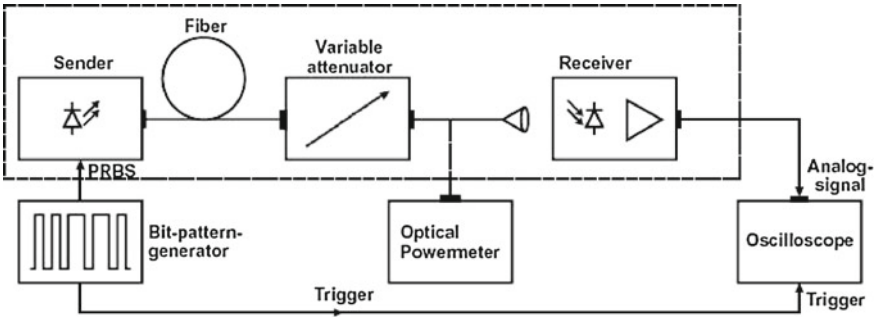

Fig. 2.10 Measurement set-up for sensitivity measurement using an eye diagram [13]

$$Q = \frac{V_{out,H} - V_{out,L}}{V_{out,noise,H} + V_{out,noise,L}} \tag{2.17}$$

After calculating the quality factor the BER can be calculated using the expression (2.12) for binary signals.

The quality factor Q can be measured using communication analyzers or digital sampling oscilloscopes. First the eye diagram is plotted using the set-up shown in Fig. 2.10. The communication analyzer determines the histograms shown in Fig. 2.11. The mean values $V_{out,H}$ and $V_{out,L}$ as well as their standard deviations ($\sigma_H = V_{out,noise,H}$ and $\sigma_L = V_{out,noise,L}$) can be read from the display of the communication analyzer. Finally the quality factor Q is calculated and displayed by the communication analyzer. You can calculate it manually using relation (2.17). The BER can be calculated using expression (2.12) for binary signals.

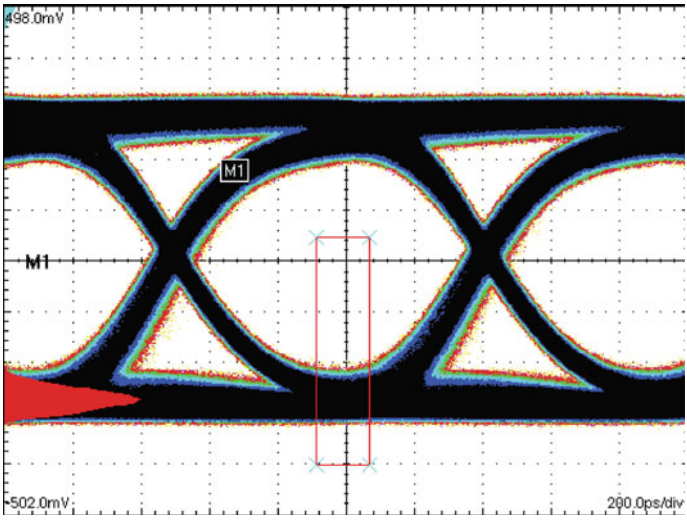

Fig. 2.11 Eye diagram with measured histogram

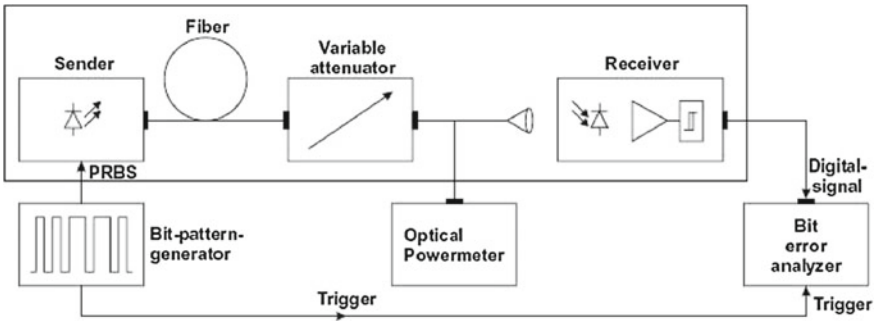

Fig. 2.12 Set-up for sensitivity measurements using a bit error rate tester [13]

A more accurate characterization of optical receivers with respect to BER can be done by using a bit error rate tester (BERT) than by the determination of Q via eye diagrams with a communication analyzer. The set-up shown in Fig. 2.12 can be used for a bit error rate measurements using a BERT. The BERT digitally compares the received bits with the sent bits and counts errors. A BERT can be used to calculate the BER only when the receiver under test has a digital output or with analog output voltage much larger than the sensitivity of BERT.

It is usual to calculate the sensitivity in dBm, the equation for calculating sensitivity for a certain BER is as follows:

$$Sensitivity = 10Log\frac{Pav,\,in}{1\,mW}\,dBm \tag{2.18}$$

For a complete analysis of the receiver sensitivity, the BER is measured for each average input optical power and plotted as in Fig. 2.13. The BER is decreasing by a factor of 1000 for each 1 dBm increase of the input optical power when there is no ISI. If there is ISI due to limited bandwidth or baseline wander, the BER decreases with a smaller rate than 1000/1 dB. BER may decrease with 100/1 dB or even 10/1 dB for strong ISI.

2.7 Noise Models

In this section noise models for photodiode, resistor noise, and MOSFET are presented. The main sources of noise are shot noise, thermal noise, and flicker noise.

Shot noise: is always coupled to a direct current flow and is present in diodes (photodiodes) and bipolar transistors (BJTs). There is no shot noise involved with MOSFETs. The external current through a diode seems to be a steady flow of current, but it is in fact the sum of different independent current pulses with an average current I_{av}. The fluctuation I is termed shot noise and is generally specified in terms of its

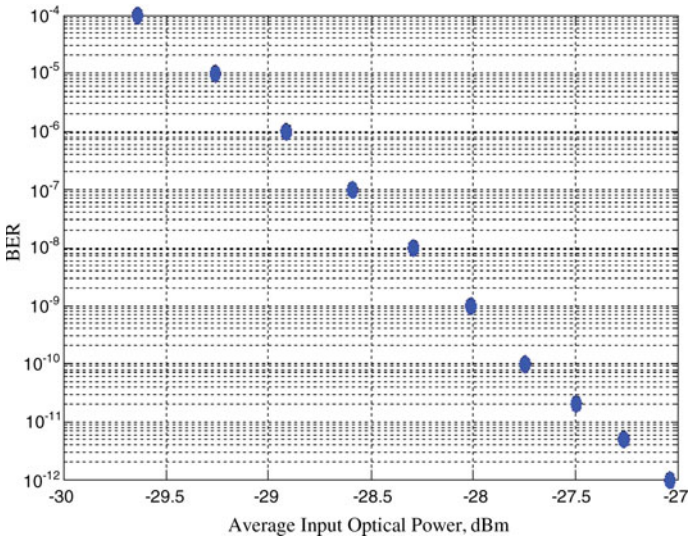

Fig. 2.13 BER as function of average input optical power

mean-square variation about the average value [14]. The shot noise current density has the average value:

$$\frac{{i_n}^2}{\Delta f} = 2q I_{av} \tag{2.19}$$

Thermal noise: is generated due to the random thermal motion of charge carriers in conventional conductors. It is unaffected by an existing or non existing direct current since the drift velocities in a conductor are much lower than thermal velocities of electrons. Therefore it is directly dependent on the temperature T. Thermal noise is diminished only if the temperature drops to absolute zero. Monolithic and thin-film resistors show thermal noise. The spectral noise current density for a resistor R can be calculated by:

$$\frac{{i_n}^2}{\Delta f} = \frac{4KT}{R} \tag{2.20}$$

The MOSFET channel's thermal spectral noise current density can be calculated by:

$$\frac{{i_n}^2}{\Delta f} = 4KT\gamma g_m \tag{2.21}$$

where g_m is the MOSFET transconductance and γ is the excess noise factor which equals to 2/3 for long-channel and increases to 2.5 for short-channel MOSFETs.

Flicker noise (1/f noise): is found in active devices like MOSFETs and BJTs. Flicker noise is caused mainly by traps related to contamination, crystal defects,

and due to Si/gate-insulator interface states in MOSFETs. Traps catch and release electrons in a random process and generate a noise with high levels at low frequencies [14]. Flicker noise is always related to direct current flow and shows a spectral density dependence on 1/f.

$$\frac{{i_n}^2}{\Delta f} = K_1 \frac{I^a}{f^b} \tag{2.22}$$

I is a direct current, K_1, a, and b are constants for a particular device. The values of these constants depend on different devices and technologies.

2.8 Bandwidth and Rise/Fall Times

The receiver bandwidth is the frequency at which the final gain will decreased by 3 dB. In practical applications the receiver's low cutoff frequency should be very low (to decrease the baseline wander); so it can be neglected compared to the high cutoff frequency. Thus the receiver bandwidth will equal to the high cutoff frequency, see Fig. 2.14.

The rise time (t_r) is the time the signal needs to rise (or fall for fall time(t_f)) from 10 to 90 % of its final value, see Fig. 2.15.

The pulse width (T) used to measure rise/fall time should be $1/(10 f_L) > T > 10/f_H$.

The maximum data rate for a certain rise and fall time is given by:

$$DR = \frac{1}{t_r + t_f} \tag{2.23}$$

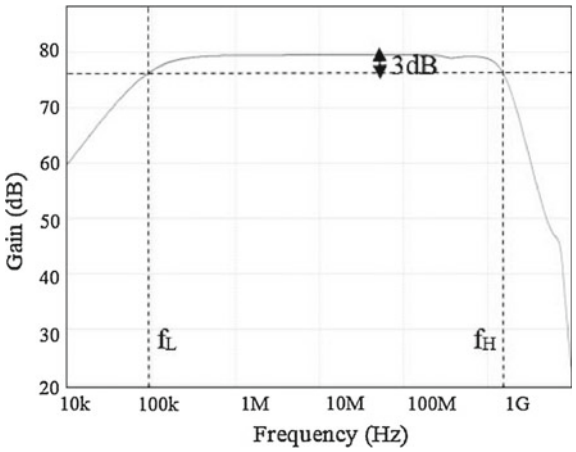

Fig. 2.14 Frequency response indicating low and high cutoff frequencies

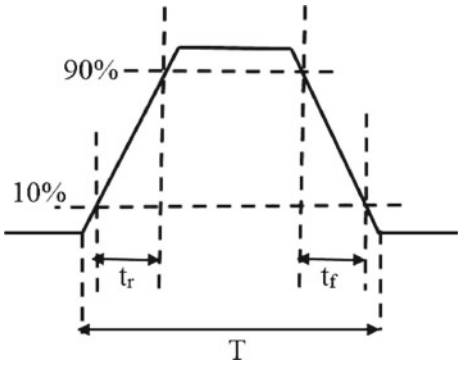

Fig. 2.15 Transient response indicating rise and fall times

The optimum system bandwidth to have a compromise between the required low noise and small ISI because of the bandwidth limitation is 0.76 of the DR for a first order circuit (one pole). So a useful relationship between rise and fall times and bandwidth is given by:

$$BW = \frac{0.76}{t_r + t_f} \tag{2.24}$$

In the case where the signal passes through cascaded transfer functions such as laser source, optical receiver under test, and measurement scope, the overall bandwidth can be approximated by:

$$BW_{total}{}^{-2} \approx BW_{laser}{}^{-2} + BW_{receiver}{}^{-2} + BW_{scope}{}^{-2} \tag{2.25}$$

The total rise/fall-time for cascaded transfer functions can be rewritten as:

$$t_{total}{}^{2} = t_{laser}{}^{2} + t_{receiver}{}^{2} + t_{scope}{}^{2} \tag{2.26}$$

This relation (2.26) is useful to calculate the actual rise/fall times of the receiver under test by subtracting the scope and laser diode rise/fall times.

2.9 Intersymbol Interference (ISI)

If the circuit bandwidth is lower than the bandwidth of the input signal, then the output signal rise/fall time and the amplitude will be affected. The rise/fall time will increase and the fast bits will have a tail that will affect the amplitude of the adjacent bits. The influence of adjacent bits on the amplitude of the output signal is called intersymbol interference (ISI). The ISI will reduce the eye opening and so will

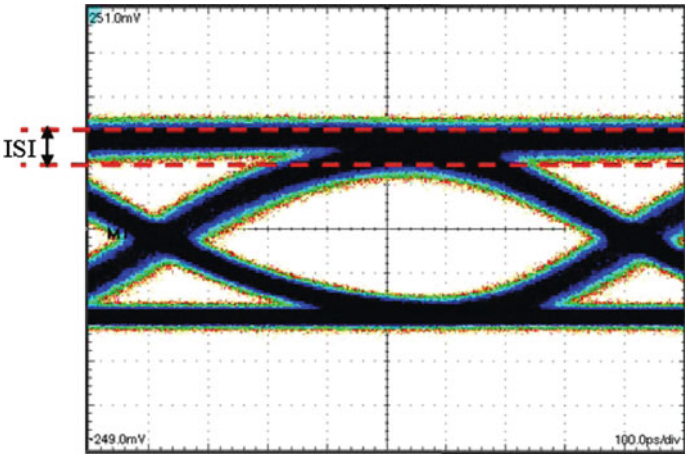

Fig. 2.16 Effect of intersymbol interference on eye diagram opening

increase the BER, see Fig. 2.16. To have zero ISI the circuit bandwidth should be equal to the input signal data rate, but this will be on the expense of more integrated noise. The optimum circuit bandwidth to get the lowest ISI and noise combination is 0.76 times the signal data rate [15].

The low cutoff frequency also introduces ISI because of the baseline wander effect. For large low-cutoff frequency the baseline wander will increase, which will reduce the eye diagram's opening because of the introduced ISI. The circuit's low-cutoff frequency should be five orders of magnitude lower than the signal data rate.

2.10 Jitter

The decision threshold voltage slices the eye diagram horizontally. Jitter is the deviations of the threshold voltage crossings from their ideal position in time, see Fig. 2.17. Jitter can influence the optimal sampling instant of the decision circuit. The jitter in the time domain is called phase noise in the frequency domain. As the jitter increases the horizontal eye opening is decreasing and a bit error can occur. There are two main jitter categories deterministic jitter and random jitter.

Deterministic Jitter: There are many types of deterministic jitter like data-dependent, bounded uncorrelated jitter, and duty-cycle distortion jitter. These type of jitter cause ISI and increase the BER. Data-dependent jitter is produced when the signal edge is shifted slightly in time, depending on the values of the adjacent bits. Data-dependent jitter is caused by an insufficient bandwidth, an insufficient phase linearity, baseline wander due to a large low-cutoff frequency, reflections on cables due to an impedance mismatch, and an overloaded TIA or post amplifier. The periodic jitter and bounded uncorrelated jitter, which can arise as a result of crosstalk from adjacent signal or power lines. The duty-cycle distortion jitter occurs if the

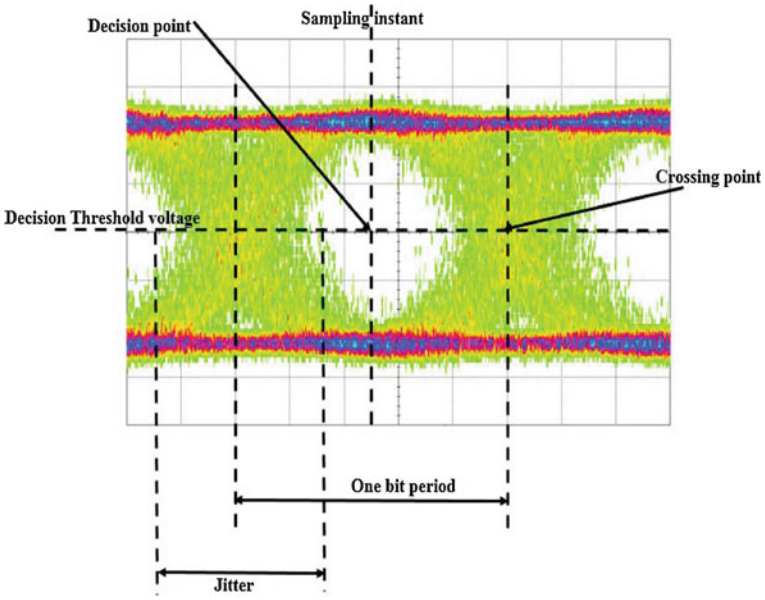

Fig. 2.17 Eye diagram with jitter

rising and falling edges do not cross each other at the decision threshold voltage, see Fig. 2.17 [1].

Random Jitter: Random jitter, is a random operation, which is not related to the data pattern. Noise in the time axis is known as random jitter. Random jitter is produced by noise on signal edges with a finite slew rate. The finite slew rate translates the signal voltage uncertainty into a timing (zero-crossing) uncertainty. Random jitter also is caused by carrier mobility variations due to instantaneous temperature fluctuations [1].

2.11 Nonlinearity

In optical communication systems employing directly modulated laser LD there is a distortion due to the LD nonlinearity in its input current versus output power curve. This nonlinearity has no effect on the performance when binary modulation is used, so nonlinearity compensation is not needed. A different situation appears when using systems based on analog signals or multilevel modulations, such as M-PAM. The effect of LD nonlinearity is compensated in the transmitter part, to compensate different LDs in different operating conditions.

For high input photocurrent the TIA can be overloaded and the output signal will be distorted. An AGC is needed to improve the TIA's linearity. For large input signals the gain is reduced to keep the TIA working in the linear region.

Post amplifiers normally show a linear transfer function for input signals below a critical input signal amplitude. For input signals above the critical input signal amplitude there appear nonlinearities. These nonlinearities make the decision operation very difficult for multilevel signals. For binary signals the limiting effect of the LA appears naturally and no special design is necessary. Nevertheless, limiting effects like pulse-width distortion and delay variations must be minimized by a controlled limiting function. These nonlinearity effects like gain compression, harmonic distortions, and intermodulation distortions are the main effects used to measure the nonlinearity of a certain circuit.

Gain compression: is the gain loss experienced by large signals relative to the small-signal gain. The broadband large-signal gain is decreasing by increasing the large-signal input value (gain compression). An important gain compression point is the 1 dB gain compression point. This is the input amplitude for which the gain is reduced by 1 dB compared to the small-signal gain [1].

Harmonic distortions: are caused by the harmonics generated because of the nonlinearities. The higher the generated harmonics level the higher the system nonlinearities. For a single-tone input signal as a sine wave with frequency f_o and amplitude A_o, the circuit output signal contains harmonic components at f_o, 2 f_o, 3 f_o, …etc., see Fig. 2.18.

The total nonlinearity can be described by total harmonic distortion (THD) [1]:

$$THD = \sqrt{HD_2{}^2 + HD_3{}^2 + \ldots} \tag{2.27}$$

where HD_2 is the second harmonic distortion, HD_3 is the third and so on.

The THD can be expressed as a percentage value of the fundamental amplitude or in dB using the conversion rule 20log(THD). The input dynamic range of an amplifier can be specified as the maximum input amplitude for which the THD is less than a certain value [1].

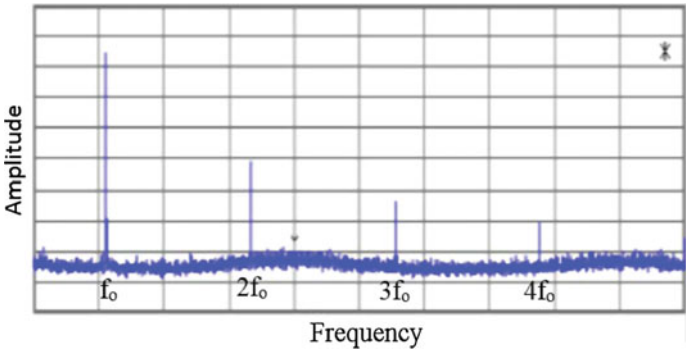

Fig. 2.18 Harmonics

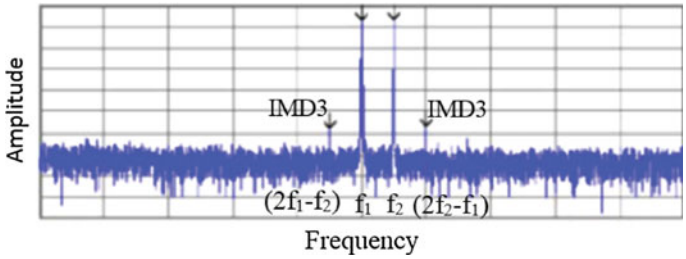

Fig. 2.19 Intermodulation distortions

Intermodulation distortions: is the amplitude modulation of signal with different frequencies due to system nonlinearities. Intermodulation distortions is important to study as the input signal to a certain circuit contains many frequency components. For a two-tone input signal with frequencies f_1 and f_2 the output signal will contain additional frequency components. These frequencies are not just at harmonic frequencies of f_1 and f_2, but also at the sum and difference frequencies of f_1 and f_2 and at multiples of those sum and difference frequencies. The third-order intermodulation products (IMD3), $2f_l - f_2$ and $2f_2 - f_l$, are of interest as they are inside the frequency working band and of interest (as f_l and f_2 are closely spaced), see Fig. 2.19. The other intermodulation products will be at higher out of band frequencies and can be ignored. The IMD3 is measured and plotted for different input levels and the IMD3 curve is extrapolated. The point at which the extrapolated IMD3 is equal to the input signal level is the third order intercept point (IP3) [1].

2.12 Power Penalty

The ratio between maximum optical power and minimum optical power is called extinction ratio, see Sect. 9.3.3. In the ideal case the extinction ratio would be infinite. Due to the finite extinction ratio, the laser source has a power penalty compared to the ideal case with the same modulation swing. The power penalty in dB can be calculated from relation (2.28), and gives a power penalty of 1.76 dB for an extinction ratio of 5 [16].

$$ER_{power\ penalty} = 10Log\left(\frac{ER+1}{ER-1}\right) \tag{2.28}$$

2.13 Dynamic Range

One important feature of the TIA is its dynamic range (DR). The dynamic range is the working range for the TIA for the input optical power going from its minimum (sensitivity at certain BER) to its maximum input optical power (overload optical power), which is limited by TIA saturation.

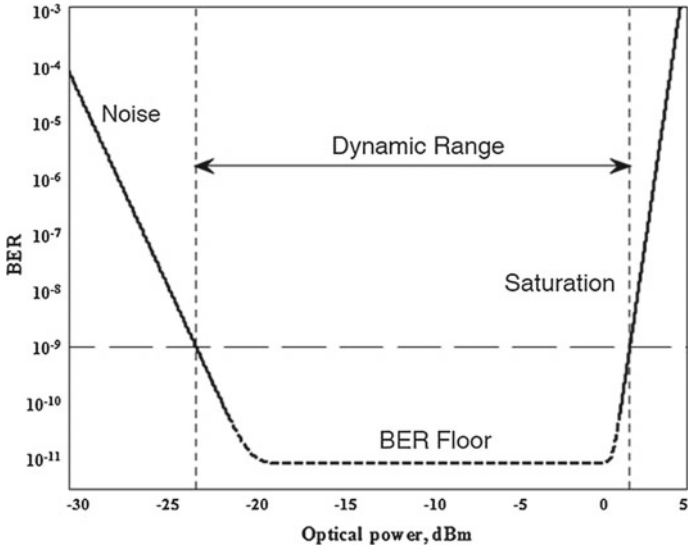

Fig. 2.20 BER as a function of the optical power defining the dynamic range region and its limits

The dynamic range can be calculated as the ratio between the maximum and minimum input optical power:

$$DR = 10log\frac{P_{max}}{P_{min}}dB \tag{2.29}$$

The dynamic range can be increased by increasing the sensitivity (TIA with lower noise) or increasing the maximum received input optical power where the TIA reaches saturation. The maximum limit can be increased by using a TIA with AGC or gain compression stage. Figure 2.20 shows the BER as a function of the optical input power defining the dynamic range region and its limits. In this figure, for BER $= 10^{-9}$ the DR is 25 dB.

References

1. E. Säckinger, *Broadband Circuits for Optical Fiber Communication* (Wiley, New Jersey, 2005)
2. J. Gao, *Optoelectronic Integrated Circuit Design and Device Modeling*, 3rd edn. (Wiley, China, 2011)
3. S. Nagel, Optical fibre-the expanding medium. IEEE Circ. Devices Mag. **5**(2), 36–45 (1989)
4. G.P. Agrawal, *Fiber-Optic Communication Systems*, 3rd edn. (Wiley, 2002)
5. M. Johnson (ed.), *ITU-T Handbook: Optical Fibres Cables and Systems* (International Telecommunication Union, ITU-T, Switzerland, 2009)
6. O. Ziemann, J. Krauser, P.E. Zamzow, W. Daum, *POF Handbook—Optical Short Range Transmission Systems*, 2nd edn. (Springer, Berlin, 2008)

7. M. Atef, R. Swoboda, H. Zimmermann, Real-time 1.25-gb/s transmission over 50-m si-pof using a green laser diode. IEEE Photonics Technol. Lett. **24**, 1331–1333 (2012)
8. M. Atef, H. Zimmermann, *Optical Communication over Plastic Optical Fibers: Integrated Optical Receiver Technology* (Springer, Berlin, 2013)
9. M.S. Filip Tavernier, *High-Speed Optical Receivers with Integrated Photodiode in NanoscaleC-MOS*. (Springer, New York, 2011)
10. J.K. Pollard, Multilevel Data Communication over Optical Fibre, in *IEE Proceedings -I Communications, Speech and Vision*, vol. 138(3) (1991), pp. 162–168
11. A.X. Widmer, P.A. Franaszek, A DC-balanced, partitioned-block, 8B/10B transmission code. IBM J. Res. Develop. **27**(5), 440–451 (1983)
12. S. Benedetto, E. Biglieri, *Principles of Digital Transmission with Wireless Applications* (Kluwer Academic Publishers, New York, 1999)
13. K. Schneider, H. Zimmermann, *Highly Sensitive Optical Receivers* (Springer, Berlin, 2006)
14. P.R. Gray, P.J. Hurst, S.H. Lewis, R.G. Meyer, *Analysis and Design of Analog Integrated Circuits* (Wiley, New York, 2001)
15. B. Razavi, *Design of Integrated Circuits for Optical Communications* (McGraw-Hill, New York, 2003)
16. B. Chomycz, *Planning Fiber Optics Networks* (McGraw-Hill, New York, 2008)

Chapter 3
Basics of Photodiodes

The integrated photodiode is the first stage in the monolithically integrated optical receiver. The photodiode converts the optical power into an electrical current. The photodiode should convert photons into electron and hole pairs with maximum efficiency and to transport them to the electrodes as fast as possible. The main speed limitations of standard nanometer CMOS photodiodes are coming from the slow diffusion of carriers in the substrate. Also the capacitance of the photodiode must be as low as possible to keep the high frequency loss of the photocurrent low to reach highest possible sensitivity. These demands are partially conflicting; so a trade-off is necessary, especially due to the wavelength dependence of the penetration depth. Many photodiode structures are introduced to increase the photodiode's intrinsic speed in nanometer CMOS technology. More details about theory of photodiodes can be found in [1–3].

3.1 Optical Absorption and Photocurrent Generation

Light travels with the velocity c in a medium with the optical refractive index n,

$$c = \frac{c_o}{n} \tag{3.1}$$

where c_o as the velocity of light in vacuum.

Photons are characterized by their wavelength in vacuum λ_0, which can be calculated from the frequency f being independent of the medium:

$$\lambda_0 = \frac{c_o}{f} \tag{3.2}$$

M. Atef and H. Zimmermann, *Optoelectronic Circuits in Nanometer CMOS Technology*, Springer Series in Advanced Microelectronics 54,
DOI 10.1007/978-3-319-27338-9_3

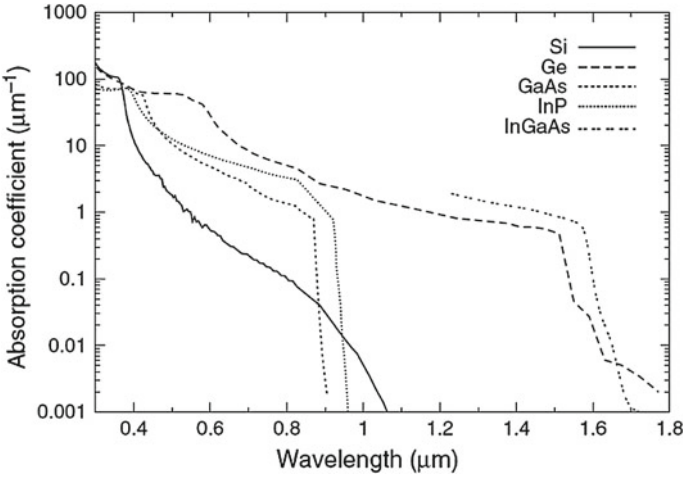

Fig. 3.1 Absorption coefficient of important semiconductor materials versus wavelength [3]

Every single photon owns its energy E:

$$E = hf = \frac{h \cdot c_o}{\lambda_0} \tag{3.3}$$

Photons will produce electron-hole pairs in silicon when the photon energy E is larger than the band gap energy ($E_g = 1.1\,\text{eV}$ for silicon). For silicon the critical wavelength is $\lambda_c = 1.1\,\mu\text{m}$, above this value the silicon is transparent and there is no absorption and Ge or InGaAs photodiode is used [3]. So silicon photodiodes are usable for wavelengths from the near-infrared up to the ultraviolet wavelength range. The absorption coefficient (α) of Si is much lower than that of the direct bandgap semiconductors like Ge and GaAs in this spectral range, see Fig. 3.1. Therefore, Si detectors need a much thicker absorption zone ($>1/\alpha$) than the direct bandgap semiconductors. In spite of the lower optical absorption of silicon, silicon is still the cheapest semiconductor material and it is worth to implement cheap optoelectronic devices and integrated circuits in silicon.

The incident photons with wavelengths smaller than the critical wavelength λ_c have enough energy to transfer an electron from the valence band to the conduction band. This process absorbs the photon and an electron-hole pair is generated. The optical absorption coefficient α defines the penetration depth $1/\alpha$ of the light into the semiconductor in accordance to Lambert-Beer's law:

$$P(x) = P_0(x)e^{(-\alpha x)} \tag{3.4}$$

The photo-generation rate per volume $G(x)$ can be expressed by:

$$G(x) = \frac{P(x) - P(x + \Delta x)}{\Delta x} \frac{1}{A \cdot h \cdot f} \tag{3.5}$$

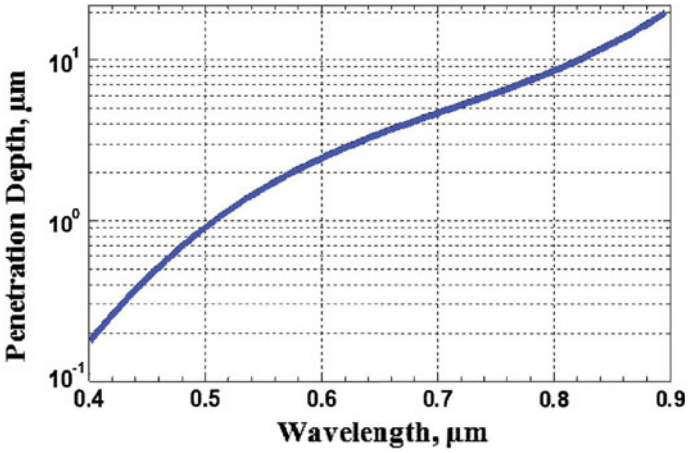

Fig. 3.2 Silicon penetration depth versus wavelength

which results in:

$$G(x) = \frac{\alpha P_o}{A \cdot h \cdot f} e^{(-\alpha x)} \tag{3.6}$$

The penetration depth $(1/\alpha)$ is defined as the material thickness required to absorb 1/e (63 %) of the incident light. The penetration depth for 650 nm is 3.3 µm. The relative absorbed power length for 90 % ($d_{90\,\%}$) of the total power is $(2.3/\alpha)$ which equals 7.6 µm for 650 nm.

The penetration for silicon can be calculated with the following relation [4]:

$$depth = \frac{1}{\alpha} = \frac{1}{10^{(13.2131 - 36.7985\lambda + 48.1893\lambda^2 - 22.5562\lambda^3)}} \tag{3.7}$$

Figure 3.2 plots the calculated penetration depth of silicon versus wavelength for visible light.

3.2 Carrier Drift and Diffusion

Carrier drift is a fast process and carrier diffusion is a slow process. Carrier drift occurs when an electric field is present. Then the charge carriers can reach a high velocity. This is a wanted effect since it improves the speed of the photodetector. In contrast, carrier diffusion occurs when no electric field is present and movement of the charge carriers is slow, which reduces the bandwidth of the whole detector. This means for a fast photodiode, most of the charge carriers must be photo-generated in a zone where a strong electric field exists. Electron-hole pairs generated in semiconductor with zero electrical field (region 1 and 2 in Fig. 3.3) move slowly due to diffusion effect. In

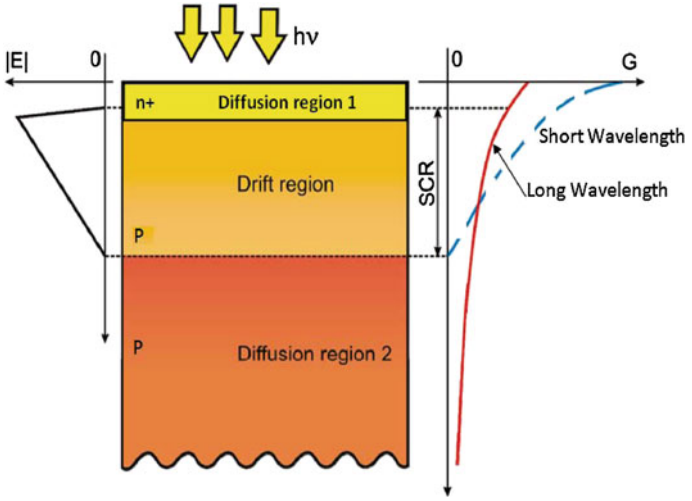

Fig. 3.3 Drift and diffusion regions with electric field and photo-generation distribution in photodiode

a typical photodiode, like a n+/p-substrate photodiode, there are two such regions. The first region is the n+ heavily doped region at the surface of the photodiode and the second diffusion region is below the space-charge region (SCR). The n+ surface region (region 1) is in not very thick so that the generated electron-hole pairs travel a small distance to the drift SCR. The diffusion effect can be considered in region 1 for short wavelengths (blue and green light). The P-substrate diffusion region (region 2) which is below the SCR has more diffusion because it is much thicker than the n+ surface region. Therefore many electron-hole pairs might be generated deep in the diffusion region 2 far away from the SCR (especially for long wavelengths). So, the carriers need a long time to diffuse to the SCR.

3.2.1 Carrier Diffusion

Figure 3.3 shows the heavily doped surface region n+ (region 1), the drift SCR region, and the lower doped p-substrate region (region 2) of the photodiode. The diagram on the left side shows the distribution of the electrical field inside the photodiode and the diagram on the right side shows the photo-generation rate G of electron-hole pairs in dependence of different wavelengths of light. Shorter wavelengths cause a higher generation rate G in shallow region because of its small penetration depth while longer wavelengths cause a higher generation rate G in the deeper region because of its higher penetration depth.

The carrier diffusion coefficients D_n and D_p depend on the carrier mobilities according to the Einstein relation:

$$D_{(n,p)} = \mu_{(n,p)} \frac{KT}{q} \tag{3.8}$$

The time $t_{diff,(n,p)}$ for the diffusion of holes through a n+ region cathode (region 1) with thickness d_n or for electron diffusion in the P-substrate anode (region 2) with d_p diffusion distance toward SCR can be calculated by [5]:

$$t_{diff,(n,p)} = \frac{d^2_{(n,p)}}{2D_{(n,p)}} = \frac{d^2_{(n,p)} q}{2\mu_{(n,p)} KT} \tag{3.9}$$

A much slower contribution to the photocurrent usually results from carriers being generated in the P-substrate (region 2); because the carriers have to diffuse a much longer distance than d_n from region 2 to the space-charge region. The diffusion distance of electrons in the p-substrate d_p can be defined as the distance at which 90 % of the incident optical power is absorbed ($d_p = 2.3/\alpha$).

The time $t_{diff,n}$ for the diffusion of electrons in the P-substrate anode(region 2) with d_p diffusion distance toward SCR can be calculated by:

$$t_{diff,n} = \frac{d_p^2}{2D_n} = \frac{d_p^2 q}{2\mu_n KT} = \frac{(\frac{2.3}{\alpha})^2 q}{2\mu_n KT} \tag{3.10}$$

The cut-off frequency of the photodiode due to carrier diffusion can be estimated by [6]:

$$BW_{diff} = \frac{2.4}{2\pi t_{diff,n}} \tag{3.11}$$

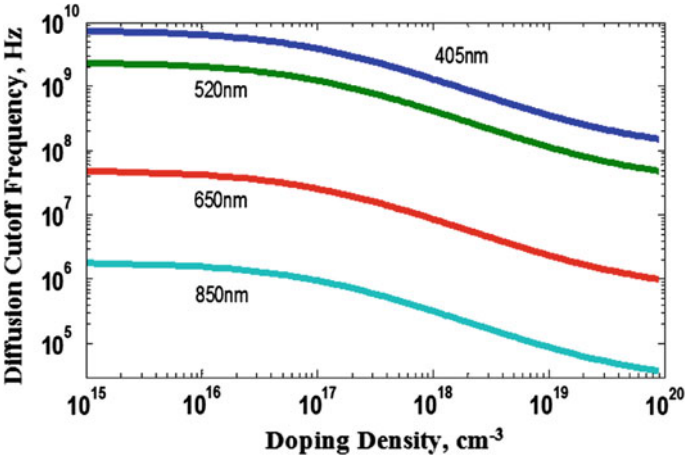

Fig. 3.4 Electron diffusion cutoff frequency (intrinsic bandwidth) as a function of the reverse biasing voltage

Figure 3.4 shows the electron diffusion bandwidth as function of doping for different wavelengths. The longer the wavelength the deeper the photocarriers are generated and more time is needed to diffuse from the neural region toward the SCR.

Generally spoken for short optical wavelengths (e.g. $\lambda = 410\,\text{nm}$) surface effects are dominant. For short wavelengths diffusion in the surface electrode (region 1) can dominate, which should then be thin to reduce carrier diffusion. For long optical wavelengths (e.g. $\lambda = 850\,\text{nm}$) charge carriers are generated deep in the substrate (region 2), where no electric field is present. Diffusion should be avoided where possible, since the effect causes the photodiode to slow down. Often a photodiode must be optimized for a particular wavelength to suppress carrier diffusion to get a higher intrinsic bandwidth. To decrease the diffusion carriers and to increase the drifting carriers a lower doping and a large bias voltage is necessary to create a wider space-charge region with strong electric field.

As the doping increases the mobility decreases and the diffusion time increases leading to smaller intrinsic bandwidth. The carrier diffusion should be reduced to obtain high speed photodiodes. One possibility is to use light with a short wavelength (large absorption coefficient) to avoid photogeneration in the substrate (region 2). Another possibility is to replace Si by other materials like Ge or GaAs which have larger absorption coefficients. A larger reverse voltage should be applied across the photodiode or the doping concentration in the photodiode should be reduced, in order to obtain a thicker space-charge region to reduce the diffusing carriers and increase the carriers' drift contribution. Different photodiode structures like spatially modulated photodiode (SML) [7], fingered photodiode [8], PIN-PD with P-epi layer [9], and photodiode with body contacts [10] are used to increase the photodiode speed by reducing the diffusion current.

3.2.2 Carrier Drift

The photogenerated carriers will drift in the electric field of the depletion region.

$$v = \mu E \tag{3.12}$$

The carrier mobilities in the drift equation give an indication for photodiode speed. An empirical expression for mobility as a function of total impurity concentration at room temperature is [11, 12]:

$$\mu_{0,(n,p)} = \mu_{(n,p),\ min} + \frac{\mu_{(n,p),max} - \mu_{(n,p),min}}{1 + (N_{total}/N_{ref,(n,p)})^{\alpha_{ref,(n,p)}}} \tag{3.13}$$

where N_{total} is the sum of the concentrations of acceptors and donors. The other parameters and constants for silicon are listed in [11, 12].

The dependence of the carrier mobilities in silicon on the electric field can be approximated by [11]:

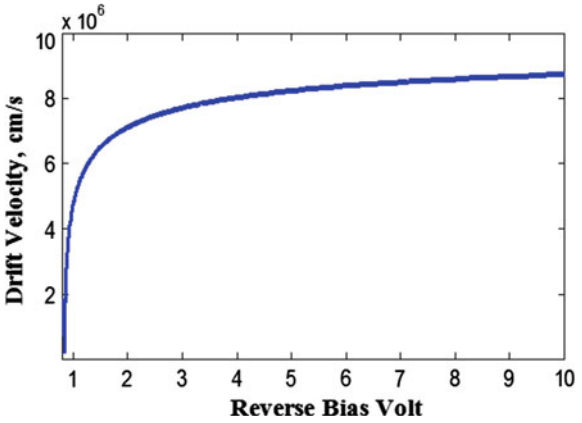

Fig. 3.5 Electron drift speed as a function of the reverse bias voltage for $N_A = 10^{15}\,\text{cm}^{-3}$ and $N_D = 10^{20}\,\text{cm}^{-3}$

$$\mu_{(n,p)} = \frac{\mu_{0,(n,p)}}{1 + \left(\mu_{0,(n,p)} E / v_{(n,p)}^{sat}\right)^2} \tag{3.14}$$

The value for the saturation velocities of electrons and holes is 10^7 cm/s. The carrier mobilities degrade for high electric fields.

By assume full-depletion approximation the depletion region around the junction has well-defined edges with an abrupt transition. The neutral region has a carrier density close to the doping density. The electric field is zero outside the depletion region and maximum value at x = 0:

$$E(x=0) = E_{max,abrupt} = \sqrt{\frac{2q}{\varepsilon_o \varepsilon_r} \frac{N_D N_A}{N_A + N_D} (V_{bi} - V_R)} \tag{3.15}$$

The drift velocities are proportional to the electric field only for smaller values of the electric field. For a large reverse voltage, which results in a large value of the electric field, the drift velocities may be approximated by the saturation velocity v_s, Fig. 3.5.

The drift time is the time taken by carriers to drift through the whole width W of the space-charge region.

$$W_{Abrupt} = \sqrt{\frac{2\varepsilon_o \varepsilon_r}{q} \frac{N_A + N_D}{N_A N_D} (V_{bi} - V_R)} \tag{3.16}$$

The linearly graded profile is a more realistic approximation for junctions formed by deep diffusions (NW or PW) into moderate or heavily doped substrates. The linearly graded profile is mathematically modeled by:

$$N(x) = N_D - N_A = \alpha X \tag{3.17}$$

The junction width for linearly graded profile can be given by [13, 14]:

$$W_{graded} = \sqrt[3]{\frac{12\varepsilon_o\varepsilon_r}{q\alpha}(V_{bi} - V_R)} \tag{3.18}$$

and the maximum electric field at the junction interface is:

$$E(x=0) = E_{max,graded} = \frac{q}{8\varepsilon_o\varepsilon_r}W_{graded}^2 \tag{3.19}$$

The drift time (t_d) is determined by the drift velocity v and by the width of the drift zone W. The drift velocity v depends on μ and E. Furthermore, μ depends on the doping concentration and electric field. The doping concentration and the reverse bias voltage V_R applied to the photodiode determine the electric field values.

$$t_d = \frac{W}{v} = \frac{W(V_R, N_D, N_A)}{\mu(V_R, N_D, N_A) \cdot E(V_R, N_D, N_A)} \tag{3.20}$$

The cut-off frequency of the photodiode due to carriers drift can be estimated by:

$$f_{c,drift} = \frac{2.4}{2\pi t_d} \tag{3.21}$$

The cut-off frequency of the photodiode due to carrier drift is plotted in Fig. 3.6.

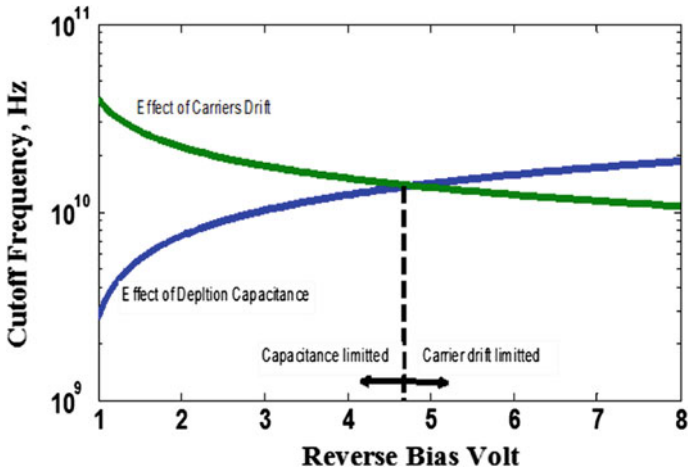

Fig. 3.6 Photodiode electron drift cutoff frequency and extrinsic cut off frequency as a function of the reverse bias voltage for $R_{in} = 50\,\Omega$. C_{PD} was calculated for an abrupt junction (N_D is infinite) from (3.22) for $N_A = 10^{15}\,\text{cm}^{-3}$

3.3 Photodiode Capacitance

The capacitance of a photodiode can be considered as the capacitance of the SCR and the boundary capacitance. The capacitance can then approximately be calculated with (3.22), which is the formula for a plate capacitor, neglecting the boundary capacitance. Silicon possesses a relative permittivity of $\varepsilon_r = 11.9$.

For PN junctions, the capacitance as a function of the reverse bias voltage can be calculated by:

$$C = A\frac{\varepsilon_r \varepsilon_o}{W} = \frac{A}{\sqrt{\frac{2}{q\varepsilon_o\varepsilon_r}\frac{N_A+N_D}{N_A N_D}(V_{bi} - V_R)}} \tag{3.22}$$

Figure 3.7 shows the junction capacitance as a function of the reverse bias voltage for different substrate doping levels. Decreasing the substrate doping or increasing the bias voltage will increase the junction width which has the effect of decreasing the junction capacitance.

The extrinsic cut off frequency of the photodiode due to its capacitance is determined by the input resistance of the next stage (TIA), see Fig. 3.6. The cut-off frequency of the photodiode due to carriers drift calculated by (3.21) is also plotted in Fig. 3.6.

$$fc_{ex} = \frac{1}{2\pi R_{in} C_{PD}} \tag{3.23}$$

Model parameters can be calculated to build a scalable photodiode model. The model parameters for photodiode capacitance model can be approximated by a constant value, which represents the short connection of the anode to the outer edge of the photodiode, plus a perimeter capacitance and an area capacitance. Since the capacitance has a constant, linear and a quadratic dependence from the photodiode

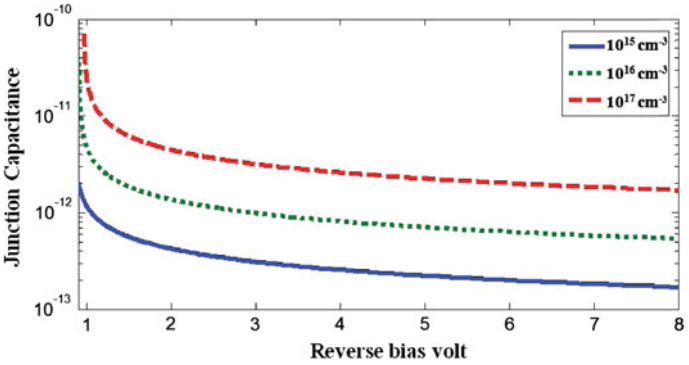

Fig. 3.7 The junction capacitance as a function of the reverse biasing volt for different substrate doping level

diameter; the following basic formula can be used to build a model for the capacitance of a photodiode [15]:

$$C_{PD} = a_1 + a_2 \times d_{PD} + a_3 \times d_{PD}^2 \tag{3.24}$$

Since there are three model parameters, the capacitance C_{PD} of three photodiodes with different dimensions has to be measured. The three equations generated from the measured photodiode capacitances will be solved to calculate the model parameters.

3.4 Photodiode Speed

To get the highest speed, the thickness of the SCR of the photodiode must be maximized. For a good trade-off, the main parameters that must be known are the maximum data rate, the operating wavelength, type of photodiode and the maximum voltage available for the photodiode. A main characteristic of integrated optical receivers is the achievable sensitivity, which depends strongly on the responsivity of the photodiode and also on the capacitance, so minimum capacitance and maximum responsivity lead to maximum sensitivity. This could be achieved by maximizing the thickness of the SCR by increasing the reverse bias voltage. But this contradicts to speed, since increasing the SCR layer thickness increases the distance the generated carriers have to drift and so the drift transition time of the carriers increases, see Fig. 3.6. As in nanometer CMOS technology the substrate doping level and the wells are in the range of 10^{15} and 10^{18} cm^{-3}, respectively, the SCR width will be small and the diffusion current will dominate over the drift current. So, the drift time will be negligible to the diffusion time. As a result the maximization of SCR is the optimum way to get the maximum speed and responsivity.

Often a certain data rate is given, so the photodiode has to achieve a certain rise and fall time. In a first order approximation the necessary 10–90 % rise- and fall-time of the photodiode can be calculated from the formula (3.20) with t_{diff} as the diffusion time through the substrate.

$$\frac{t_r + t_f}{2} \approx t_{diff} = \frac{1}{2DR} \tag{3.25}$$

When rise- and fall-time are nearly equal, then the formula (3.25) simplifies to formula (3.26) and gives approximately the diffusion time of the charge carriers through the substrate region. In both cases the factor 3 indicates a conservative approximation and for an aggressive approximation the factor can be two.

$$t_r = t_f \approx t_{diff} = \frac{1}{2DR} \tag{3.26}$$

$$DR = \frac{1}{t_r + t_f} \tag{3.27}$$

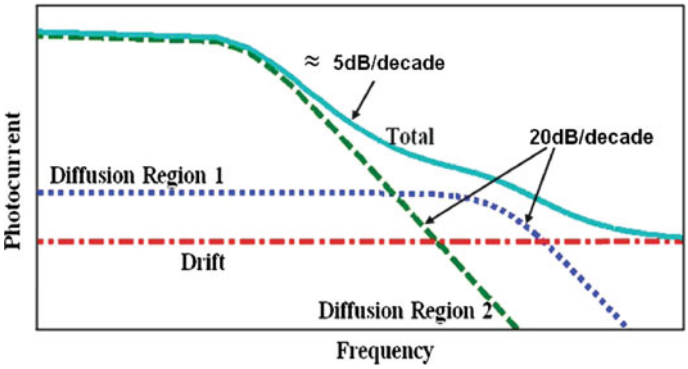

Fig. 3.8 Frequency response of the photocurrent with drift and diffusion

The intrinsic bandwidth of the photodiode due to the diffusion of electrons in the P-substrate can be estimated by:

$$BW_{c,in} \approx \frac{2.4}{\pi(t_{diff,n})} = \frac{2.4}{\pi(t_{diff,n})} \tag{3.28}$$

where $t_{diff,n}$ is given by (3.11).

The extrinsic bandwidth of the photodiode due to its capacitance is determined by the input resistance of the next stage (TIA), see Fig. 3.6.

$$BW_{ex} = \frac{1}{2\pi R_{in} C_{PD}} \tag{3.29}$$

where C_{PD} is given by (3.22).

The total bandwidth of the photodiode can be calculated by:

$$BW_{total}^{-2} = BW_{ex}^{-2} + BW_{in}^{-2} \tag{3.30}$$

The influence of carrier diffusion on the frequency response of photodetectors is shown in Fig. 3.8. The total intrinsic photodiode frequency response in Fig. 3.8 shows a small roll-off (e.g. 5 dB) due to the combination of the photocurrent components from diffusion regions 1 and 2. The total photocurrent response roll-off is smaller than the diffusion roll-off in region 2, due to the larger influence of the faster response of diffusion region 1 and the drift region currents.

3.5 Quantum Efficiency

The quantum efficiency gives the percentage of generated electron-hole pairs per incident photons, which are transported to the electrodes. It can be calculated by another form of relation (3.39) with:

$$\eta = \frac{R \cdot h \cdot c_o}{q\lambda} \tag{3.31}$$

The quantum efficiency in relation (3.32) can be divided into optical quantum efficiency η_o and an internal quantum efficiency η_i.

$$\eta = \eta_i \eta_o \tag{3.32}$$

3.5.1 Internal Quantum Efficiency

The internal quantum efficiency η_i defines how effectively the optical power is converted to electron-hole pairs, which contributes to the photocurrent at the connection terminals. As there is no recombination in the drift region; all carriers photogenerated in drift regions contribute to the photocurrent. So the space-charge regions of photodiodes do not reduce the internal quantum efficiency. The recombination of photogenerated carriers in regions 1 and 2 (see Fig. 3.3), reduces the internal quantum efficiency. In region 1 (n+ cathode), the carrier lifetime is reduced considerably because of the high doping. This reduces the internal quantum efficiency for short wavelengths. The recombination of photogenerated carriers in region 1 is not very important for long wavelengths due to the small portion of photogenerated carriers in region 1 because of its large penetration depth. Light with long wavelengths penetrates deep into the silicon substrate (region 2) and the recombination can reduce the internal quantum efficiency.

η_i has a frequency dependent behavior, since two transport effects are involved, carrier drift and carrier diffusion. Carriers being photogenerated in diffusion regions do not have enough time to diffuse to the drift region before the next light pulse. The diffusion tails of the photocurrent of consecutive light pulses overlap. The higher the pulses data rate is the smaller the (dynamic) quantum efficiency becomes until the minimum is reached. This minimum is set by the portion of carriers being generated in the space-charge region. The following expression introduced in [1] describes the dynamical internal quantum efficiency:

$$\eta_i = (1 - exp[-(d_p + d_I)])exp(-d_p) \tag{3.33}$$

Figures 3.9 and 3.10 show the effect of junction depth and depletion region width on the internal efficiency of silicon photodiodes, respectively.

3.5.2 Optical Quantum Efficiency

The reflection behavior at integrated CMOS photodiode surface layers is more difficult to evaluate since there are several layers involved. The situation gets worse

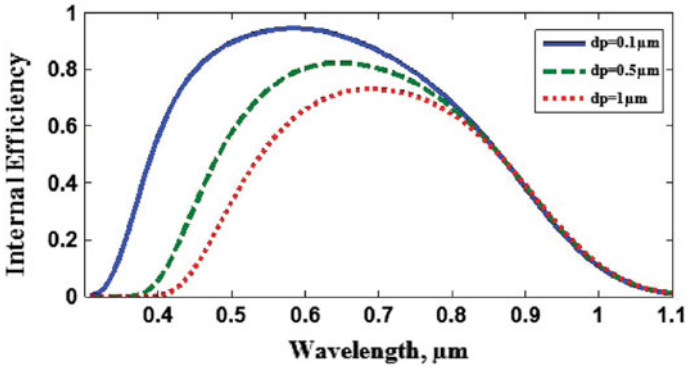

Fig. 3.9 Effect of junction *depth* on the internal efficiency of silicon photodiode

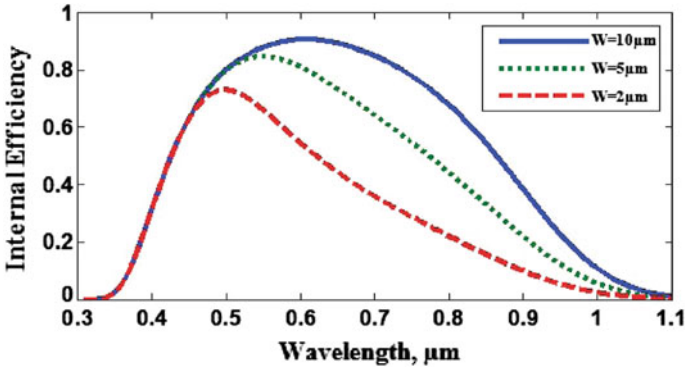

Fig. 3.10 Effect of SCR *width* on the internal efficiency of silicon photodiode

in each newer generation of CMOS technology, because the number of metal and dielectric layers increases. For example, in 40 nm CMOS technology, there are eleven levels of metals, and thus eleven levels in the dielectric stack. In addition to these dielectric layers, there are two passivation layers on top of the last metal layer. This makes the optical behavior by far more critical since the variation of transmission T over wavelength becomes significantly higher.

For low cost integrated photodiodes in certain technologies, photodiodes are realized by using standard layers available in the used technology process. These layers are not optimized to reduce light reflection. The coupling of light through different stacked oxide and nitride surface layers (used in metalization process) into silicon is a critical performance issue. The light reflection coefficient R_{opt} of photodiode surface layers significantly influences the overall photodiode sensitivity. Reflections on an integrated photodiode can occur due to different surface layers like silicon oxide (SiO2) or silicon nitride (Si3N4) with different optical parameters (n and k)

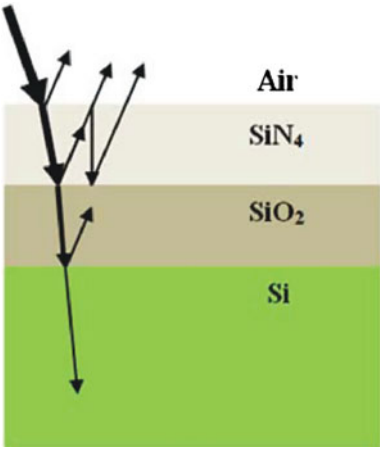

Fig. 3.11 Transmission and reflection of incident light at typical CMOS photodiode nitride and oxide surface layers

compared to silicon. The absorption in SiO2 or Si3N4 can be neglected ($k = 0$) and the light reflection only depends on the refractive index n of the two different layers [16]:

$$R_{opt} = \frac{(n_1 - n_2)^2}{(n_1 + n_2)^2} \tag{3.34}$$

In almost every CMOS technology the uppermost layer is silicon nitride (Si_3N_4) for passivation that protects the fabricated devices from environmental influences. Below the nitride layer there are different oxide layers (SiO_2) used as inter-metal oxide to isolate the metal lines. A typical layer combination for such photodiodes is drawn in Fig. 3.11. Since there is a certain amount of reflection on every layer interface, the superposition of multiple reflected waves with the incident wave must be considered. The phase shift of a wave entering a layer from top side and reflected back at bottom side interface is given by [17]:

$$\sigma = \frac{4\pi}{\lambda} n_i d_i \tag{3.35}$$

Depending on layer thickness d_i and refractive index n_i the superposition of all reflected wave contributions shows minima and maxima at certain wavelengths λ. To calculate the R_{opt} of multiple layer structures, different methods were developed by using matrix algebra mathematics [17].

The transmission coefficient for different wavelengths is plotted in Fig. 3.12. The spectral behavior shows a strong variation of the transmission coefficient ($T = 1 - R_{opt}$) over wavelength from very high values of 1 (=100 %) down to low values of 0.4 (=40 %). This can be explained by the refractive indices which

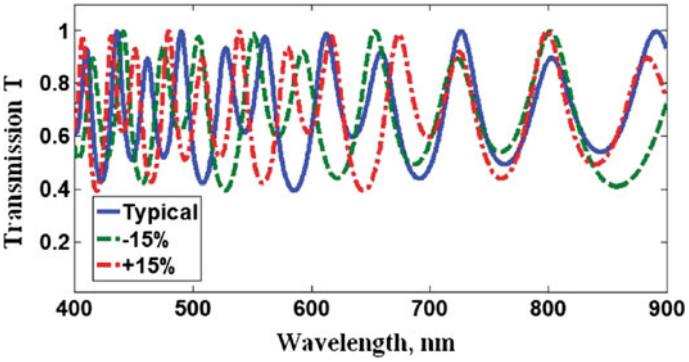

Fig. 3.12 Light transmission at a typical CMOS surface layer stack with nitride and oxide on silicon

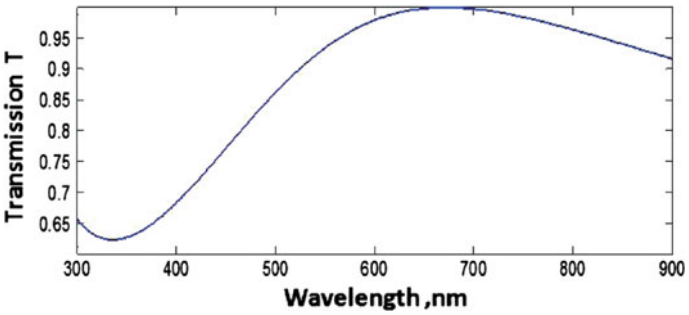

Fig. 3.13 Light transmission at an optimized surface layer stack with oxide on a defined nitride layer on silicon

change from n = 1 for air to a high value of $n_{sn} = 2.0$ for nitride, back to a lower value $n_{ox} = 1.46$ for oxide and again to a high value of $n_{si} = 4.18$ for silicon.

To improve the transmission, the refractive indices should increase continuously from one layer to another (in the direction from air towards silicon). So the nitride layer has to be removed from top and instead inserted between oxide and silicon. A risk by doing this is the missing passivation surface. So humidity might permeate into the silicon substrate and change the transistors behavior over lifetime. This effect has to be investigated during photodiode qualification process. The simulation of an improved photodiode surface stack is shown in Fig. 3.13.

The nitride layer with defined thickness acts as an anti-reflection-coating (ARC) on the photodiode and can improve the light transmission. The drawback is that additional process steps are necessary resulting in a non-standard IC process flow with higher production costs.

A part of the total optical power incident to the photodiode (P_{OPT}) is reflected from the surface. The optical power P_0 directly below the surface in the semiconductor can be calculated with the reflection factor R_{OPT} by:

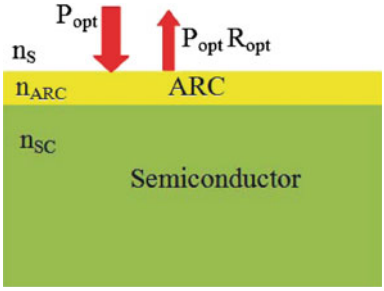

Fig. 3.14 Reflection on the semiconductor surface with antireflection coating

$$P_0 = (1 - R_{OPT})P_{OPT} \tag{3.36}$$

The reflection coefficient can be minimized with an ARC at the surface of the semiconductor for a certain wavelength or a more or less small wavelength range. In Fig. 3.14 the principle structure of the antireflection coating and the distribution of the optical power below the semiconductor surface can be seen. The optimum values of the thickness of the ARC d_{ARC} and its refractive index n_{ARC} can be calculated with the formulas (3.37) and (3.38) with the refractive index of the surrounding n_S and the refractive index of the semiconductor n_{SC}. The materials SiO_2 ($n_{ARC} = 1.45$) and Si_3N_4 ($n_{ARC} = 2.0$) are most adequate for the ARC coating.

$$n_{ARC} = \sqrt{n_S n_{SC}} \tag{3.37}$$

$$d_{ARC} = \frac{\lambda_0}{4n_{ARC}} \tag{3.38}$$

3.6 Photodiode Responsivity

The conversion factor from optical power P_{OPT} to electrical current is defined by the responsivity R, it can be calculated by:

$$R = \frac{I_{PH}}{P_{OPT}} = \frac{\eta\lambda[nm]}{1240} \tag{3.39}$$

The responsivity is an important link between the optical power and the current. It is necessary to evaluate the sensitivity (minimum optical power for data transmission for a certain bit-error ratio) of an optical receiver. Figure 3.15 shows the responsivity of silicon photodiode as a function of wavelength considering diffusion, junction depth and width effects. The dashed line for 100 % efficiency is the theoretical limit. The line of maximum efficiency is given by the relation (3.39).

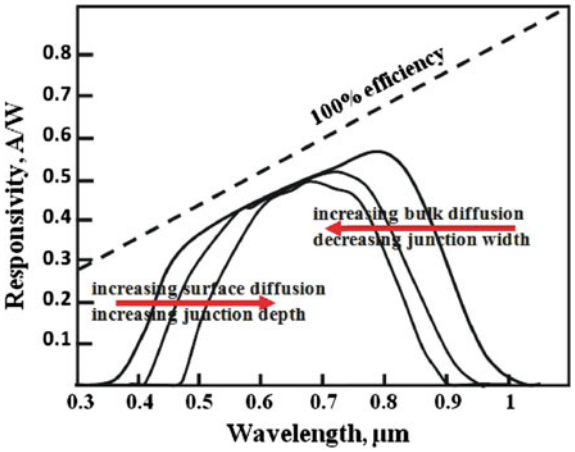

Fig. 3.15 Responsivity of silicon photodiodes as a function of wavelength considering diffusion, junction *depth* and *width* effects

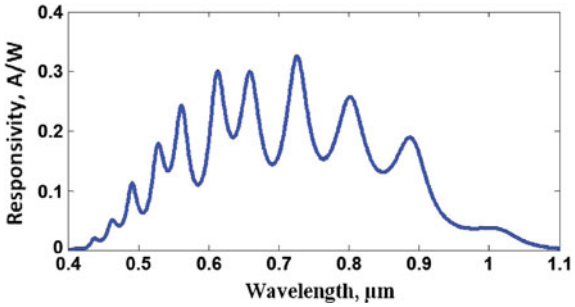

Fig. 3.16 Responsivity of silicon photodiode as a function of wavelength after considering the optical efficiency

For short wavelengths the responsivity goes to zero because of surface recombination of carriers being photogenerated close to the silicon surface. Surface recombination is due to surface states that come from the crystal imperfection and the oxidized silicon that grows at the surface of silicon. Also, the short wavelengths responsivity decreases due to recombination when the junction distance from the surface (junction depth) increases. Long wavelengths responsivity decrease if the recombination in the substrate increases. The responsivity increases for long wavelengths with increasing SCR width. Figure 3.16 shows the photodiode responsivity calculated by (3.39) after considering the optical efficiency (Sect. 3.5.2) and internal quantum efficiency (Sect. 3.5.2). The strong fluctuation in the responsivity due to reflectance from the dielectrics stack in 90 nm CMOS process was also reported in [18]. A simulation analysis for the optical transmission coefficient in 130 nm CMOS process and its effect on the responsivity was introduced in [19, 20].

3.7 Photodiode Dark and Noise Currents

The photodiode dark current is the leakage current that flows through the photodiode in the absence of light. The photodiode dark current comes from thermally generated carriers at a reverse voltage. The dark current, as a function of biasing voltage, can be expressed by:

$$I_{dark} = I_s \left(\exp\left(\frac{qV}{KT}\right) - 1 \right) \tag{3.40}$$

where I_s is the reverse saturation current, q is the electron charge, V is the applied bias voltage, k is the Boltzmann constant, and T is the absolute temperature. The dark current of a photodetector is usually too low and its influence on receiver sensitivity is negligible. However, in high sensitivity optical receivers (small bandwidth and/or with APDs), the dark current can be a limitation. The reverse biasing voltage must be kept below the breakdown voltage. As the reverse biasing voltage reaches the breakdown voltage the dark current increased dramatically.

The major noise source in photodiodes is the shot noise. Shot noise originates from the statistical fluctuation in both dark current and photocurrent. The mean square of the total photodiode noise current is given by the following expression [21]:

$$< i^2_{shot,PD} > = 2q \left(I_p + I_d\right) \Delta f \tag{3.41}$$

For avalanche PDs the shot noise is expressed by [22]:

$$< i^2_{shot,APD} > = 2qM^2F \cdot \left(I_p + I_d\right) \Delta f \tag{3.42}$$

where M is the avalanche multiplication gain and F is the excess noise which is also a function of the multiplication gain M [22]:

$$F = k_A \cdot M + (1 - K_A) \cdot \left(2 - \frac{1}{M}\right) \tag{3.43}$$

where k_A is the ionization coefficient ratio. If only one type of carrier participates in the avalanche process, then $k_A = 0$ and the excess noise factor is minimized. However, if electrons and holes both are participating, then $k_A > 0$ and more excess noise is produced. For an InGaAs APD, $k_A = 0.5{-}0.7$ and the excess noise factor increases almost proportional to M for a silicon APD, $k_A = 0.02{-}0.05$ and the excess noise factor increases much more slowly with M [23, 24]. Thus from a noise point of view, the silicon APD is preferable, but silicon cannot detect long wavelengths commonly used in telecommunication applications(1.35 and 1.55 μm).

The PD series resistance R_s, which originates from the contact and anode/cathode resistance, introduces another source of noise. The PD's series resistance R_s causes the mean squared noise current:

$$< i^2_{R_s} > = \frac{4K_{BT}}{R_s} \Delta f \tag{3.44}$$

3.8 Photodiode Small-Signal and Noise Equivalent Circuit Model

Figure 3.17 shows the photodiode's small-signal and noise equivalent circuit model. The current source I_{Ph} represents the photocurrent. C_{PD} is the capacitance of the space-charge region of the photodiode. The shunt resistor R_D, which models the reverse, leakage, or dark current of a photodiode usually is very large (hundreds of MΩ) and can be neglected in most cases. The series resistance R_S (contact and anode/cathode resistance) cannot be neglected when the photocurrent has to flow through low-doped regions in the photodiode. R_S can be neglected (few ohms) for many photodiodes with highly doped anode and cathode regions. The noise of the junction itself is represented by the shot noise current source (3.42). The resistor R_D is noiseless, resistor R_S generates thermal noise given by (3.44).

The PD should be connected to TIA with shunt feedback resistor R_f to amplify the weak photocurrent, see Fig. 3.18. By neglecting the shunt resistor R_D due to its large value and the serial resistor R_s because of its small value; the receiver bandwidth can be approximated by:

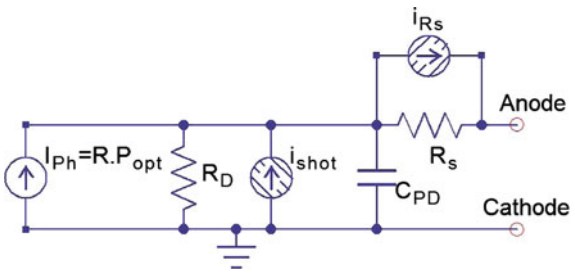

Fig. 3.17 The photodiode's small-signal and noise equivalent circuit model

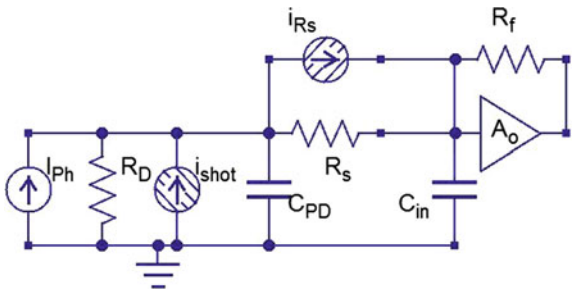

Fig. 3.18 The photodiode small-signal and noise equivalent circuit model connected to a TIA with shunt feedback resistor

$$\frac{A_o}{2\pi R_f (C_{PD} + C_{in})} \tag{3.45}$$

where A_o is the amplifier's voltage gain and C_{in} is the TIA's input capacitance.

In Chap. 6, (6.96) shows that the equivalent input noise current spectral density rises proportional to ${C_{PD}}^2$ and f^2 at high frequencies.

So C_{PD} should be minimized to get a higher bandwidth and to get a lower noise at high frequency.

When the PD is connected to the TIA with shunt feedback resistor R_F, the output noise power spectral density due to the PD's shot noise and series resistor is shown in Fig. 3.19 [25].

At low frequencies $<1/(2\pi R_D C_{PD})$, the minimum value of the noise power is proportional to *Rs* : the higher *Rs*, the higher the noise contribution [25].

$$v_{n,Rs} = KTR_s \, R_F^2 / R_D^2 \tag{3.46}$$

The low frequency noise is dominated by shot noise.

$$v_{n,shot} = 2qR_f^2 i_{shot} \tag{3.47}$$

For high frequencies $>1/(2\pi R_s C_{PD})$ the R_s is dominant and the maximum value of the noise power is proportional to *Rs* : the higher *Rs*, the higher the noise contribution [25].

$$v_{n,max} = 4KTR_s \left(\frac{A_o R_f}{R_{out} + R_f} \right)^2 \tag{3.48}$$

The series resistance R_s should be minimized to minimize the PD's noise.

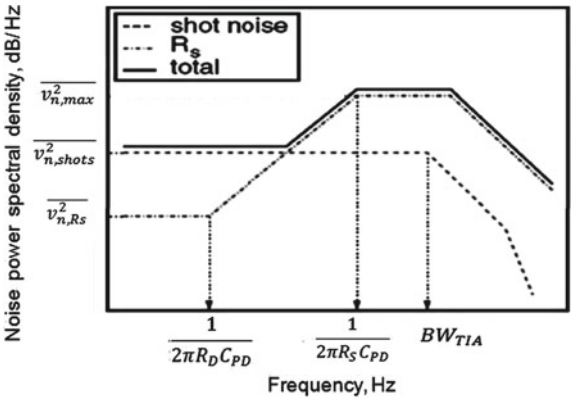

Fig. 3.19 Photodiode output noise power spectral density [25]

References

1. H. Zimmermann, *Integrated Silicon Optoelectronics*, second edn. (Springer, Berlin, 2010)
2. K. Schneider, H. Zimmermann, *Highly Sensitive Optical Receivers* (Springer, Berlin, 2006)
3. H. Zimmermann, *Silicon Optoelectronic Integrated Circuits* (Springer, Berlin, 2004)
4. W.J. Liu, O.T.-C. Chen, L.-K. Dai, F.-W. Jih, A CMOS Photodiode Model, in *IEEE International Workshop on Behavioral Modeling and Simulation, Santa Rosa*, Chung Cheng, California (2001)
5. C.W.G. Winstel, *Optoelektronik II* (Springer, Berlin, 1986)
6. S. Sze, *Physics of Semiconductor Devices* (Wiley, New York, 1981)
7. S.-H. Huang, W.-Z. Chen, Y.-W. Chang, Y.-T. Huang, A 10-Gb/s OEIC with meshed spatially-modulated photodetector in 0.18 μm CMOS technology. IEEE J. Solid-State Circuits **46**(5), 1158–1169 (2011)
8. H. Zimmermann, H. Dietrich, A. Ghazi, P. Seegebrecht, Fast CMOS-Integrated Finger Photodiodes for aWide Spectral Range, ESSDERC, pp. 435–438 (2002)
9. B. Ciftcioglu, L. Zhang, J. Zhang, J.R. Marciante, J. Zuegel, R. Sobolewski, Wu Hui, Integrated silicon PIN photodiodes using deep N-well in a standard 0.18 μm CMOS technology. J. Lightwave Technol. **27**(15), 3303–3323 (2009)
10. M.-J. Lee, W.-Y. Choi, Performance comparison of two types of silicon avalanche photodetectors based on N-well/P-substrate and P+/N-well junctions fabricated with standard CMOS technology. J. Opt. Soc. Korea **15**(1), 1–3 (2011)
11. D.M. Caughey, R.E. Thomas, Carrier mobilities in silicon empirically related to doping and field. Proc. IEEE **55**, 2192–2193 (1967)
12. S. Selberherr, Process and device modeling for VLSI. Microelectron. Reliab. **24**(2), 225–257 (1984)
13. U. Mishra, J. Singh, *Semiconductor Device Physics and Design* (Springer, Netherlands, 2008)
14. M.D. Hussein Ballan, *High Voltage Devices and Circuits in Standard CMOS Technologies* (Springer, 1999)
15. R. Swoboda, High-Speed Integrated Optical Receivers, Ph.D. thesis, Vienna University of Technology, 2007
16. E. Palik, *Handbook of Optical Constants of Solids* (Academic, Orlando, 1985)
17. J. Sturm, Photodiode Modeling and Optoelectronic Integrated Circuits for Optical -Data - Storage Applications, Ph.D. thesis, Vienna University of Technology, 2006
18. A. Polzer, K. Schneider-Hornstein, J. Dong, P. Kostov, H. Zimmermann, Investigation of triple-junction photodetector in 90 nm CMOS technology, *Proceedings EUROSENSORS XXV, Procedia Engineering*, vol. 25, pp.864–867 (2011)
19. F. Tavernier, M. Steyaert, High-speed optical receivers with integrated photodiode in 130 nm CMOS. IEEE J. Solid-State Circuits **44**, 2856–2867 (2009)
20. B. Nakhkoob, S. Ray, M. Hella, High speed photodiodes in standard nanometer scale cmos technology: a comparative study. Opt. Express **20**, 11256–11270 (2012)
21. J. Gao, *Optoelectronic Integrated Circuit Design and Device Modeling*, 3rd edn. (Wiley, China, 2011)
22. R.J. McIntyre, Multiplication noise in uniform avalanche diodes. IEEE Trans. Electron Devices **13**(1), 164–168 (1966)
23. G.P. Agrawal, *Fiber-Optic Communication Systems*, 2nd edn. (Wiley, New York, 1997)
24. E. Säckinger, *Broadband Circuits for Optical Fiber Communication* (Wiley, New Jersey, 2005)
25. C. Hermans, M. Steyaert, *Broadband Opto-Electrical Receivers in Standard CMOS* (Springer, Netherlands, 2007)

Chapter 4
Discrete Photodiodes

Low production costs are important for large volume production. For POF communications large-area PDs are needed which will increase the chip costs if integrated in the same expensive nanometer CMOS technology. This is because an external photodiode requires one sixth to one eighth of mask and processing steps compared to an integrated circuit [1]. Additionally the photodiode and the circuitry can be optimized separately and therefore a better performance compared to an OEIC solution can be achieved especially for large-area photodiodes and for long wavelengths. So, an external large-diameter photodiode (fabricated in cheap old silicon technologies) bonded to small-area circuit chip (fabricated in nanometer CMOS technology) is more cost-efficient, offers a higher speed, and lower noise compared to an OEIC solution in the same nanometer CMOS technology. As it was stated in Sect. 1.5 that the large volume of optical interconnects (PCB to PCB, chip to chip, on chip) will be the main driver to reduce the OEIC costs in the coming years.

4.1 Discrete Photodiodes for Visible Light

Silicon 0.6 μm BiCMOS and 0.35 μm silicon-germanium technologies can be used to implement high performance external PDs in the visible light spectrum range in arbitrary geometry and in arrays as full-custom ASICs. These BiCMOS technologies are considered cheap compared to the recent nanometer CMOS technologies like 40 nm.

A PIN PD was implemented in 0.6 μm silicon BiCMOS process which provides a P-epitaxial layer with doping concentration of about 10^{13} cm^{-3}. A mask for defining the optical window with antireflection coating and a mask for blocking out the wells is available in this technology. The native N-type epitaxial layer with 10^{14} cm^{-3} doping concentration, which exists in regions where no P-well or N-well is generated,

M. Atef and H. Zimmermann, *Optoelectronic Circuits in Nanometer CMOS Technology*, Springer Series in Advanced Microelectronics 54,
DOI 10.1007/978-3-319-27338-9_4

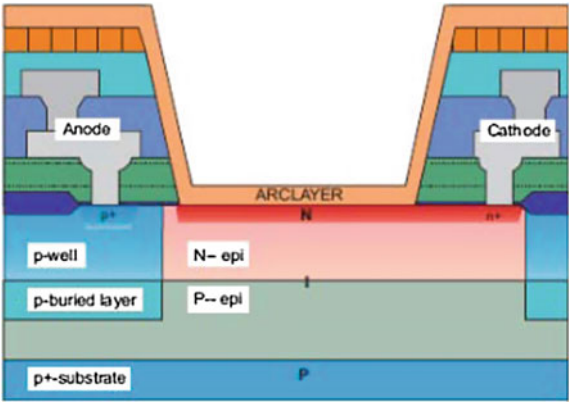

Fig. 4.1 Vertical PIN photodiode in 0.6 µm BiCMOS technology [3]

is part of the PIN photodiode. A vertical PIN photodiode is created by placing an N+ source/drain region in an area excluded from automatic P-well generation. This results in a thick almost intrinsic layer reaching from the N+ cathode to the P+ substrate, Fig. 4.1. The PIN photodiode has an antireflection coating optimized for red light and a measured quantum efficiency of 96 % for a wavelength of 660 nm. The rise and fall times of 516 and 545 ps resulted at a wavelength of 660 nm. The thick depletion region reduced the area capacitance to 0.0158 fF/µm^2 [2].

Fingered PIN PDs were fabricated in 0.35 µm SiGe BiCMOS technology [4]. Owing to the thick low doped intrinsic epitaxial layer, high bandwidths and high dynamic quantum efficiencies result for a wide spectrum of optical wavelengths from ultraviolet and blue to red and near-infrared, see Fig. 4.2. The capacitance of a finger diode with cathode stripes as wide as 0.6 µm and separated by 4.6 µm is 0.022 fF/µm^2. The dark currents are below 2 pA at room temperature for reverse biases to above 5 V. The cathode finger structure results in a high responsivity of 0.2 A/W (quantum efficiency 61 %) for 410 nm light and a bandwidth of 1.253 GHz

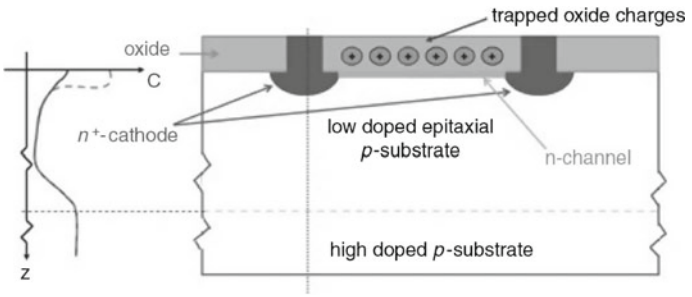

Fig. 4.2 Cross-section of an interdigitated SiGe BiCMOS PIN PD. *Left dashed curve* n+ cathode, and *solid curve* p-type doping [4]

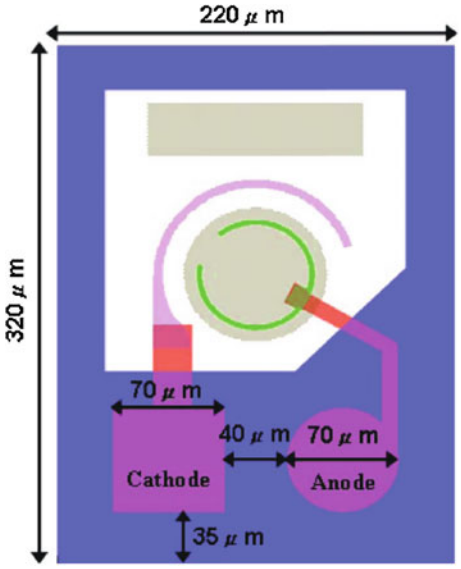

Fig. 4.3 Top view of commercial GaAs PIN PD [5]

at a reverse bias voltage of 3 V, and the rise and fall times of 250 and 310 ps resulted at a wavelength of 410 nm. For 675 nm the measured rise and fall times are 460 ps and 470 ps, respectively, and the responsivity is 0.4 A/W [4].

GaAs is interesting for the realization of high-speed and high-responsivity photodiodes in the visible range. Due to its higher absorption coefficient compared to silicon; the GaAs has a smaller diffusion region and most of the light is absorbed in the depletion/intrinsic zone. Due to the higher mobility of GaAs; the diffusion and drift speeds are higher than those of silicon. As a result the GaAs PDs show a higher speed and responsivity than the silicon counter part. A commercial 10 Gbps GaAs PIN photodiode chip shows a responsivity of 0.65 A/W at 850 nm, see Fig. 4.3 [5]. The PD sensitive area is typical 70 μm in diameter which fits to multi-mode GOF with 62 μm diameter. It has a dark current of 0.1 nA at 5 V reverse bias, a capacitance of 0.22 pF at 1.5 V reverse bias, and a bandwidth of 9 GHz.

4.2 Discrete Photodiodes for Infrared Light

GaAs, InGaAs, InP, and Ge and related compound materials are interesting for the realization of high-speed and high-responsivity photodiodes. In particular high speed photodetectors consisting of III–V materials can be combined with CMOS integrated circuits in hybrid OEICs. This combination seems advantageous due to the following aspects: III–V materials and technologies are not as highly developed as silicon material and silicon technologies. Only 10^5 transistors can be integrated on one III–V chip compared to 10^8 and even more transistors on Si chips [6]. III–V photodetectors

allow much higher data rates and responsivity than Si photodetectors. For IR optical receivers the III–V photodiodes are a must.

A lateral Ge PIN PD was selectively grown at the end of silicon waveguides. The Ge PIN-PD has a responsivity of 0.8 A/W at 1.55 μm [7]. This high value, close to published record, demonstrates the good coupling efficiency between the Si waveguide and the lateral Ge photodetector and the efficiency of complete light absorption after a 10 μm propagation length thanks to the butt coupling configuration used here. The dark current of the Ge photodetector is about 4 μA at a bias voltage of −1 V leading to a high dark current density of 80 A/cm^2. This value, rather high compared to [8], was probably due to dopant diffusion during thermal annealing. However, this dark current only increased by a factor of 2 with temperature increased from 25 to 90 C . An open eye diagram at 40 Gb/s was demonstrated under zero-bias at a wavelength of 1.55 μm [7] .

InGaAs photodiodes are quite useful for practical optical communications systems; as the performance of PIN PDs can be improved considerably by using direct bandgap semiconductors, such as InGaAs. For indirect bandgap semiconductors such as Si, typically intrinsic layer width must be in the range 20–50 μm to ensure high quantum efficiency. The bandwidth of such photodiodes is then limited by a relatively long transit time (200 ps). By contrast, intrinsic layer width can be as small as 3–5 μm for photodiodes that use direct-bandgap semiconductors, such as InGaAs, where the transit time for such photodiodes is around 10 ps [9].

A PIN PD commonly used for optical communications systems uses InGaAs for the middle layer and InP for the surrounding p-type and n-type layers. Figure 4.4 shows such an InGaAs PIN PD. InP is transparent for light whose wavelength exceeds 0.92 μm, whereas the InGaAs PD has a cutoff wavelength of 1.65 μm. The middle InGaAs layer thus absorbs strongly in the wavelength region 1.3–1.6 μm. The diffusion component of the detector current is eliminated completely in such a

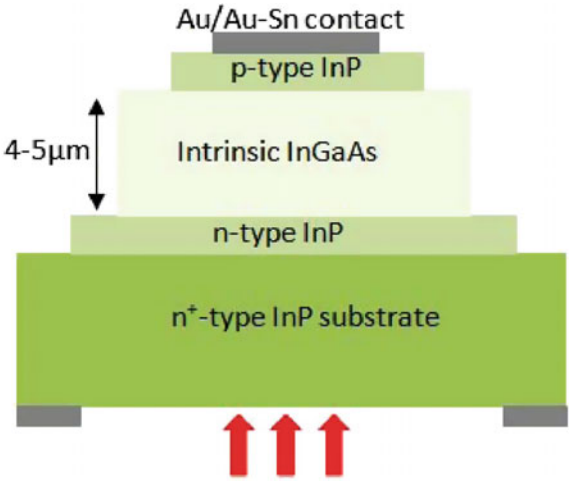

Fig. 4.4 InGaAs PIN photodiode for IR light detection [9]

heterostructure photodiode because photons are absorbed only inside the depletion region. The front facet is often coated using suitable dielectric layers to minimize reflections. The quantum efficiency can reach almost 100 % by using an InGaAs layer width of 4–5 μm [9].

4.3 External Photodetector Connected with Bond Wires

The III/V photodetector chip can be connected via bond-wires with the CMOS circuitry chip, see Fig. 4.5. Usually the bond pads, which are used on both chips for providing the area where the bond wires are bonded to, have a size of approximately $70 \times 70\,\mu m^2$. These large bond pads introduce a large parasitic capacitance and the bond wire represents an inductance compared to the integrated solution (Fig. 1.4a and c). For high-speed optical receivers and data rates of more than several Gb/s, the bond-wires and bond-pads deteriorate the optical receiver performance. The so-called flip-chip interconnection technique can be used instead.

4.4 External Photodetector Connected Using Flip-Chip Technique

This flip-chip technique uses gold bumps with a diameter of approximately 30 μm on bond pads of approximately the same size and thereby reduces stray capacitance and inductance, when the chips are carefully aligned face-down (Fig. 4.6). The pads were located directly over the junction area of the photodiode in order to realize a minimum parasitic capacitance. In this flip-chip technique, the bumps had to cover the pads perfectly and there were strict requirements for the alignment to an accuracy of only a few micrometers [6].

An improved flip-chip interconnection technique, which permits loose chip alignment because of the self-alignment effect of the molten solder was developed [10].

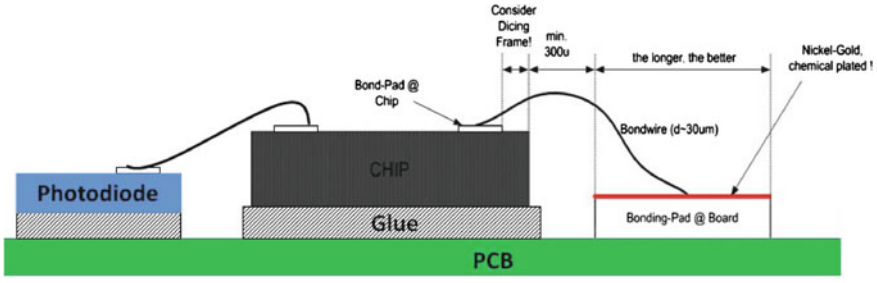

Fig. 4.5 Connecting the III/V photodiode chip via bond wires with the Si chip

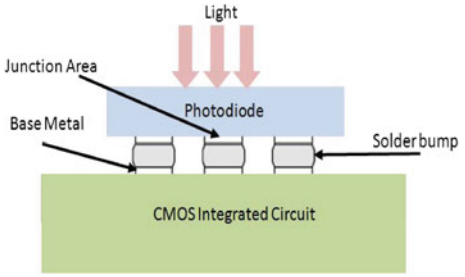

Fig. 4.6 Photodetector chip and a receiver chip which are flip-chip bonded

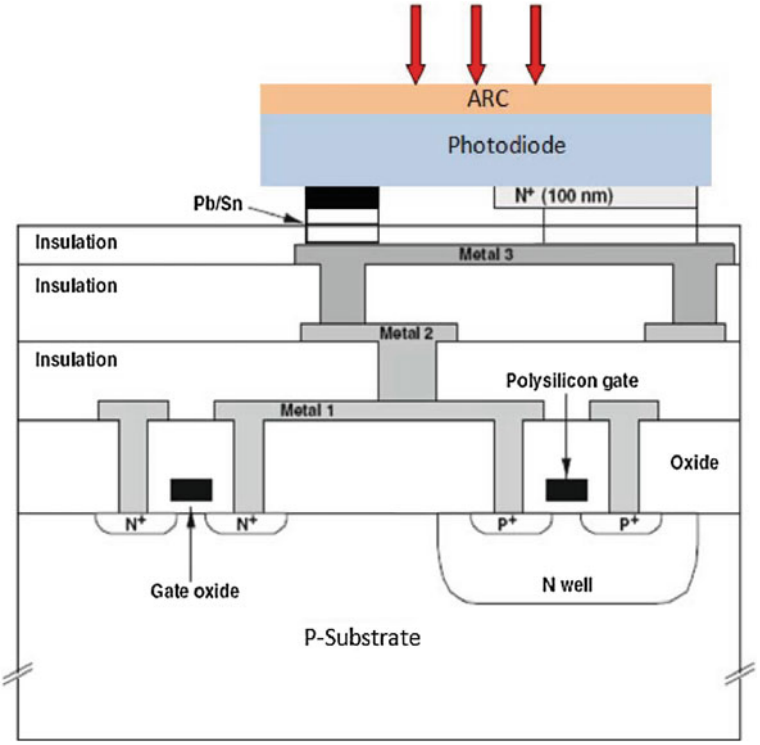

Fig. 4.7 Cross section of a hybrid integrated AlGaAs photodiode and CMOS circuitry [11]

The solder, moreover, enables bonding at a lower temperature (300 °C) than the gold bump bonding technique and, therefore, reduces damage to the devices, like an increased dark current of InGaAs photodiodes for instance. In [10], solder bumps with a diameter of 26 µm were used. The solder bump, for contacting the photodiode was fabricated over the junction area of the photodiode (Fig. 4.6).

The surface tension of the molten solder moves the chip and accomplishes precise chip positioning. An alignment accuracy of approximately 10 μm is sufficient due to the self-alignment effect during the reflow process. A positioning error of less than 1 μm, which is almost equal to the tolerance of the bump diameter, was reported. Figure 4.7 shows the flip-chip integration of a GaAs-AlGaAs PIN photodiode, to a CMOS chip. The multi-quantum-well (MQW) GaAs-AlGaAs PIN-PD had a responsivity of 0.45–0.5 A/W for $\lambda = 850$ nm. The MQW PIN-PD had a size of 2045 μm^2 and a dark current of about 10 nA at a reverse bias of 10 V. The advantage of the GaAs-AlGaAs PIN-PD was its large absorption coefficient, enabling a thinner absorption layer and a larger bandwidth compared to a silicon photodiode [11].

References

1. A. Marchlewski, W. Gaberl, H. Zimmermann, A SiGe Optical Receiver with Large-area Photodiode, in Proceedings of SPIE, Optical Sensing Technology and Applications, vol. 6585 (2007), pp. 65851Q1–Q9
2. M. Atef, H. Zimmermann, *Optical Communication over Plastic Optical Fibers: Integrated Optical Receiver Technology* (Springer, Berlin, 2013)
3. http://www.xfab.com/home/
4. A. Marchlewski, H. Zimmermann, I. Jonak-Auer, G. Meinhardt, E. Wachmann, Improvement of universal PIN Photodetectors in 0.35 μ m SiGe BiCMOS technology. Electron. Lett. **45**(13), 705–706 (2009)
5. http://www.truelight.com.tw
6. H. Zimmermann, *Integrated Silicon Optoelectronics*, 2nd edn. (Springer, Berlin, 2010)
7. L. Vivien, A. Polzer, D. Marris-Morini, J. Osmond, J.M. Hartmann, P. Crozat, E. Cassan, C. Kopp, H. Zimmermann, J.M. Fdli, Zero-bias 40 Gbit/s germanium waveguide photodetector on silicon. Opt. Express **20**(2), 1096–1101 (2012)
8. L. Vivien, J. Osmond, J.-M. Fdli, D. Marris-Morini, P. Crozat, J.-F. Damlencourt, E. Cassan, Y. Lecunff, S. Laval, 42 GHz p.i.n germanium photodetector integrated in a silicon-on-insulator waveguide. Opt. Express **17**(8), 6252–6257 (2009)
9. G.P. Agrawal, *Fiber-Optic Communication Systems*, 3rd edn. (Wiley, NewYork, 2002)
10. K. Katsura, T. Hayashi, F. Ohira, S. Hata, K. Iwashita, A novel flip-chip interconnection technique using solder bumps for high-speed photoreceivers. IEEE J. Lightwave Technol. **8**(9), 1323 (1990)
11. A.V. Krishnamoorthy, D.A.B. Miller, Scaling optoelectronic-VLSI circuits into the 21st century: a technology roadmap. IEEE J. Sel. Top. Quantum Electron. **2**(1), 55 (1996)

Chapter 5
Integrated Photodiodes in Nanometer CMOS Technologies

This chapter contains aspects like technology selection and scaling of photodiode performance. Classical PN junction, double-junction photodiodes, finger photodiodes, PIN photodiodes, a spatially modulated light detector, a triple-junction photodetector and avalanches photodiodes in nanometer CMOS are described. A comparison completes this chapter.

5.1 Effects of Technology Selection and Scaling on Photodiode Performance

BiCMOS processes combine the advantages of both bipolar (BJT) and MOSFET transistors. A kind of PIN-PD can be constructed using the buried N+ layer as PD cathode, the n collector epitaxial layer can serve as the intrinsic region of a PIN-PD, and the anode can be constructed from the P-base or P+ source/drain implant. The main advantage of such a kind of a PIN diode is its high electric field if the I-region is depleted. In this region, carrier transport is drift, which results in a fast photodiode, but the disadvantage is a low responsivity due to the thin collector layer. A true PIN photodiode with a thick intrinsic layer can be integrated in BiCMOS and bipolar technologies with certain process modifications [1]. The intrinsic region width of such true PIN-PDs can be optimized for a certain wavelength penetration depth. Such a large intrinsic layer thickness requires a low doped epitaxial layer to spread the electric field over the entire intrinsic region. However, this can destroy the performance of the BJT. The integrated photodiode needs a thick lowly doped intrinsic layer whereas the BJT needs a thin higher doped layer for the n collector. Therefore, process modifications are needed to optimize the performance of both PIN-PD and BJT. BiCMOS PIN-PDs achieve higher performance than the CMOS PDs at the expense of a more complex technology [1, 2].

Silicon on Insulator (SOI) technology builds the transistors on a very thin layer of silicon. An insulating layer is placed between the thin layer of silicon

M. Atef and H. Zimmermann, *Optoelectronic Circuits in Nanometer CMOS Technology*, Springer Series in Advanced Microelectronics 54,
DOI 10.1007/978-3-319-27338-9_5

and the substrate. The buried oxide layer prevents the slowly diffusing carriers generated in the substrate to reach the photodiode implemented on the thin silicon layer. Photodetectors implemented in SOI have a higher speed compared to standard CMOS photodetectors. On the other hand, a low responsivity is expected, especially at long wavelengths, due to the large percentage of substrate generated carriers which will not contribute to the photodiode current. Also SOI technology involves higher costs than their standard CMOS counterparts.

The low cost is the main motivation for choosing silicon CMOS technology for integrated photodiodes. The new nanometer CMOS technologies like 40 nm, however, are expensive. But in a few years, the high manufacturing cosst will be reduced by the large production volumes and due to the introduction of even smaller-structure size technologies. Also, the trend towards optical interconnects for rack-to-rack and chip-to-chip connects instead of wire communications will increase the mass production volumes and reduce the production costs. The photodiode integration with the receiver circuitry reduces the number of external components, resulting in lower system costs. Furthermore, an integrated photodiode reduces the total input capacitance by eliminating the parasitic due to the diode's package, the PCB wiring, the IC package and two bond pads plus bond wire(s). In such a way the transimpedance-bandwidth product and the EMI immunity are increased.

Nanometer CMOS technologies allow the integration of high-performance digital signal processing and complex logic functions within a much smaller chip area than in larger structure-size CMOS or BiCMOS technologies. The high switching speed of nanometer-size MOSFETs allows for high clock frequencies and therfore also for high signal frequencies to be processed digitally. Therefore, a cost advantage of nanometer CMOS OEICs arises for complex analog-digital optical receivers and sensors.

Standard CMOS technology is the cheapest silicon technology because of the lower number of production masks and large volume production compared to BiCMOS and SOI technologies. The main challenge of choosing standard nanometer CMOS for light detection is that the technology is not optimized for optoelectronic devices. Moreover, as the technology scaled down the layers' doping levels were increased and the supply voltage is decreased. The technology's down scaling will increase the disadvantages of the integration of photodiodes in newer CMOS technologies because of four reasons. First, the width of the space-charge region is small due to the high doping and the limited bias voltage which does not match the penetration depth of light in silicon. This will lead to slow photodiodes due to the large diffusion components in the photocurrent. As the maximum junction width decreases by technology down scaling the dynamic responsivity will be reduced because the internal dynamic quantum efficiency will decrease [1]. Second, the limited biasing voltage and the high doping have the effect to increase the junctions' capacitance and as a result the bandwidth will be decreased. Third, as the technology scaled down the number of metal and insulation layers increased which increased the optical reflection and optical interference within the insulator stack; this decreased the optical quantum efficiency and so the responsivity as well as increased the spectral fluctuation in it. Finally, the design rules for metal density per area become more

restricted as the technology scaled down. As a result unwanted metal layers must be distributed over the photodiode to fulfill the metal density rules. This reduces the photodiode's active area and part of the light will be reflected from the undesired metal regions and the responsivity will decrease. The above limitations result in integrated photodiodes in nanometer CMOS with smaller responsivity and lower speed compared to commercially available discrete diodes or photodiodes integrated in other technologies as e.g. OPTO ASIC BiCMOS technologies. We should overcome these CMOS photodiode limitations to implement cheap but high performance integrated optical receivers.

In the coming section different structures will be introduced to enhance the speed and responsivity problems in nanometer CMOS technologies. The fingered PD can be used to increase the electric field near the surface and so the diffusion current will be decreased. Spatially modulated light (SML) detectors can be used to eliminate the substrate diffusion current. Both fingered and SML PDs increase the PD speed and also solve the problem of the metal density; the metals are distributed over all the PD area. Avalanche PDs (APDs) can be implemented to achieve a high responsivity and an extended bandwidth due to the involved inductive behavior due to the avalanche effect. An anti-reflection coating (ARC) can be used to reduce the light reflection from the insulator stack and it can be optimized for a certain wavelength.

5.2 Classical PN Junctions

The simplest way to build CMOS photodiodes is to use the classical PN junctions available in CMOS processes: source or drain/substrate, source or drain/well, and well/substrate diodes. Figure 5.1 shows the structure of the proposed PDs fabricated with a standard 40 nm CMOS process from Taiwan Semiconductor Manufacturing Company (TSMC). In Fig. 5.1a the N+ source/drain implant with the P-substrate are used to implement the PN photodiode. In Fig. 5.1b the N-well with the P-substrate implement another PN photodiode structure. Due to the shallow trench isolation (STI), the breakdown voltages of the N+/P-substrate photodiode is 18.5 V and the breakdown voltage of the NW/P-substrate photodiode is 16.3 V. The STI prevents premature edge breakdown for N+/P-substrate PD but cannot do this for the NW/P-substrate PD because the NW is deeper than the STI. Therefore, the maximum bias voltages of 18 and 14.8 V were chosen for the N+/P-substrate and the NW/P-substrate photodiode, respectively. The leakage currents for these diodes with an area of 70 μm $\times$ 70 μm were 50 nA at these bias voltages. Each one of the two PDs is designed with different dimensions 30 μm $\times$ 30 μm, 50 μm $\times$ 50 μm and 70 μm $\times$ 70 μm. Due to the strict metal density rules in this technology; the PD must have some metal inside it. This is managed by covering a part of the PD far from the PD center by some metal to fulfill the metal density requirements. A metal layer is added to surround the PDs' active area for shielding the lateral PN junction. This metal shield prevents the diffusion current coming from the surroundings of the PD in case of a large incident light spot diameter [3].

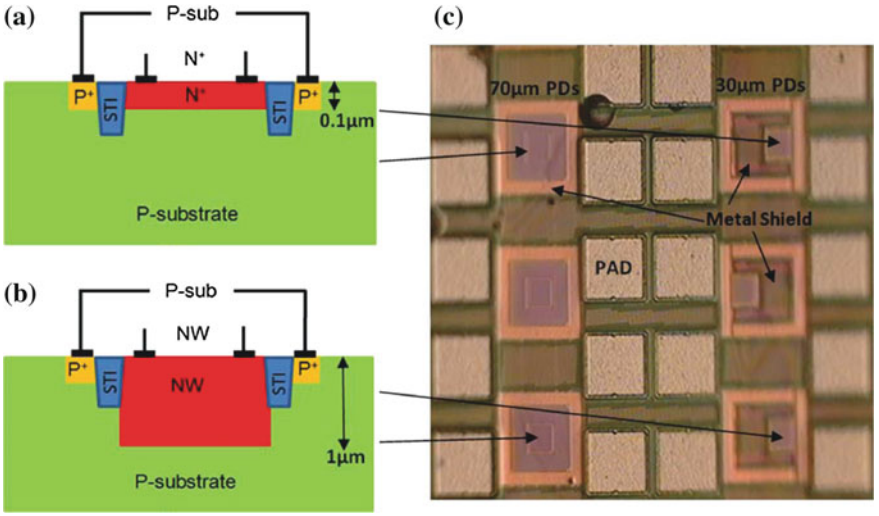

Fig. 5.1 Photodiode structures **a** N+/P-substrate, **b** NW/P-substrate; **c** die photo for the different photodiode structures with different dimensions [3]

A specific layer (NT-N) is added to prevent the automatic generation of a P-well in the photodiode area. This layer is used in the standard 40 nm CMOS process to design the so-called native NMOS on the P-substrate not inside the PW [3, 4]. A die photo of the fabricated photodiodes with different dimensions is shown in Fig. 5.1c. The responsivity of the N+/P-substrate and of the NW/P-substrate photodiodes is measured using a wavelength-variable light source, a 175 W short-arc Xenon lamp with broadband output followed by a Digikröm CM110 monochromator. The wavelength of the incident light is stepped from 400 to 900 nm in 10 nm wavelength steps and the output currents of the PDs are measured. A beam splitter is used to connect a monitor path with a calibrated reference detector.

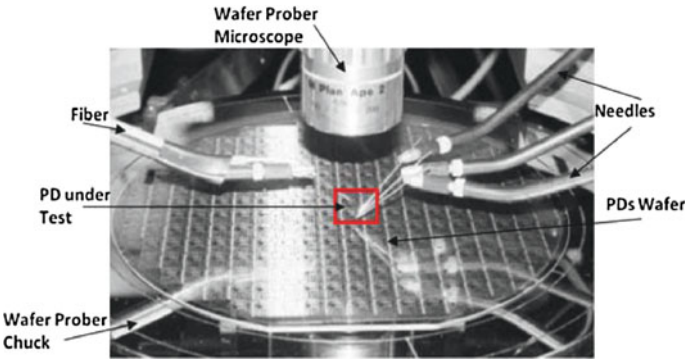

Fig. 5.2 Photodiodes measurement setup on the wafer prober

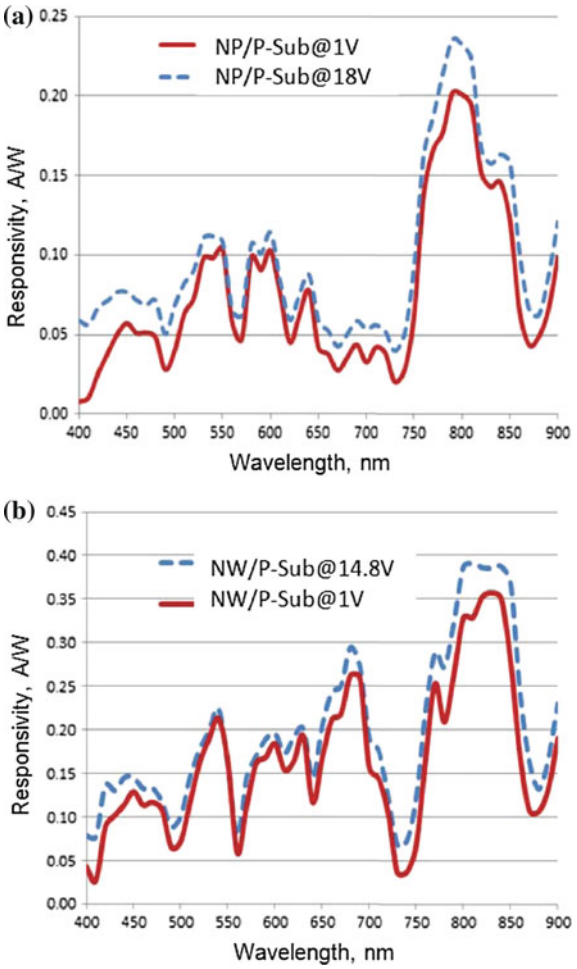

Fig. 5.3 DC responsivity for **a** N+/P-substrate and **b** NW/P-substrate photodiode [3]

The wafer containing the PDs is mounted on wafer prober and probe needles are used to connect the PD under test with the biasing, excitation sources and measurement devices, see Fig. 5.2. The laser source (660 and 520 nm) is stimulated by a rectangular signal from a pulse generator at a frequency of 80 MHz. The laser source's modulated optical signal is coupled to the PD through a tapered multimode fiber. The generated PD photocurrent is coupled through probe needles to the oscilloscope (Tektronix CSA 8000) to measure the transient response. A bias-T is used to bias the PD and pass the AC signal to the oscilloscope.

The measured responsivity curves of the N+/P-substrate and NW/P-substrate PD are shown in Fig. 5.3a, b, respectively. Both responsivity curves in Fig. 5.3 show maxima and minima because of optical interference in the dielectric stack. As there are 10 copper layers and one aluminum layer in this process, there are a lot of different

dielectric layers the light needs to pass. The distance between the top dielectric layer and the silicon surface is around 6 μm in this process. This results in a very irregular transmission characteristics as a function of wavelength. As the thickness of each dielectric layer can vary by as much as 20 %, the optical transmission coefficient can vary strongly for the process corner cases and from chip to chip as well as from wafer to wafer.

The low responsivity at short wavelengths compared to the larger responsivity around 800 nm can be explained by two reasons. First, the short wavelength light is strongly absorbed closer to the silicon surface; in general the responsivity of silicon PDs peaks nearly around 650 to 850 nm (depending on substrate doping concentration and carrier lifetime) and then gradually falls at shorter wavelengths. Second, strong reflection of light with short wavelengths as well as shallow light penetration in connection with partial absorption of the short wavelengths in the dielectric stacks occurs. By increasing the bias voltage the responsivity increases due to the increase of the quantum efficiency as a result of higher electric field in the space-charge region. The NW/P-substrate PD shows a higher responsivity than the N+/P-substrate PD because of the NW's lower doping compared to the N+ layer's doping. Lower doping increases the carriers' life-time and decreases the recombination rate and, thus more carriers can be collected in the NW compared to the N+ region. The NW region is deeper than the N+ region, which increases the responsivity for the NW PD at longer wavelengths compared to the N+/P-substrate PD's responsivity. In addition on top of N+ regions silicide is formed in order to reduce the series resistance of sources and drains. Silicide was therefore present on top of the N+ cathode, which reduced the responsivity even stronger than the recombination within the N+ layer itself.

The responsivity of the presented PDs can be increased by increasing the reverse bias voltages (more than 18 V for the N+/P-substrate and more than 14.8 V for NW/P-substrate PD) to induce avalanche multiplication but at the expense of larger leakage current and larger excess noise [5].

The capacitances of the two PN PDs with different dimensions (30 μm × 30 μm, 50 μm × 50 μm and 70 μm × 70 μm) are measured with a HP 4284A LCR meter. Figure 5.4a, b shows the capacitance of the N+/P-Substrate and NW/P-substrate PDs against the applied bias voltage. By increasing the bias voltage the depletion layer width increases so the PD capacitance decreases. By increasing the PD area its capacitance also increases. The N+/P-substrate and the NW/P-substrate PDs show comparable capacitance values. As the capacitance of the NW/P-substrate PD goes from 0.04 pF at 30 μm × 30 μm, via 0.12 pF for 50 μm × 50 μm to 0.21 pF for 70 μm × 70 μm. The expected PD capacitance for 300 μm × 300 μm is 4 pF. This kind of large-area PD is required for large core diameter POF. The RC-limited 3dB bandwidth of this 300 μm × 300 μm PD is 800 MHz (assuming 50 Ω resistance), which is enough for 1.25 Gbit/s being required for home networks using step-index (SI) POF [6, 7].

To investigate the frequency response of the PDs at different wavelengths, a 850 nm VCSEL, a 660 nm VCSEL and a 520 nm edge emitting laser diode were used with different available optical power. The PDs are biased through a bias-tee,

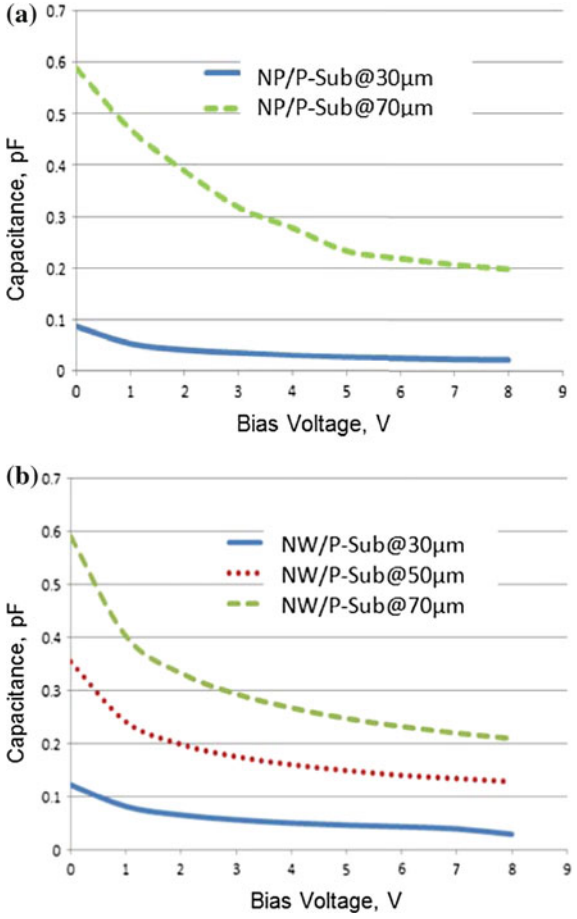

Fig. 5.4 Measured capacitance for different PDs with different sizes **a** N+/P-substrate, **b** NW/P-substrate [3]

which in turn is connected to a vector network analyzer (R&S ZVM). The 660 nm VCSEL and the 520 nm laser's bandwidth are 2.5 and 1 GHz, respectively. A calibration for the light source and cables is made using a 12 GHz NEWFOCUS photodiode to cancel the effect of these elements on the PDs under test. The light sources are modulated directly by the vector network analyzer. Then the frequency responses of each PD and each light source are recorded. Figures 5.5 and 5.6 show the measured frequency responses of the N+/P-Substrate and NW/P-substrate PD, respectively, for different biasing voltages and light sources. The frequency response of the 30 µm × 30 µm N+/P-Substrate PD introduced in Fig. 5.5a for 850 nm shows a very small bandwidth (40 MHz at 18 V) because of the larger diffusion current. The N+ region is very shallow (0.1 µm) and most of the photons for the 850 nm are absorbed deep in the substrate that is far away from the junction. Even with the

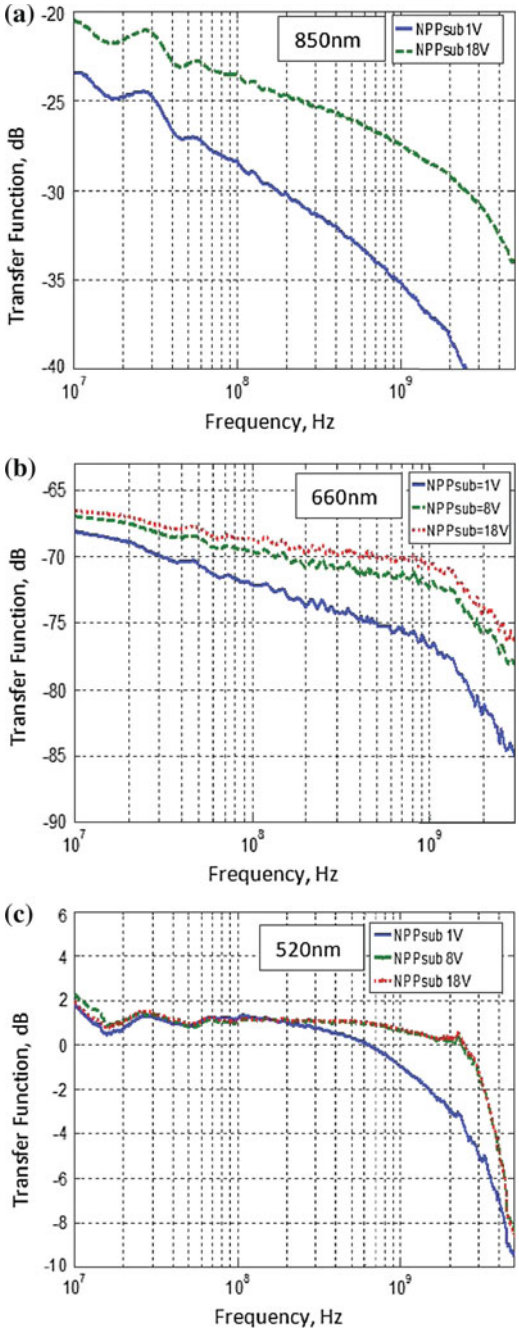

Fig. 5.5 Measured frequency response for NP/P-substrate PD using different light sources **a** 850 nm, **b** 660 nm, **c** 520 nm [3]

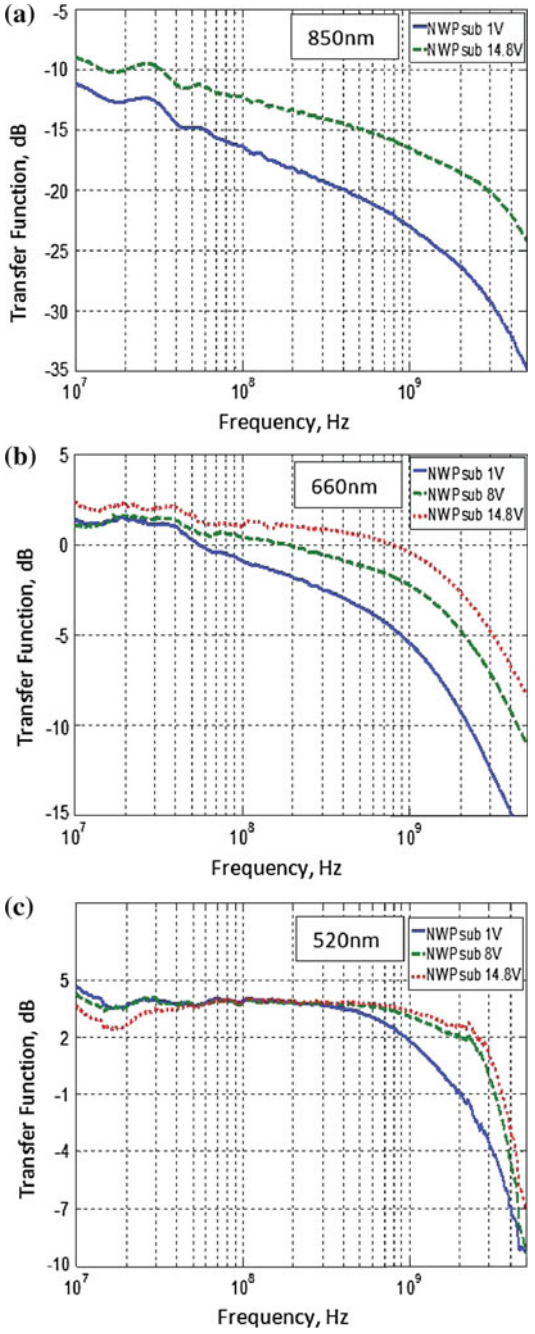

Fig. 5.6 Measured frequency response for NW/P-substrate PD using different light sources **a** 850 nm, **b** 660 nm, **c** 520 nm [3]

Table 5.1 Measured bandwidths for N+/P-substrate and NW/P-substrate photodiodes

	Bandwidth
N^+/Psub@850 nm	40 MHz@18 V
N^+/Psub@660 nm	1.0 GHz@18 V
N^+/Psub@520 nm	3.0 GHz@8 V
NW/Psub@850 nm	40 MHz@14.8 V
NW/Psub@660 nm	1.2 GHz@14.8 V
NW/Psub@520 nm	3.0 GHz@14.8 V

high bias voltage of 18 V (width of the space-charge region is 4.7 μm) carrier diffusion still dominates. Figure 5.5b for 660 nm wavelength shows a larger bandwidth of 1.0G Hz for 18 V due to less diffusion compared to the 850 nm wavelength. The frequency response for the 520 nm wavelength introduced in Fig. 5.5c shows a 1 GHz bandwidth at 1 V bias. This is because of the small penetration depth of 0.8 μm for this 520 nm wavelength. The higher the biasing voltage the higher the bandwidth is due to the higher electric field strength for carrier drift and the increased width of the depletion region. Because the width of the depletion region is increased, more photogenerated carriers drift in the depletion region instead of diffusing [3]. A bandwidth of 3 GHz was measured for 8 V which is nearly the same for voltages up to 18 V. Velocity saturation occurs at about 8 V and therefore the bandwidth does almost not increase for larger voltages.

The frequency response for the 30 μm × 30 μm NW/P-Substrate PD is introduced in Fig. 5.6a for 850 nm. The bandwidth is 40 MHz at 14.8 V, which is rather small because of the large diffusion current. The NW's depth is around 1 μm and most of the 850 nm light is absorbed in the substrate that is far away from the junction. Even for the high bias voltage of 14.8 V the diffusion current still dominates over the drift current.

The frequency response for the 660 nm wavelength introduced in Fig. 5.6b shows 200 MHz bandwidth at 1 V bias. This is because the 1/e penetration depth of 3.3 μm for 660 nm is comparable to the NW depth (about 1 μm) plus the depletion layer width of 1.5 μm (assuming a gradually doped PN junction) and most of the light is absorbed in the N-well and in the depletion region. A 800 MHz bandwidth is measured for 8 V and the bandwidth for 14.8 V is 1.2 GHz. The space-charge region widths are 3.9 and 4.8 μm for 8 and 14.8 V, respectively.

The frequency response for the 520 nm wavelength introduced in Fig. 5.6c shows a 1 GHz bandwidth at 1 V bias. A bandwidth of 2 GHz is measured for 8 V and it increased to 3 GHz for 14.8 V. Table 5.1 summarizes the measured bandwidths for N+/Psub and NW/Psub photodiodes [3].

The RC-limited −3 dB bandwidth of the 70 μm × 70 μm PDs is 15 GHz for the 50 Ω input resistance of the measurement equipment, which is higher than the measured 3 GHz. Therefore, the PD's frequency response is primarily limited by the carrier transit time in the absorption region and not by its area.

An integrated optical receiver will have the PD integrated with a transimpedance amplifier (TIA) which converts the photocurrent to a voltage signal. The TIA should have high sensitivity and low input impedance to reduce the effect of the PD's capacitance on the optical receiver bandwidth. A 70 μm × 70 μm N+/Psub PD was integrated with a TIA in a 40 nm standard CMOS technology. The p-type substrate which is the PD's anode is reverse biased with 5 V. This is possible because all circuits are placed inside a deep N-well or simple N-well. The optical receiver achieves a measured transimpedance gain of 78 dBΩ, 1.5 GHz bandwidth at 520 nm and a measured input referred noise current density of 7.2 pA/$\sqrt{Hz}$ up to 1.5 GHz [8].

The laser source (660 and 520 nm) is stimulated by a rectangular signal from a pulse generator at a frequency of 80 MHz. The measured transient responses using Tektronix CSA 8000 are shown in Figs. 5.7 and 5.8.

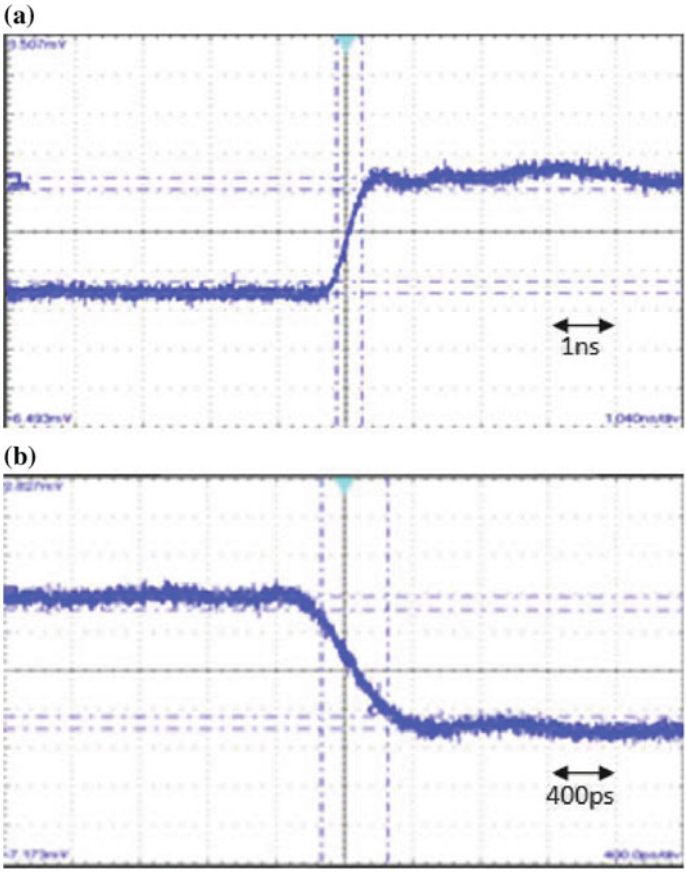

Fig. 5.7 Measured transient responses; **a** rise and **b** fall time determination for N+/Psub photodiode at 520 nm [3]

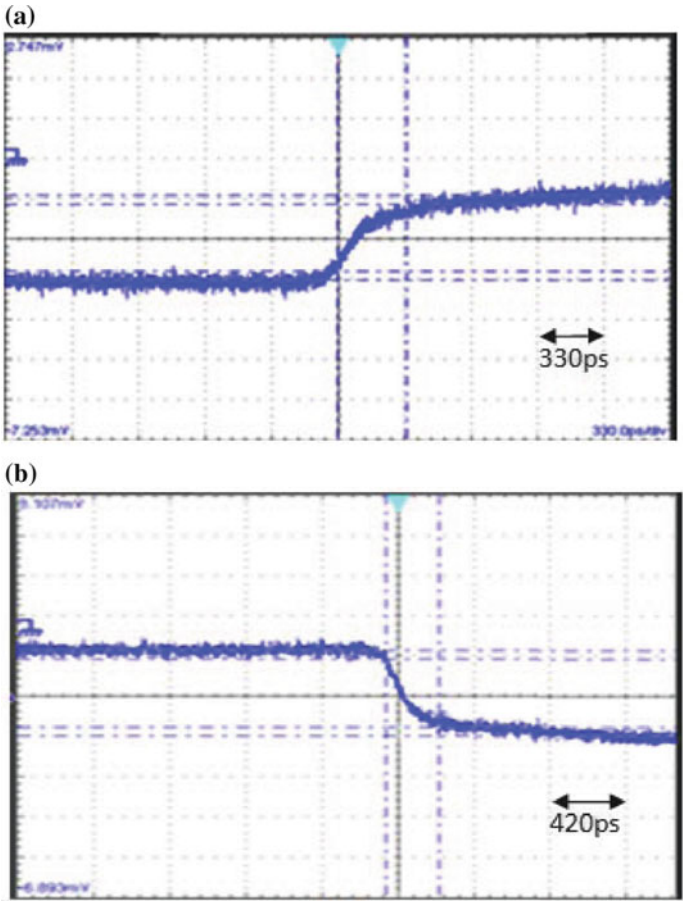

Fig. 5.8 Measured transient responses; **a** rise and **b** fall time determination for NW/Psub photodiode at 660 nm [3]

The reached 10–90 % rise and fall times and data rates are summarized in Table 5.2. For the 660 nm wavelength the N+/Psub PD shows rise and fall times of 350 and 405 ps where the NW/Psub PD has rise and fall times of 293 and 340 ps. The 520 nm laser source has rise and fall times (360 and 380 ps, respectively) in the same range as the photodiodes under test. The 520 nm laser is the limiting element in these measurements and the PDs have in fact much smaller rise and fall times. The rise and fall times and data rates are limited by the 1 GHz green laser's bandwidth. At 660 nm, the N+/Psub photodiode is slower than the NW/Psub photodiode, because the contribution of slow diffusion of photocarriers in the substrate is larger in the N+/Psub diode due to this PN-junction being closer to the silicon surface. For 520 nm, the N+/Psub photodiode is faster than the NW/Psub diode, because the 520 nm light does not penetrate so deep as the 660 nm light and the PN-junction of the NW/Psub

Table 5.2 Measured rise and fall times and data rates for N+/P-substrate and NW/P-substrate photodiodes using 660 and 520 nm light sources

	Rise time (ps)	Fall time (ps)	Data rate (Gb/s)
660 nm VCSEL	140	195	3
N^+/Psub@660 nm@18 V	350	405	1.3
NW/Psub@660 nm@14.8 V	293	340	1.6
520 nm Laser	360	380	1.35
N^+/Psub@520 nm@18 V	385	420	1.25
NW/Psub@520 nm@14.8 V	502	510	0.99

diode is in a larger depth than that of the N+/Psub diode; in addition there is carrier diffusion within the N-well of the NW/Psub diode, whereas the N+ region of the N+/Psub diode is much thinner than the N-well and the space-charge region of the N+/Psub diode within the P-substrate at 18 V is larger than the penetration depth of the 520 nm light [3].

5.3 Double-Junction Photodiodes

The double photodiode structure (DPD) can be used to increase the PD speed. The DPD was first introduced in [9] to extend the speed of photodiodes by extending the width of the depletion region. The DPD has two junctions with two vertically arranged depletion regions. One is the PW/DNW junction and the other one is the DNW/P-Substrate junction [4]. The inductive effect involved with the avalanche multiplication in the P+/NW/P-sub DPD can be used to increase the PD's bandwidth and responsivity [4, 10]. Another double junction PD with the P+/NW junction is used to detect the signal, and the substrate contacts work as guard to draw away the slowly diffusing carriers photogenerated in the substrate, if the NW cathode is connected to the amplifier input. The double junction PD with guard ring will be faster on the expense of a smaller responsivity [11, 12].

5.3.1 PW/DNW/P-Substrate Double Photodiode

A PW/DNW/P-substrate DPD was implemented in nanometer CMOS technology [4]. Figure 5.9 shows the structure of the PW/DNW/P-substrate DPD. A standard 40 nm CMOS process is used in this design without any process modifications. The DPD is formed by PW/DNW and DNW/P-substrate junctions. The breakdown voltage of the PW/DNW/P-substrate DPD is 15.22 V. The bias voltage of 15.1 V was chosen to achieve a maximum responsivity.

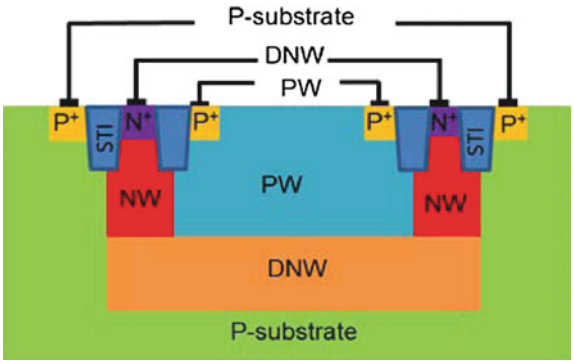

Fig. 5.9 DPD PW/DNW/P-substrate structure [4]

The leakage current for the PW/DNW/P-substrate DPD with an area of 70 µm × 70 µm at these voltages was 100 nA. Due to the strict metal density rules in this technology the PD must have some metal inside it. This was managed by covering a part of the DPD far from the DPD's center by some metal to fulfill the metal density requirements. A metal layer was added to surround the DPD's active area for shielding the lateral PN junction. This metal shield prevents diffusion currents coming from the surroundings of the DPD in case of a large incident light spot diameter. The responsivity of the PW/DNW/P-substrate DPD was measured using a wavelength-variable light source, a 175 W short-arc Xenon lamp with broadband output followed by a Digikröm CM110 monochromator. The wavelength of the incident light was stepped from 400 to 900 nm in 10 nm wavelength steps and the

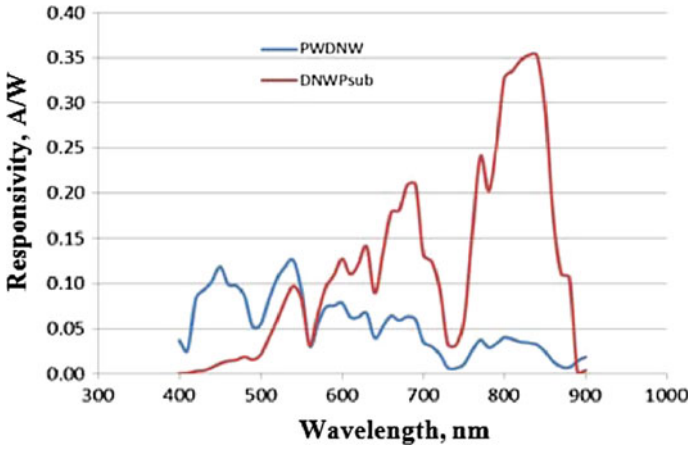

Fig. 5.10 DC responsivity for PW/DNW/P-substrate at 14.8 V [4]

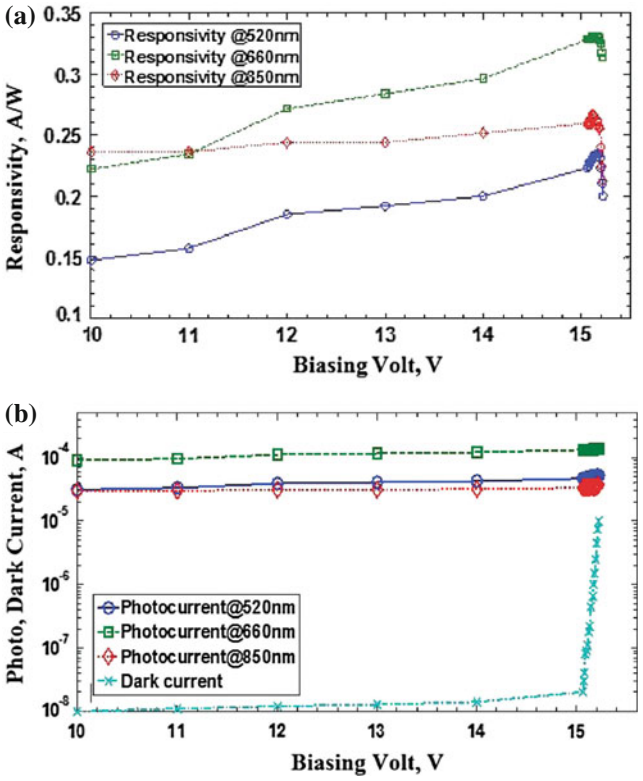

Fig. 5.11 **a** DC responsivity for PW/DNW/P-sub DPD as a function of biasing voltage for 520, 660 and 850 nm. **b** DC dark and illumination currents for PW/DNW/P-sub DPD as a function of biasing voltage for 520, 660 and 850 nm [4]

output currents of the PDs were measured. A beam splitter was used to connect a monitor path with a calibrated reference detector. For measurement of the DNW/P-sub diode's responsivity, the PW was shorted to the DNW. The measured responsivity curves of the PW/DNW/P-substrate DPD is shown in Fig. 5.10 [4].

By increasing the bias voltage the responsivity increases (Fig. 5.11a) due to the widening of the space-charge region and more drift current is collected. The PW/DNW/P-sub DPD's responsivity for 660 and 850 nm is shown in Fig. 5.11b in logarithmic scale. We see that the dark current increases dramatically above 15 V without a significant increase in the illumination current. This dark current increase is due to edge breakdown.

The capacitances of the DPD with dimension of 70 μm $\times$ 70 μm were measured with a HP 4284A LCR meter. Figure 5.12 shows the capacitance of the PW/DNW/P-substrate DPD against the applied bias voltage. By increasing the bias voltage the depletion layer width increases and the PD's capacitance decreases. The DPD's capacitance is the sum of the capacitances of two parallel junctions. The capacitance of the PW/DNW/P-substrate DPD is 1.9 pF at 8 V. To investigate the bandwidth of

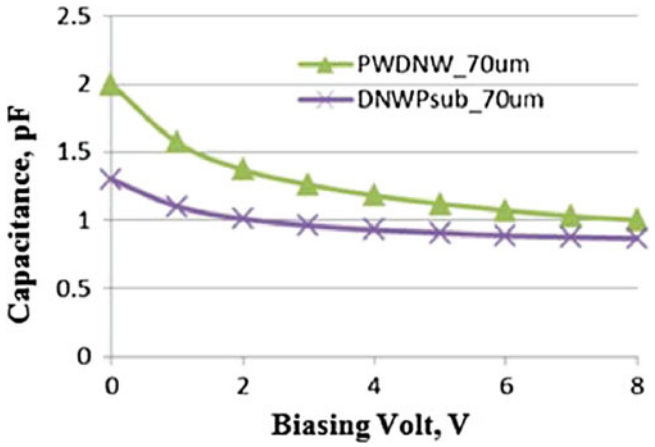

Fig. 5.12 Measured capacitance of PW/DNW/P-substrate DPD [4]

the DPDs at different wavelengths, a 850 nm VCSEL, a 660 nm VCSEL and a 520 nm edge-emitting laser diode were used. The PDs were biased through a bias-tee, which in turn was connected to a vector network analyzer (VNA). The 660 nm VCSEL and 520 nm LD bandwidths' are 2.5 and 1 GHz, respectively. A calibration for the light source and cables was made using a 12 GHz photodiode (from NEWFOCUS) to cancel the effect of these elements on the PDs under test. The light sources were modulated directly by the VNA. Then the frequency responses of each PD and each light source were recorded.

The frequency response for the 70 μm × 70 μm PW/DNW/P-Substrate DPD is measured for 850 nm, see Fig. 5.13a. The bandwidth is 4 MHz at 1 V, which is rather small because of the large diffusion current. The PW's depth is around 1 μm and most of the 850 nm light is absorbed in the substrate far away from the junction. Even with the high bias voltage of 15.1 V the bandwidth increased a little bit to 9 MHz because the diffusion current still dominates over the drift current. The frequency response for the 660 nm wavelength shows a 160 MHz bandwidth at 1 V bias, see Fig. 5.13b. This is because the 1/e-penetration depth of 3.3 m for 660 nm is comparable to the PW depth (about 1m) plus the depletion layer width of 1.5 μm (assuming a gradually doped PN junction) and most of the light is absorbed in the P-well and in the depletion region. A 370 MHz bandwidth was measured for 5 V and the bandwidth for 10 V and 15.1 V is 430 MHz and 450 MHz respectively. The space-charge region widths are 3.3 and 3.8 μm for 10 and 15 V, respectively. The frequency response for the PW/DNW/P-sub 70–70 m DPD indicates a maximum reached bandwidth of 800 MHz for 520 nm at 15.1 V. Table 5.3 summarizes the measured bandwidth for the PW/DNW/P-sub DPD at 15.1 V.

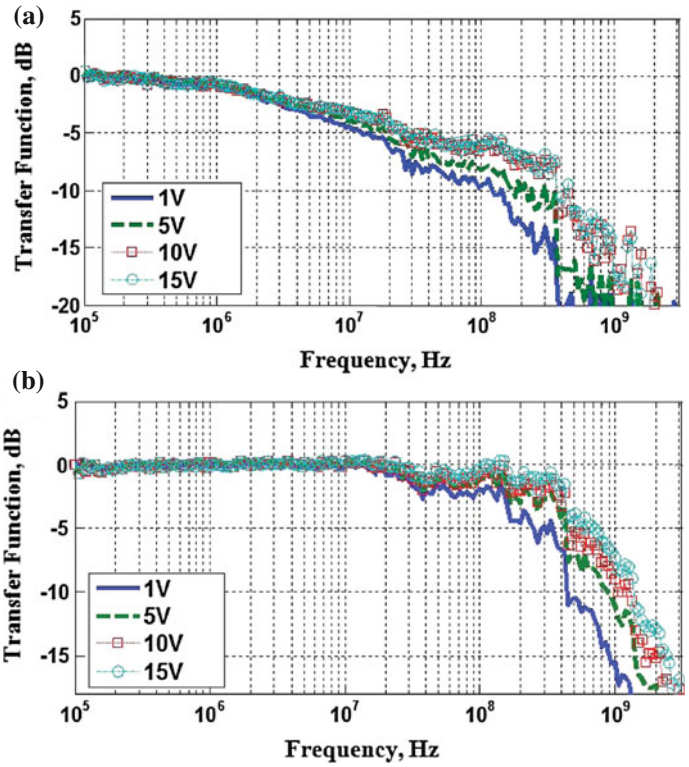

Fig. 5.13 Measured frequency response of PW/DNW/P-substrate DPD. **a** for 850 nm and **b** for 660 nm [4]

Table 5.3 Measured bandwidths of the PW/DNW/P-substrate photodiode at 15.1 V

	Bandwidth
PW/DNW/Psub@850 nm	9 MHz
PW/DNW/Psub@660 nm	450 MHz
PW/DNW/Psub@520 nm	800 MHz

5.3.2 P+/NW/P-Sub Avalanche Double Photodiode

Figure 5.14 shows the structure of a P+/NW/P-substrate DPD. A standard 40 nm CMOS process was used in this design without any process modifications. The P+ source/drain implants with NW, and the P-substrate are used to implement the avalanche double photodiode (ADPD) [4]. The basic structure of the presented ADPD is formed by P+/NW and NW/P-substrate junctions where the avalanche effect occurs at the P+/NW junction. The breakdown voltage of the P+/NW junction is 8.44 V. The shallow trench isolation (STI), prevents the premature edge breakdown for the P+/NW junction but cannot prevent it for the PW/DNW junction because the PW is deeper than the STI (Fig. 5.9). In principle, this ADPD can be connected to the

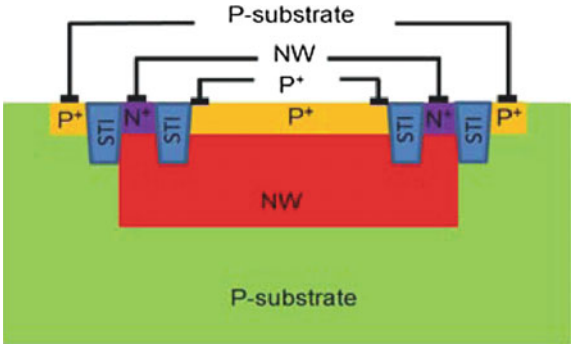

Fig. 5.14 P+/NW/P-substrate DPD structure [4]

amplifier in two ways: first, the NW cathode can be used as output and the P+ anode can be connected to a negative bias voltage and second, the P+ anode can be connected to the amplifier input and the NW cathode can be biased by a positive voltage. But, the second way should be preferred because slow diffusion currents from the substrate do not contribute. For the measurements, the bias voltage of 8.41 V was chosen for the P+/NW ADPD, to achieve a maximum responsivity. The leakage current for the P+/NW/P-substrate ADPD, with an area of 70 μm $\times$ 70 μm, is 3 μA at 8.41 V. Due to the strict metal density rules in this technology, the PD must have some metal inside it. This was managed by covering a part of the PD far from the PD's center by some metal to fulfill the metal density requirements. A metal layer was added to surround the PDs' active area for shielding the lateral PN junction. This metal shield prevents diffusion currents coming from the surroundings of the PD in case of a large incident light spot diameter [4].

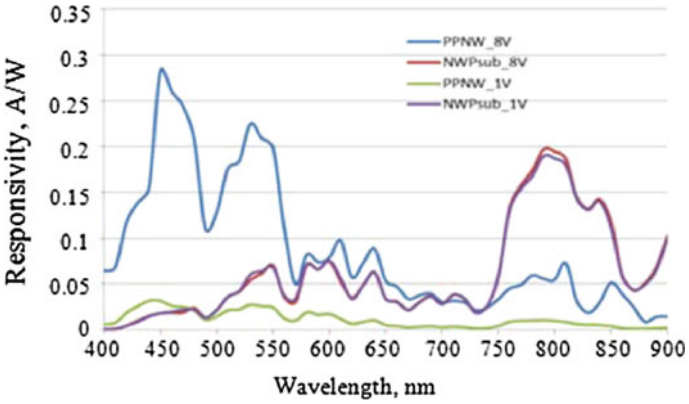

Fig. 5.15 DC responsivity for P+/NW/P-substrate photodiodes at 1 and 8 V [4]

To measure the responsivity of the P+/NW diode the substrate is shorted with the NW. For the NW/P-substrate diode's responsivity the P+ is shorted to the NW. The measured responsivity curve of the P+/NW/P-substrate ADPD is shown in Fig. 5.15. The responsivity curve shows maxima and minima because of optical interference in the dielectric stack. The low responsivity for short wavelengths at small bias voltages compared to the larger responsivity around 800 nm can be explained as in Sect. 5.3.1.

By increasing the bias voltage the responsivity increases due to the widening of the space-charge regions and more drift current will be collected. The large increase in the P+/NW/P-substrate PD's responsivity is because of the avalanche effect which occurs at the P+/NW junction. The STI prevents the premature edge breakdown for the P+/NW junction. The short wavelengths in Fig. 5.15 have higher responsivity than the long wavelengths for 8 V biasing. The short wavelengths are strongly absorbed in the p+ region. Since all photons are absorbed before the multiplication region; short wavelengths have a pure electron injection in the multiplication region. For long wavelengths electron-hole pairs generated beyond the multiplication region, only holes reach the multiplication region. The lower gain of long wavelengths results from the lower ionization coefficient of holes than electrons, so they undergo a reduced

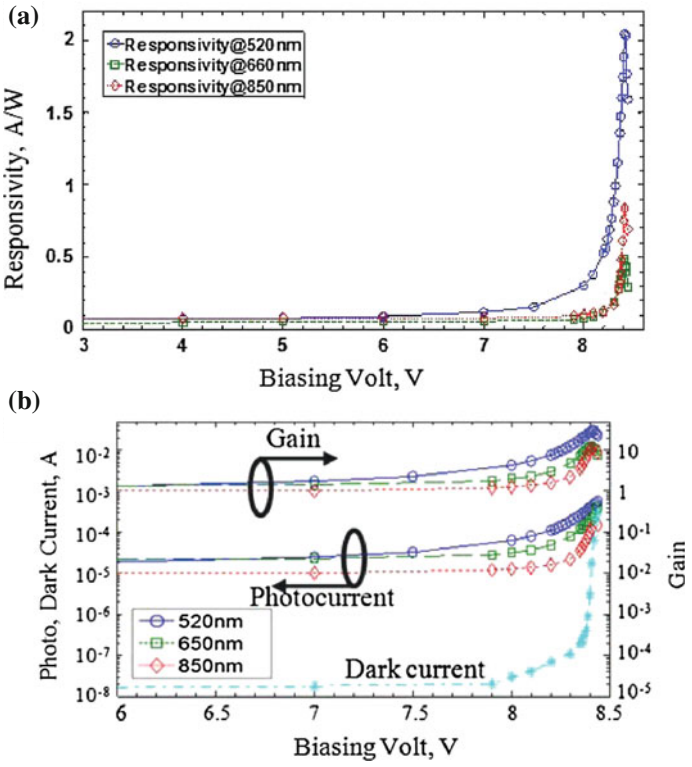

Fig. 5.16 **a** DC responsivity of P+/NW/P-sub ADPD, **b** dark and illumination currents as well as avalanche gain of P+/NW/P-sub ADPD [4]

gain. To measure the total responsivity for the P+/NW/P-sub DPD, the substrate is shorted with the P+. A 127 μW 850 nm vertical cavity surface emitting laser (VCSEL), a 405 μW 660 nm VCSEL and a 210 μW 520 nm edge-emitting laser diode (LD) were used. Figure 5.16a indicates that an avalanche multiplication occurs in the P+/NW/P-sub DPD's responsivity and peak responsivities of 0.84, 0.49 and 2.04 A/W were achieved for 850, 660 and 520 nm, respectively. The unity-gain responsivities are 0.076, 0.038 and 0.068 A/W for 850, 660 and 520 nm, respectively. Figure 5.16b shows that the dark current increases above 8 V with a significant increase in the illumination current due to avalanche multiplication. The leakage current of the P+/NW junction is 3 μA at 8.41 V. The multiplication factor is 11 for 850 nm, 13 for 660 nm and 30 for 520 nm. The multiplication factor for shorter wavelength is higher than for longer wavelengths because the avalanche effect occurs at the P+/NW shallow junction and not at the NW/P-sub deeper junction. The most of the 520 nm light is absorbed in the P+/NW diode where the avalanche multiplication occurs (see Fig. 5.15) [4]. The measured excess noise factor for the 520 nm light is $F = 12$ at a multiplication gain of 27. The ratio of hole to electron ionization probabilities is 25 % calculated by fitting the measured excess noise with the model developed by McIntyre [13].

The capacitances of the two junctions within the DPD with a dimension of 70 m × 70 m were measured with a HP 4284A LCR meter (Fig. 5.17). By increasing the bias voltage the depletion layer width increases and the PD capacitance decreases. The DPD's capacitance is the sum of the capacitances of the two parallel junctions. The capacitance is 3.12 pF for the P+/NW/P-substrate DPD at 8 V. The P+/NW/P-substrate diode has a higher capacitance than the PW/DNW/P-substrate diode (Sect. 5.3.1). The higher P+ doping concentration compared to the PW doping leads to a thinner depletion region and higher capacitance for the P+/NW junction compared to the PW/DNW junction [11].

To investigate the bandwidth of the DPDs at different wavelengths, a 850 nm VCSEL, a 660 nm VCSEL and a 520 nm LD were used. The PDs were biased

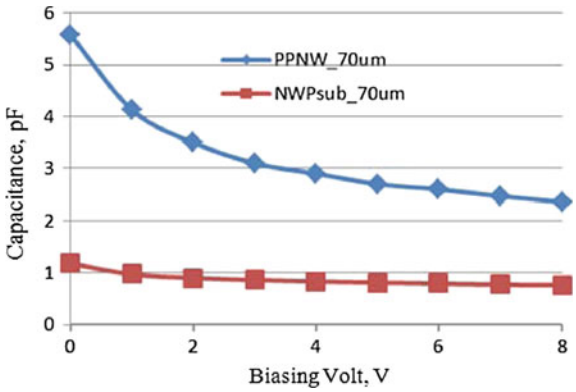

Fig. 5.17 Measured capacitance of P+/NW/P-substrate DPD [4]

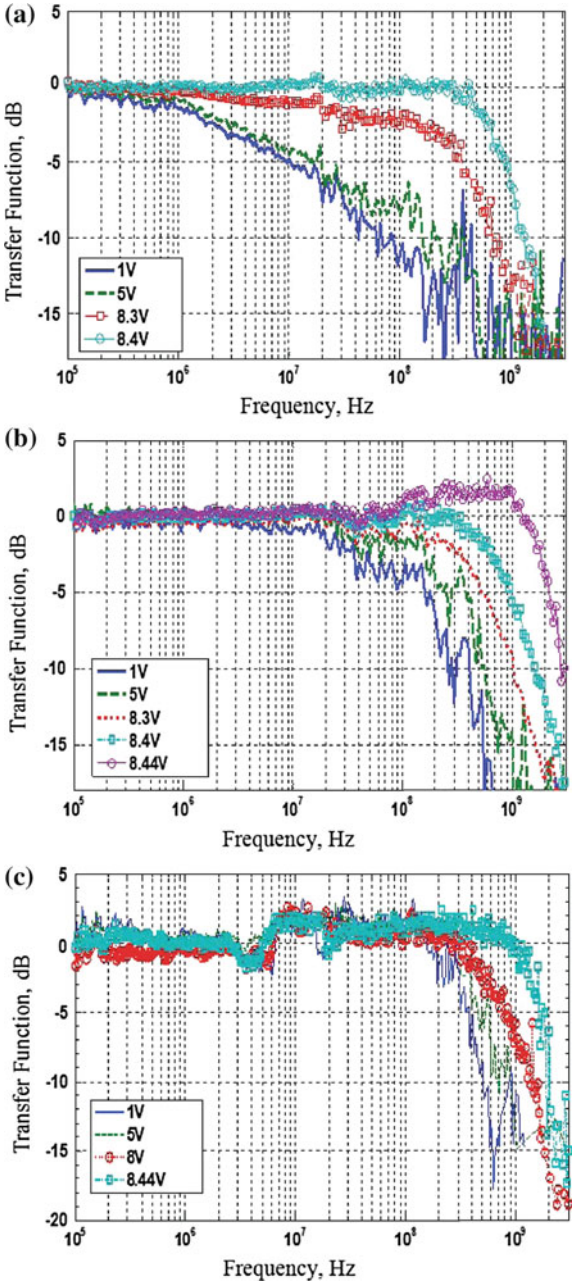

Fig. 5.18 Measured frequency response of the P+/NW/P-substrate ADPD using **a** 850 nm, **b** 660 nm and **c** 520 nm [4]

Table 5.4 Maximum measured bandwidths (bandwidths in the avalanche mode) for P+/NW/P-sub DPD

	Bandwidth
P+/NW/Psub@850 nm	700 MHz
P+/NW/Psub@660 nm	1.8 GHz
P+/NW/Psub@520 nm	1.4 GHz

through a bias-tee, which in turn was connected to a vector network analyzer (VNA). The 850 nm VCSEL, 660 nm VCSEL, and 520 nm LD bandwidths' are 8, 2.5 and 1 GHz, respectively. A calibration for the light source and cables was made using a 12 GHz photodiode (from NEWFOCUS) to cancel the effect of these elements on the PDs under test. The light sources were modulated directly by the VNA. Then the frequency responses of each PD and each light source were recorded. Figure 5.18 shows the measured frequency responses of the P+/NW/P-substrate DPD for different biasing voltages and light sources. The frequency response of the 70 μm × 70 μm P+/NW/P-substrate DPD introduced in Fig. 5.18a for 850 nm shows a very small bandwidth (3 MHz at 1 V) because of the large diffusion current. The P+ region is very shallow (0.1 μm) and most of the photons for the 850 nm are absorbed deep in the substrate far away from the junction. For higher bias voltages the bandwidth increases due to the wider space-charge region and the higher electric field in the two P+/NW and NW/P-substrate junctions. At 8.44 V bias the bandwidth increased to 700 MHz. Figure 5.18b for 660 nm wavelength shows a larger bandwidth of 150 MHz for 1 V due to less diffusion compared to the 850 nm wavelength. This is because of the small penetration depth of 3.3 μm for this 660 nm wavelength. The higher the biasing voltage the higher the bandwidth is due to the higher electric field strength for carrier drift and the increased width of the depletion region. Because the width of the depletion region is increased, more carriers drift in the electric field instead of diffusing. The bandwidth of 750 MHz is measured for 8.4 V and increased to 1.8 GHz for 8.44 V due to the inductive components introduced with the avalanche effect [5]. The frequency response for the P+/NW/P-sub 70 μm × 70 μm DPD at 520 nm is illustrated in Fig. 5.18c. The bandwidth increases by increasing the bias voltage. The maximum reached bandwidth is 1.4 GHz at 8.44 V. Table 5.4 summarizes the measured bandwidth for the P+/NW/P-sub avalanche double photodiode [4].

The RC-limited −3 dB bandwidth of the P+/NW/P-sub 70 μm × 70 μm DPD is 1.1 GHz for the 50 Ω input resistance of the measurement equipment, which is lower than the measured −3 dB bandwidth of 1.4 GHz. This is possible due to the inductive effect occurring in the avalanche mode [5].

The device under test was stimulated by a rectangular signal with a frequency of about 80 MHz. The measured transient responses are shown in Fig. 5.19. For 660 nm the P+/NW/P-sub ADPD has rise and fall times of 0.22 and 0.24 ns. The P+/NW/P-sub DPD shows a rise and fall times of 0.365 and 0.412 ns at 520 nm which is limited by the green laser's rise and fall times (0.36 and 0.38 ns, respectively) [4].

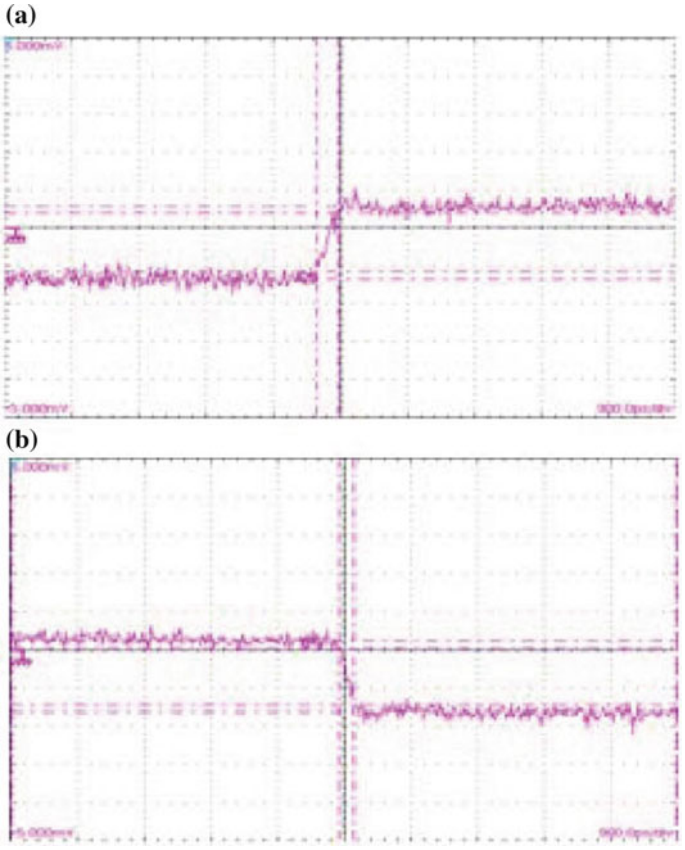

Fig. 5.19 Measured transient responses; **a** rise and **b** fall time determination for P+/ NW/Psub photodiode at 660 nm [4]

5.3.3 P+/NW/P-Substrate Photodiode with Guard

The DPDs introduced in Sects. 5.3.1 and 5.3.2 have a high capacitance; since the DPD has two junctions and the total capacitance is the sum of the two junction capacitances. Also the substrate diffusion current contributes to the total photocurrent; slowing down the PD speed. Another technique can be used to reduce the PD capacitance and to block the substrate diffusion current. The double junction PD is a P+/NW/P-sub with a P+/NW junction together with the NW/P-sub junction. The structure of P+/NW/P-substrate photodiode with guard ring is shown in Fig. 5.20a. P+/NW junction is used to detect the optical signal, and the P-sub/N-well junction acts as guard (equalizing PD) collecting the slowly diffusing carriers generated in the substrate [2, 12].

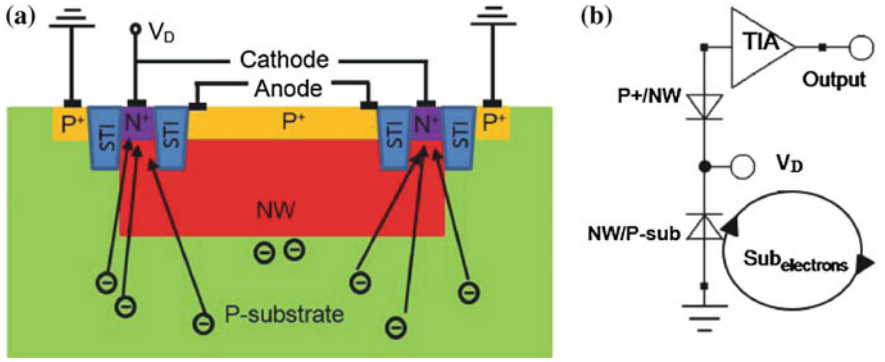

Fig. 5.20 **a** Structure of P+/NW/P-substrate photodiode with guard ring. **b** Photodiode currents and its connection to TIA [11]

The NW cathode current consists of the electrons diffusion current from the substrate as well as of a part of the drift component of the depletion region, while the P+ anode current consists of a part of the diffusing holes from the NW and of a drift current. Most of the diffusing electrons coming from the substrate will be collected by the cathode (NW). By connecting the NW cathode to a high positive voltage V_D, while connecting the anode (P+) to the subsequent amplifier, maximum speed can be achieved (see Fig. 5.20b) [11]. In other words, the P-sub/N-Well junction collects most of the diffusing photocarriers, shielding the upper photodiode junction P+/N-Well from these slow components. This technique increases the PD's speed, but on the expense of smaller responsivity. The capacitance of the P+/NW/P-sub with guard is only the capacitance of the P+/NW junction, which is smaller than the DPD capacitance.

Simulation results at 600 and 850 nm for a double junction PD with a dimension of 65 μm × 65 μm in 130 nm CMOS with guard were presented in [11]. The PD has a bandwidth of 790 MHz in response to 850 nm light and 624 MHz for 600 nm. The bandwidth of the double junction photodiode with guard is higher than the simple PN junction's bandwidth (Sect. 5.2) and the DPD bandwidth in Sect. 5.3.1, and the responsivity of the double junction photodiode with guard is less than that of the simple PN junction responsivity. This is due to shielding of the substrate current component in the double junction photodiode with guard. The double junction photodiode responsivity is 74.8 mA/W at 600 nm and 17.6 mA/W for 850 nm. For shorter wavelengths (600 nm), the responsivity is higher than at 850 nm wavelength. The reason is that for shorter wavelengths, the penetration depth of light is smaller and thus the percentage of substrate current is less. Therefore, less current is absorbed by the cathode and more current is available at short wavelengths.

A P+/NW/P-sub APD with guard was fabricated in 130 nm standard CMOS technology [14]. The PD is formed from the P+/NW junction with the P+ contacts for the substrate are tied to ground. In order to be able to exploit the avalanche effect, V_D has to be much larger than V_{dd}. An optical window having the area of

30 μm × 30 μm was formed by blocking the salicide process. The total device area was 43.8 μm × 43.8 μm. The photodetector was realized without violating any design rule for the CMOS process technology. For characterization of the APD, an 850 nm laser source was used. The avalanche breakdown voltage was about 9.8 V and the peak responsivity was 20.3 A/W. The PD had a 3-dB bandwidth of 2.5 GHz, which is higher than that of the DPD introduced in Sect. 5.3.2; due to the effect of the equalizing PD and the lower doping in 130 nm technology compared to 40 nm CMOS technology.

5.4 Finger Photodiodes

As described in Sect. 3.6, the short wavelengths (blue light) have low quantum efficiency and so responsivity. The majority of the photogenerated carriers, therefore, will strongly recombine near the surface and in the N+ region. The carriers will diffuse slowly due to the low electron and hole mobilities in the highly doped N+ region (Sect. 3.2.1). The slow carrier diffusion processes outside the space-charge region are responsible for a limited response speed.

To reduce both, recombination and diffusion, space-charge regions at the surface have to be implemented. An interdigitated network of N+ fingers is employed instead of a continuous N+ region for maximizing the depletion regions available for carrier collection, particularly near the surface of the device. An interdigitated structure is shown in Fig. 5.21. The P-sub PD is used as the anode and the N+ fingers are used as the cathode of the photodiode. The N+ fingers are connected by metal outside of the photo-sensitive region. Photogenerated carriers can be separated in the space-charge region, electrons drift to the N+ fingers and holes drift to the P-sub. For a small number of fingers the distance between fingers is so wide that carrier diffusion should be expected to contribute and to slow down the response speed. The photodiode speed can be increased by increasing the number of fingers. The space-charge regions of

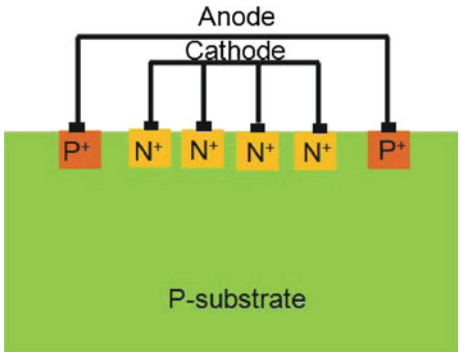

Fig. 5.21 Cross section of an interdigitated finger N+/P-sub photodiode

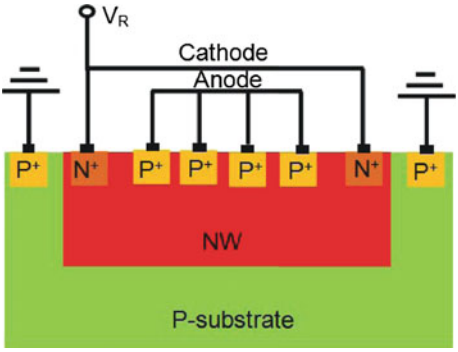

Fig. 5.22 Cross section of an interdigitated finger P+/NW/P-sub photodiode

neighboring N+ regions overlap and there is no carrier diffusion (outside the N+ regions) involved for blue light.

The N+/P-/P-sub fingered PD was introduced in [15]. The PD responsivity increased to 0.148 A/W at 400 nm (efficiency = 46 %) using 9 N+ fingers. A reference photodiode without interdigitated N+ region showed a lower responsivity of 0.045 A/W (efficiency = 13.9 %), both without ARC.

A P+/NW/P-sub interrupted P+ finger photodiode fabricated in 0.35 µm CMOS process, was introduced in [16, 17], see Fig. 5.22. The P+/NW junction forms the PD. The slow diffusion of carriers generated by light with a wavelength of 850 nm in the P-sub, therefore, does not contribute to the photocurrent, if the NW cathode is connected to a (positive) supply voltage and the anode is connected to the amplifier input. This slow current is collected by the NW and shorted to the power supply. To enhance the speed of the photodiode, an interdigitated network of P+ fingers is employed instead of a continuous P+ region for maximizing the depletion regions available for carrier collection, particularly near the surface of the device. The NW had a size of 16.5 µm × 16.5 µm. With a detector bias of 10 V a bit rate of 1 Gbit/s was reported. The responsivity of the photodiode with the reported values of 0.01–0.04 A/W was very low due to the shallow P+/NW junction. It should be mentioned that this wide range of responsivity is due to optical interference in the CMOS isolation and passivation stack (Sect. 3.5.2).

In [18], a PD with multiple PW/NW structure in standard 180 nm CMOS technology has been presented. The PW together with the P-sub and the NW form the avalanche structure, see Fig. 5.23. In addition, the standard shallow trench isolation (STI) is also implemented in between wells to reduce surface leakage and enhance breakdown characteristics. The PW is deeper than the NW and the PW is higher doped than the NW. The total active area of this PD is 50umx50um , which consists of seven PN diodes. For each diode, the width for PW (W_P), NW (W_N), STI oxide (W_{STI}), and contact metal strip is 1.95, 3.45, 0.36, and 0.45 µm, respectively. The distance that the generated carriers drift is determined by the width of each diode and thus affects the drift time directly. In this design, the width of each well and STI oxide is minimized and limited by design rules. At suitable reverse bias, the

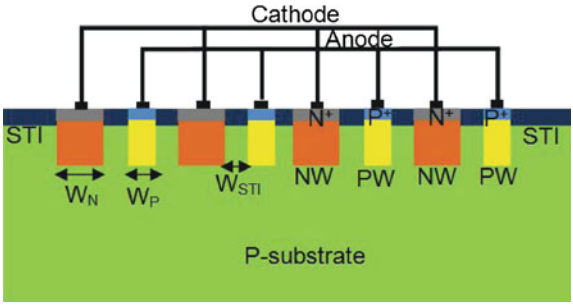

Fig. 5.23 Cross section of an interdigitated finger PW/NW/P-sub photodiode

NW and P-sub region are depleted and converted into a wide absorption region. By using STI oxide, it not only improves the breakdown voltage but also extends the depletion region. Therefore, the fully depleted region is wider than published Si PD using P-sub as P-region only. Accordingly, it would improve the responsivity. The PD shows a high responsivity of 0.37 A/W at zero bias. At a reverse bias of 14.3 V, PD demonstrates a high responsivity of 0.74 A/W, a 3-dB bandwidth of 1.6 GHz.

5.5 PIN Photodiode

A PIN-PD consists of a P (anode) and N (cathode) with an intrinsic (I) layer in between. The PIN-PD is reverse biased, to achieve a strong electric field in the intrinsic layer. Electron-hole pairs photogenerated in the intrinsic region, are immediately separated by the high electric field in the intrinsic region. The internal efficiency and the speed of the PIN-PD depend on the intrinsic region width. The wider the intrinsic region is, the higher is the internal efficiency (see Fig. 3.10). Also, the PD's capacitance is decreasing by increasing the intrinsic region's width (see 3.22) and a higher bandwidth can be obtained (3.29). On the other hand, the wider the intrinsic region's width is, the longer the electrons and holes need to drift through the intrinsic region. Figure 5.24 shows the effect of the intrinsic layer's width on the PIN-PD's bandwidth. Therefore a tradeoff between efficiency and speed should be found.

The P- epi-layer, which has lower doping than the heavily doped substrate, is used to pair with NW to form the intrinsic region (I) in a vertical PIN PD, Fig. 5.25. In such a PIN-PD structure, the electric field is largely confined inside the P- epi-layer, between the P+ substrate (Anode) and N+ region (cathode). One advantage of epi-CMOS PDs is that when the epi-layer is thin enough, a low bias voltage can deplete the entire epi-layer, because of its low doping level. Second advantage, the minority carrier recombination is faster in the P+ substrate than in the bulk CMOS process, and hence most photogenerated electrons inside the substrate recombine during diffusion before reaching the depletion region. This will reduce the diffusion current effect and a faster PD is expected but at lower responsivity. A 120 μm × 120 μm vertical PIN-PD was fabricated in 180 nm CMOS technology with epi-layer. The vertical PIN-

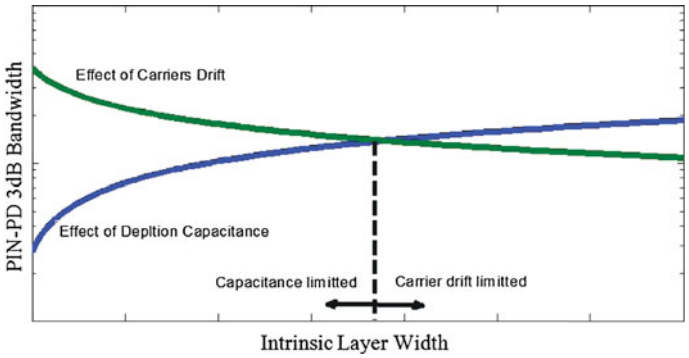

Fig. 5.24 Effect of intrinsic layer width on the PIN-PD bandwidth

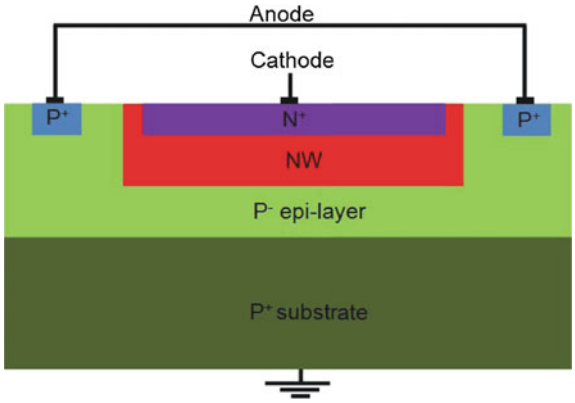

Fig. 5.25 Cross section of a vertical PIN photodiode [19]

PD showed a responsivity of 0.135 A/W, 0.77 pF capacitance, and a bandwidth of 1.15 GHz at 5 V reverse bias and for 850 nm light. At 15.5 V the bandwidth increased to 1.61 GHz [19].

A lateral interleaved P+/P-/N+ PIN detector was designed in [20]. Figure 5.26 illustrates the cross sectional view of this lateral PIN-PD. The PIN-PD is surrounded by NW and DNW. Compared to conventional N+/P substrate junctions, the electron-hole pairs are mainly generated in the laterally depleted regions. The DNW is connected to the highest positive voltage, which protects the PIN diode against substrate diffusion current. The finger spacing affects the drift time of the photogenerated carriers in the depletion region between anode and cathode fingers. The smaller the fingers' spacing is, the faster electrons and holes drift to the electrodes. On the other hand, the smaller the fingers' spacing is, the higher the PD's capacitance, and hence the smaller the bandwidth. The lateral PIN-PD is shielded from the substrate diffusion current by the DNW. When the spacing between the fingers is properly selected, the lateral PIN-PD can be faster than a vertical PIN-PD but at a lower responsivity.

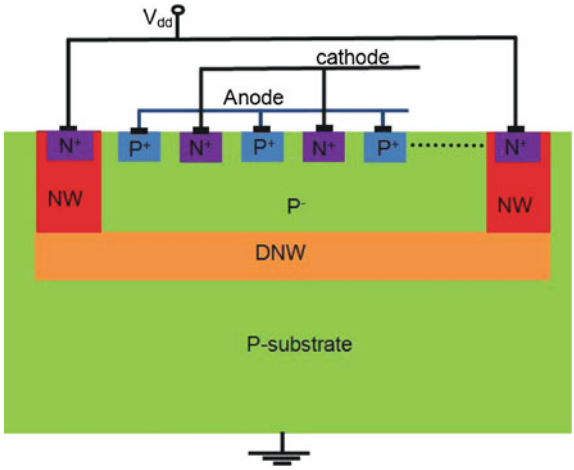

Fig. 5.26 Cross section of a lateral interleaved finger PIN photodiode

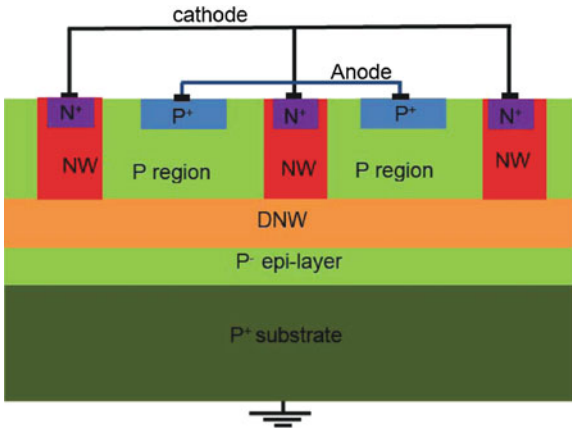

Fig. 5.27 Cross section of PIN photodiode with DNW [19]

A lateral PIN-PD with an area of 50 μm by 50 μm, consisting of 13 interleaved fingers was fabricated with a 180 nm CMOS process. The P+ and N+ strips were 1.45 μm wide, separated by a 0.5 μm wide P- region. A reverse bias of 6 V was sufficient to deplete the P- regions between the cathodes (N+) and anodes (P+) of the lateral PIN-PD. An optical receiver based on the lateral PIN-PD was connected to RGC-TIA achieved data rate of 2.5 Gb/s at −4.5 dBm sensitivity for 850 nm wavelength without using an equalizer [20].

A kind of PIN-PD structure can be implemented by adding inside a DNW a PW region as shown in Fig. 5.27 [19]. The DNW forms the cathode of the upper vertical PIN P+/PW/DNW PD. The NW is the cathode of the lateral P+/PW/NW PIN PD.

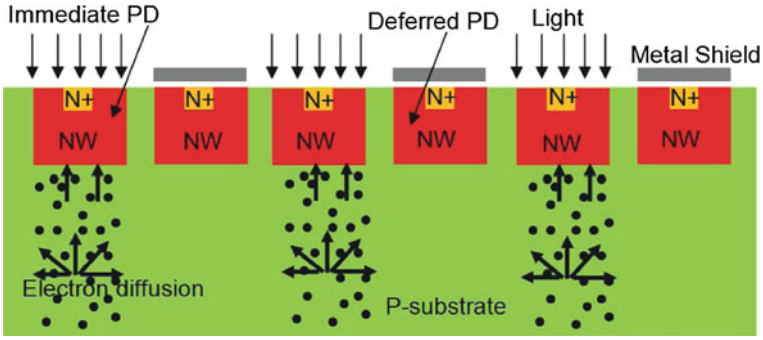

Fig. 5.28 Cross section of spatially modulated light detector

The bottom DNW/P-epi/P-sub junction, with the substrate connected to ground, can be used to isolate the PIN PD from the substrate and to block the substrate diffusion current. The PD structure with the DNW combines the advantages of vertical and lateral PIN devices to enhance the electric field inside the epi-layer. Therefore, it can improve the PD's bandwidth without sacrificing too much responsivity. A DNW PIN-PD with an area of 70 μm × 70 μm achieved a 3 dB bandwidth of more than 3.13 GHz at 15.5 V bias. At 5 V bias, the same DNW PD achieved 2.19 GHz bandwidth, versus about 1.15 GHz for a conventional vertical PIN PD [19]. At 850 nm, its responsivity was measured to be 0.14 A/W corresponding to a quantum efficiency of 20 % when biased below 10 V. The responsivity increased to 0.4 A/W or a quantum efficiency of 58 % at 16.2-V bias. Its dark current was 0.46 nA at 5 V bias [19].

5.6 Spatially Modulated Light Detector

The spatially modulated light (SML) detector was first introduced in [21] to increase the PD speed by subtracting a part of the diffusion current. Figure 5.28 shows the cross section of the SML detector. Half of the fingers are shielded from the incident light by floating metal strips. The shielded fingers connected together constitute the "deferred detector". The other unshielded fingers connected together constitute the "immediate detector". When light is incident on the detector, the metal mask blocks the incident light on the deferred detector, and light is absorbed in the immediate detector. Photocarriers are generated below the immediate detector region and not under the deferred region. The incident light is spatially modulated according to the metal covered and uncovered regions.

The carriers generated close to an illuminated junction will have a higher probability to be collected by the immediate detector junctions. Bulk generated carriers (diffusion) have the same probability to reach the immediate or differed detector junctions. The deferred current (slow diffusion only) is subtracted from the immedi-

ate current (both slow diffusion and fast drift components), to obtain the "effective detector" current. The effective detector current has a faster response; because part of the slow diffusive carriers is canceled by the subtraction.

The diffusion component in the P-sub determines the SML photodiode bandwidth. The smaller the distance between the adjacent NWs is, the better the diffusion cancellation works, and the higher the bandwidth is [22]. The speed is improved on the expense of a lower responsivity. Half of the light is reflected from the shielded fingers, and more deferred detector current is subtracted from the immediate detector current by increasing the number of fingers. The SML detector has a smaller capacitance compared to a reference PD, without using SML, because the SML is split into two PDs. The receiver sensitivity increases somewhat due to the low SML detector's capacitance decreasing the input capacitance. The low responsivity of the SML detector is partially compensated by the lower capacitance of the device.

The subtraction is performed by connecting the SML PDs to a differential transimpedance amplifier (TIA). The differential TIA has the advantages of higher rejection of supply noise, and better linearity due to the suppression of even harmonics compared to a single ended TIA.

The N+/NW/P-Sub SML photodetector introduced in [23], fabricated in 0.8 μm CMOS, has a DC responsivity of 0.05 A/W at 860 nm and 0.132 A/W at 635 nm for a reverse bias of 3.8 V. The photodetector shows a data rate of 300 Mb/s for 860 nm and up to 510 Mb/s for 635 nm. The responsivity for 635 nm light is larger than for 860 nm, which is not the case for classical PDs (Sect. 5.2) where 850 nm has a higher responsivity than 650 nm. The shorter wavelength has a smaller penetration depth; so less diffusing carriers are generated, and a smaller part is canceled by subtraction at shorter wavelengths.

In [24] a 75 μm × 75 μm SML N+/NW/P-sub photodetector was integrated in 180 nm CMOS technology with TIA and an equalizer. The PD had a net responsivity of 0.052 A/W at 2.6 V bias and a −3 dB bandwidth of 700 MHz. A metal layer over the circuits was used to block incident light. A higher metal layer would have resulted in a lower parasitic capacitance, but the incident light is blocked more effectively by using a metal close to the active circuitry.

A SML photodetector was integrated with TIA, equalizer, LA, and output buffer also in 130 nm CMOS process [25]. The SML PD had 0.05A/W responsivity at 1.5 V supply voltage. Simulations showed that the TIA with the integrated SML PD achieves the −3 dB bandwidth of 1.2 GHz at 850 nm light. An adaptive equalizer was integrated to increase the optical receiver bandwidth. The fully integrated optical receiver achieved a measured bandwidth of 5.9 GHz at 850 nm by implementing an equalizer. An optical sensitivity of −3.2 dBm for a BER of 10^{-12} was reported.

Medici simulations were performed in [2] for a 80 μm × 80 μm differential PD (same as SML PD) in 90 nm CMOS technology. The DC responsivity varied between 0.052 A/W for 18 fingers and 0.034 A/W for 34 fingers. More fingers correspond to more interconnection overhead, and a slightly lower responsivity. The intrinsic −3 dB bandwidth ranged from 1 GHz for 18 fingers to 4.4 GHz for 34 fingers. More fingers per diode mean less distance for the diffusing carriers resulting in a faster response.

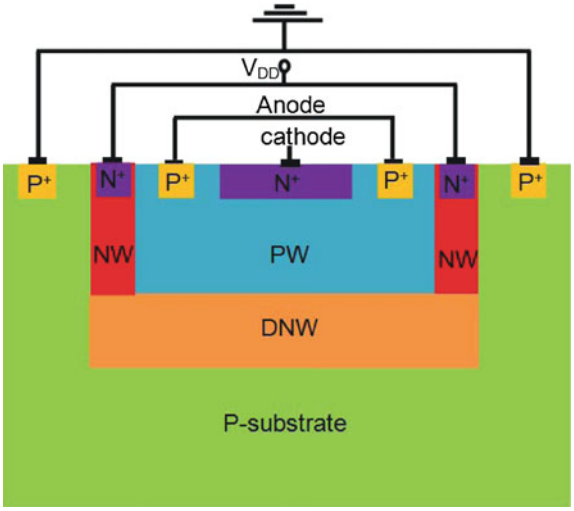

Fig. 5.29 Cross section of triple junction photodetector

5.7 Triple Junction Photodetector

To cancel out the slow diffusion components, N+/PW photodiode can be created inside a DNW as shown in Fig. 5.29. The electrons photogenerated in the P-sub diffusion region are collected by the DNW and shorted to the supply contact of DNW. The holes photogenerated inside the DNW diffusion region are collected by the PW and shorted to the ground contact of the PW [11]. A part of the electrons generated in the PW diffusion region is collected by the DNW and shorted to the supply contact of DNW. A part of the electrons generated in the PW diffusion region is collected by the N+/PW photodiode [11]. Since the distance traveled by these electrons by diffusion before being collected by the N+ cathode is short (PW depth $\leq 1\,\mu m$) a high bandwidth ($\geq 2\,GHz$) is expected, from relation (3.10), where $D_n = 10\,cm^2/s$ inside the PW.

The structure of the interrupted finger can be mixed with the triple junction PD to obtain a triple well interrupted N-Finger photodiode [11]. The N+/PW/DNW/P-sub photodiode with interrupted N+ finger anode with a total size of $65\,\mu m \times 65\,\mu m$ was implemented in 130 nm CMOS technology. Using Medici, the photodiode achieves a simulated bandwidth of 9.7 GHz while having a responsivity of 2.3 mA/W for 850 nm. Due to the efficient cancellation of the diffusive carriers in this type of the photodiode; the bandwidth is increased on the expense of a very low responsivity.

Furthermore, a triple-junction photodetector can be used for color detection in the wavelength range from 400 to 900 nm. Three vertically stacked photodiodes in the N+/ PW/ DNW/P-sub detector can be adopted to accurately determine the color of the incident light. In addition, the photodetector can be used as wavelength division demultiplexer for three different wavelengths in optical data receivers

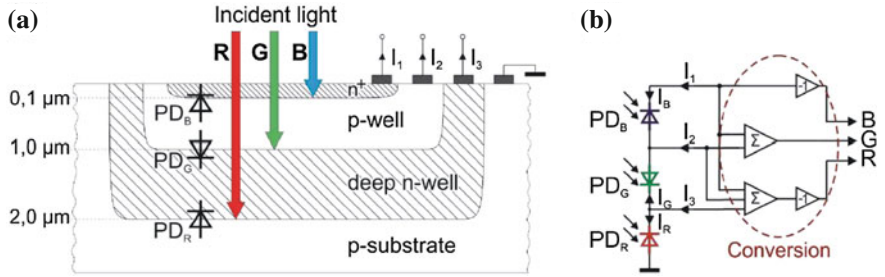

Fig. 5.30 **a** Cross section of triple-junction photodetector, and **b** equivalent circuit of triple Junction photodetector [26]

without needing beam splitters or optical filters [26]. The top junction has its maximum responsivity at a wavelength of around 480 nm (blue) while the middle diode has the highest responsivity at 550 nm (green) and the deep diode at 750 nm (IR), see Fig. 5.30. So, the three photodiodes are corresponding to different wavelengths. The top photodiode corresponds to a wavelength of around 450 nm, the middle junction corresponds to a wavelength of 580 nm and the deep photodiode corresponds to a wavelength of 700 nm. Thus the output currents of the three photodiodes are proportional to the blue, the green and the red part of the incident light. A vertically stacked triple-junction photodetector (N+/PW/DNW/P-sub) was fabricated in standard 90 nm CMOS technology with an area of 80 µm × 80 µm. Measurements showed that the 3 dB-cutoff frequency of the N+/PW top diode was 264 MHz and the middle and deep photodiode achieved data rates of 150 Mb/s [26].

5.8 Avalanche Photodiodes

Avalanche photodiodes (APDs) are operated above the breakdown voltage, i.e. with high electric field strength in the space-charge region. This high electric field accelerates photogenerated charge carriers to such high energies that they can impact-ionize Si atoms in the crystal lattice and generate secondary electron-hole pairs. These secondary electrons and holes are also accelerated in the high electric field and the impact ionization goes on leading to avalanche. In this way, the photocurrent of an APD is strongly amplified by the avalanche effect and the resulting photocurrent is much larger than that of PN PDs, which usually are operated at much lower electric field strengths. In old CMOS technologies, to avoid edge breakdown in P+/NW/P-substrate APDs, a gap in the drawn NW was necessary. This gap reduces the electric field strength at the edge of the P+ anode by reducing the N-type doping at this place [27]. The P+/NW/P-substrate APD shown in Fig. 5.31 was realized in a 0.8 µm CMOS process. An optimal value for the gap width was a value of 0.6 µm, and the breakdown voltage was increased to 19.5 V instead of the edge breakdown voltage

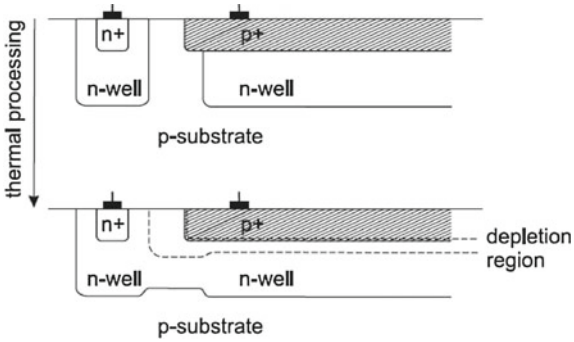

Fig. 5.31 Cross section of an UV avalanche photodiode [27]

of only 13.5 V for a zero gap width. For wider gaps than 1 μm a punchthrough was observed indicating that the N-type regions did not diffuse together and therefore a P-type region was present in the N-well gap. A very low dark current of 400 pA/mm^2 for an avalanche gain of 20 was reported. The unity gain responsivity was 0.18 A/W for 470 nm [10]. With a reverse voltage of 19.1 V, the responsivity increased to 4.6 A/W for 470 nm. The excess noise factor was found to be 7 for a mean gain of 20 at 400 nm. The technique of a gap in the drawn NW of the device breaks the standard design rules for the layout, but it does not modify the process.

In nanometer CMOS technologies from 250 nm and lower, the shallow trench isolation (STI) is used. STI uses dielectrics to fill trenches which are etched in the silicon between active areas to reduce the leakage currents between different active layers (P+ and N+). STI between P+ and N+ regions allows much higher field strengths across junctions before breakdown occures which prevents premature edge breakdown. STI is shallow and cannot prevent premature edge breakdown for the PW created inside DNW because the PW is deeper than the STI [4]. So for the design of an APD an N+ or P+ region should be used as the top APD terminal. Section 5.3.2 shows the structure and the detailed measured results of the P+/NW/P-substrate ADPD fabricated in standard 40 nm CMOS without any process modifications. The P+ source/drain implants with NW and the P-substrate are used to implement the ADPD. The basic structure of the presented ADPD is formed by P+/NW and NW/P-substrate junctions where the avalanche effect occurs at the P+/NW junction. The breakdown voltage of the P+/NW junction is 8.44 V. The STI prevents the premature edge breakdown for the P+/NW junction. The measured responsivities were 0.84 A/W, 0.49 A/W and 2.04 for 850, 660, and 520 nm, respectively. The multiplication factor is 11 for 850 nm, 13 for 660 nm, and 30 for 520 nm, see Fig. 5.16. The multiplication factor for shorter wavelength is higher than for longer wavelengths because the avalanche effect occurs at the P+/NW shallow junction and not at the NW/P-sub deeper junction. The most of the 520 nm light is absorbed in the P+/NW diode where the avalanche multiplication occurs. The measured excess noise factor for the 520 nm light is $F = 12$ at a multiplication gain of 27. The maximum

reached bandwidth is 750 MHz, 1.8, and 1.4 GHz at 8.44 V for 850, 660, and 520 nm, respectively [4].

A P+/NW/P-sub APD was fabricated in 130 nm standard CMOS technology [14]. An optical window having the area of 30 μm × 30 μm was formed by blocking the salicide process. The total device area was 43.8 μm × 43.8 μm. The photodetector was realized without violating any design rule for the CMOS process technology. For characterization of the APD, an 850 nm laser source was used. The avalanche breakdown voltage was about 9.8 V and the peak responsivity was 20.3 A/W. The PD had a −3 dB bandwidth of 2.5 GHz, which is higher than that of the DPD introduced in Sect. 5.3.2; due to the lower doping in 130 nm technology compared to 40 nm CMOS technology.

A triple junction structure (see Fig. 5.29) was exploited in an APD. The APD was designed and fabricated with 65 nm standard CMOS technology without any process modification. The reverse bias voltage was applied to the N+/PW junction, and photocurrents were extracted from the N+ (cathode) contact located inside the PW (anode). STI between P+ and N+ regions prevents premature edge breakdown. An optical window having the area of 30 μm by 30 μm was formed by blocking the salicide process. It achieved the responsivity of 2.94 A/W for 850 nm light and a −3 dB bandwidth of 3.2 GHz at the reverse bias voltage of 10.6 V [5].

5.9 Comparison of Photodiodes

For comparison, a Figure of Merit (FoM) is defined for the photodiodes using their performance parameters such as junction capacitance per area (C, fF μm^2), leakage current (I_r, μA), bandwidth (BW, MHz) and responsivity (R, A/W) as well as the used technology node (L, nm) [11].

$$FoM_1 = \left(\frac{BW(MHz) \cdot R(A/W)}{C\left(\frac{fF}{\mu m^2}\right) \cdot L^2\,(nm) \cdot I(\mu A)} \right) \cdot 1000 \tag{5.1}$$

FoM_1 (5.1) is inversely proportional to the square of the technology feature length; because of the effect of higher doping concentration, which increases the junction capacitance. Also as technology scales down, the number of dielectric layers in the stack on top of the die increase, this increases optical interference and reduces somewhat the quantum efficiency [11]. Table 5.5 shows a comparison between PD structures for 850 nm wavelength. FoM_2 is calculated without the effect of dark current. FoM_3 is calculated without considering the effect of dark current and technology scaling, i.e. without considering L. Disregarding a scientific consideration, for the application FoM_3 may be the most relevant one. By arranging Table 5.5 in a way ordered with respect to FoM_2, the top (best) three PDs are avalanche PDs. Also by using FoM_3 the avalanche PDs still have the first and second best performance. The APDs have high responsivity and high bandwidth on the expense of high dark

Table 5.5 Comparison between different photodiodes structures

	BW(MHz)	C(fF)/μm^2	R(A/W)	L(nm)	I_r(μA)	FoM_1	FoM_2	$FoM_3(M)$
P+/NW/P-sub APD with guard [14]	2500	0.3[a]	20.3	130	3	3336.3^6	10009	169.16
N+/PW/DNW/P-sub APD [5]	3200	0.46[a]	2.94	65	0.3	16134^4	4840.2	20.45
P^+/NW/P-sub ADPD [4]	700	0.63	0.840	40	3	194.4^{10}	583.2	0.933
NW-PW/P-sub [18]	1600	0.138	0.74	180	0.1	2648^7	264.8	8.579
NW/P-sub APD [14]	100	0.138[a]	4.68	130	2	100.3^{11}	200.6	3.391
NW/P-sub [3]	40	0.048	0.34	40	0.05	3541^5	177.05	0.283
N+/NW/P-epi/Psub PIN [19]	1610	0.053	0.135	180	0.00048	263690^1	126.57	4.1
N+/P-sub [3]	40	0.039	0.16	40	0.05	1923^8	102.5	0.164
P+/P-epi/DNW/P-sub PIN [19]	3130	0.163	0.141	180	0.00046	180369^2	82.97	2.7
Lateral P+/P-/N+ PIN [19]	1850	0.171	0.137	180	0.00043	106385^3	45.74	1.48
N+/NW/P-sub SML [24]	700	0.088	0.052	180	–	–	12.76	0.413
PW/DNW/P-sub DPD [4]	9	0.387	0.33	40	0.1	47.9^{12}	4.79	0.0076

[a]There is no value in the reference and this value is from the literature for a similar PD structure

current. The high dark current involved with the APD will increase the PD noise and will reduce the optical receiver sensitivity. By considering the dark current (FoM_1), the best three PDs are PIN-PDs [19] because of high bandwidth and the low dark current.

References

1. H. Zimmermann, *Integrated Silicon Optoelectronics*, 2nd edn. (Springer, Berlin, 2010)
2. C. Hermans, M. Steyaert, *Broadband Opto-Electrical Receivers in Standard CMOS* (Springer, Netherlands, 2007)
3. M. Atef, A. Polzer, H. Zimmermann, High-speed photodiodes in 40 nm standard CMOS technology. Sens. Actuators A Phys. (2013)
4. M. Atef, A. Polzer, H. Zimmermann, Avalanche double photodiode in 40nm standard CMOS technology. J. Quantum Electron. (Submitted) (2013)
5. M.-J. Lee, W.-Y. Choi, A silicon avalanche photodetector fabricated with standard CMOS technology with over 1 THz gain-bandwidth product. Opt. Express **18**(23), 24189–24194 (2010)
6. ETSI TS 105 175-1 V1.1.1(2010-01), Access, Terminals, Transmission and Multiplexing (ATTM); Plastic Optical Fibre System Specifications for 100 Mbit/s and 1 Gbit/s (2010). http://www.etsi.org/WebSite/homepage.aspx
7. M. Atef, R. Swoboda, H. Zimmermann, 1.25 Gbit/s over 50 m step-index plastic optical fiber using a fully integrated optical receiver with an integrated equalizer. J. Lightwave Technol. **30**(1), 118–122 (2012)
8. M. Atef, H. Zimmermann, Optical receiver using noise cancelling with an integrated photodiode in 40 nm cmos technology. IEEE Trans. Circuits Syst. I (TCAS I) (2013)
9. H. Zimmermann, K. Kieschnick, T. Heide, A. Ghazi, Integrated high-speed, high-responsivity Photodiode in CMOS and BiCMOS technology, in *Proceeding of the 29th. European Solid-State Device Research Conference*, (1999), pp. 332–335
10. A, Rochas, A.R. Pauchard, P.-A. Besse, D. Pantic, Z. Prijic,R.S. Popovic, Low-Noise silicon avalanche photodiodes fabricated in conventional CMOS technologies. IEEE Trans. Electron Devices **49**(3), 387–394 (2002)
11. B. Nakhkoob, S. Ray, M. Hella, High speed photodiodes in standard nanometer scale cmos technology: a comparative study. Opt. Express **20**, 11256–11270 (2012)
12. T.K. Woodward, A.V. Krishnamoorthy, 1 Gbit/s CMOS photoreceiver with integrated photodetector operating at 850 nm. IEE Electron. Lett. **34**(12), 1252–1253 (1998)
13. R.J. McIntyre, Multiplication noise in uniform avalanche diodes. IEEE Trans. Electron Devices **13**(1), 164–168 (1966)
14. M.-J. Lee, W.-Y. Choi, Performance comparison of two types of silicon avalanche photodetectors based on N-well/P-substrate and P+/N-well junctions fabricated with standard CMOS technology. J. Opt. Soc. korea **15**(1), 1–3 (2011)
15. H. Zimmermann, H. Dietrich, A. Ghazi, P. Seegebrecht, Fast CMOS-Integrated finger photodiodes for a wide spectral range. ESSDERC 435–438 (2002)
16. T.K. Woodward, A.V. Krishnamoorthy, 1 Gbit/s CMOS photoreceiver with integrated detector operating at 850 nm. Electron. Lett. **34**(12), 1252–1253 (1998)
17. T.K. Woodward, A.V. Krishnamoorthy, 1 Gbit/s integrated optical detectors and receivers in commercial cmos technologies. IEEE J. Sel. Top. Quantum Electron **5**(2), 146–156 (1999)
18. W.-K. Huang, Y.-C. Liu, Y.-M. Hsin, A high-speed and high-responsivity photodiode in standard CMOS technology. IEEE Photonics Technol. Lett. **19**(4), 197–199 (2007)
19. B. Ciftcioglu, L. Zhang, J. Zhang, J.R. Marciante, J. Zuegel, R. Sobolewski, W. Hui, Integrated silicon PIN photodiodes using deep N-Well in a standard 0.18 um CMOS technology. J. Lightwave Technol. **27**(15), 3303–3323 (2009)

20. W.-Z. Chen, S.-H. Huang, A 2.5 Gbps CMOS fully integrated optical receicer with lateral PIN detector, in *IEEE 2007 Custom Intergrated Circuits Conference (CICC)*, (2007), pp. 293–296
21. D. Coppee, W. Pan, R. Vounckx, M. Kuijk, The spatially modulated light detector. Opt. Fiber Conf. OSA Tech. Dig. Ser. pp. 315–316 (1998)
22. J. Genoe, D. Coppee, J.H. Stiens, R.A. Vounckx, M. Kuijk, Calculation of the current response of the spatially modulated light CMOS detector. IEEE Trans. Electron. Dev. **48**(9), 1892–1902 (2001)
23. M. Kuijk, D. Coppee, R. Vounckx, Spatially modulated light detector in CMOS with sense-amplifier receiver operating at 180 Mb/s for optical data link applications and parallel optical interconnects between chips. IEEE J. Sel. Top. Quantum Electron. **4**(6), 1040–1045 (1998)
24. T.-C. Kao, F. Musa, A. Carusone, A 5-gbit/s cmos optical receiver with integrated spatially modulated light detector and equalization. IEEE Trans. Circuits Syst. I: Regul. Pap. **57**, 2844–2857 (2010)
25. T.-C. Kao, F. Musa, A. Carusone, A 5-gbit/s cmos optical receiver with integrated spatially modulated light detector and equalization. IEEE Trans. Circuits Syst. I: Regul. Pap. **57**, 2844–2857 (2010)
26. A. Polzer, K. Schneider-Hornstein, J. Dong, P. Kostov, H. Zimmermann, Investigation of triple-junction photodetector in 90 nm CMOS technology. PROC EUROSENSORS XXV, Procedia Eng. **25**, 864–867 (2011)
27. A. Pauchard, A. Rochas, Z. Randjelovic, P. Besse, R. Popovic, Ultraviolet avalanche photodiode in cmos technology. Int. Electron Devices Meet. (IEDM '00) 709–712 (2000)

Chapter 6
Transimpedance Amplifiers

Current-to-voltage converters are necessary in optical receivers in order to convert and amplify the weak photocurrent delivered by the photodiode into a strong output voltage signal which is proportional to the input current. The transimpedance amplifier (TIA) is the most suitable preamplifier configuration used for optical receivers. For high performance optical receivers TIAs need to have a high gain, high bandwidth, and low input referred noise. An optical communication system like Ethernet Passive Optical Network (10G-EPON) needs a broadband TIA with a high sensitivity. In this chapter the effect of technology scaling and the main characteristics of transimpedance amplifiers will be discussed. Also different traditional TIA topologies, new topologies and their performance will be introduced.

6.1 Transimpedance Gain, Bandwidth, and Noise

The transimpedance gain of the TIA is the ratio of the AC output voltage to the AC input current. The higher the transimpedance gain value, the higher the output voltage is for a given input photocurrent. The TIA's AC output voltage value should be large enough (tens of mV) to drive the post-amplifier. It is important to have a high TIA gain to get s small input referred noise current at the TIA input.

As transimpedance gain increases the TIA's bandwidth will decrease. A limited bandwidth introduces ISI and reduces the eye opening and so the receiver's sensitivity. On the other hand using a TIA with a bandwidth higher than the required data rate will integrate more noise and also decrease the TIA's sensitivity. An optimum value for the TIA bandwidth is equal to 0.7 times the data rate as was suggested by [1].

The total integrated input referred noise current is a calculated value to compare the sensitivity for amplifiers with different gains. First the noise power spectral density at the output is calculated taking account of each component's noise sources. By using the superposition theorem, the total output noise power is the sum of each noise source's output noise power. The input-referred noise current spectral density

M. Atef and H. Zimmermann, *Optoelectronic Circuits in Nanometer CMOS Technology*, Springer Series in Advanced Microelectronics 54,
DOI 10.1007/978-3-319-27338-9_6

is calculated by dividing the total output noise power by the transimpedance gain transfer function:

$$\overline{i_{n,in}^2(f)} = \frac{\overline{v_{n,out}^2(f)}}{|Z_{TIA}(f)|^2} \tag{6.1}$$

It is clear from this equation, that the higher the gain the lower the input referred noise current. The total integrated input referred noise current of the TIA is determined by dividing the root mean square of the output noise voltage by the midband transimpedance gain. The root mean square output noise voltage is calculated by integrating the average squared output noise voltage and then taking the square root [2]:

$$i_{n,in} = \frac{\sqrt{\int \overline{v_{n,out}^2(f)} df}}{Z_{TIA}(0)} \tag{6.2}$$

For an easy analytical calculation, the integration is carried out from zero to infinity. The calculated noise using simulations tools can be easily integrated up to practical bandwidths (not infinity). The practical integrated noise bandwidth is 1.57 times the TIA bandwidth as an upper limit for a first order (one pole) TIA. For second order TIAs with two poles the noise should be integrated up to 1.22 times the TIA bandwidth [2]. As the TIA has more poles; the frequency response role-off become sharper, and hence the noise bandwidth is decreasing up to the TIA bandwidth for a third order TIA.

6.2 Effect of Technology Scaling

The scaling of MOSFETs requires thinning of the gate oxide and reduction of the supply voltage in order to reduce the electric field to prevent the transistors' breakdown. Furthermore, the threshold voltages of the nanometer MOSFETs are reduced at smaller rate compared to the supply voltages which limits the intrinsic gain of the nanometer MOSFET transistors. The decreasing voltage headroom in deep submicron and nanometer CMOS technologies limits the number of transistors that can be stacked. The MOSFETs transconductance to drain current efficiency (g_m/i_d) is increasing slowly with technology scaling and the output conductance (g_{ds}) is increased at higher rate. This is due to the reduction of carrier mobility and the halo implants required reducing short-channel effects and controlling punch-through in nanometer CMOS processes. The MOSFETs' intrinsic voltage gain (g_m/g_{ds}) for different nanometer CMOS technologies is shown in Fig. 6.1 [3]. Although the transit frequency (f_T) is increasing by scaling of the CMOS processes, the intrinsic gain reduction is an important factor. This reduction in the intrinsic MOSFET gain is a problem for analog designers to get high-speed high-gain amplifier. A high overdrive voltage ($V_G - V_{T0}$) is needed to fulfill the high-speed requirements. For high overdrive voltage the intrinsic MOSFET gain decreases dramatically compared to the maximum gain, see Fig. 6.1 [3]. The technology choice gives an upper limit to the

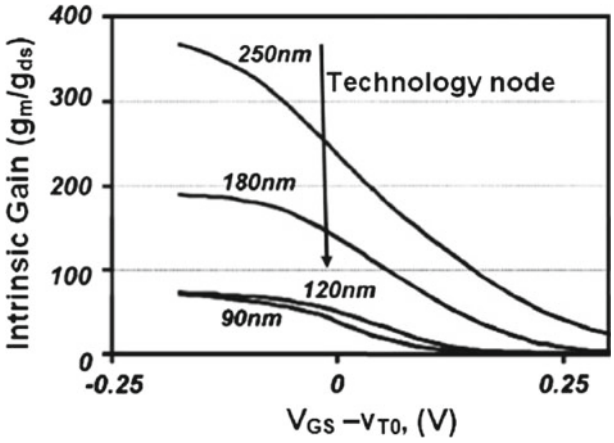

Fig. 6.1 MOSFET intrinsic gain for four nanometer CMOS process generation [3]

required amplifier speed. The amplifier data rate should not be larger than one forth of the transit frequency for $V_{DS} = V_{DD}/2$ over all temperature and process corners [4]. This is close to the commonly accepted criteria that f_T of the technology must be ten times larger than the data rate of the receiver [5].

A multi-stage amplifier can be used to increase the TIA core amplifier voltage gain A_0. By increasing the amplifier voltage gain of the TIA core, the TIA bandwidth and/or sensitivity is increased. However, due to stability requirements the voltage gain of a multi-stage amplifier is limited by the ratio of the bandwidth of the TIA and the f_T of a certain technology. For small-bandwidth TIAs with a bandwidth $\leq 0.05 f_T$, a multi-stage amplifier is the best design choice. It gives more gain without decreasing the stability; as there is enough space in the frequency domain to put the poles apart to ensure the stability. For high-bandwidth TIAs with a bandwidth $\geq 0.1 f_T$, the maximum achievable voltage gain of a single-stage amplifier is larger than the maximum achievable voltage gain of a multi-stage amplifier; as there is not enough spacing in frequency domain to add extra poles coming from the multi-stage approach [6, 7].

6.3 Simplest Preamplifier

The preamplifier is used to convert the incoming photocurrent into an output voltage, which is amplified by the following stages. The simplest way to do this conversion is a resistor between the PD output and the supply voltage as shown in Fig. 6.2. The preamplifier is the most effective stage in determining the sensitivity and bandwidth of an optical receiver. The sensitivity mainly depends on the responsivity of the PD and the input referred noise current of the circuit. Due to the fact that the output current of the PD is very weak; the preamplifier input node is the most sensitive node to noise in an optical receiver. Therefore, the noise of the preamplifier is the

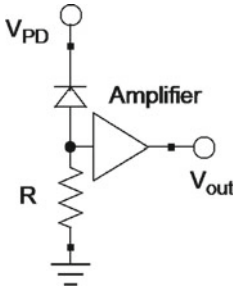

Fig. 6.2 Simplest possibility of the preamplifier

dominating part of the input referred noise current. So, for the simplest preamplifier shown in Fig. 6.2 the noise of the resistor R and the first amplifying stage are the deciding factors.

The second interesting characteristics of the preamplifier is the bandwidth of the circuit. For the simple receiver shown in Fig. 6.2, the bandwidth is indirectly related to the capacitance of the input node and the resistor R:

$$BW = \frac{1}{2\pi C_{in} R} \tag{6.3}$$

The capacitance of the input node consists of the capacitance of the PD, and the input capacitance of the amplifier. For a given PD capacitance, to achieve a high bandwidth therefore the resistance R has to be small. On the other hand, the noise of the circuit shown in Fig. 6.2 depends also on the thermal noise current density of the resistor R:

$$\frac{\overline{i_{n,R}^2}}{\Delta f} = \frac{4K_B T}{R} \tag{6.4}$$

The thermal noise current density of R is increasing by decreasing its resistance value, (6.4). To achieve high bandwidths the resistor R must be small (6.3) and therefore its noise current dominates the sensitivity of the optical receiver. By using a better preamplifier topology, for example a TIA, a better performance can be achieved. Due to its high effect on the optical receiver's bandwidth and noise, the TIA is the most critical building block at the optical receiver side in an optical communication system. In the next sections we will discuss the state of the art of most of the TIA topologies.

6.4 Open Loop TIAs

The TIA can be categorized into two main topologies: the open loop TIAs and the shunt-shunt feedback TIAs. The open loop TIAs like common-gate (CG), regulated cascode (RGC), and inverter common drain feedback (ICDF) TIAs will be introduced

first in this section. The common-gate configuration relaxes the effect of large input parasitic capacitance on the bandwidth. However, the poor device characteristics of nanometer MOSFET cannot totally isolate the parasitic capacitance. The CG-TIA's noise performance is worse than for the CS-TIA with shunt feedback. The CG-TIA's noise performance cannot be optimized without scarfing the power consumption. The regulated cascode (RGC) TIA has a very low input impedance which can support a high bandwidth at low power consumption but the noise performance is still worse than that of the common-source TIA with shunt feedback. The ICDF-TIA shows a higher performance compared to the RGC-TIA. By using the ICDF-TIA an inductorless optical receiver with a high data rate and a high sensitivity can be designed without deteriorating the power consumption.

6.4.1 Common Gate Input Stage

The common-gate TIA's input stage circuity is shown in Fig. 6.3 and a practical implementation follows in Fig. 6.4a. The common-gate input stage's small-signal equivalent circuit is shown in Fig. 6.4b. A common-gate TIA is an open-loop TIA which has low input impedance. For short-channel MOSFETs the channel length

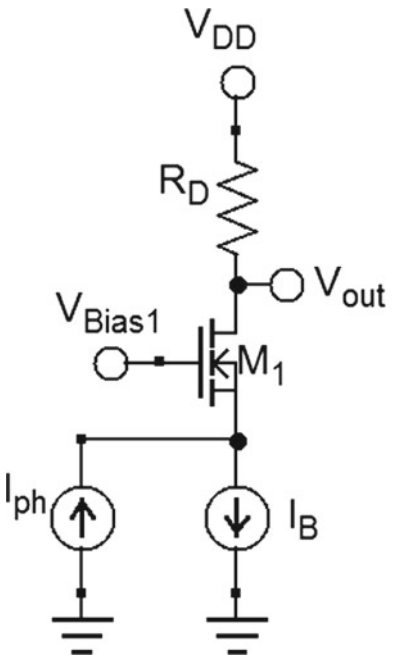

Fig. 6.3 Common-gate transimpedance amplifier circuitry

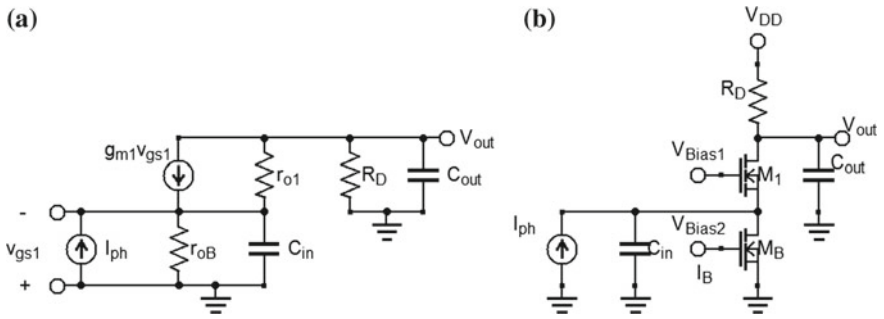

Fig. 6.4 Practical realization for CG-TIA (**a**), and its small-signal model (**b**)

modulation cannot be neglected and the input impedance is given by:

$$R_{in} \approx \left(\frac{r_{o1} + R_D}{1 + g_{m1} r_{o1}} \right) // \, r_{oB} \tag{6.5}$$

For short channels r_{o1} cannot be neglected and R_D affects the input impedance. Increasing the transimpedance gain (R_D) has the effect of increasing the input impedance and so decreasing the TIA's bandwidth. For long-channel MOSFETs the channel length modulation can be neglected (r_o is in the order of tens of kΩ) and the input impedance (6.5) is approximated to:

$$R_{in} \approx \frac{1}{g_{m1}} \tag{6.6}$$

The input impedance is only dependent on the properties of the device, and is independent of the load resistance R_D.

The low-frequency transimpedance gain is given by:

$$A_T\,(0) = R_D \tag{6.7}$$

By considering the input capacitance and the output capacitance, the transimpedance gain is given by:

$$A_T = \frac{R_D}{\left(1 + \left(\frac{1}{g_{m1}}\right) C_{in} S\right) (1 + R_D C_{out} S)} \tag{6.8}$$

As $C_{in}/g_{m1} > C_{out} R_D$ the dominant pole is g_{m1}/C_{in} which determines the TIA's bandwidth:

$$BW = \frac{g_{m1}}{2\pi C_{in}} \tag{6.9}$$

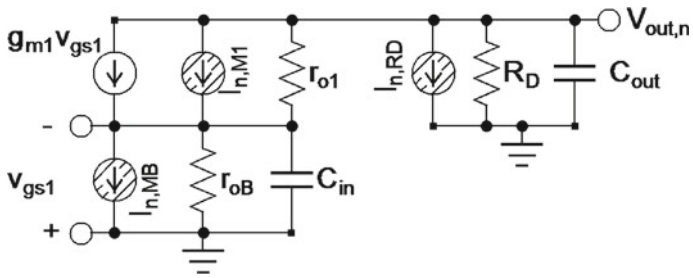

Fig. 6.5 Common gate TIA equivalent circuit noise model

The maximization of g_{m1} by choosing a proper biasing current I_B and a large width of M_1 size can yield a low input impedance and a maximum bandwidth can be achieved.

It is important to study the noise performance of the CG-TIA to see the effect of each element on the CG-TIA's noise behavior. The noise sources of each element in the CG stage were added in Fig. 6.5. In the first calculation the channel length modulation will be neglected for long-channel MOSFETs and for simplicitly.

The contributions of all noise sources are computed and then the output noise contributions are added by applying the superposition theorem to calculate the total output noise.

The output noise density $V_{n,out,MB}$ due to MOSFET M_B noise $I_{n,MB}$ is:

$$V_{n,out,MB} = \frac{R_D}{\left(1 + \frac{1}{g_{m1}} C_{in} S\right)(1 + R_D C_{out} S)} I_{n,MB} \tag{6.10}$$

The output noise density $V_{n,out,M1}$ due to MOSFET M_1 noise $I_{n,M1}$ is:

$$V_{n,out,M1} = \frac{\frac{1}{g_{m1}} R_D \, C_{in} S}{\left(1 + \frac{1}{g_{m1}} C_{in} S\right)(1 + R_D C_{out} S)} I_{n,M1} \tag{6.11}$$

The output noise density $V_{n,out,RD}$ due to the load resistor R_D noise $I_{n,RD}$ is:

$$V_{n,out,RD} = \frac{R_D}{1 + R_D C_{out} S} I_{n,RD} \tag{6.12}$$

The output noise power spectral density is computed by adding (6.10), (6.11), and (6.12) using the superposition theorem.

$$\overline{V^2_{n,out}} = \frac{R^2_D}{\left(1+\left(\frac{2\pi}{g_{m1}}C_{in}f\right)^2\right)\left(1+(2\pi R_D C_{out}f)^2\right)}\overline{I^2_{n,MB}} + \frac{\left(\frac{2\pi}{g_{m1}}R_D\ C_{in}f\right)^2}{\left(1+\left(\frac{2\pi}{g_{m1}}C_{in}f\right)^2\right)\left(\left(1+(2\pi R_D C_{out}f)^2\right)\right)}\overline{I^2_{n,M1}} + \frac{R^2_D}{1+(2\pi R_D C_{out}f)^2}\overline{I^2_{n,RD}} \quad (6.13)$$

The input referred noise current spectral density is calculated by dividing the output noise power spectral density (6.13) by the transimpedance gain $|A_T\ (S)|^2$, where A_T is given by (6.8).

$$\overline{I^2_{n,in}} = \overline{I^2_{n,MB}} + \left(\frac{2\pi C_{in}(2\pi f)}{g_{m1}}\right)^2\overline{I^2_{n,M1}} + \left(1+\left(\frac{2\pi}{g_{m1}}C_{in}f\right)^2\right)\overline{I^2_{n,RD}} \quad (6.14)$$

$$\overline{I^2_{n,in}} = 4KT\gamma g_{mB} + 4KT\gamma\frac{C^2_{in}}{g_{m1}}(2\pi f)^2 + \left(1+\left(\frac{2\pi}{g_{m1}}C_{in}f\right)^2\right)\frac{4KT}{R_D} \quad (6.15)$$

It is clear from the squared input referred noise current density given by (6.15) that the noise contributed by M_1 increases with C_{in} and frequency; because as C_{in} increases, a larger part of M_1 noise flows from the output node to the input instead of circulating through M_1. The noise contribution of R_D to the input noise current also increases with C_{in}f. This is because the transimpedance gain falls as the frequency reaches the bandwidth [1].

To calculate the total input referred noise current, the total noise should be calculated at the output and then divided by the low frequency transimpedance gain $A_T(0) = R_D$ (6.2). The total output noise is calculated by integrating the output noise power spectral density (6.13) from f = 0 to infinity.

By using partial fraction, the two poles noise equations can be spliced into the sum of one pole equation by using the relations (6.16) and (6.17) [1].

$$\frac{1}{\left(a^2x^2+b^2\right)\left(c^2x^2+d^2\right)} = \frac{1}{a^2d^2-b^2c^2}\left(\frac{a^2}{a^2x^2+b^2}+\frac{-c^2}{c^2x^2+d^2}\right) \quad (6.16)$$

and

$$\frac{x^2}{\left(a^2x^2+b^2\right)\left(c^2x^2+d^2\right)} = \frac{1}{a^2d^2-b^2c^2}\left(\frac{-b^2}{a^2x^2+b^2}+\frac{d^2}{c^2x^2+d^2}\right) \quad (6.17)$$

Then the second order functions can be integrated using relation (6.18).

$$\int\frac{1}{a^2+x^2}\,dx = \frac{1}{a}tan^{-1}\left(\frac{x}{a}\right) \quad (6.18)$$

$$\overline{V^2_{n,out,B}} = \frac{R^2{}_D}{\left(1+\left(\frac{1}{g_{m1}}C_{in}\omega\right)^2\right)\left(1+(R_DC_{out}\omega)^2\right)}\,\overline{I^2_{n,MB}} \tag{6.19}$$

After decomposition using partial fraction:

$$\overline{V^2_{n,out,B}} = \left(\frac{\left(\frac{2\pi}{g_{m1}}C_{in}\right)^2}{\left(1+\left(\frac{2\pi}{g_{m1}}C_{in}f\right)^2\right)} + \frac{-(2\pi R_DC_{out})^2}{\left(1+(2\pi R_DC_{out}f)^2\right)}\right) \frac{R^2{}_D\,\overline{I^2_{n,MB}}}{\left(\frac{2\pi}{g_{m1}}C_{in}\right)^2-(2\pi R_DC_{out})^2} \tag{6.20}$$

After integration:

$$\begin{aligned}\overline{V^2_{n,out,B,total}} &= \frac{\pi}{2}\left(\frac{2\pi C_{in}}{g_{m1}}-2\pi R_DC_{out}\right)\frac{4KT\gamma g_{mB}R^2{}_D}{\left(\frac{2\pi}{g_{m1}}C_{in}\right)^2-(2\pi R_DC_{out})^2}\\ &= \frac{KT\gamma g_{mB}R^2{}_D}{\left(\frac{1}{g_{m1}}C_{in}+R_DC_{out}\right)}\end{aligned} \tag{6.21}$$

$$\overline{V^2_{n,out,M1}} = \frac{\left(\frac{2\pi}{g_{m1}}R_{D}\,C_{in}\right)^2 f}{\left(1+\left(\frac{2\pi}{g_{m1}}C_{in}f\right)^2\right)\left(\left(1+(2\pi R_DC_{out}f)^2\right)\right)}\,\overline{I^2_{n,M1}} \tag{6.22}$$

After decomposition using partial fraction:

$$\overline{V^2_{n,out,M1}} = \left(\frac{-1}{1+\left(\frac{2\pi}{g_{m1}}C_{in}f\right)^2} + \frac{1}{\left(1+(2\pi R_DC_{out}f)^2\right)}\right)\frac{\left(\frac{2\pi}{g_{m1}}R_{D}\,C_{in}\right)^2\overline{I^2_{n,M1}}}{\left(\frac{2\pi}{g_{m1}}C_{in}\right)^2-(2\pi R_DC_{out})^2} \tag{6.23}$$

After integration:

$$\begin{aligned}\overline{V^2_{n,out,M1,total}} &= \frac{\pi}{2}\left(\frac{g_{m1}}{2\pi C_{in}}-\frac{1}{2\pi R_DC_{out}}\right)\frac{\left(\frac{2\pi}{g_{m1}}R_{D}\,C_{in}\right)^2 4KT\gamma g_{m1}}{\left(\frac{2\pi}{g_{m1}}C_{in}\right)^2-(2\pi R_DC_{out})^2}\\ &= \frac{R_D\,C_{in}KT\gamma}{C_{out}\left(\frac{1}{g_{m1}}C_{in}+R_DC_{out}\right)}\end{aligned} \tag{6.24}$$

$$\overline{V^2{}_{n,out,RD}} = \frac{R^2{}_D}{1 + (2\pi R_D C_{out} f)^2} \overline{I^2{}_{n,RD}} \tag{6.25}$$

After integration:

$$\overline{V^2{}_{n,out,RD,total}} = \frac{\pi}{2} \frac{R^2{}_D}{2\pi R_D C_{out}} \frac{4KT}{R_D} = \frac{KT}{C_{out}} \tag{6.26}$$

After summing all the noise components (6.21), (6.24), and (6.26) to get the total output noise power, the total output noise is then divided by $A_T^2(0) = R_D^2$ to get the total squared input referred noise current:

$$\overline{I^2{}_{n,in,total}} = \frac{KT\gamma g_{mB}}{\left(\frac{1}{g_{m1}} C_{in} + R_D C_{out}\right)} + \frac{C_{in} KT\gamma}{C_{out} R_D \left(\frac{1}{g_{m1}} C_{in} + R_D C_{out}\right)} + \frac{KT}{R^2{}_D\, C_{out}} \tag{6.27}$$

$$\overline{I^2{}_{n,in,total}} \approx \frac{KT\gamma g_{mB} g_{m1}}{C_{in}} + \frac{g_{m1}\, KT\gamma}{C_{out} R_D} + \frac{KT}{R^2{}_D\, C_{out}} \tag{6.28}$$

The total input referred noise current is calculated by taking the square root of (6.28). The noise contributions from R_D and M_B trade-off with each other. The noise current produced by the load resistor R_D and by the bias transistor M_B are directly referred to the input. The load resistor R_D can be increased in order to reduce its noise contribution. As R_D increases the biasing current will be reduced which would increase the noise contribution of M_B. Improving the noise performance of the CG-TIA will come at the price of increased biasing current and power supply voltage.

For short-channel MOSFETs the channel length modulation cannot be neglected (r_o is only in the order of hundreds of ohms) and the input impedance is given by (6.5). Due to including the MOSFET's output resistance the effect of R_D on the input impedance of the common-gate TIA cannot be neglected. R_D should be maximized to reduce the noise (6.28). On the other hand increasing R_D has the effect of increasing the input impedance (6.5) and so decreasing the TIA's bandwidth.

The low frequency transimpedance gain is given by:

$$A_T(0) = g_{m1}\left[\left(\frac{r_{o1+} R_D}{1 + g_{m1} r_{o1}}\right) // r_{oB}\right] R_D < R_D \tag{6.29}$$

The effect of including the MOSFET's output resistance is the decreasing of the low frequency transimpedance gain of the common-gate TIA compared to R_D.

By considering the input capacitance and the output capacitance, the transimpedance gain is given by:

$$A_T = \frac{g_{m1}\left[\left(\frac{r_{o1+} R_D}{1+g_{m1} r_{o1}}\right) // r_{oB}\right] R_D}{\left(1 + \left[\left(\frac{r_{o1+} R_D}{1+g_{m1} r_{o1}}\right) // r_{oB}\right] C_{in} S\right)(1 + R_D C_{out} S)} \tag{6.30}$$

The bandwidth is given by:

$$BW = \frac{1}{\left[\left(\frac{r_{o1}+R_D}{1+g_{m1}r_{o1}}\right)//r_{oB}\right]C_{in}} \tag{6.31}$$

The effect of including the MOSFET's output resistance is the decreasing of the bandwidth compared to the simple (6.9).

The noise analysis is repeated again here after considering the effect of the channel length modulation.

$$\overline{V^2_{n,out,M1}} = \frac{\left(2\pi\left(\frac{r_{o1}+R_D}{1+g_{m1}r_{o1}}\right)R_D C_{in} f\right)^2}{\left(1+\left(2\pi\left(\frac{r_{o1}+R_D}{1+g_{m1}r_{o1}}\right)C_{in} f\right)^2\right)\left(1+(2\pi R_D C_{out} f)^2\right)}\overline{I^2_{n,M1}} \tag{6.32}$$

$$\overline{V^2_{n,out,RD}} = \frac{R_D{}^2}{1+(2\pi R_D C_{out} f)^2}\overline{I^2_{n,RD}} \tag{6.33}$$

$$\overline{V^2_{n,out,B}} = \frac{\left(g_{m1}\left(\frac{r_{o1}+R_D}{1+g_{m1}r_{o1}}\right)R_D\right)^2}{\left(1+\left(2\pi\left(\frac{r_{o1}+R_D}{1+g_{m1}r_{o1}}\right)C_{in} f\right)^2\right)\left(1+(2\pi R_D C_{out} f)^2\right)}\overline{I^2_{n,MB}} \tag{6.34}$$

$$\overline{I^2_{n,in}} = \overline{I^2_{n,MB}} + \frac{1+\left(2\pi\left(\frac{r_{o1}+R_D}{1+g_{m1}r_{o1}}\right)C_{in} f\right)^2}{\left(g_{m1}\left(\frac{r_{o1}+R_D}{1+g_{m1}r_{o1}}\right)\right)^2}\overline{I^2_{n,RD}} + \frac{(2\pi C_{in} f)^2}{g^2{}_{m1}}\overline{I^2_{n,M1}} \tag{6.35}$$

$$\overline{I^2_{n,in}} = 4KT\gamma g_{mB} + \frac{1+\left(2\pi\left(\frac{r_{o1}+R_D}{1+g_{m1}r_{o1}}\right)C_{in} f\right)^2}{\left(g_{m1}\left(\frac{r_{o1}+R_D}{1+g_{m1}r_{o1}}\right)\right)^2}\frac{4KT}{R_D} + \frac{(2\pi C_{in} f)^2}{g_{m1}}4KT\gamma \tag{6.36}$$

The effect of the MOSFET's limited output resistance on the input noise current density is clear in the second term (R_D noise: compare (6.36), (6.15)). As R_D has to be decreased to get the required bandwidth, the R_D thermal noise increases.

By using partial fractions of (6.32), (6.33), and (6.34) and then the integration using (6.18), the integrated output noise of each noise source can be given by:

$$\overline{V^2_{n,out,B}} = \frac{KT\gamma g_{mB} R^2{}_D}{\left(\left(\frac{r_{o1}+R_D}{1+g_{m1}r_{o1}}\right)C_{in} + R_D C_{out}\right)} \tag{6.37}$$

$$\overline{V^2_{n,out,M1}} = \frac{\left(\frac{r_{o1}+R_D}{1+g_{m1}r_{o1}}\right) R_D\, C_{in} KT\gamma g_{m1}}{C_{out}\left(\left(\frac{r_{o1}+R_D}{1+g_{m1}r_{o1}}\right) C_{in} + R_D C_{out}\right)} \tag{6.38}$$

$$\overline{V^2_{n,out,RD}} = \frac{\pi}{2}\frac{R^2{}_D}{2\pi R_D C_{out}}\frac{4KT}{R_D} = \frac{KT}{C_{out}} \tag{6.39}$$

By summing all the noise components (6.37), (6.38), and (6.39) to get the total output noise power and dividing it by:

$$A_T^2(0) = \left(g_{m1}\left[\left(\frac{r_{o1}+R_D}{1+g_{m1}r_{o1}}\right)//r_{oB}\right]R_D\right)^2$$

to get the total squared input referred noise current:

$$\begin{aligned}\overline{I^2_{n,in,total}} = {} & \frac{KT\gamma g_{mB}}{\left(\left(\left(\frac{r_{o1}+R_D}{1+g_{m1}r_{o1}}\right)//r_{oB}\right)C_{in} + R_D C_{out}\right)\left(g_{m1}\left(\left(\frac{r_{o1}+R_D}{1+g_{m1}r_{o1}}\right)//r_{oB}\right)\right)^2} \\ & + \frac{C_{in}KT\gamma}{C_{out}g_{m1}\left[\left(\frac{r_{o1}+R_D}{1+g_{m1}r_{o1}}\right)//r_{oB}\right]R_D\left(\left(\left(\frac{r_{o1}+R_D}{1+g_{m1}r_{o1}}\right)//r_{oB}\right)C_{in} + R_D C_{out}\right)} \\ & + \frac{KT}{\left(g_{m1}\left[\left(\frac{r_{o1}+R_D}{1+g_{m1}r_{o1}}\right)//r_{oB}\right]R_D\right)^2 C_{out}}\end{aligned} \tag{6.40}$$

$$\begin{aligned}\overline{I^2_{n,in,total}} \approx {} & \frac{KT\gamma g_{mB}}{g^2{}_{m1}\left(\left(\frac{r_{o1}+R_D}{1+g_{m1}r_{o1}}\right)//r_{oB}\right)^3 C_{in}} \\ & + \frac{KT\gamma}{C_{out}R_D g_{m1}\left[\left(\frac{r_{o1}+R_D}{1+g_{m1}r_{o1}}\right)//r_{oB}\right]^2} \\ & + \frac{KT}{\left(g_{m1}\left[\left(\frac{r_{o1}+R_D}{1+g_{m1}r_{o1}}\right)//r_{oB}\right]R_D\right)^2 C_{out}}\end{aligned} \tag{6.41}$$

The total input referred noise current is calculated by taking the square root of (6.41).

The effect of the MOSFET's limited output resistance on the total input referred noise current is that the noise current becomes a strong function of R_D, compared to (6.28). As R_D has to be decreased to get the required bandwidth, the total input noise current increases.

6.4.2 Regulated-Cascode TIA

Figure 6.6 shows the circuitry of a regulated-cascode TIA (RGC-TIA). The photocurrent is converted to a voltage at the drain of M_1. The stage consisting of $M_2$and R_2 operates as a local feedback and thus reduces the input impedance by the amount of its own voltage gain. It will be shown that the RGC-TIA has a very low input impedance which can support a high bandwidth at low power consumption [8]. The circuit structure using the RGC configuration can isolate the capacitance effect.

Figure 6.7 depicts the small-signal equivalent circuit of the RGC-TIA taking into consideration the effect of channel length modulation, the gate-drain, gate-source and load capacitances.

Where $C_{in} = C_{PD} + C_{ESD} + C_{PAD}$, $C_T = C_{in} + C_{gs2}$, $C_1 = C_{gd2} + C_{gs1}$, and $\grave{R}_2 = R_2 // r_{ds2}$

The transimpedance gain of RGC-TIA is given by:

$$A_T = \frac{g_{m1}\left(g_{m2} + \frac{1}{\grave{R}_2}\right)}{\left(\frac{1}{R_1} + C_{out}S\right)\left[\left(g_{m2} + \frac{1}{\grave{R}_2}\right)(g_{m1} + C_1 S) + \left(\frac{1}{\grave{R}_2} + C_1 S\right)\left(\frac{1}{r_{o,C}} + C_T S\right)\right]} \tag{6.42}$$

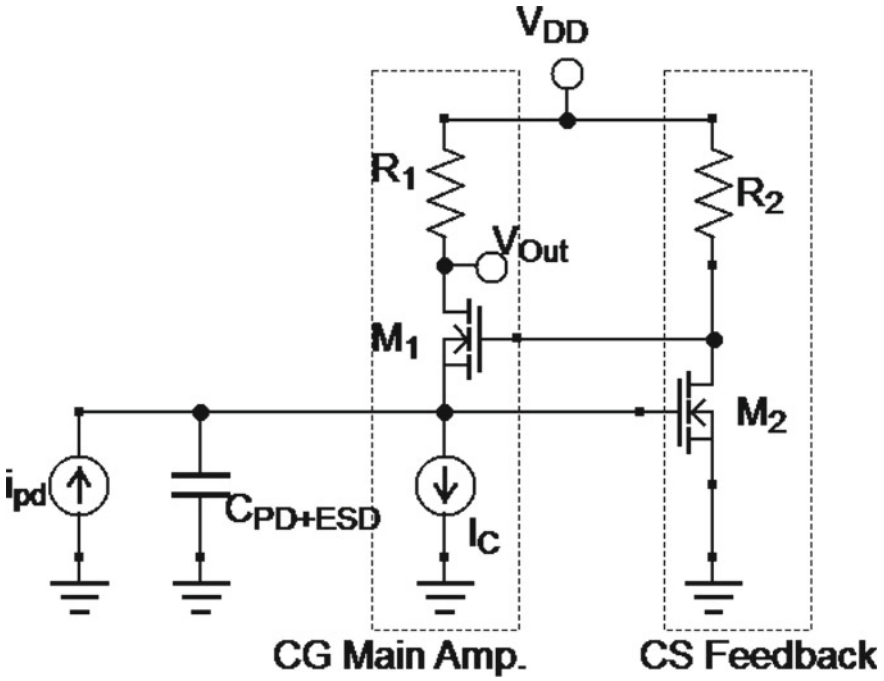

Fig. 6.6 Regulated cascode transimpedance amplifier circuitry

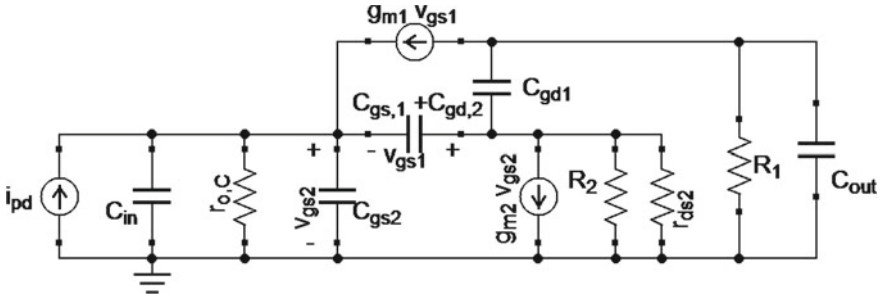

Fig. 6.7 Regulated cascode transimpedance amplifier small-signal equivalent circuit

$$A_T = \frac{\frac{R_1 r_{o,C} g_{m1}\left(g_{m2}\grave{R}_2+1\right)}{r_{o,C} g_{m1}\left(g_{m2}\grave{R}_2+1\right)+1}}{(1+R_1 C_{out} S)\left[1+\left(\frac{r_{o,C}\left(g_{m2}\grave{R}_2+1\right)C_1+r_{o,C}C_T+\grave{R}_2 C_1}{r_{o,C} g_{m1}\left(g_{m2}\grave{R}_2+1\right)+1}\right)S+\left(\frac{r_{o,C}\grave{R}_2 C_1 C_T}{r_{o,C} g_{m1}\left(g_{m2}\grave{R}_2+1\right)+1}\right)S^2\right]} \tag{6.43}$$

By assuming $r_{o,C}\left(g_{m2}\grave{R}_2+1\right)C_1 > r_{o,C}C_T$ and $> \grave{R}_2 C_1$, and $1/g_{m2} \approx R_2$, the transimpedance gain can be approximately given by:

$$A_T \approx \frac{R_1}{(1+R_1 C_{out} S)\left(1+\left(\frac{C_1}{g_{m1}}\right)S\right)\left(1+\frac{C_T}{g_{m1}\left(g_{m2}\grave{R}_2+1\right)}S\right)} \tag{6.44}$$

The low-frequency gain is given by:

$$A_T(0) \approx R_1 \tag{6.45}$$

$$\omega_{P1} = \frac{g_{m1}\left(g_{m2}\grave{R}_2+1\right)}{C_T}, \omega_{P2} = \frac{1}{R_1 C_{out}},\ \omega_{P3} = \frac{g_{m1}}{C_1} \tag{6.46}$$

As $C_T \gg C_1$ and C_L then ω_{P1} is the dominant pole and the RGC bandwidth is:

$$\text{BW} = \frac{g_{m1}\left(g_{m2}\grave{R}_2+1\right)}{2\pi C_T} = \frac{g_{m1}(A_{cs}+1)}{2\pi C_T} \tag{6.47}$$

where $(1+A_{cs})$ is the gain of the local feedback loop of the CS stage (M_2, R_2) and the CS voltage gain is:

$$A_{cs} = g_{m2}.(R_2//r_{ds2}) = g_{m2}\grave{R}_2 \tag{6.48}$$

Compared with the CG-TIA's bandwidth (6.9), the RGC bandwidth is increased by the factor $(1 + A_{cs})$ for the same g_{m1}.

According to the small-signal analysis, the input resistance of the RGC circuit is given by:

$$\mathrm{Z}_{in}\,(0) \approx \frac{1}{g_{m1}\left(g_{m2}\grave{R}_2 + 1\right)} = \frac{1}{g_{m1}(A_{cs} + 1)} \tag{6.49}$$

It is clearly seen that the input impedance is $1 + g_{m2}\grave{R}_2$ times smaller than that of a CG input TIA. That is, the RGC input behaves qualitatively as a CG TIA with a large transconductance of $g_{m1}(1 + g_{m2}\grave{R}_2)$.

The main advantage of the RGC-TIA is its low input impedance which can isolate the effect of a larger photodiode capacitance. If it is assumed again that the dominant pole is located at the input. Compared to the bandwidth of the common-gate (or common-base) stages, the RGC bandwidth can be increased for the same transconductance of M_1. For the same bandwidth a CG-stage needs a higher g_{m1} than the RGC-stage which means a larger current is needed through M_1. The power consumption of the RGC-TIA is lower than that of the CG-TIA.

The RGC transimpedance gain is limited by R_1 which cannot be increased due to limited low supply voltage. The input impedance can be decreased by increasing the CS voltage gain A_{cs}. But this gain is limited also by the low supply voltage. Increasing the current through M_2 will increase g_{m2} but r_{ds2} will decrease. The maximum current is also limited by:

$$(V_{DD} - V_{GS1} - V_{GS2})/R_2 \approx (V_{DD} - 2.V_{TH})/R_2 \tag{6.50}$$

Increasing the current will be on the expense of a smaller R_2 value. As a result the A_{cs} value will be limited. The conventional RGC-TIA is difficult to be implemented at a low supply voltage. A modification in the RGC-TIA was introduced in [9, 10] for low voltage operation.

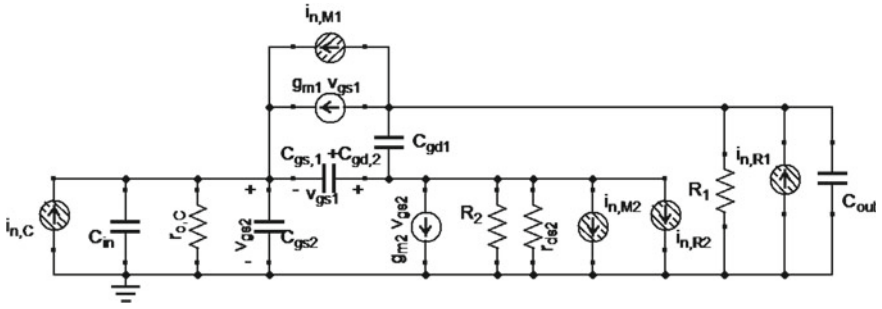

Fig. 6.8 Regulated cascode transimpedance amplifier small-signal equivalent noise

It is important to study the noise performance of the RCG TIA to see the effect of each element on the noise behavior. The noise sources of each element in the RGC stage were added in Fig. 6.8. The output noise density $V_{n,out,M1}$ due to the noise $I_{n,M1}$ of MOSFET M_1 is [8]:

$$v_{n,out,M1} = \frac{i_{n,M1}\left[\left(g_{m2}+\frac{1}{\grave{R}_2}\right)C_1S+\left(\frac{1}{\grave{R}_2}+C_1S\right)\left(\frac{1}{r_{o,C}}+C_TS\right)\right]}{\left(\frac{1}{R_1}+C_{out}S\right)\left[\left(g_{m2}+\frac{1}{\grave{R}_2}\right)(g_{m1}+C_1S)+\left(\frac{1}{\grave{R}_2}+C_1S\right)\left(\frac{1}{r_{o,C}}+C_TS\right)\right]} \tag{6.51}$$

$$\approx \frac{i_{n,M1}\left(g_{m2}+\frac{1}{\grave{R}_2}\right)C_1S}{\left(\frac{1}{R_1}+C_{out}S\right)\left[\left(g_{m2}+\frac{1}{\grave{R}_2}\right)(g_{m1}+C_1S)+\left(\frac{1}{\grave{R}_2}+C_1S\right)\left(\frac{1}{r_{o,C}}+C_TS\right)\right]} \tag{6.52}$$

The output noise density $V_{n,out,C}$ due to noise $I_{n,C}$ from the current source I_C is:

$$v_{n,out,C} = \frac{i_{n,C}g_{m1}\left(g_{m2}+\frac{1}{\grave{R}_2}\right)}{\left(\frac{1}{R_1}+C_{out}S\right)\left[\left(g_{m2}+\frac{1}{\grave{R}_2}\right)(g_{m1}+C_1S)+\left(\frac{1}{\grave{R}_2}+C_1S\right)\left(\frac{1}{r_{o,C}}+C_TS\right)\right]} \tag{6.53}$$

The output noise density $V_{n,out,R1}$ due to R_1 noise $I_{n,R1}$ is:

$$v_{n,out,R1} = \frac{i_{n,R1}}{\left(\frac{1}{R_1}+C_{out}S\right)} \tag{6.54}$$

The output noise density $V_{n,out,R2}$ due to R_2 noise $I_{n,\grave{R}2}$ is:

$$v_{n,out,R2} = \frac{i_{n,\grave{R}2}g_{m1}\left(\frac{1}{r_{o,C}}+C_TS\right)}{\left(\frac{1}{R_1}+C_{out}S\right)\left[\left(g_{m2}+\frac{1}{\grave{R}_2}\right)(g_{m1}+C_1S)+\left(\frac{1}{\grave{R}_2}+C_1S\right)\left(\frac{1}{r_{o,C}}+C_TS\right)\right]} \tag{6.55}$$

The output noise density $V_{n,out,M2}$ due to MOSFET M_2 noise $I_{n,M2}$ is:

$$v_{n,out,M2} = \frac{i_{n,M2}g_{m1}\left(\frac{1}{r_{o,C}}+C_TS\right)}{\left(\frac{1}{R_1}+C_{out}S\right)\left[\left(g_{m2}+\frac{1}{\grave{R}_2}\right)(g_{m1}+C_1S)+\left(\frac{1}{\grave{R}_2}+C_1S\right)\left(\frac{1}{r_{o,C}}+C_TS\right)\right]} \tag{6.56}$$

The output noise density is computed by adding (6.53), (6.51), (6.54), (6.55), and (6.56) using the superposition theorem. The input referred noise current density $\overline{I^2_{in}}$ is calculated by dividing the output noise density by the squared magnitude of the transimpedance gain $|A_T(S)|^2$, where A_T is given by (6.42):

$$\overline{I^2_{n,in}} \approx \overline{I^2_{n,C}} + \frac{C^2{}_1\omega^2}{g_{m1}{}^2}\overline{I^2_{n,M1}} + \frac{g_{m1}{}^2 + C^2{}_1\omega^2}{g_{m1}{}^2}\overline{I^2_{n,R1}} + \frac{\left(\frac{1}{r^2_{o,C}} + C^2{}_T\omega^2\right)}{\left(g_{m2} + \frac{1}{\grave{R}_2}\right)^2}\overline{I^2_{n,M2}} + \frac{\left(\frac{1}{r^2_{o,C}} + C^2{}_T\omega^2\right)}{\left(g_{m2} + \frac{1}{\grave{R}_2}\right)^2}\overline{I^2_{n,\grave{R}2}} \quad (6.57)$$

$$\overline{I^2_{n,in}} \approx 4KT\gamma g_{m,C}(1+M) + \frac{4KT\gamma C^2{}_1}{g_{m1}}\omega^2 + \frac{4KT\left(g_{m1}{}^2 + C^2{}_1\omega^2\right)}{R_1 g_{m1}{}^2} + \frac{\left(\frac{1}{r^2_{o,C}} + C^2{}_T\omega^2\right)}{\left(g_{m2} + \frac{1}{\grave{R}_2}\right)^2}4KT\left(\gamma g_{m2} + \frac{1}{\grave{R}_2}\right) \quad (6.58)$$

where the noise density of the current source is given by: $4KT\gamma g_{m,C}(1+M)$, γ is the noise factor of the MOSFET, and M is the current mirror ratio.

The squared input referred noise current density equation (6.58) can be rearranged to be:

$$\overline{I^2_{n,in}} \approx 4KT\gamma g_{m,C}(1+M) + \frac{4KT}{R_1} + \frac{4KT\left(\gamma g_{m2} + \frac{1}{\grave{R}_2}\right)}{r^2{}_{o,C}\left(g_{m2} + \frac{1}{\grave{R}_2}\right)^2} + \left(\frac{4KT\gamma C^2{}_1}{g_{m1}} + \frac{4KTC^2{}_1}{R_1 g_{m1}{}^2} + \frac{4KT\left(\gamma g_{m2} + \frac{1}{\grave{R}_2}\right)C^2{}_T}{\left(g_{m2} + \frac{1}{\grave{R}_2}\right)^2}\right)\omega^2 \quad (6.59)$$

It is seen from (6.59) that the thermal noise contributions from M_c and R_1 are dominant at low frequencies and the dominant high-frequency noise comes due to the large input capacitance C_{in}. The dominant high-frequency noise is divided by $\left(g_{m2} + \frac{1}{\grave{R}_2}\right)$, of the local feedback rather than the input transconductance g_{m1}, which is the case of the CG TIA (6.36). In the RGC TIA, g_{m1} is related to C_{gs1} (first and second terms of the high- frequency component) that is much smaller than the input capacitance C_{in}. Thus, the noise contribution from M_1 can be significantly reduced for the same value of g_{m1} as in the CG TIA. The denominator $\left(g_{m2} + \frac{1}{\grave{R}_2}\right)$ in the -third term of the high-frequency components is controllable by sizing M_2 and R_2 appropriately, along with large enough gate-source voltage of M_2. Therefore, the dominant high-frequency noise contribution due to the large input capacitance C_{in} can be efficiently reduced without deteriorating the TIA's bandwidth. For that the value of g_{m2} should be designed to be larger than that of g_{m1} [8].

Because the regulated-cascode TIA contains a local feedback loop (CS amplifier R_2 and M_2), stability needs to be guaranteed. This can be done by ensuring that the open-loop gain is smaller than one when the phase open loop gain has shifted by 180°. Because there are three poles the loop gain, and the maximum phase shift can reach 270°, the instability of the RGC-TIA is possible. The RGC-TIA needs a

carefully design to ensure the stable operation while keeping the required bandwidth, gain, and noise.

6.4.3 Inverter Based Common-Drain Feedback TIA

The common-gate configuration relaxes the effect of large input parasitic capacitance on the bandwidth [11]. However, the poor device characteristics of nano-scale MOSFETs cannot totally isolate the parasitic capacitance. By modifying the conventional CG-TIA to a regulated cascode (RGC) [8] or with an active feedback TIA [12] which have lower input impedance, a higher bandwidth can be obtained. The RGC-TIA's noise performance is worse than that of the common-source configuration [13]. The circuit structure using the RGC configuration can isolate the effect of the capacitance but it is difficult to be implemented at a low supply voltage. A modification in the RGC-TIA was introduced in [9, 10] for low voltage operation. In this section a new TIA topology employs an inverter with an active common-drain feedback (ICDF-TIA) is introduced. The ICDF-TIA achieves a high bandwidth like the conventional RGC-TIA with a better sensitivity and a lower power consumption.

A TIA employs an inverter with an active common-drain feedback (ICDF-TIA) to achieve a higher bandwidth at lower power consumption than with the normal CS-TIA and a better sensitivity than with the conventional RGC-TIA introduced in [14]. Figure 6.9 shows the circuitry of the ICDF-TIA. The inverter stage with M_{n2} and M_{p2} has a voltage gain A_{inv}:

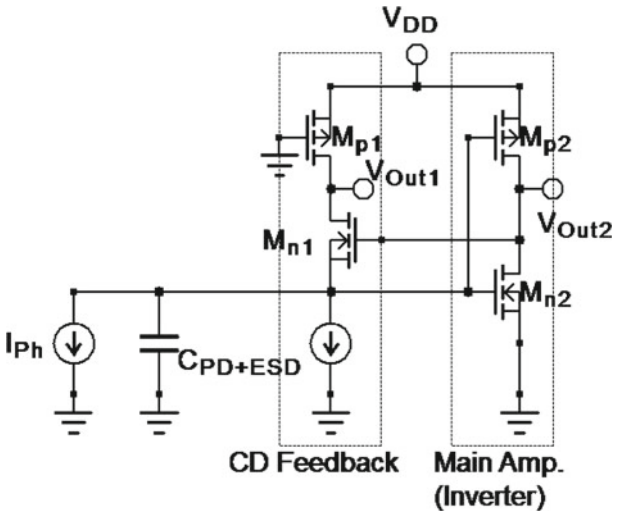

Fig. 6.9 Inverter based common-drain feedback [14]

$$A_{inv} = (g_{m,n2} + g_{m,p2}) \times (r_{ds,n2} // r_{ds,p2}) \tag{6.60}$$

$$= g_{m2,eff} \times (r_{ds,n2} // r_{ds,p2}) \tag{6.61}$$

where

$$g_{m2,eff} = g_{m,n2} + g_{m,p2} \tag{6.62}$$

The input impedance becomes:

$$Z_{in}(0) = \frac{r_{ds,p1} + r_{ds,n1}}{1 + g_{m,n1} \cdot r_{ds,n1}(1 + (g_{m,n2} + g_{m,p2}) r_{ds,n2} // r_{ds,p2})}$$

$$= \frac{r_{ds,p1} + r_{ds,n1}}{1 + g_{m,n1} \cdot r_{ds,n1}(1 + A_{inv})} \tag{6.63}$$

The input impedance of the proposed ICDF-TIA can be decreased by increasing the sum of $g_{m,n2}$ and $g_{m,p2}$. Due to the small power supply value, the width of Mp1 should be large to obtain a small saturation voltage. Therefore, $r_{ds,p1}$ will be small. This leads to the circumstance that $r_{ds,n1}$ will be larger than $r_{ds,p1}$ and the exact (6.63) can be simplified to:

$$Z_{in}(0) \approx \frac{1}{g_{m,n1}[1 + A_{inv}]} \tag{6.64}$$

The ICDF-TIA's input impedance can be lower than the RGC-TIA's input impedance due to the increase of the effective g_m by adding the inverter stage (M_{n2} and M_{p2}) which in general has a higher gain than the common-source amplifier stage (M_{n2} and R_2) of the RGC-TIA. The smaller input impedance of the ICDF-TIA has the effect of increasing the TIA bandwidth at the same current consumption as for the RGC-TIA. By using the ICDF-TIA the same bandwidth can be obtained at a lower current consumption.

The main difference between the RGC and the ICDF TIAs is the output node and which stage is the main amplifier or the feedback. The RGC output Fig. 6.6 is taken from the drain of the CG amplifier (M_{n1}, R1) V_{out1}. The common-source amplifier

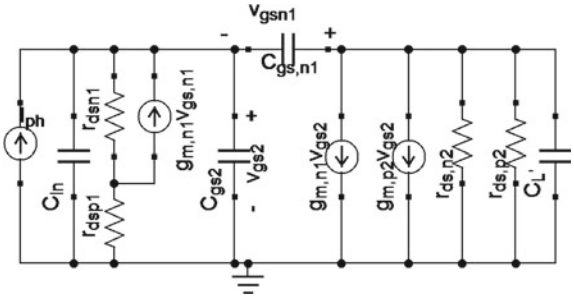

Fig. 6.10 Inverter based common-drain feedback small-signal equivalent circuit

(M_{n2}, R2) is working as local feedback stage. Whereas the ICDF output node V_{out2} (Fig. 6.9) is the drain of the inverter amplifier (M_{n2}, M_{p2}). M_{n1} and M_{p1} are working as local feedback stage. The feedback input is the gate of M_{n1} and the feedback output is the source of M_{n1}.M_{n1} and M_{p1} work as a common-drain feedback (CDF) for the main amplifier (the inverter M_{n2}, M_{p2}).

The ICDF-TIA's transimpedance gain with respect to the output V_{out2} can be calculated from the small-signal model. Figure 6.10 shows the small-signal equivalent circuit of the ICDF-TIA. From the small-signal analysis of the equivalent circuit, the transimpedance gain of the ICDF-TIA is given by (6.65):

$$A_T(S) = \frac{-(r_{ds,n1}+r_{ds,p1})\,A_{inv}}{1+g_{m,n1}r_{ds,n1}(1+A_{inv})+\left[r_{ds2,eq}\overline{C}_L\,A_{inv}\left(1+g_{m,n1}r_{ds,n1}\right)+\left(r_{ds,n1}+r_{ds,p1}\right)\overline{C}_{in}\right]S+ \left[r_{ds2,eq}\left(r_{ds,n1}+r_{ds,p1}\right)\overline{C}_{in}\overline{C}_L\,A_{inv}\right]S^2} \tag{6.65}$$

where $C_{in} = C_{PD} + C_{ESD} + C_{PAD}$, $\overline{C}_{in} = C_{in} + C_{gs2} + (1 + A_{inv})C_{gs1}$, $\overline{C}_L = C_L + (1 - 1/A_{inv})C_{gs1}$, $r_{ds2,eq} = r_{ds,n2}//r_{ds,p2}$.

From (6.65) the transimpedance gain at low frequency is given by:

$$A_T(0) = \frac{-(r_{ds,n1} + r_{ds,p1})\,A_{inv}}{1 + g_{m,n1}\cdot r_{ds,n1}(1 + A_{inv})} \tag{6.66}$$

The actual output voltage node of the proposed ICDF-TIA is V_{out2} because of its higher gain compared to V_{out1}. The output V_{out1} has a transimpedance gain of $r_{ds,p1}$ which is limited by the low power supply voltage. The transimpedance gain of V_{out2} is larger than $1/g_{m,n1}$ which can be maximized even at low supply voltage operation. The multi-threshold technology enables to use V_{out2} with higher gain and enough voltage headroom (M_{n2} and M_{p2} are selected to be low-threshold transistors). By comparing simulations for V_{out1} and V_{out2} in the ICDF-TIA, a higher transimpedance gain results for using V_{out2} [14].

The ICDF-TIA's transimpedance gain in (6.66) is larger than that of the RGC-TIA calculated by (6.45) for the same power consumption. The ICDF-TIA and the RGC-TIA were investigated in [14] and the values of the TIA elements were optimized to reach the same bandwidth at the largest possible transimpedance gain to minimize the noise. The transistor's sizing, resistor's values, and biasing currents were swept to reach the desired specifications. At the same input impedance the biasing current of the ICDF shows a 28 % reduction in the current consumption compared to the RGC-TIA. The input impedance and transimpedance gain are compared for the RGC and ICDF topologies in Fig. 6.11. The two TIAs have the same input impedance but the ICDF-TIA has a 5 dB larger transimpedance gain than the RGC. The higher transimpedance of the ICDF-TIA enhances the simulated sensitivity by 2 dB over the RGC sensitivity.

The first and second pole in the second order system in (6.65) can be calculated by [13]:

$$\omega_{p1} \approx \frac{1 + g_{m,n1}r_{ds,n1}(1 + A_{inv})}{\left(r_{ds,n1} + r_{ds,p1}\right)\overline{C}_{in}} \tag{6.67}$$

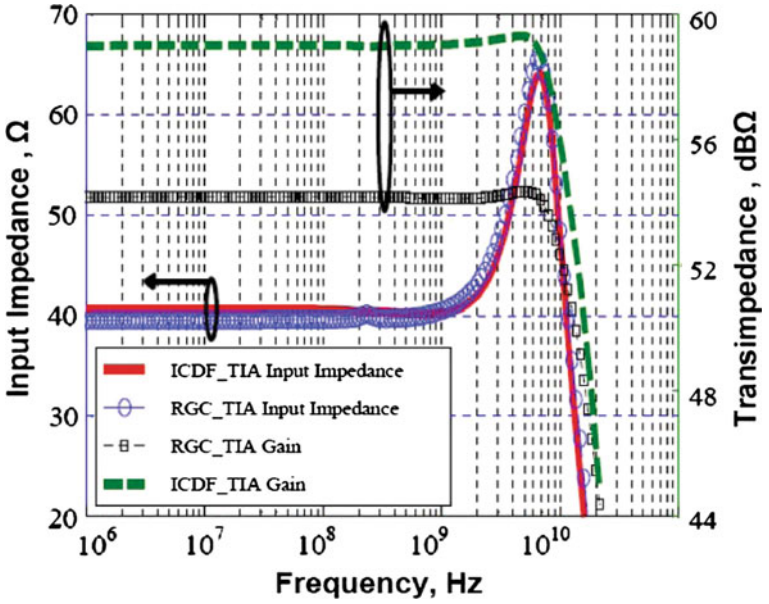

Fig. 6.11 Comparison of transimpedance gain and input impedance simulated for the ICDF and RGC TIAs [14]

$$\omega_{p2} \approx \frac{1}{r_{ds2,eq}\overline{C}_L} \tag{6.68}$$

$\overline{C}_{in} \gg \overline{C}_L$; so the effect of the second pole can be neglected compared to the dominant first pole.

The −3 dB bandwidth of the ICDF-TIA can be approximately given by:

$$BW \approx \frac{1 + g_{m,n1} r_{ds,n1} \left(A_{inv} + 1\right)}{2\pi \left(r_{ds,n1} + r_{ds,p1}\right) \overline{C}_{in}} \tag{6.69}$$

By substituting from (6.64) into (6.47) the bandwidth can be calculated by:

$$\mathrm{BW} = \frac{1}{2\pi Z_{in}(0)\,\overline{C}_{in}} \tag{6.70}$$

Equation (6.69) indicates that by increasing the inverter gain (A_{inv}) of the ICDF-TIA the bandwidth will increase. Due to the Miller effect the C_{gs1} and C_{gd2} values are multiplied by A_{inv}. By increasing C_{gs2} and the Miller capacitance with A_{inv} the input capacitance is increased and the bandwidth will decrease. As results, A_{inv} should be optimized to obtain the largest TIA bandwidth.

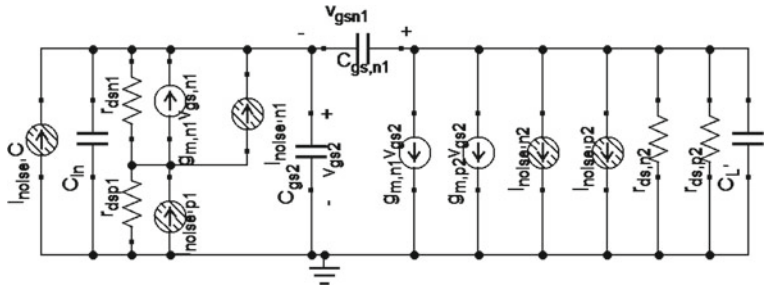

Fig. 6.12 Inverter based common-drain feedback small-signal equivalent circuit with noise generators

Figure 6.12 shows the noise equivalent circuit of the ICDF-TIA including the channel thermal noise sources of the MOSFETs. The output noise voltage v_{o1} due to the thermal noise of M_{p1} is given by (6.72). Also, the output noise voltage v_{o2} due to the channel thermal noise of $M_{n1}$is given by (6.73). The output noise voltage v_{o3} due to the thermal noise of M_{n2} and M_{p2} is given by (6.74). Finally, The output noise voltage v_{o4} due to the noise of the current source is given by (6.75).

$$i_{noise,C} = 4KT\gamma g_{m,C}\,(1+M)\,\Delta f \tag{6.71}$$

where γ is the noise factor of the MOSFET and M is the current mirror ratio.

$$v_{o1}(S) = \frac{-r_{ds,p1}A_{inv}i_{noise,p1}}{1+g_{m,n1}r_{ds,n1}(1+A_{inv})+\left[r_{ds2,eq}\overline{C}_L A_{inv}\left(1+g_{m,n1}r_{ds,n1}\right)+\left(r_{ds,n1}+r_{ds,p1}\right)\overline{C}_{in}\right]S+\left[r_{ds2,eq}\left(r_{ds,n1}+r_{ds,p1}\right)\overline{C}_{in}\overline{C}_L A_{inv}\right]S^2} \tag{6.72}$$

$$v_{o2}(S) = \frac{-r_{ds,n1}A_{inv}i_{noise,n1}}{1+g_{m,n1}r_{ds,n1}(1+A_{inv})+\left[r_{ds2,eq}\overline{C}_L A_{inv}\left(1+g_{m,n1}r_{ds,n1}\right)+\left(r_{ds,n1}+r_{ds,p1}\right)\overline{C}_{in}\right]S+\left[r_{ds2,eq}\left(r_{ds,n1}+r_{ds,p1}\right)\overline{C}_{in}\overline{C}_L A_{inv}\right]S^2} \tag{6.73}$$

$$v_{o3}(S) = \frac{-r_{ds2,eq}[1+(r_{ds,n1}+r_{ds,p1})\overline{C}_{in}S+g_{m,n1}r_{ds,n1}](i_{noise,n2}+i_{noise,p2})}{1+g_{m,n1}r_{ds,n1}(1+A_{inv})+\left[r_{ds2,eq}\overline{C}_L A_{inv}\left(1+g_{m,n1}r_{ds,n1}\right)+\left(r_{ds,n1}+r_{ds,p1}\right)\overline{C}_{in}\right]S+\left[r_{ds2,eq}\left(r_{ds,n1}+r_{ds,p1}\right)\overline{C}_{in}\overline{C}_L A_{inv}\right]S^2} \tag{6.74}$$

$$v_{o4}(S) = \frac{-(r_{ds,n1}+r_{ds,p1})A_{inv}\,i_{noise,C}}{1+g_{m,n1}r_{ds,n1}(1+A_{inv})+\left[r_{ds2,eq}\overline{C}_L A_{inv}\left(1+g_{m,n1}r_{ds,n1}\right)+\left(r_{ds,n1}+r_{ds,p1}\right)\overline{C}_{in}\right]S+\left[r_{ds2,eq}\left(r_{ds,n1}+r_{ds,p1}\right)\overline{C}_{in}\overline{C}_L A_{inv}\right]S^2} \tag{6.75}$$

By summing the squares of the noise voltages from (6.72)–(6.75) we get the output noise power density. The squared input noise current density (6.76) is calculated by

dividing the sum of the output noise power density by $|A_T\,(S)|^2$, where A_T is given by (6.65). The input referred noise current density is the square root of (6.76).

$$\overline{i_{in,noise}^2} = \left(\frac{r_{ds,p1}}{r_{ds,n1}+r_{ds,p1}}\right)^2 4K_BT\gamma g_{m,p1} + \left(\frac{r_{ds,n1}}{r_{ds,n1}+r_{ds,p1}}\right)^2 4K_BT\gamma g_{m,n1} + 4K_BT\gamma g_{m,C}(1+M) + \left(\left(\frac{1+g_{m,n1}\cdot r_{ds,n1}}{r_{ds,n1}+r_{ds,p1}}\right)^2 + \left(2\pi\cdot\overline{C}_{in}\,f\right)^2\right)\frac{4K_BT\gamma}{(g_{m,n2}+g_{m,p2})} \tag{6.76}$$

We can see from (6.76) that the low frequency input noise can be decreased by decreasing $g_{m,n1}$ and $g_{m,p1}$ due to increasing the transimpedance gain (6.66). On the other hand decreasing $g_{m,n1}$ has a negative effect by reducing the bandwidth. The enhanced input transconductance of the inverter stage $g_{m,n2} + g_{m,p2}$ reduces the low frequency and high-frequency noise contribution. Increasing the inverter transconductance has the effect of increasing the bandwidth by increasing the inverter gain A_{inv}.

The sizes and biasing current of transistors M_{n1}, M_{n2} and M_{p2} can optimized to get the minimum input noise current for the required TIA bandwidth.

6.5 Shunt-Shunt Feedback TIA

A TIA with a core based on a CS amplifier with shunt-shunt feedback resistor can improve the bandwidth and reduces the input-referred noise. However, the bandwidth improvement forces a trade-off with the gain (low in nanometer technology) and power consumption. Further increase in the transimpedance gain to decrease the input referred noise requires a higher power consumption which is undesirable for low power optical receivers. New TIA topologies like noise canceling TIA and Inverter Based Cascode TIA will be introduced to increase the bandwidth and sensitivity of optical receivers without scarifying the power consumption.

Figure 6.13 shows the basic circuit of a shunt feedback TIA. In the TIA circuit C_T consists of the capacitance of the PD and the input capacitance of the TIA.

Let us compare the small-signal transfer function of the two structures in Fig. 6.2 for the simplest TIA (simple resistor) and in Fig. 6.13, which is the transimpedance gain v_o/i_{in}. For Fig. 6.2 with the shunt feedback TIA we have:

$$\frac{v_o}{i_{in}} = \frac{R}{1+j\omega C_T R} \tag{6.77}$$

For Fig. 6.13 we have:

$$\frac{v_o}{i_{in}} = \frac{R_{fb}\frac{A_o}{1+A_o}}{1+j\omega C_T R_{fb}\frac{1}{1+A_o}} \tag{6.78}$$

The DC transimpedance gain for the simple TIA in Fig. 6.13 is the transimpedance resistor value R_{fb}. For high amplifier gains A_o this is also the case for Fig. 6.13. Comparing the frequency behavior, the dominant poles which define the small-signal bandwidth for the two circuits are defined by:

From Fig. 6.2:

$$BW = \frac{1}{2\pi C_T R} \tag{6.79}$$

From Fig. 6.13:

$$BW = \frac{1 + A_o}{2\pi C_T R_{fb}} \tag{6.80}$$

Equation (6.79) shows, that the bandwidth of the simple resistor TIA is completely defined by a given transimpedance and photodiode capacitance. Equation (6.80) shows, that the shunt-shunt feedback TIA has an approximately A_0 times higher bandwidth compared to a simple resistor TIA. The amplifier gain A_0 is design dependent and higher than one. For high data rate TIAs, the frequency dependent gain $A(\omega)$ has to be used. Therefore the increased bandwidth for the same transimpedance value is one of the most important advantages of a shunt-shunt feedback TIA compared to a simple resistor solution. The advantage of a shunt-shunt feedback TIA compared to the simple circuit described before is the fact that the bandwidth is indirectly related to the resistor R_{fb} divided by the open-loop gain A_o of the TIA (R_{fb}/A_o). Therefore the noise can be decreased for a given bandwidth, because of a large feedback resistor R_{fb}/A_o.

The smaller C_T and the larger R_{fb}, the higher the sensitivity is for a given responsivity (more discussion will be presented in the next sections). For a high bandwidth it is important to have small C_T and small R_{fb}. In an OEIC, C_T is dominated by the PD capacitance, a trade off between sensitivity and bandwidth has to be found for the feedback resistor.

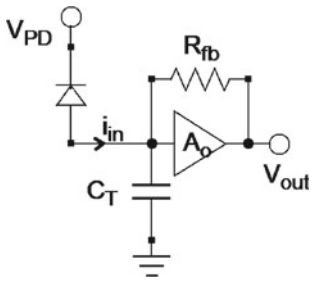

Fig. 6.13 Basic circuit of shunt feedback TIA

6.5.1 Frequency Response

The operational amplifier is characterized by its frequency dependent amplification:

$$A(\omega) = \frac{A_o}{1 + j\frac{\omega}{\omega_T} A_o} \tag{6.81}$$

Application of Kirchhoff's law onto the circuit shown in Fig. 6.14 gives:

$$V_i + I_2 Z_{fb} + V_o = 0 \tag{6.82}$$

Where $Z_{fb} = R_{fb} // C_{fb}$, and

$$I_{ph} + I_1 - I_2 = 0 \tag{6.83}$$

Inserting $I_1 = V_i / Z_{pd}$,where $Z_{pd} = R_{pd} // C_{pd}$, and $V_i = V_o/(A(\omega))$ into (6.82) and (6.83) leads to:

$$I_{ph} + \frac{V_o}{(A(\omega) Z_{pd}} - I_2 = 0 \tag{6.84}$$

and

$$V_o(1 + 1/(A(\omega)) - I_2 Z_{fb} = 0 \tag{6.85}$$

Multiplying (6.84) by Z_{fb} and adding it to formula (6.85) results in the transfer function $G(\omega)$ which is the transimpedance and possesses the dimension Ω:

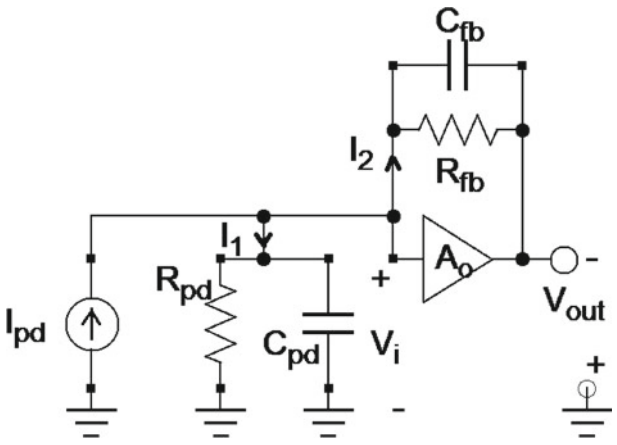

Fig. 6.14 Transimpedance amplifier circuit with photodiode model

$$G(\omega) = \frac{V_o}{I_{pd}} = -\frac{Z_{fb}}{1 + \frac{1}{A(\omega)}(1 + \frac{Z_{fb}}{Z_{pd}})} \tag{6.86}$$

with $A(\omega) = \frac{A_o}{1+j\frac{\omega}{\omega_T}A_o}$ (6.86) can be written as [15]:

$$G(\omega) = \frac{V_o}{I_{pd}} = -\frac{R_{fb}}{1 - R_{fb}(C_{pd} + C_{fb})\frac{\omega^2}{\omega_T} + j\omega(\frac{1}{\omega_T} + \frac{R_{fb}}{R_{pd}\omega_T} + R_{fb}C_{fb} + R_{fb}\frac{C_{pd}}{A_o})} \tag{6.87}$$

For an ideal operational amplifier with A_o and ω_T equal to ∞ the gain in (6.87) becomes:

$$G\,(\omega) = -\frac{R_{fb}}{1 + j\omega\left(R_{fb}C_{fb}\right)} \tag{6.88}$$

and the -3 dB bandwidth can be calculated from the following equation:

$$BW = \frac{1}{2\pi \cdot \left(R_{fb}C_{fb}\right)} \tag{6.89}$$

The assumption of the ideal amplifier is too simple for the approximation of the real behavior of transimpedance amplifiers. The gain peaking (overshoot) which occurs often in practice, for instance, is not included in the ideal transfer function. To study this gain peaking, the magnitude of the transfer function of (6.87) can be rearranged in the new formula (6.90) implementing the damping factor D:

$$G\,(\omega) = -\frac{\omega_N^2 R_{fb}}{\omega_N^2 + j2\omega D\omega_N - \omega^2} \tag{6.90}$$

The magnitude of the transfer function can be determined from formula (6.90) to be:

$$|G\,(\omega)| = -\frac{\omega_N^2 R_{fb}}{\sqrt{\omega_N^4 - 2\omega_N^2\omega^2 + \omega^4 + 4\omega^2 D^2\omega_N^2}} \tag{6.91}$$

With the characteristic angular frequency ω_N:

$$\omega_N = \sqrt{\frac{\omega_T}{R_{fb}(C_{pd} + C_{fb})}} \tag{6.92}$$

and the damping factor D:

$$D = 0.5\sqrt{\frac{\omega_T}{R_{fb}(C_{pd} + C_{fb})}}.(\frac{1}{\omega_T} + \frac{R_{fb}}{R_{pd}\omega_T} + R_{fb}C_{fb} + R_{fb}\frac{C_{pd}}{A_o}) \tag{6.93}$$

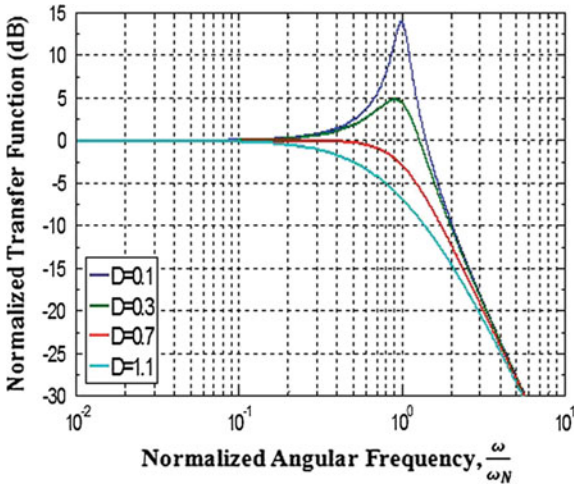

Fig. 6.15 Normalized magnitude of the transfer function $G(\omega)$ for different values of the damping factor D

In Fig. 6.15 the normalized magnitude of the transfer function is plotted for different values of the damping factor D. Equation (6.91) can have a resonance (gain-peak) at $\omega_{GP} = \omega_N\sqrt{1-2D^2}$, provided that the damping factor D <0.7. The value of the normalized transfer function at the gain-peak frequency is:

$$\frac{|G(\omega)|}{R_{fb}} = -\frac{1}{2D\sqrt{1-D^2}} \tag{6.94}$$

Gain-peaking is a possibility to increase the bandwidth of the transimpedance amplifier at the cost of a higher group delay variation and longer settling time due to a higher overshoot in the step response [15]. A shunt feedback capacitance C_{fb} can be used to control the damping factor D (6.93) to decrease the peaking and ensure stability.

6.5.2 Noise Analysis of Shunt Feedback TIA

In this section noise models of transimpedance amplifiers with a shunt feedback resistor will be presented. First a TIA with an ideal amplifier will be discussed. A TIA with CS input stage is described in detail subsequently. The input referred noise current of the TIA depends on the input node capacitance C_T, the feedback resistor R_{fb}, and the transistor noise of the amplifier's input stage. Different shunt feedback TIAs based on different amplifiers topologies like inverter based TIA, inverter based

cascode TIA, and noise canceling TIA can be used to enhance the performance. In the next sections different TIA topologies will be introduced and their performance will be discussed.

6.5.3 Noise of Ideal TIA

First the general model of the ideal TIA is considered. In Fig. 6.16 the noise sources of the individual TIA components are included. The noise of the TIA depends on the technology and topology of the circuit. As a first estimation we define a very simple circuit which also converts an input current into an output voltage, a photodiode and transimpedance amplifier. In the easiest case this amplifier is a common-source circuit. With more complicated amplifier designs better results can be achieved, but as an upper bound for noise we consider the circuit shown in Fig. 6.17. The main noise sources are the thermal noise of the resistor R_F and the equivalent input noise sources of the amplifier ($\overline{i^2_{n.amp}}$ and $\overline{v^2_{n.amp}}$). We can calculate the input referred squared noise current spectral density as follows [16]:

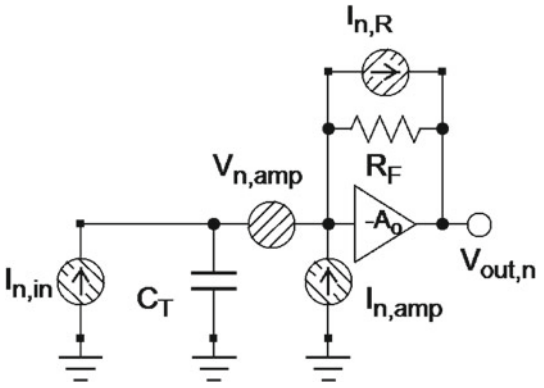

Fig. 6.16 Basic circuit of a shunt feedback TIA including general noise sources

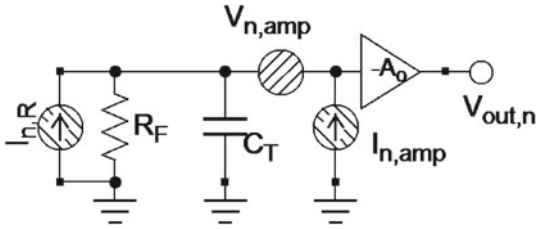

Fig. 6.17 Basic circuit representing an upper bound for noise analysis

$$\overline{i_{n.in}^2} = \overline{i_{n.amp}^2} + \frac{\overline{v_{n.amp}^2}}{\left|\frac{R_F}{1+j\omega C_T R_F}\right|^2} + \overline{i_{n.R}^2} \tag{6.95}$$

The TIA's input referred current noise $i_{n.amp}$ can be neglected for MOS transistors.

Which leads to:

$$\overline{i_{n.in}^2} = \overline{v_{n.amp}^2}\left(\frac{\left(1 + 4\pi^2 f^2 C_T^2 R_F^2\right)}{R_F^2}\right) + \overline{i_{n.R}^2} \tag{6.96}$$

Equation 6.96 shows that the squared input referred noise current spectral density raises proportional to C_T^2 and f^2 at high frequencies. The total input noise current is obtained by integrating equation (6.96) over 1.57 times the bandwidth as stated in Sect. 6.1 for first order circuits.

6.5.4 TIA with Common-Source Input Stage

For a MOSFET common-source (CS) input stage the amplifier circuit is shown in Fig. 6.18. Again the noise sources of the amplifier equal the equivalent input noise sources of the FET $\overline{v_{n.M}^2}$ and $\overline{i_{n.M}^2}$. The noise of the load resistor R_A again is neglected, because it is attenuated by the transconductance of the MOSFET, if it is referred to the input.

The equivalent input noise voltage of the MOSFET can be calculated by using the following equation [17]:

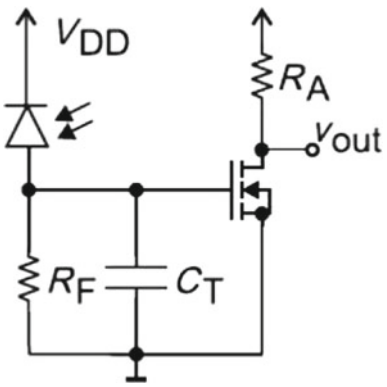

Fig. 6.18 PIN-FET amplifier circuit for small-signal FET-TIA noise analysis

$$\frac{\overline{v_{n.M}^2}}{\Delta f} = 4K_B T \frac{2}{3g_m} + K_f \frac{I_D^a}{g_m^2 f} \tag{6.97}$$

A second noise source is the 1/f dependent flicker noise, which can be significant for low frequencies up to a few MHz. It is proportional to a constant K and the drain current I_D. For high frequency applications, the 1/f-noise can be neglected.

$$\frac{\overline{v_{n.M}^2}}{\Delta f} \approx 4K_B T \frac{\gamma}{g_m} \tag{6.98}$$

Substituting the noise generators shown in (6.98) in (6.96) leads to:

$$\frac{\overline{i_{n.in}^2}}{\Delta f} = \frac{4K_B T}{R_F} + \frac{1 + 4\pi^2 f^2 C_T^2 R_F^2}{R_F^2} \left(4K_B T \frac{\gamma}{g_m}\right) \tag{6.99}$$

For low-noise applications, the g_m of the input MOSFET and therefore its drain current I_D should be as high as possible.

6.5.5 *Multistage Inverter Based CMOS TIA*

The transimpedance amplifier can be represented by a voltage amplifier with open loop gain A and resistive feedback resistance R_{FB}. The input capacitance C_{in} is large as it includes the photodiode, pad, and ESD capacitance. The dominant pole of this structure is at the input node, so its bandwidth can be approximated by (6.80).

For a given transimpedance R_{FB} a higher bandwidth requires a larger loop gain. For a given amplifier bandwidth the transimpedance should be maximal and vice versa. To do so, gain A should be maximized.

The maximum achievable gain is restricted by stability requirements. An increased gain A shifts the higher order poles of the amplifier down closer to the first pole and thus reduces the phase-margin. To obtain a stable structure, the maximal open-loop gain A is limited.

In nanometer CMOS technology the Early voltage is small and so r_{ds} is smaller than R_c. In the traditional CS circuit (Fig. 6.19a) the open-loop gain of one stage is given by:

$$A = g_m(R_C // r_{ds}) = \frac{g_m}{g_{ds} + g_c} \approx \frac{g_m}{g_{ds}} \tag{6.100}$$

The CS single stage open loop gain A is nearly equal to the intrinsic gain of the MOS transistor and depends on biasing and process parameters. In nanometer CMOS technology the intrinsic gain is low, and it is strongly process dependent. To obtain a high data rate in combination with a large amplification factor, a single-stage amplifier is not sufficient. A three-stage amplifier is needed (Fig. 6.19a). Another disadvantage

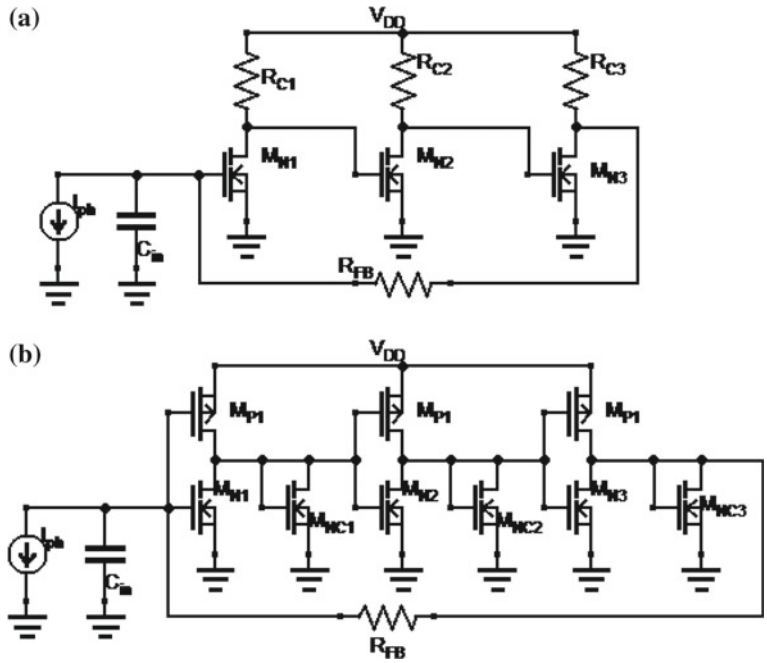

Fig. 6.19 **a** Traditional multi-stage CS TIA, **b** multi-stage inverter based TIA with measures to achieve stability

is that the stability of this circuit with feedback is also strongly dependent on g_{ds}. A system with feedback is stable if the second pole's frequency is more than 2 times larger than the gain-bandwidth product, which requires a large safety margin in the design. This results in circuits with either limited speed or transimpedance gain.

The circuit shown in (Fig. 6.19b) is a single-ended TIA in pure CMOS technology, which means a low-cost implementation [18]. This circuit uses a 3-stage inverter based amplifier to achieve a high gain. The output of the 3rd stage is connected to the TIA input via the transimpedance resistor R_{fb}. Every stage of the three amplifier stages consists of 3 MOSFET transistors. The gain of every stage is defined by the transconductance of the three transistors, as for the first stage:

The design in [18] uses a three-stage inverter (Fig. 6.19b) and the amplification of a single stage A_1 is given by:

$$A_1 = \frac{g_{mN} + g_{mP}}{g_{mc} + g_{dsN} + g_{dsP} + g_{dsc}} \approx \frac{g_{mN} + g_{mP}}{g_{mc}}$$

$$A_1 \approx \frac{\left(\frac{W}{L}\right)_{M_N}}{\left(\frac{W}{L}\right)_{M_c}} + \sqrt{\left(1 + \frac{\left(\frac{W}{L}\right)_{M_N}}{\left(\frac{W}{L}\right)_{M_c}}\right) \cdot \frac{\left(\frac{W}{L}\right)_{M_P}}{\left(\frac{W}{L}\right)_{M_c}}} \cdot \sqrt{\frac{\mu_p}{\mu_n}} \tag{6.101}$$

The open-loop gain is lowered by the diode-connected MOSFETs' ($M_{NC1,2,3}$) transconductance g_{mc} to ensure stability. The process dependent ratio of the mobilities appears only as a square root. The rest of the formula forms a purely geometrical parameter. The optimization of transistor sizing can limit the effect of the mobility variation.

The disadvantage of this topology is that the additional diode-connected MOSFETs $M_{NC1,2,3}$ are always ON and always lower the gain of the amplifier. This increases the input referred noise due to lowering the gain and the additional noise added by $M_{NC1,2,3}$ itself. At the same time M_{NC} is needed to ensure stability by lowering the amplifier open loop gain. The stability problem becomes more critical with automatic gain control (AGC) or gain compression circuitry used to reduce the transimpedance gain with increasing the input optical power. As the gain bandwidth product is constant; the bandwidth is increasing, and the fundamental pole will be closer to the second pole. So, the stability will be reduced at high input optical power. The stability will be analyzed in the end of this section.

The circuitry of multi-stage inverter based TIA with compression and stability compensation stages fabricated in 90 nm CMOS [19] is shown in Fig. 6.20. The MOSFETs M_{NC1} and M_{NC2} are the gain compression transistors used to increase the dynamic range. The diode-connected MOSFET M_{NC} in Fig. 6.19b is modified in Fig. 6.20 and two MOSFET switches M_{NC4} and M_{NC5} are added. At low input optical power the control signal V_C is small; so the switches $M_{NC4,5}$ are OFF, thus M_{NC3} and M_{NC6} are disconnected from the amplifier. The amplifier gain is higher and no additional noise comes from M_{NC3} or M_{NC6}. At higher input optical power the control signal V_C has to be increased to connect M_{NC3} and M_{NC6} to the amplifier. This reduces the open loop gain to ensure stability. At high input optical power the added noise from M_{NC3} and M_{NC6} has no effect on the signal quality. As a result the dynamic range can be increased with a higher sensitivity and with a larger maximum optical input power but without stability problems.

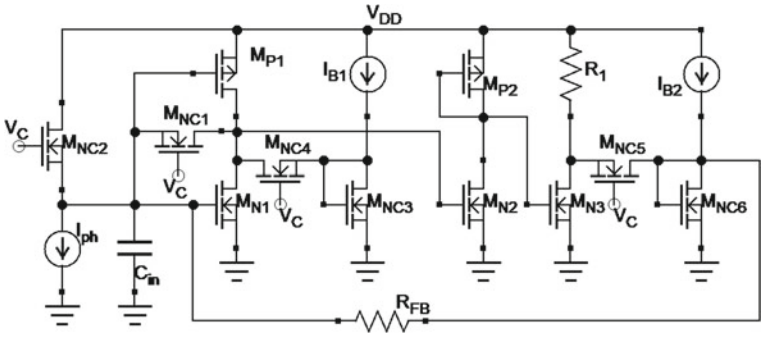

Fig. 6.20 Circuitry of multi-stage inverter based TIA with compression and stability compensation stages [19]

The circuitry of multi-stage inverter based TIA with stability compensation stages, shown in Fig. 6.20, is formed by a three-stage inverting amplifier (M_{N1}, M_{P1}, M_{N2}, M_{P2}, M_{N3}, R_1) and a fixed shunt feedback resistor $R_{FB} = 1.5\,k\Omega$.

Three stages are needed for the inverting amplifier to achieve enough open loop gain A. The amplifier has three stages, the first is an inverter, second one is a common-source stage with active PMOS load, and the third stage is a simple common-source amplifier. These amplifier stages are suitable for low-voltage operation. In this design, an inverter (M_{N1}, M_{P1}) is used as the first stage, because it shows the highest gain, hence optimizing the overall noise performance.

The second stage (M_{N2}, M_{P2}) implements a diode-connected PMOS transistor resulting in a lower noise and a smaller area than the traditional resistive load. The third (M_{N3}, R_1) stage is a common-source circuit with a load resistor R_1. Minimal length is used in all MOS transistors to achieve a high transconductance (g_m) and to optimize frequency and noise response. The widths of the NMOS transistors are chosen for a good noise-power trade-off and the widths of the PMOS transistors and the value of the resistor R_1 is designed to keep a biasing voltage of 0.6 V over the whole signal path. A parametric sweep for the transistor widths is performed. The input noise current, gain, bandwidth and power consumption are calculated for different device widths. The optimum widths are selected for the best performance.

Transistor M_{NC1} creates a current feedback path which avoids input photocurrent. M_{NC2} forms a further current path for high photocurrents in order to be able to use a small bias current through M_{N1} and M_{P1}. Both effects enhance the input dynamic range. At low input photocurrents ($<160\,\mu A$) V_C is lower than 0.6 V and is increased by increasing the input optical power. V_C reaches its maximum (1V) at the maximum input average photocurrent ($500\,\mu A$). The operation of $M_{NC1,2}$ transistors depends on the values of their gate voltage V_C: when $V_C \leq 600$ mV, the transistors are OFF, in the denominated inactive region; when $V_C > 600$ mV they start to conduct and are working in active region. In this way, the TIA works in two different regions of operation depending on the control voltage V_C. To ensure stability over the whole input dynamic range, an additional circuitry is needed to lower the open loop gain at high input optical power. At the same time, as shown in the schematic, V_C drives transistors M_{NC4} and M_{NC5}, which constitute a compensation circuit together with $M_{NC3,6}$. This compensation circuit helps to dynamically reduce the open-loop gain of the first and third stages ensuring stability over the whole input dynamic range.

From the small signal model the maximum transimpedance gain is given:

$$A_{Tmax} \approx \frac{R_{FB} \cdot \left(g_{m,eq} \cdot r_{eq}\right) \cdot \left(g_{m,N2} \cdot r_{ds,P2} // r_{ds,N2}\right) \cdot \left(g_{m,N3} \cdot R_1 // r_{ds,N3}\right)}{1 + \left(g_{m,eq} \cdot r_{eq}\right) \cdot \left(g_{m,N2} \cdot r_{ds,P2} // r_{ds,N2}\right) \cdot \left(g_{m,N3} \cdot R_1 // r_{ds,N3}\right)} \quad (6.102)$$

where $g_{m,eq} = g_{m,n1} + g_{m,n2}$ and $r_{eq} = r_{ds,n1} // r_{ds,p1}$

The transimpedance gain is reduced when $V_C > 0.6$V and reaches its minimum when $V_C = 1$V.

$$A_{Tmin} \approx \left(\frac{r_{ds,NC2}}{R_{in}+r_{ds,NC2}} \right) \cdot \frac{R_{FB} \cdot \left(g_{m,eq}r_{eq}//r_{ds,NC1}//r_{ds,NC3}\right) \cdot \left(g_{m,N2} \cdot r_{ds,P2}//r_{ds,N2}\right) \cdot \left(g_{m,N3}R_1//r_{ds,N3}//r_{ds,NC6}\right)}{1+ \left(g_{m,eq}r_{eq}//r_{ds,NC1}//r_{ds,NC3}\right) \cdot \left(g_{m,N2} \cdot r_{ds,P2}//r_{ds,N2}\right) \cdot \left(g_{m,N3}R_1//r_{ds,N3}//r_{ds,NC6}\right)} \tag{6.103}$$

where the input resistance of the TIA is:

$$R_{in} \approx \frac{R_{FB}}{1+ \left(g_{m,eq}r_{eq}//r_{ds,NC1}//r_{ds,NC3}\right) \cdot \left(g_{m,N2} \cdot r_{ds,P2}//r_{ds,N2}\right) \cdot \left(g_{m,N3}R_1//r_{ds,N3}//r_{ds,NC6}\right)} \tag{6.104}$$

and the resistance of the MOSFETs as a function of V_{GS} is calculated by:

$$r_{ds,NC1} = \frac{1}{\left(\frac{W}{L}\right) \mu C_{ox} \left(V_{GS} - V_T\right)} \text{(works in triode region)}$$

$$r_{ds,NC2} = \frac{1}{\lambda \left(\frac{W}{L}\right) \mu C_{ox} \left(V_{GS} - V_T\right)^2} \text{(works in saturation region)}$$

The circuit is shown in Fig. 6.20. The first amplifier stage is built by M_{N1} and M_{P1}. The inverter structure was chosen to minimize the input-noise current of the amplifier. M_{N2} / M_{P2} and R_1 / M_{N3} build the second and third amplifier stage, respectively. The noise of the second and third stage is not critical, because it is not directly at the input node and therefore is attenuated by the gain of the previous stage, when referred to the input. At low input optical power the switches $M_{NC4,5}$ are OFF, so $M_{NC3,6}$ is disconnected from the amplifier. The amplifier gain is higher and no additional noise comes from M_{NC3}.

The output noise contributions of these noise sources are computed then and the input referred noise current in (6.105) is calculated by dividing the output noise by the transimpedance gain.

$$\frac{\overline{i_{n.in}^2}}{\triangle f} = 4K_B T \left[\frac{1}{R_{FB}} + 4\pi^2 f^2 C_T^2 \cdot \left(\frac{\gamma}{g_{m,n1} + g_{m,p1}} \right) + \frac{1}{R_{FB}^2} \cdot \left(\frac{\gamma}{g_{m,n1} + g_{m,p1}} \right) \right] \tag{6.105}$$

The advantage of using the inverter based amplifier is the reduction in the input noise current by increasing the transconductance to be the sum of the two input transistors' transconductance. The loads $M_{NC3,6}$ have no effect on the noise performance as the switches $M_{NC4,5}$ are OFF at low input optical power.

For stability analysis, the loop gain A_{loop} is calculated and the bode diagram of the open loop system is plotted. Figure 6.21 depicts the opened loop of the system. The cutting site between X_1 and X_2 is chosen, to minimize the influence of the opening

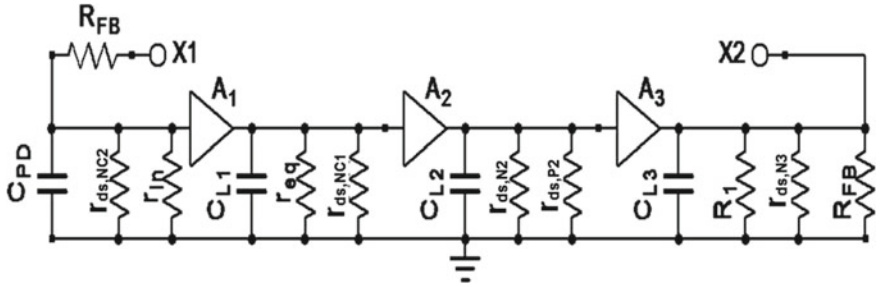

Fig. 6.21 The opened loop of the system for stability analysis

Table 6.1 Phase margin of the system with decreasing $r_{ds,CN1,2}$

$r_{ds,NC1,2}$	Phase margin
10 kΩ	45
67 Ω	47
20 Ω	68
10 Ω	90

on the behavior of the circuit. To keep the load of the feedback network also at the output, R_{FB} is inserted again as load.

To investigate stability, the feedback resistor $R_{FB} = 1.5\,\text{k}\Omega$ was selected to ensure the stability when $r_{ds,NC1,2} = \infty$. The loop gain $A_{loop}(\text{s})$ of the system is calculated by:

$$\mathrm{A_{Loop(S)}}=\frac{\left(\mathrm{g_{m,eq}r_{eq}//r_{ds,NC1}//r_{ds,NC3}}\right)\left(\mathrm{g_{m,N2}r_{P2}//r_{ds,N2}}\right)\left(\mathrm{g_{m,N3}R_1//r_{ds,N3}//r_{ds,NC6}}\right)}{\left(1+\mathrm{S}\cdot\mathrm{C_{PD}.R_{FB}//r_{ds,NC2}//r_{in}}\right)\left(1+\mathrm{S}\cdot\mathrm{C_{L1}.r_{eq}//r_{ds,NC1}}\right)\ \left(1+\mathrm{S}\cdot\mathrm{C_{L2}r_{P2}//r_{ds,N2}}\right)\left(1+\mathrm{S}\cdot\mathrm{C_{L3}.R_1//r_{ds,N3}//R_{FB}}\right)} \tag{6.106}$$

where

$$\mathrm{r_{in}}=\frac{\mathrm{r_{ds,NC1}}}{1+\mathrm{g_{m,eq}r_{eq}//r_{ds,NC1}//r_{ds,NC3}}} \tag{6.107}$$

The circuit instability can be avoided by varying the voltage gain of the amplifier stages by reducing the values of the load resistors. This gain reduction is done by connecting the diode loads formed by transistors (M_{CN3} and M_{CN6}). The reduction of the load resistances has two consequences. On the one hand the voltage gain of the amplifier stage is reduced and on the other hand, the cutoff frequencies of the amplifier stages move towards higher frequencies. This leads to an increasing phase margin and therefore a stable system for every $r_{ds,NC1,2}$. The Bode diagram of the loop gain with decreasing $r_{ds,NC1,2}$ is plotted in Fig. 6.22. Looking at the phase margin of the loop gain for reducing the compression transistors' resistance $r_{NC1,2}$

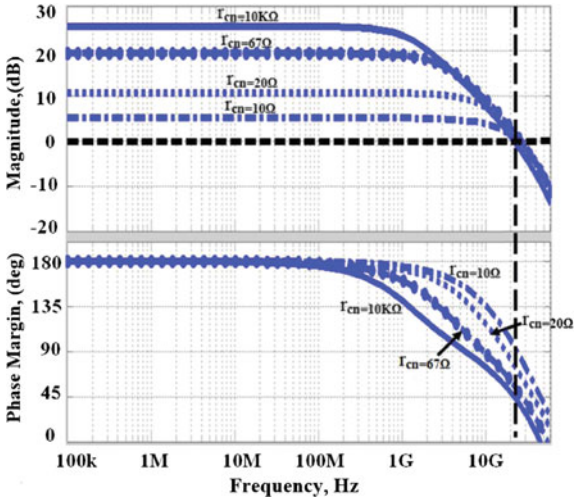

Fig. 6.22 Bode diagram of the loop gain with decreasing $r_{ds,CN1,2}$

(see Fig. 6.22 and Table 6.1), it can be clearly seen, that the phase margin is increasing and stability of the system is ensured.

6.5.6 Noise Canceling TIA

The shunt-shunt feedback TIA with common-source/inverter amplifier is the most popular TIA topology used in optical communication; as it improves the bandwidth and reduces the input-referred noise. The bandwidth improvement forces a trade-off with the gain, noise and power consumption. Further increase in the transimpedance gain to decrease the input referred noise requires higher power consumption which is undesirable. A noise canceling technique can be used with the multistage CMOS inverter based TIA [20] to enhance the sensitivity of the conventional multistage inverter based amplifier.

Noise cancelling was used for a low noise amplifier (LNA) [21]. This allows for designing wide-band impedance matching amplifiers with small noise figure, without suffering from instability. The input stage in the LNA was designed to make a matching for 50 Ω impedance to the antenna. In the optical applications the input stage is a TIA with its input connected to a PD which has a very large shunt resistance R_{sh}. The first trial (simulation only) with noise canceling for optical receivers was introduced in [22]. Noise canceling introduced in [22] used the fact that at a point inside the feedback resistance a virtual ground for the input current is located. This node was used to cancel the noise of the TIA.

The TIA in [20] employs a noise canceling technique to decrease the input referred noise to achieve a high sensitivity with a high bandwidth at low power consumption.

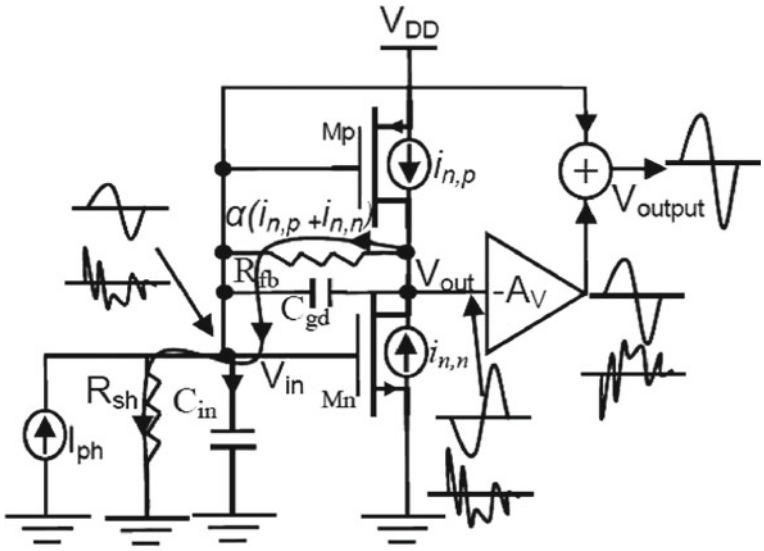

Fig. 6.23 Inverter based transimpedance amplifier using a noise canceling technique

The TIA is followed by two stages of differential amplifiers and the last stage is a 50 Ω differential output driver as interface to the measurement setup.

To understand the principle of noise canceling, consider the inverting TIA stage of Fig. 6.23. Let us now analyze the signal and the noise voltages at the input node V_{in} and output node V_{out}, both with respect to ground, due to the noise current $i_{n,n}$ and $i_{n,p}$ of the input stage CMOS inverter (M_n,M_p). The noise current causes two noise voltages at nodes V_{in} and V_{out}, which have the same sign. On the other hand, the signal voltages at nodes V_{in} and V_{out} have opposite sign, see Fig. 6.23. This difference in sign for noise and signal makes it possible to cancel the noise of the input devices and at the same time adding the two signals coming from V_{in} and V_{out} constructively. This is done by creating a new output V_{output}, where the voltage at node V_{in} is added to a scaled negative value replica (scaling factor A_V) of the voltage at node V_{out}. The optimum value for this scaling factor (A_V) makes a noise canceling at the output node V_{output}.

Any noise source that can be modeled by a current source between the drain and source of the input device is canceled as well (e.g., 1/f noise, thermal noise of the distributed gate resistance, and the bias noise current injected into node V_{out}). However, the noise of R_{fb} will not be canceled.

By circuit inspection, the input device noise voltages at node V_{in} and V_{out} are

$$V_{in,n} = \alpha \cdot (i_{n,n} + i_{n,p}) \cdot Z_{sh},$$

where $Z_{sh} = R_{sh}\ (1/j\omega C_{in})$ and C_{in} is the sum of C_{PD}, $C_{ESD+PAD}$ and C_{gs}.

Depending on the relation between $Z_{in}(g_m,R_{fb},R_{sh})$ of the TIA and R_{sh} a part of the input MOSFET noise current $\alpha \cdot I_n$ flows out of the input MOSFET through R_{fb} and R_{sh}.

$$V_{out.n} = \alpha \cdot (i_{n,n} + i_{n,p}) \cdot (Z_{sh} + Z_{fb}),$$

where $Z_{fb} = R_{fb}\ (1/j\omega C_{gd})$ and C_{gd} is the total gate drain capacitance for M_n, M_p. and $\alpha < 1$ is a function of the input MOSFET's transconductance ($g_{m,n}$, $g_{m,p}$) and of the PD's shunt resistance R_{sh}.

After the scaling factor A_V and the addition of $V_{in,n}$ and $V_{out,n} \cdot A_v$, the final output noise voltage is then equal to:

$$V_{output.n} = V_{in,n} - V_{out,n} \cdot A_v = \alpha \cdot (i_{n,n} + i_{n,p}) \cdot (Z_{sh} - A_V \cdot (Z_{sh} + Z_{fb}))$$

A real implementation for noise canceling TIA (NC-TIA) is introduced in Fig. 10.6. In this circuitry the noise currents $i_{n,n}$ and $i_{n,p}$ are amplified by A_{v2} as well as A_{v3}:

$$V_{output.n} = \alpha . (i_{n,n} + i_{n,p}) . A_{v2} . A_{v3} . (Z_{sh} - A_{V2} . (Z_{sh} + Z_{fb}))$$

Total output noise canceling is achieved when $V_{output.n} = 0$ for a gain A_{v2} equal to A_{vc} :

$$A_{vc} = Z_{sh}/(Z_{sh} + Z_{fb})$$

$$A_{vc} = \frac{R_{sh} + jR_{sh}R_{fb}C_{gd}!}{R_{sh} + R_{fb} + jR_{sh}R_{fb}(C_{gd} + C_{in})!} \tag{6.108}$$

R_{sh} is the internal shunt resistance of the PD which is infinity in an ideal case and tens of mega ohms for the integrated PD. R_{fb} is the feedback resistor of the TIA which in our case is 2 kΩ.

Because of the high value of R_{sh} compared to R_{fb}, the gain value A_{vc} (6.108) is approximately equal to unity for low frequencies. Because of the very small value of $R \cdot_{fb} C_{gd}$ and $R \cdot_{fb} C_{in}$ (C_{in} is small for a small-area integrated PD); the terms $\omega \cdot R \cdot_{fb} C_{gd}$ and $\omega \cdot R \cdot_{fb} C_{in}$ will be less than one for frequencies up to 1 GHz. So the absolute values of the denominator and the numerator will be close to R_{sh} and the A_{vc} will be close to unity.

Circuit parasitic capacitances not only limit the signal bandwidth but also degrade noise and distortion cancellation. Equation 6.108 shows that exact noise cancellation occurs only at low frequency for $A_V = A_{VC}$. As the frequency increases, the cancellation degrades mainly because of C_{in}. However, this effect can be modest up to relatively high frequencies because of the low input-node resistance ($Rs/2 = 25\ \Omega$) and low C_{in} in the LNA of [21]. In the proposed integrated optical receiver the input-node resistance is higher (65 Ω) and C_{in} is also higher due to the PD's capacitance. As a result the deviation from the perfect noise canceling should be higher in the presented optical receiver than in the case of the LNA in [21].

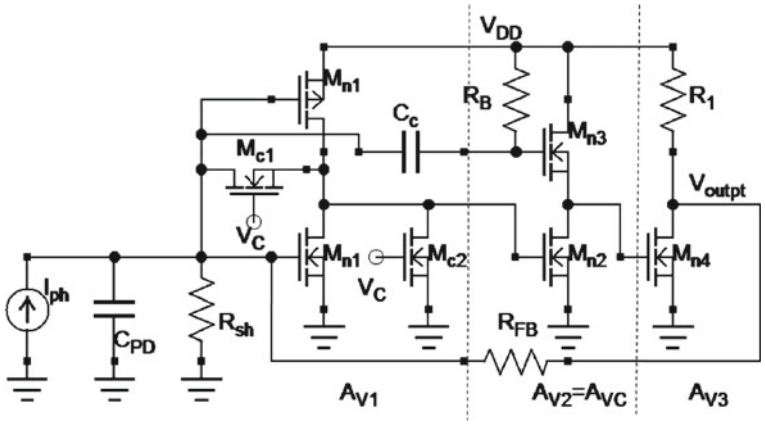

Fig. 6.24 Circuitry of the presented transimpedance amplifier using a noise cancellation technique (NC-TIA)

A TIA designed in 40 nm standard CMOS according to the introduced noise canceling (NC-TIA) concept will be introduced. The proposed NC-TIA, shown in Fig. 6.24, is formed by a three-stage inverting amplifier (M_{n1}, M_{p1}; M_{n2}, M_{n3}; M_{n4}, R_1) with a fixed shunt feedback resistor $R_{fb} = 2\,k\Omega$. The input stage exploits shunt feedback (R_{fb}) across a CMOS inverter (M_{n1}, M_{p1}) to provide high gain and to achieve a low input impedance of about 65 Ω.

The low input impedance compensates for the large capacitance C_{PD} of the large-area PD to obtain enough bandwidth for a data rate of 2.5 Gbit/s.

The input node of the input inverter stage (M_{n1}, M_{p1}) is ac coupled to the gate of M_{n3} via Cc = 0.18 pF. M_{n3} is biased from the voltage supply through $R_B = 77\,k\Omega$. The common-source stage M_{n2} and M_{n3} implements the adder used to subtract the two noise signals coming from the input (V_{in}) and the output (V_{out}) nodes, see Fig. 6.23. Transistor M_{n3} also acts as a source follower, copying the voltage at the input node of the (M_{n1}, M_{p1}) stage to the output. The output of the (M_{n2}, M_{n3}) stage delivers the sum of signal voltages and the difference of the noise voltages.

The sizing of M_{n2} and M_{n3} is selected to have an optimum value for $A_{V2} = A_{VC} \approx 1$ to minimize the input referred noise. The practical value for this voltage amplification A_{V2} is nearly equal to one, as expected for the approximation of relation (6.108).

The last stage is a common-source amplifier with the N-channel MOSFET Mn4 and a load resistance R_1 which is selected to have a potential of 0.55 V (half of V_{DD}) at the output and at all TIA stages as well as to achieve enough gain for the TIA to have a low input impedance.

The main noise contributions are coming from the feedback resistor R_{fb} (48 %) and input inverter stage M_{n1} and M_{p1} (33 %). The second stage M_{n2} and M_{n3} contributes by 6 % of the noise and the last stage M_{n4} and R_1 by 3 %. Due to the low noise

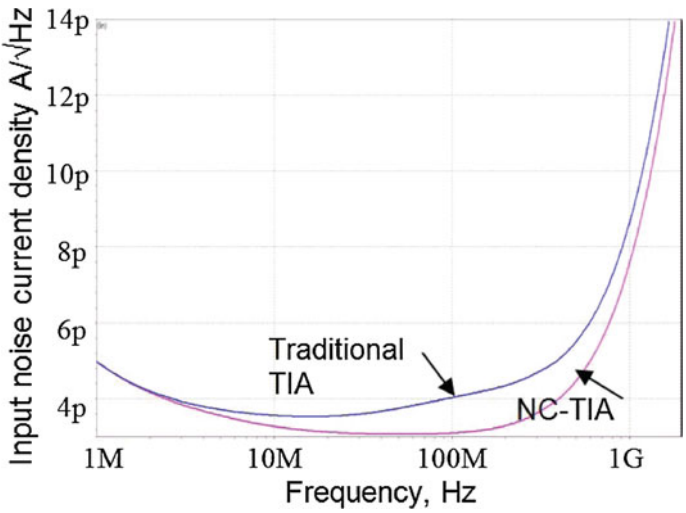

Fig. 6.25 Simulated total input noise current density [20]

contribution from the second and third stages their effect on the noise cancelling can be neglected.

The M_{c1} and M_{c2} N-MOSFETs are used to increase the input optical power dynamic range. The M_{c1} and M_{c2} are off at low input optical power. When the input optical power increases, the control signal V_c switches on M_{c1} and M_{c2}. M_{c1} then creates a secondary path for the generated photocurrent in addition to R_{fb}. When M_{c1} and M_{c2} are off they have the effect to decrease the bandwidth compared to the TIA without M_{c1} and M_{c2} due to their parasitic capacitances. However, this technique succeeds to increase the maximum input optical dynamic range.

A comparison was made by post layout circuit simulations between the presented NC-TIA and the same TIA with the same topology (three inverter stages) but without applying noise canceling, i.e. without C_C (both in the same 40 nm CMOS technology, same biasing conditions and same photodiode model). The NC-TIA shows 15 % reduction in the integrated input referred noise current (256 nA) compared to the traditional three-stage inverter TIA (302 nA). There is only a 15 % reduction in the noise because the noise cancelling method can cancel only thermal noise of the input MOSFETs but cannot cancel the thermal noise of R_{fb}. The spectral input noise current density for the proposed noise cancelling TIA and for the conventional TIA without noise cancelling is illustrated in Fig. 6.25.

The effect of the process variation on the NC-TIA noise performance was studied using Monte-Carlo simulation. The NC-TIA input noise was calculated for every Monte-Carlo run up to 1000 runs. Then the mean value for the 1000 values of the input referred noise currents and standard deviations are calculated to be 256 nA with a standard deviation of 30 nA, respectively. Without noise canceling, the mean value for the input referred noise is 302 nA with 60 nA standard deviation [20].

6.5.7 Inverter Based Cascode TIA

To obtain a high sensitivity despite the high input-node capacitance, the transconductance of the input transistors and therefore the current through them has to be maximized for minimum noise. TIAs which use CMOS inverters were studied extensively in the past years. The advantage is the use of N- and P-MOSFETs at the input is to achieve higher gain due to larger effective transconductance [23]. The inverter structure shows a better noise behavior than a simple common-source amplifier (CS). The input referred noise of the amplifier, which mainly consists of the thermal drain current noise of the input transistors, is divided by the sum of the transconductances of the input transistors (of N- and P-MOSFETs) in the case of the inverter, whereas the drain current noise of the CS amplifier, consisting of the thermal noise of the transistor and the resistor noise, is only divided by the transconductance of the single transistor [24].

To achieve a high amplifier gain this requires devices that have large size to provide a sufficient transconductance. Unfortunately, large parasitic capacitance is also associated with large devices. The gate-source capacitances of the devices add directly to the input capacitance. Also the gate–drain capacitance appears directly across the feedback resistor and due to the Miller effect; this capacitance can be the limiting factor with regards to bandwidth and can dominate over the photodiode capacitance [25].

A one-stage TIA has a too low gain; also, the already noted disadvantage of the Miller capacitance occurs. Using more than one amplifier stage, the circuits become more unstable and difficult to use [23, 26]. A system with feedback is stable if the second pole's frequency is more than 2 times larger than the gain-bandwidth product which requires a large safety margin in the design. This results in circuits with either limited speed or limited transimpedance gain.

The open-loop gain is lowered by the diode-connected MOSFETs to ensure stability [16, 18]. The process dependent ratio of the mobilities appears only as a square root. The optimization of transistor sizing can limit the effect of the mobility variation [18].

The disadvantages of this topology are that the additional diode-connected MOSFETs are always ON and always lower the gain of the amplifier. This increases the input referred noise due to lowering the gain and adding their noise.

The diode-connected MOSFETs are modified and MOSFET switches are added in [27]. At low input optical power the switches are OFF, so the diode-connected MOSFETs are disconnected from the amplifier. The amplifier gain is higher and no additional noise comes from diode-connected MOSFETs. At higher input optical power the control signal has to be increased to connect the diode-connected MOSFETs to the amplifier. This reduces the open-loop gain to ensure stability.

A voltage gain stage employing the cascode configuration can be used to boost the voltage-gain with no further degradation of bandwidth from the parasitic capacitance [28].

A folded cascode amplifier was introduced in [26]. The principle of this circuit has been known from operational amplifiers. Such a circuit has the advantage to work at low voltages and it has the advantage of a cascode stage increasing the bandwidth. Therefore, it is suitable for high-speed CMOS circuits at low supply voltages.

A three stage inverter based TIA where only one of the two transistors in the first stage was cascoded was introduced [29]. This amplifier benefits from the high output resistance of the cascode and the sum of the inverter MOSFETs' transconductance, resulting in a higher gain. The TIA in [29] has a bandwidth of 280 MHz, maximum transimpedance gain of 2.5 kΩ, 85 dB dynamic range, 220 nA equivalent input noise current and a high maximum input current (4 mA) overdrive capability for use with a monitor photodiode.

A three-stage inverter-based TIA with a photodiode and pad capacitance of 60fF, and a photodiode responsivity of 0.7 A/W, is designed for a transimpedance gain of 4 kΩ, a 7 GHz bandwidth, and is sized to achieve a BER of 10^{-12} [30].

A TIA employing an inverter based cascode can achieve a higher bandwidth at lower power consumption than with the normal CS-TIA, and can achieve a better sensitivity than with the conventional inverter based TIA.

Figure 6.26a shows the circuitry of the conventional inverter based TIA. From the small-signal model the transimpedance gain is given by:

$$Z_{T,inv}(0) \approx \frac{A_{inv} \cdot {R_{FB}}^2}{\left(R_{FB} + r_{ds,p1}//r_{ds,n1}\right)\left(A_{in`v} + 1\right)} \tag{6.109}$$

where

$$A_{inv} = (g_{m,n1} + g_{m,n2}).r_{ds,p1}//r_{ds,n1} \tag{6.110}$$

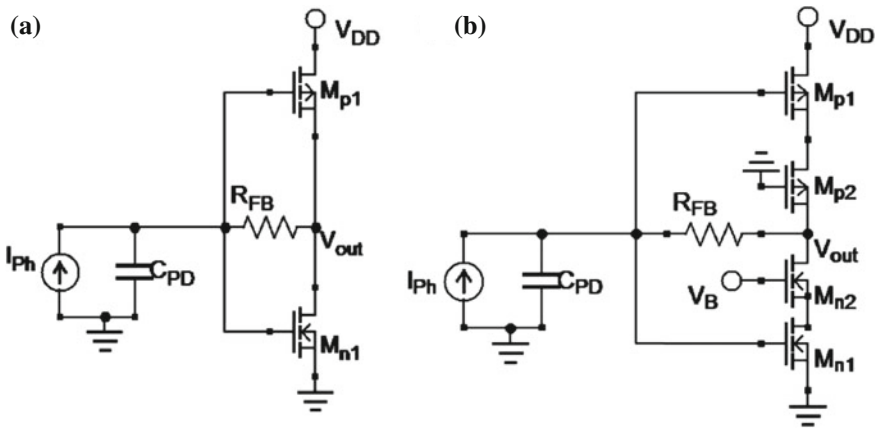

Fig. 6.26 **a** Conventional inverter based TIA circuitry and **b** circuitry of the proposed Inv-Cascode-TIA with CS post amplifier

$$BW_{inv} \approx \frac{1 + A_{inv}}{2\pi R_{FB}(C_{PD} + C_{PAD} + C_{gs,n1,p1} + C_{gd,n1,p1} \cdot (1 + A_{inv}))} \tag{6.111}$$

The main advantage of the inverter based TIA is its higher gain A_{inv} compared to the TIA with CS amplifier. Figure 6.26b shows the circuitry of the proposed Inv-Cascode-TIA where two transistors NMOS and PMOS (M_{n2} and M_{p2}) are added in series with the inverter transistors M_{n1} and M_{p1}, respectively where V_B is selected to be 0.9 V for optimum performance.

The transimpedance gain is given by:

$$Z_{T,invC}(0) \approx \frac{A_{invC} \cdot R_{FB}}{(A_{invC} + 1)} \tag{6.112}$$

The voltage gain of the cascode inverter becomes:

$$\begin{aligned} A_{invC} = {} & (g_{m,n1} + g_{m,n2}) \times (r_{o1} // r_{o2}) = \left(g_{m,n1} + g_{m,n2}\right) \\ & \cdot \left[\left(g_{m,n2} \cdot r_{ds,n2} \cdot r_{ds,n1}\right) // \left(g_{m,p2} \cdot r_{ds,p2} \cdot r_{ds,p1}\right) \right] \end{aligned} \tag{6.113}$$

$$BW_{invC} \approx \frac{1 + A_{invC}}{2\pi R_{FB}\left(C_{PD} + C_{PAD} + C_{gs,n1,p1} + C_{gd,n1} + C_{gd,p1}\right)} \tag{6.114}$$

Due to the high output resistance of the cascode implementation the voltage gain of the cascode structure is higher than that of the regular inverter, compare (6.110) and (6.113). The transimpedance gain of the Inv-Cascode TIA $Z_{T,invC}$, (6.112) is a little bit higher than the transimpedance gain $Z_{T,inv}$ in (6.109) because of its higher voltage gain.

The bandwidth of the inverter based cacode TIA BW_{invC} in (6.114) is much higher than the BW_{inv} introduced in (6.111) because of two reasons. The first is the higher cascode voltage gain. The second is the Miller capacitance effect for C_{gd} which is amplified by A_{inv} at the input node, see (6.111).

A comparison by simulation was performed to confirm the better performance of the Inv-Cascode-TIA compared to the inverter based TIA. Both inverter based TIA and Inv-Cascode-TIA were implemented in the 40 nm CMOS technology and the transistor sizing was optimized to reach the maximum achievable bandwidth at using the same feedback resistor $R_{FB} = 635\ \Omega$. The maximum achievable bandwidth for the inverter based TIA is 4.2 GHz and transimpedance gain is 510 Ω at 2.1 mA biasing current. Whereas the Inv-Cascode-TIA achieves 8 GHz bandwidth and 585 Ω transimpedance gain at 2.51 mA biasing current.

The inverter based cascode TIA has a higher gain and a lower input capacitance compared to the conventional inverter based TIA. The drawback of the cascode configuration is that it requires a higher supply voltage to maintain the required gain and bandwidth, and there will be a small noise contribution coming from the cascode transistor.

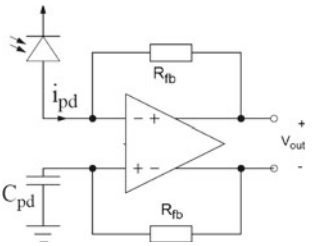

Fig. 6.27 Block diagram of the differential TIA topology

6.6 Differential TIA

All TIAs presented up to here in this book are single-ended. The single ended topology is convenient since usually only one photodiode is connected. A differential TIA has improved immunity to power-supply and substrate noise as well as a higher voltage swing compared to single-ended TIA. A differential TIA typically has a single-ended input and differential outputs, as shown in Fig. 6.27. Since the input signal is only single ended, for a fully differential amplifier the second input must be connected properly. The first possibility is to connect the second input to a dummy photodiode which has the same structure and area like the active photodiode. This makes the differential inputs fully balanced. This makes the circuit fully symmetrical which means a significantly optimized PSRR. The power supply and the substrate noise are coupled equally to both inputs of the differential TIA and thus these noises are suppressed as a common-mode signal. Another way is to connect the capacitor C_{PD} having the same capacitance as the photodiode.

On the other hand, the differential TIA has the same transimpedance gain like single-ended TIAs, i.e. nearly R_{fb}, but the area and power consumption are doubled. Another significant disadvantage is the noise performance; both feedback resistors and input transistors contribute to the noise current, due to the small capacitance value C_{PD}. This means, that the input noise current for this circuit is actually by a factor $\sqrt{2}$ higher compared to a single-ended TIA.

We can summarize that a differential TIA has larger input-referred noise, higher power consumption, and needs more silicon area. On the other hand, a differential TIA has improved immunity to power-supply and substrate noise as well as a higher voltage swing compared to single-ended TIAs. So, differential TlAs are needed in noisy environments where the power supply and substrate noise are strong. On-chip mixed-signal systems where digital noise can be coupled to front-end low-noise analog circuits (TIA) is an example where differential TIAs are preferred. Differential TIAs give a high immunity against common-mode and supply noise coming from high levels of electromagnetic interference in cars. Immunity against common-mode noise is also required in case a charge pump is used to increase the PD's reverse biasing voltage [31]. Also in low-voltage systems a differential output signal is needed to provide a larger dynamic range compared to the single ended one. A differential

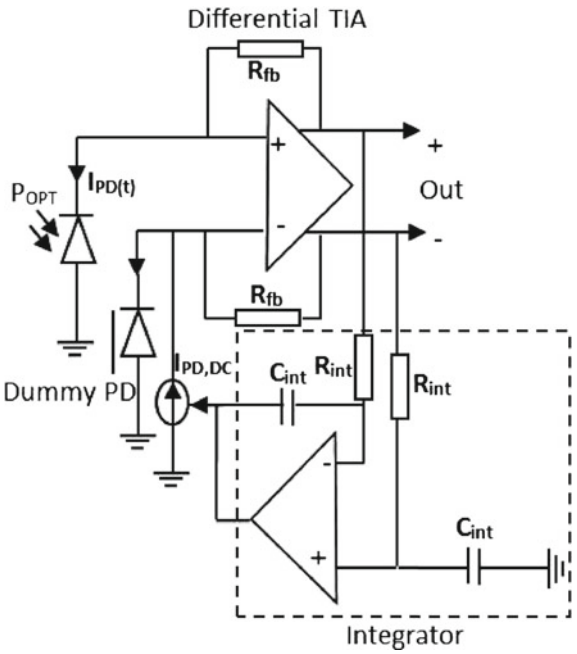

Fig. 6.28 Block diagram of the differential TIA using dummy photodiode

TIA can be directly connected to a differential post amplifier; there is no need for an additional single-ended to differential conversion stage [2].

The optical signal, i.e. the photocurrent, always has a positive value. To have a fully balanced output, a current equal to the average photocurrent generated in the active PD should be injected in the dummy PD, see Fig. 6.28. The control loop injects a current into the dark side of the TIA equal to the average photocurrent. The current source is steered by an integrator, which senses the differential output voltage of the TIA. Therefore the common-mode differential voltage at the TIA output is held at zero [32].

6.7 TIA with Gain Control

The dynamic range of a TIA is defined by its overload current and its sensitivity. For a shunt-feedback TIA, both quantities are related to the value of the feedback resistor, and thus the dynamic range of the optical input power can be extended by making this resistor variable with the input optical signal power, Fig. 6.29. The variable feedback resistor can be implemented with a MOSFET operating in the linear region, connected in parallel to a fixed resistor to improve the linearity and to limit the maximum resistance (Fig. 6.29). The AGC output V_c is a DC signal with

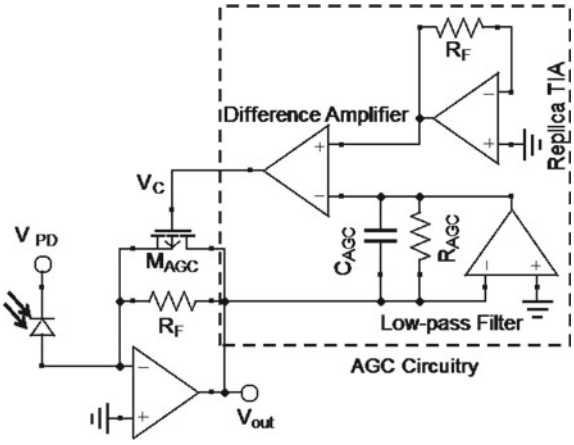

Fig. 6.29 Transimpedance amplifier with gain control [1]

a value depending on the input signal. The automatic gain control (AGC) can be implemented with a circuit that determines the output signal strength (peak detector or average power) to control the gate voltage of the MOSFET (a low pass filter will make the averaging process). The replica TIA will generate a reference voltage for the difference amplifier. The difference amplifier subtracts the reference voltage from the averaged signal to give the control signal V_c [1].

The stability of the AGC-TIA is an important issue. When the feedback resistance (R_F) varies while keeping the open-loop gain fixed, both the bandwidth and the quality factor will change. If R_F is reduced, the frequency of the first open-loop pole at the input $1/(R_{fb}.C_T)$ increases and it will come closer to the second pole at the output $1/(R_{out}.C_{out})$, which may lead to peaking at a given fixed loop gain (see Sect. 6.5.1 and Fig. 6.15). The bandwidth of the AGC-TIA tends to increase with the magnitude of the input signal. The low-cut off frequency is determined by the integration time of the integrator (low pass filter bandwidth) $1/(2\pi(R_{AGC}C_{AGC}))$. The low-pass filter bandwidth should be low as tens of kHz to meet the optical communications standards. So, the capacitor C_{AGC} and the resistor R_{AGC} should have a high value which is limited by the chip area.

6.8 TIA with Gain Compression

The optical receiver used to monitor the optical power of a laser diode or the sensor circuit in the pulsed time-of-flight (TOF) based laser ranging system experience high amplitude variation of the monitored laser power/reflected optical pulse. The received optical power depends on the laser optical power, the measurement distance, reflectivity, and an angle of the reflection. The dynamic range of the monitored/reflected

signal depends on the application, and may be as high as one to a million or even more. Such a kind of the high dynamic range optical signal can easily saturate conventional receivers based on a linear TIA, thus making it inefficient. An optical receiver overcomes this problem by using a TIA with an input current compression technique to increase the dynamic range. For small photocurrents below the break point current $I_{in,B}$, the output voltage rises linearly with the small input current, see Fig. 6.30. As the input current increases further above $I_{in,B}$, the output voltage is compressed and rises very slowly with the input current. This causes the output-voltage pulse width to appear wider in the compression region, see Fig. 6.31. The effect of the pulse-width variation is more apparent by pulses with short rise/fall times [33].

A TIA circuit based on a square root compression is shown in Fig. 6.32. The TIA consists of a fixed feedback resistor R_{fb} across a three-stage voltage amplifier: inverter (M_{1n}, M_{1p} and M_2 cascode) at the input and two common-source stages (M_5, R_l, and M_7, R_2). A cascode inverter is used as first stage as it gives a higher gain than the CS stage and the conventional inverter, see Sect. 6.5.7. The open loop voltage gain A_0 is further increased by two resistive load common-source amplifiers (M_5, R_l, M_7, and R_2) [34]. When the TIA is operating in the linear region ($I_{in} < I_{in,B}$), the output voltage V_{out} is linearly proportional to the input current through the fixed feedback resistor R_{fb}:

$$v_{out} \approx i_{in} R_{fb} + v_{gs,M1n} \tag{6.115}$$

The TIA operates in the linear region as long as transistor M_C is off. As the input current increases the drain voltage of M_2, transistor M_C becomes (more) conductive, thus creating a parallel current path to the feedback resistor R_{fb}. This parallel path

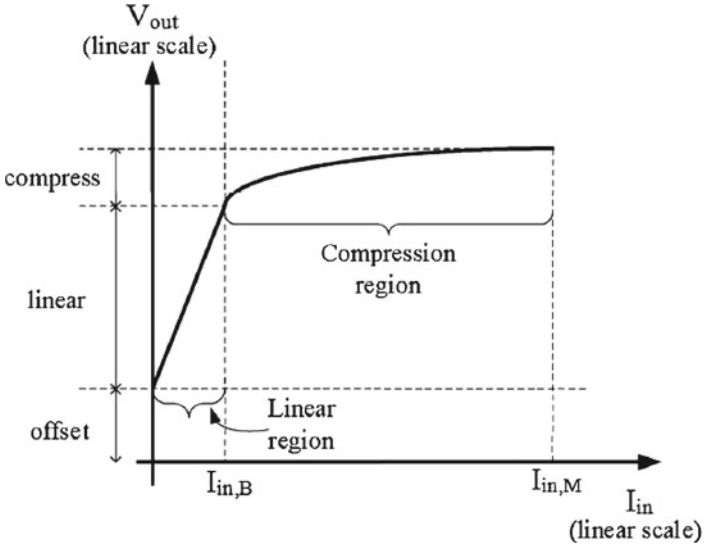

Fig. 6.30 Compression TIA principle of operation [34]

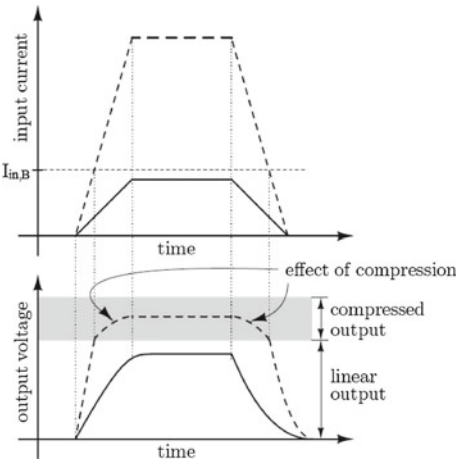

Fig. 6.31 Effect of the pulse width variation caused by the input current compression [33]

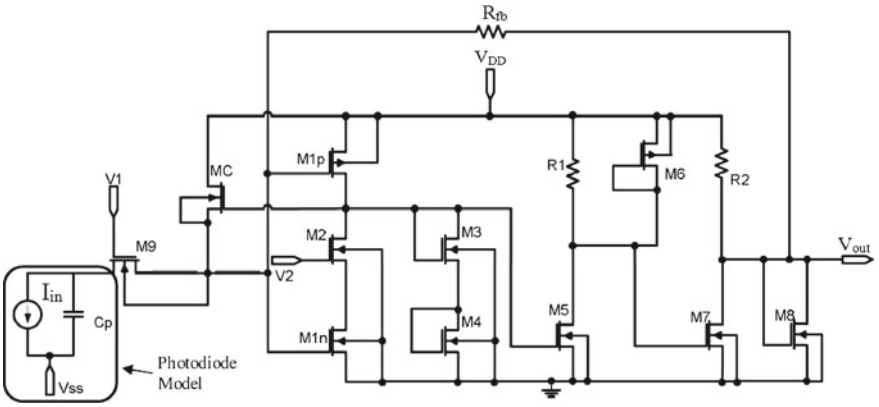

Fig. 6.32 Square-low compression TIA circuitry [34]

lowers the overall transimpedance gain of the TIA significantly in response to the increasing input current signal preventing the TIA from entering into saturation. The output voltage in the compression region is given by [34]:

$$v_{out} \approx \sqrt{\frac{2\, i_{in}\, L_{MC}}{K\, W_{MC}}} + v_{gs,M1n} \tag{6.116}$$

Important is that the phase margin of the loop gain reduces with the input current. This reduction causes oscillations and a potential stability problem at some input-current pulses. This is the reason for the need of the dynamic compensation, which reacts to the input current and maintains the phase margin of the loop gain. In response

to the input current, the dynamic compensation rapidly reduces the dc open-loop gain A_0, and moves the non-dominant pole frequency towards higher values. The dynamic compensation is active only in the compression region and does not affect the behavior in the linear region. Diode connected transistors M_3, M_4, M_6 and M_8 create a compensation circuit, which is necessary for stability over the full range of the input current. The compensation circuit lowers the output resistance of each stage, resulting in a lower overall open-loop gain and hence maintains the stability of the TIA at high input currents. The compensation circuit works only in the compression region [34].

6.9 Bandwidth Enhancement Techniques for TIAs

The transimpedance amplifier bandwidth can be extended by decreasing the TIA input impedance or by introducing a peaking to compensate for the undesired parasitics or input capacitances.

The CG, RGC, ICDF TIAs, and super-Gm TIA can extend the bandwidth by introducing a small input impedance which can isolate the effect of a large input capacitance compared to the conventional CS TIA with shunt-shunt feedback resistor. In feedback TIAs with shunt-shunt resistor the input impedance can be reduced by increasing the amplifier's voltage gain. An inverter based TIA increases the transconductance and hence the voltage gain. The inverter based cascode TIA can further increase the amplifier's voltage gain and also cancel the Miller capacitance's effect, and thus increases the TIA bandwidth. As most of the above techniques were studied in the previous sections, only the super Gm-TIA will be discussed here.

The second approach of introducing a zero (peaking) to cancel the effect of the amplifier's pole can be realized by several methods. The inductive peaking, capacitive source degeneration, and negative capacitance are examples of these techniques.

6.9.1 Super-Gm

In order to provide a low input impedance with low DC current consumption and high sensitivity, a super-Gm TIA (SGm-TIA) can be used, shown in Fig. 6.33 [35]. M_1, M_2 and R_1, R_2 form a cross-coupled CG auxiliary amplifier that drives the gates of M_3, M_4. C_1, C_2 AC-couple the input signals to the gates of M_1, M_2, doubling their g_m at high frequencies. R_5through R_8 are used to set the gate bias voltages for M_1, M_2. C_D is a decoupling capacitor. M_3, M_4 and R_3, R_4 are CG amplifiers which drive the gates of M5, M6, making the input transistors M5, M6 absorb most of the input signal currents. The g_m of M_5, M_6 is increased by twice the squared voltage gains compared to without using the auxiliary amplifier. It should be noted that the final proposed differential TIA has four individual single-ended signal paths that all

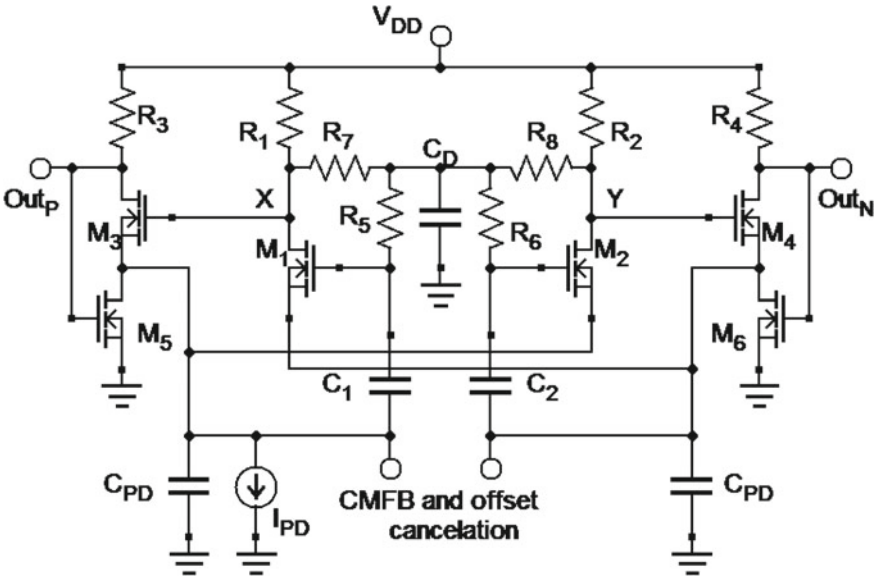

Fig. 6.33 Super-Gm transimpedance amplifier circuitry [35]

enhance the effective g_m seen at the input nodes; so it is called super-Gm TIA. A common-mode feedback (CMFB) and an offset-cancellation are needed [35].

The high-frequency differential input impedance can be written as [35]:

$$Z_{in}(S) = \frac{(1 + C_x R_1 S)\left(1 + C_{outp} R_3 S\right)}{(1 + 2g_{m1}R_1)\, g_{m3} R_3 g_{m5} + (1 + 2g_{m1}R_1)\, g_{m3} + 2g_{m1}} \tag{6.117}$$

$$Z_{in}(0) = \frac{1}{(1 + 2g_{m1}R_1)\, g_{m3} R_3 g_{m5} + (1 + 2g_{m1}R_1)\, g_{m3} + 2g_{m1}} \tag{6.118}$$

Compared to the low frequency input impedance of the RGC (6.49) the super-Gm low-frequency input impedance (6.118) is much lower. More terms are added to the dominator which reduces the input impedance.

The transimpedance gain of the proposed SGm-TIA shown in Fig. 6.33, can be derived as:

$$A_T = \frac{V_{outn} - V_{outp}}{I_{PD}} \approx \frac{R_3}{1 + g_{m5} R_3} \cdot \frac{1}{1 + Z_{in}(S) C_{PD} S} \tag{6.119}$$

A trade-off can be made between power consumption, stability, bandwidth, and noise performance to get the optimum performance [35].

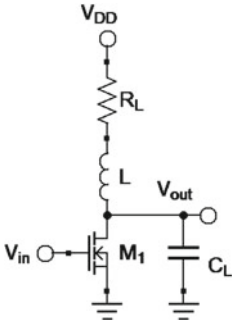

Fig. 6.34 Shunt inductive peaking

6.9.2 Inductive Peaking

Inductive peaking is an efficient way for bandwidth extension. For a certain technology, increasing the amplifier bandwidth requires a higher power dissipation. The advantage of the shunt peaking technique is that the bandwidth can be increased without additional power dissipation.

Shunt inductive peaking uses an inductor in series with the load resistor (the total branch being in shunt with the output node), see Fig. 6.34. This will add a zero in the voltage gain transfer function to maintain a constant gain over a wider frequency range by reducing the effect of the pole caused by a load capacitance at the output [1]:

$$A_V = \frac{-g_m R_L \left(1 + \frac{L}{R_L} S\right)}{LC_L S^2 + R_L C_L S + 1} \tag{6.120}$$

This strategy helps to minimize the effect of the capacitive load at the amplifier's output node and therefore extends the amplifier's bandwidth.

The parasitic capacitance and Eddy currents in the substrate affect the inductor's intrinsic resonant frequency and its quality factor. Also, large inductors tend to consume a large chip area. Fortunately, the peaking inductors do not require a high quality factor, and thus can be optimized for low parasitic capacitances and small area [28]. The inductor's serial resistance can be considered as a part of the amplifier's resistive load during the design. So the inductor's turns, width, and spacing can be minimized to permit the desired inductance to be realized while minimizing the spiral's area and capacitance.

The relationship between the amount of peaking (m) and the required inductance (L) for the simple shunt-peaked common-source amplifier (Fig. 6.34) is given by [36, 37]:

$$m = \frac{\frac{L}{R_L}}{R_L C_L} \tag{6.121}$$

The 3 dB bandwidth increases as m increases. The maximum bandwidth is obtained when $m = 0.71$ and yields an extended bandwidth of 1.85 times the bandwidth without peaking. A maximally flat response may be obtained for $m = 0.41$ with a still impressive bandwidth being 1.72 times the old bandwidth. The optimum group delay case occurs when $m = 0.32$ giving a new bandwidth of 1.6 times the bandwidth without shunt peaking [36].

For the given load resistor R_L and the amount of peaking m, the smaller the original bandwidth (or the higher $R_L C_L$) the higher the required peaking inductance value is. Thus for amplifiers with small bandwidth (less than 5 GHz), the necessary inductance value has to be very large unless R_L is very small. For bandwidths less than 5 GHz the required inductance from (6.121) has to be tens of nH. Such an inductor is generally impractical as an on-chip implementation because of the required large area. As a result the conventional inductive shunt peaking technique tends to be ineffective below 5 GHz unless R_L is small [37]. Unfortunately, to achieve enough gain R_L should have a large value; due to the high conductance (small output resistance) of the nanometer CMOS technology.

As a modification of shunt-peaking a new inductance amplification was introduced in [37]. Figure 6.35 shows a modification of the conventional shunt-peaking topology using a load with inductive source degeneration. The new equivalent inductance can make the shunt peaking effective at frequencies below 5 GHz using a smaller inductor. The output impedance Z_{out} of the load is given by [37]:

$$Z_{out} = r_o + L_{new} S \ , where \ L_{new} = L\,(1 + G_m r_o) \tag{6.122}$$

Where r_o is the output resistance of transistor Mp, and G_m is the overall transconductance of the cascode amplifier stage. Comparing (6.122) with (6.121), for the same

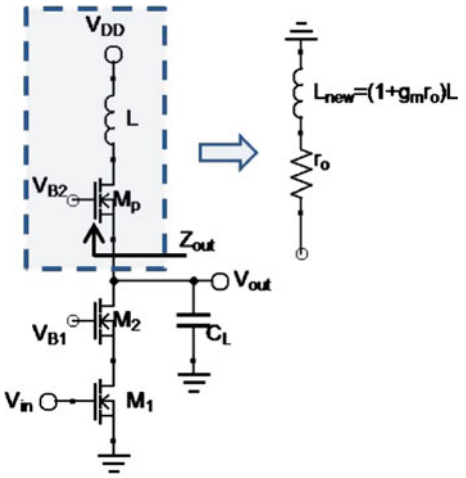

Fig. 6.35 Shunt peaked cascode amplifier with the inductance amplification technique [37]

amount of peaking and assuming r_o and C_L are equal to R_L and C_L, respectively, a much smaller sized inductor is required with the shunt peaked amplifier cascode shown in Fig. 6.35 in comparison with the conventional shunt-peaked amplifier in Fig. 6.34.

The parasitic gate-source capacitance of M_P decreases the inductor's effective value. More accurately, including the gate-source capacitance $C_{gs,MP}$ of the active load, the output impedance can be expressed as [37]:

$$Z_{out} = r_o + \frac{L\,(1 + G_m r_o)\,S}{1 + LC_{gs,MP}S} \tag{6.123}$$

Therefore, the sizing of M_P and the biasing voltages should be selected with care so that the $C_{gs,MP}$ capacitance does not turn the load impedance into a capacitive load at the frequency of interest.

6.9.3 Active Inductive Peaking

In the inductor peaking approach the spiral inductor occupies a large chip area making it unsuitable for cheap optical receivers. Also the substrate coupling increases through the inductors, resulting in higher crosstalk. Another technique is the capac-

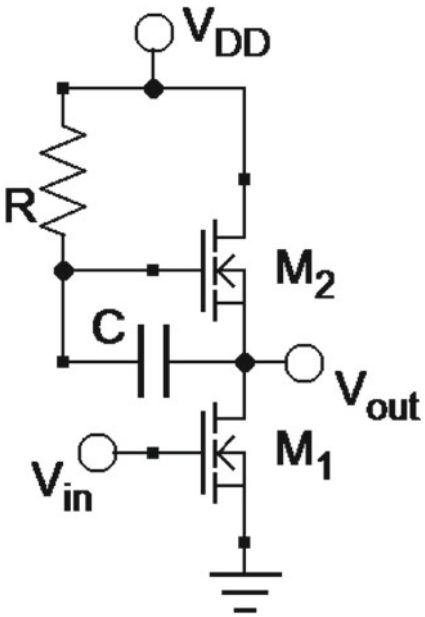

Fig. 6.36 Active inductor circuitry

itive degeneration peaking, but it suffers from lower DC gain. The inductor can be replaced by an active one having the same behavior by means of devices that are easier to integrate. The active inductor occupies much less area compared to an integrated spiral inductor and hence has a smaller parasitic substrate capacitance, thus a higher bandwidth is expected.

A common-source amplifier with active inductor is shown in Fig. 6.36. The voltage gain transfer function of the circuit in Fig. 6.36 is given by the following relation [1, 38]:

$$A_v = g_{m1} Z_{out} \ , where \ Z_{out} = \frac{1 + RCS}{g_{m2}\left(1 + \frac{C}{g_{m2}} S\right)} \tag{6.124}$$

For low frequency the output impedance Z(0)= $1/g_{m2}$ and the high frequency output impedance $Z(\infty) = R$. There is an inductive peaking when R > $1/g_{m2}$ [1].

For operating frequencies less than $\omega_p = g_{m2}/C$, the pole can be neglected and the output impedance can be approximated by [38]:

$$Z_{out} \approx \frac{1}{g_{m2}} + \frac{RC}{g_{m2}} S \tag{6.125}$$

Compared to (6.122) $r_o = 1/g_{m2}$ and $L_{new} = RC/g_{m2} = R/\omega_p$ for a small value of the pole frequency a high inductance can be obtained. On the other hand, this will limit the maximum achievable bandwidth. At small values of the pole frequency, however, the effect of the pole cannot be neglected and the input impedance will be capacitive and not inductive. The active inductive load needs a higher voltage headroom (due to $V_{gs,M2}$) compared to the passive inductive peaking introduced in Fig. 6.34.

6.9.4 Negative Capacitance

A negative capacitance can be used to compensate an undesired parasitic capacitance of the TIA to enhance its bandwidth. Figure 6.37 shows the schematic of the core amplifier of the TIA based on a negative capacitance (NC) technique. High feedback resistor values and amplifier gains are required to achieve high transimpedance gain and high sensitivity. However, the high feedback resistor value and the high amplifier gain will reduce the bandwidth of the optical receiver. For maintaining high sensitivity without sacrificing the TIA bandwidth; the negative capacitance is adopted as shown in Fig. 6.37. The negative capacitance induces the peaking effect, which enhances the optical receiver's bandwidth [39].

A negative impedance converter (NIC) consisting of M_3 and M_4 in Fig. 6.37 transforms Cc to a negative capacitance between the drains of $M_3$and M_4. If the gate-drain capacitance of M_3 and M_4 is neglected, the impedance seen looking into the drains of M_3 and M_4 is expressed as [40]:

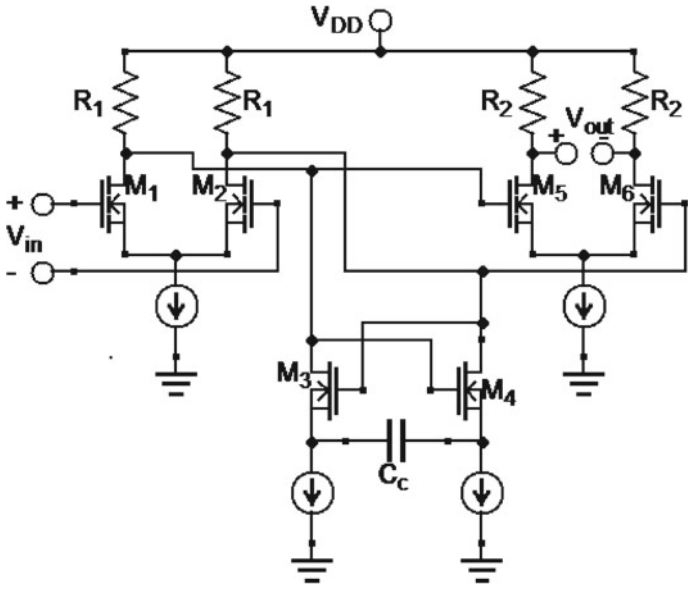

Fig. 6.37 The schematic of the core amplifier of the TIA based on negative capacitance (NC) [39]

$$Z_{INC} = -\frac{1 + \frac{(C_{gs}+2C_c)}{g_m} S}{\left(1 - \frac{C_{gs}}{g_m} S\right) C_c S} \tag{6.126}$$

For frequencies lower than g_m/C_{gs}, Z_{INC} can be approximated to:

$$Z_{INC} = -\frac{1}{C_{cS}} - \left(\frac{C_{gs}}{C_c} + 2\right) \frac{1}{g_m} \tag{6.127}$$

The impedance is equivalent to a negative capacitance C_c in series with a negative resistance $(C_{gs}/C_c + 2)/g_m$. The circuit stability can be reduced due to the NC technique, and a careful selection of the C_c value is necessary to ensure stability. The upper limit for the value of C_c is that which places the circuit at the edge of oscillation. For practical designs C_c must be far below this upper limit to ensure minimal ringing and ISI [40].

References

1. B. Razavi, *Design of Integrated Circuits for Optical Communications* (McGraw-Hill, New York, 2003)
2. E. Säckinger, *Broadband Circuits for Optical Fiber Communication* (Wiley, New Jersey, 2005)
3. A.-J. Annema, B. Nauta, R.V. Langevelde, H. Tuinhout, Analog circuits in ultra-deep-submicron CMOS. IEEE J. Solid-State Circuits **40**(1), 132–143 (2005)

4. S. Voinigescu et al., Circuits and technologies for highly integrated optical network IC-s at 10 Gb/s to 40 Gb/s, in *Custom Integrated Circircuit Conference*, pp. 331–338 (2001)
5. P. Muller, Y. Leblebici, *CMOS Multichannel Single-Chip Receivers for Multi-Gigabit Optical Data Communications* (Springer, Netherlands, 2007)
6. C. Hermans, M. Steyaert, *Broadband Opto-Electrical Receivers in Standard CMOS* (Springer, Netherlands, 2007)
7. M. Ingels, M. Steyaert, *Integrated CMOS Circuits for Optical Communications* (Springer, New York, 2004)
8. S.M. Park, H. Yoo, 1.25-Gb/s regulated cascode CMOS transimpedance amplifier for gigabit ethernet applications. IEEE J. Solid-State Circuits **39**(1), 112–121 (2004)
9. C. Chan, O.T. Chen, Inductor-less 10Gb/s CMOS transimpedance amplifier using source-follower regulated cascode and double three-order active feedback, in *Proceedings—IEEE International Symposium on Circuits and Systems* pp. 5487–5490 (2006)
10. H. Chen, C. Chen, W. Yang, J. Chiang, Inductorless CMOS receiver front-end circuits for 10-Gb/s optical communications. Tamkang J. Sci. Eng. **12**(4), 449–458 (2009)
11. K. Park, W.S. Oh, B. Choi, J. Han, S.M. Park, A 4-Channel 12.5Gb/s common-gate transimpedance amplifier array for DVI/HDMI applications, in *Proceedings—IEEE International Symposium on Circuits and Systems* pp. 2192–2195 (2007)
12. J. Borremans, P. Wambacq, C. Soens, Y. Rolain, M. Kuijk, Low-area active-feedback low-noise amplifier design in scaled digital CMOS. IEEE J. Solid-State Circuits **43**, 2422–2433 (2008)
13. E. Säckinger, The transimpedance limit. IEEE Trans. Circuits Syst. I(TCAS I) **57**, 1848–1856 (2010)
14. M. Atef, H. Zimmermann, 10Gbit/s 2mW inductorless transimpedance amplifier, IEEE Int. Symp. Circuits Syst. (ISCAS), Seoul, Korea (South), pp. 1728–1731 (2012)
15. H. Zimmermann, *Integrated Silicon Optoelectronics*, second edn. (Springer, Berlin, 2010)
16. K. Schneider, H. Zimmermann, Highly sensitive wide-dynamic range optical burst-mode receivers for ultra-fast gain switching. Analog Integr. Circuit Signal Process. **49**, 141–149 (2006)
17. K. Schneider, H. Zimmermann, *Highly Sensitive Optical Receivers* (Springer, Berlin, 2006)
18. M. Ingels, G. Van Der Plas, J. Crols, M. Steyaert, A CMOS 18 THz Ω 248 Mb/s transimpedance amplifier and 155 Mb/s LED-driver for low cost optical fiber links. IEEE J. Solid-State Circuits **29**, 1552–1559 (1994)
19. M. Atef, F. Aznar, S. Schid, A. Polzer, W. Gaberl, H. Zimmermann, 8 Gbit/s inductorless transimpedance amplifier in 90 nm CMOS technology. Analog Integr. Circuit Signal Process. **79**(1), 27–36 (2014)
20. M. Atef, H. Zimmermann, 2.5 Gbit/s transimpedance amplifier using noise cancelling for optical receivers, in *IEEE International Symposium on Circuits Systems (ISCAS)*, Seoul, Korea (South), pp. 1740–1743 (2012)
21. F. Bruccoleri, E. Klumperink, B. Nauta, Wide-band cmos low-noise amplifier exploiting thermal noise canceling. IEEE J. Solid-State Circuits **39**, 275–282 (2004)
22. D.J.A. Groeneveld, Bandwidth extension and noise cancelling for TIAs, M.Sc. thesis, University of Twente (2010)
23. C. Kromer, G. Sialm, T. Morf, M.L. Schmatz, F. Ellinger, E. Daniel, H. Jackel, A low-power 20-GHz 52-dB transimpedance amplifier in 80 nm CMOS. IEEE J. Solid-State Circuits **39**(6), 885–894 (2004)
24. K. Schneider, H. Zimmermann, A. Wiesbauer, Optical receiver in deep-sub-micrometre CMOS with −28.2 dBm Sensitivity at 1.25 Gbit/s. Electron. Lett. **40**(4), 262–263 (2004)
25. A. Vilches, R. Loga, M. Rahal, K. Fobelets, C. Papavassiliou, T.J. Hall, Monolithic large-signal transimpedance amplifier for use in multi-gigabit, short-range optoelectronic interconnect applications. IEEE Trans. Circuits Syst. II **52**(2), 102–106 (2005)
26. K. Schrödinger, J. Stimma, M. Mauthe, A fully integrated CMOS receiver front-end for optic gigabit ethernet. IEEE J. Solid-State Circuits **37**(7), 874–880 (2002)
27. F. Aznar, W. Gaberl, H. Zimmermann, A highly sensitive 2.5 gb/s transimpedance amplifier in cmos technology, in *IEEE International Symposium on Circuits and Systems(ISCAS 2009)*, Taipei, pp. 189–192 (May 2009)

28. J. Tak, H. Kim, J. Shin, J. Lee, J. Han, S.M. Park, A low-power wideband transimpedance amplifier in 0.13 m CMOS, in *IEEE International Microwave Workshop Series on Intelligent Radio for Future Personal Terminals (IMWS-IRFPT)*, Daejeon, pp. 1–2 (2011)
29. M. Hassan, H. Zimmermann, An 85 dB dynamic range transimpedance amplifier in 40 nm CMOS technology, NORCHIP, pp. 1–4 (2011)
30. F. Liu, D. Patil, J. Lexau, P. Amberg, M. Dayringer, J. Gainsley, H.F. Moghadam, X. Zheng, J.E. Cunningham, A.V. Krishnamoorthy, E. Alon, R. Ho, 10 Gbps, 530 fJ/b optical transceiver circuits in 40 nm CMOS, in *Symposium on VLSI Circuits (VLSIC)* pp. 290–291 (2011)
31. T. De Ridder, P. Ossieur, X. Yin, B. Baekelandt, C. Melange, J. Bauwelinck, X.Z. Qiu, J. Vandewege, BiCMOS variable gain transimpedance amplifier for automotive applications. Electron. Lett. **44**(4), 287–288 (2008)
32. R. Swoboda, M. Frtsch, H. Zimmermann, 3 Gbps-per-Channel Highly-Parallel Silicon Receiver OEIC, in *33rd European Conference and Ehxibition of Communication (ECOC)* pp. 1–2 (2007)
33. D. Micusik, Design of hybrid optical receiver with wide dynamic input range, Ph.D. thesis, Vienna University of Technology, 2008
34. M. Hassan, H. Zimmermann, An 85 dB dynamic range transimpedance amplifier in 40 nm CMOS technology, in *NORCHIP, Lund* pp. 1–4 (2011)
35. Y. Dong, K.W. Martin, A high-speed fully-integrated POF receiver with large-area photo detectors in 65 nm CMOS. IEEE J. Solid-State Circuits **47**(9), 2080–2092 (2012)
36. S.S. Mohan, M.D.M. Hershenson, S.P. Boyd, T.H. Lee, Bandwidth extension in CMOS with optimized on-cip inductors. IEEE J. Solid-State Circuits **35**, 346–355 (2000)
37. O. Yong-Hun, S.-G. Lee, An inductance enhancement technique and its application to a shunt–peaked 2.5 Gb/s transimpedance amplifier design. IEEE Trans. Circuits Syst. II **51**(11), 624–628 (2004)
38. M. Atef, D. Abd-elrahman, 2.5 Gbit/s compact transimpedance amplifier using active inductor in 130 nm CMOS technology, *The 21st International Conference in Mixed Design of Integrated Circuits Systems (MIXDES)* pp. 103–107 (June 2014)
39. J.-S. Youn, H.-S. Kang, M.-J. Lee, K.-Y. Park, W.-Y. Choi, High-speed CMOS integrated optical receiver with an avalanche photodetector. IEEE Photonics Technol. Lett. **21**(20), 1553–1555 (2009)
40. S. Galal, B. Razavi, 10-Gb/s limiting amplifier and laser/modulator driver in 0.18 μm CMOS technology. IEEE J. Solid-State Circuits **38**(12), 2138–2146 (2003)

Chapter 7
Equalizers

Most of the nanometer CMOS photodetectors presented in Chap. 5 have a small bandwidth due to a slow diffusion current. The photodiodes' bandwidth can be extended to several Giga hertz by applying equalization techniques like introduced in [1, 2]. The fiber channel can also limit the required data rate due to its low bandwidth, like for the step-index (SI) POF channel, see Sect. 2.1.3.1. A pre/post equalizer is needed to equalize the low-pass filter due to the dispersion effect of the multi-mode SI-POF as reported in [3, 4]. The term equalization can be used to describe any circuit or signal processing operation that minimizes ISI. The purpose of an equalizer is to reduce ISI as much as possible to minimize the probability of wrong decisions. The ideal response requires an equalizer that responds as the inverse of the channel. The ideal response is typically not practical; it needs many components, a large chip area and a complex design. The cost and power constraints are the limits for such equalizers. The equalization target is to reduce distortions and BER to acceptable levels. The equalizers can be classified as:

1. Linear equalizers and non-linear equalizers: if the equalizer does not have a feedback path, the equalization is linear. If there is a feedback path, the equalization is nonlinear like for the Decision Feedback Equalizer (DFE) and for the Maximum Likelihood Sequence Estimator (MLSE) [5]. The linear equalization techniques are easy to implement, but largely amplify the noise power because they work by inverting the hannel's frequency response. The non-linear equalization techniques are more difficult to implement, but have much less noise enlargement than linear equalizers. The linear equalization is faster than the nonlinear equalization but the nonlinear equalizers can achieve a better equalization on the expense of complexity and loss of speed.
2. Pre-equalizers and post-equalizers: the equalizer may be placed anywhere in the channel, at the transmitter side (pre-equalizer), at the receiver side (post-equalizer), or at both sides. Pre-equalization adjusts the magnitude of the transmitter output by adding certain peaking based on the channel response. This reduces the maximum swing, but produces an open eye.

M. Atef and H. Zimmermann, *Optoelectronic Circuits in Nanometer CMOS Technology*, Springer Series in Advanced Microelectronics 54,
DOI 10.1007/978-3-319-27338-9_7

3. Passive equalizers and active equalizers: The passive equalizer introduces only attenuation at certain frequencies. The active equalizers can attenuate or amplify certain frequencies to achieve the required equalization.
4. Fixed equalizers and adaptive equalizers: the fixed equalizer has a fixed coefficient suited for a time invariant channel. The adaptive equalizer's coefficients can be adapted to equalize for a time variant channel.

A combination between the above equalization types can be made to get the one which fits the required specifications best. The nanometer CMOS photodiode and the fiber channel have a limited bandwidth with a frequency response like a low-pass filter, so the required equalizer has to respond like a high-pass filter.

7.1 Passive Equalizer

A passive equalizer where the low frequency components are attenuated as in a passive R-C filter (see Fig. 7.1) is a high pass filter.

$$A(\omega) = \frac{Z_3}{R_1 + Z_2 + Z_3} \tag{7.1}$$

where

$$Z_2(\omega) = \frac{R_2}{1 + J\omega R_2 C_2}, Z_3(\omega) = \frac{R_3}{1 + J\omega R_3 C_3} \tag{7.2}$$

The branch $C_2//R_2$ introduces a zero to the equalizer's transfer function and the branch $C_3//R_3$ introduces a pole. The resistance R_2 introduces attenuation to the low frequency signal. By controlling the position of the zero and pole as well as the attenuation, the equalization can be done.

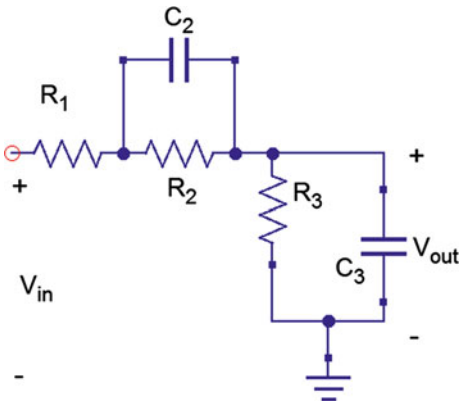

Fig. 7.1 Passive R-C equalizer

7.2 Active Equalizer

Passive filters attenuate the signal and they do not have appropriate input and output impedances that the active filters have. Active filters have an operating frequency range and power range where they can be used successfully to achieve better performance than passive filters. The low frequency components are attenuated and the high frequency components can be amplified by using an active high pass filter. In the next sections different types of active high-pass filters will be introduced being based on source degeneration and inductive load amplifiers.

7.3 Source Degeneration

A common-source amplifier with a degenerated source as an active high-pass filter is shown in Fig. 7.2 [6]. Low-frequency signals cannot pass through the branch consisting of R and C because of the high impedance value of C at low frequencies. The signal can pass through R at high frequencies because the capacitor C then will have a low impedance value.

The voltage gain transfer function of this amplifier has a high-pass filter characteristic:

$$A(\omega) = \frac{J\omega g_m R_L C}{g_m + J\omega C(g_m R + 1)} \tag{7.3}$$

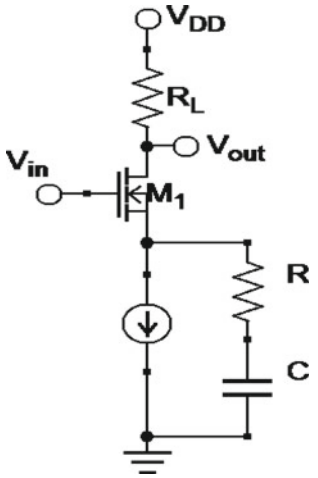

Fig. 7.2 Source degeneration high pass filter [6]

where

$$A(\infty) = \frac{g_m R_L}{1 + g_m R}, \omega_z = \frac{1}{R_L C}, \omega_p = \frac{g_m}{C(1 + g_m R)} \tag{7.4}$$

$A(\infty)$ is the high-frequency gain. ω_z (zero) and ω_p (pole) are the low and high corner frequency of the high-pass filter.

The drain-source resistance of the current source (simple current mirror) limits the maximum resistance in parallel with the R and C branch. A larger value of C is needed to reach the required corner frequency consuming more silicon area. In nanometer CMOS technologies the output resistance is smaller and this effect becomes severe. Also the PVT variations have a large effect on the deviations of the source degeneration equalizer corner frequencies from their nominal values.

7.4 Continuous Time Linear Equalizer (CTLE) with Multi-Shunt-Shunt Feedbacks

A CTLE with multi-shunt-shunt feedbacks is proposed in Fig. 7.3. the proposed CTLE uses multiple LPF feedback branches to produce a peaking with well-defined peaking frequency and rising slope. The CTLE includes an active feedback branch without any LPF to extend its high-frequency bandwidth. As frequency increases, the feedbacks via G_{m3} and G_{m4} are cut off which produces a gradual increase in the feedforward gain. The signal transfer function of the ideal CTLE shown in Fig. 7.3, ignoring the loading of R_1 and R_2, can be approximated as [7]:

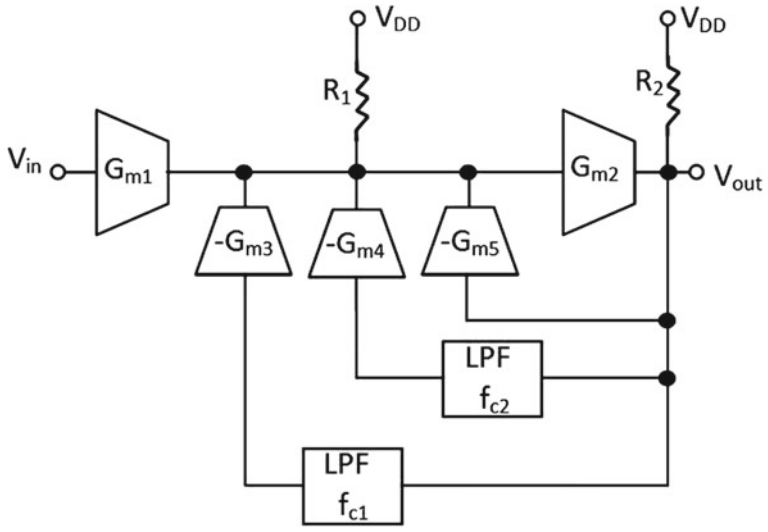

Fig. 7.3 Block diagram of the CTLE with multi-shunt-shunt feedbacks, [7]

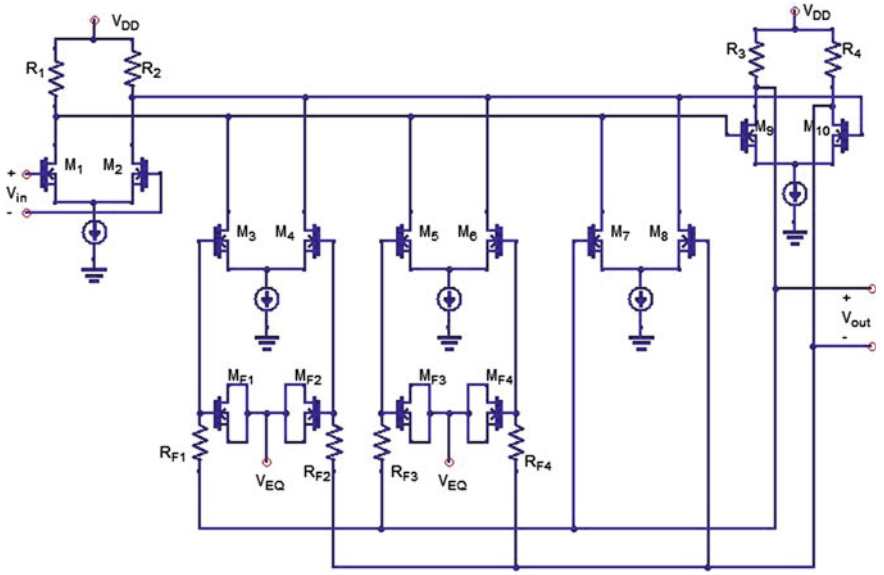

Fig. 7.4 Circuitry of the CTLE with multi-shunt-shunt feedbacks, [7]

$$A(\omega) = \frac{G_{m1}}{G_{m3}/\left(1 + \frac{S}{2\Pi f_{c1}}\right) + G_{m4}/\left(1 + \frac{S}{2\Pi f_{c2}}\right) + G_{m5}} \tag{7.5}$$

The starting frequency of the peaking can be changed by varying f_{C1} and f_{C2}, while the peaking slope can be adjusted by modifying the ratio between G_{m3}, G_{m4}, and G_{m5}. Figure 7.4 shows the transistor implementation of the multi-shunt-shunt feedback CTLE. G_{m3}, G_{m4}, and G_{m5} in Fig. 7.3 are implemented by M_3/M_4, M_5/M_6, and M_7/M_8. The LPFs inside the feedback branches are realized by poly-silicon resistors R_{F1} to R_{F4} and controlled NMOS capacitors $M_{F1}\ rmto\ M_{F4}$ [7]. The cut-off frequencies of the LPFs are controlled by V_{EQ}.

7.5 Inductive Load Equalizer

A common-source amplifier with an inductive load is shown in Fig. 7.5a. The equivalent impedance of the inductor L in shunt with the resistor R results in a low voltage gain at low frequencies and a high voltage gain at high frequencies.

The voltage gain transfer function is [9]:

$$A(\omega) = \frac{J\omega L g_m}{1 + J\omega \frac{L}{R}} \tag{7.6}$$

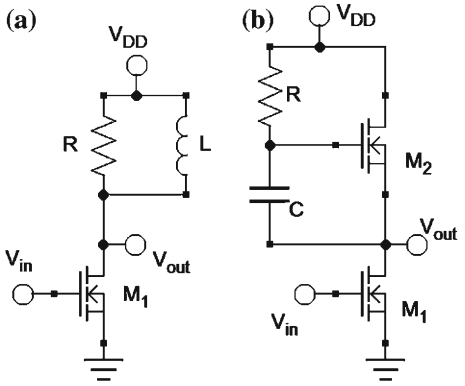

Fig. 7.5 Inductive load amplifier high-pass filter with **a** passive inductor **b** active inductor, [8]

where

$$A(\infty) = g_m R, \omega_p = \frac{R}{L} \tag{7.7}$$

where $A(\infty)$ is the high frequency gain and ω_p is the corner frequency of the high pass filter.

An integrated inductor occupies a large chip area, which increases the costs. It suffers from a low quality factor due to the presence of a highly doped substrate [10]. The inductor can be replaced by an active one having the same behavior by means of devices that are easier to integrate A common-source amplifier with active inductor is shown in Fig. 7.5b. The MOSFET M_2 makes a path for the low frequency signal where the R-C branch has high impedance due to C. At high frequencies the branch R-C creates a second signal path in parallel to M_2 and most of the signal current flows through the low impedance path R-C. At high frequencies, the load impedance is determined by R. If R is lower than $1/g_{m2}$ of M_2, the voltage gain at high frequencies is significantly higher than at low frequencies. The voltage gain transfer function of the circuit in Fig. 7.5b is given by the following relation [8]:

$$A(\omega) = \frac{g_{m1}(1 + J\omega RC)}{g_{m2}\left(1 + J\omega \frac{C}{g_{m2}}\right)} \tag{7.8}$$

where

$$A(0) = \frac{g_{m1}}{g_{m2}}, A(\infty) = g_{m1} R, \omega_p = \frac{g_{m2}}{C} \tag{7.9}$$

$A(0)$ is the low-frequency gain, $A(\infty)$ the high-frequency gain, and ω_p is the corner frequency of the high-pass filter. To have a high pass filter effect $A(\infty) > A(0)$ leading to $R_1 > \frac{1}{g_{m2}}$.

7.6 Adaptive Equalization

If the channel is frequency selective, the equalizer enhances the frequency components with small amplitudes and attenuates the strong frequencies in the received frequency response. For a time-varying channel, an adaptive equalizer is needed to track the channel variations. Ideally, tap coefficients are tuned to each system to account for operational (V, T) and manufacturing variation. This is done using adaptive algorithms. Perfect adaptation is practically not limited by things like update rate and coefficient resolution. A filter with N delay elements has N + 1 taps, and N + 1 tunable complex weights. These weights are updated continuously by an adaptive algorithm. The object is to adapt the coefficients to minimize the noise and intersymbol interference (depending on the type of equalizer) at the output. The adaptation of the equalizer is driven by an error signal.

7.6.1 Continuous Time Adaptive Equalizer

Figure 7.6 shows the block diagram of the adaptive analog equalizer introduced in [11]. The received input signal first goes through the equalization filter, which has two signal paths: the low-frequency path and the high-frequency path. The high-frequency path compensates for the high-frequency loss of the channel so that the overall loss becomes equal for all frequencies. The output of the equalization filter is then fed into a comparator. The servo loop determines the need for more boosting by comparing the high-frequency contents of the comparator's input and output signals.

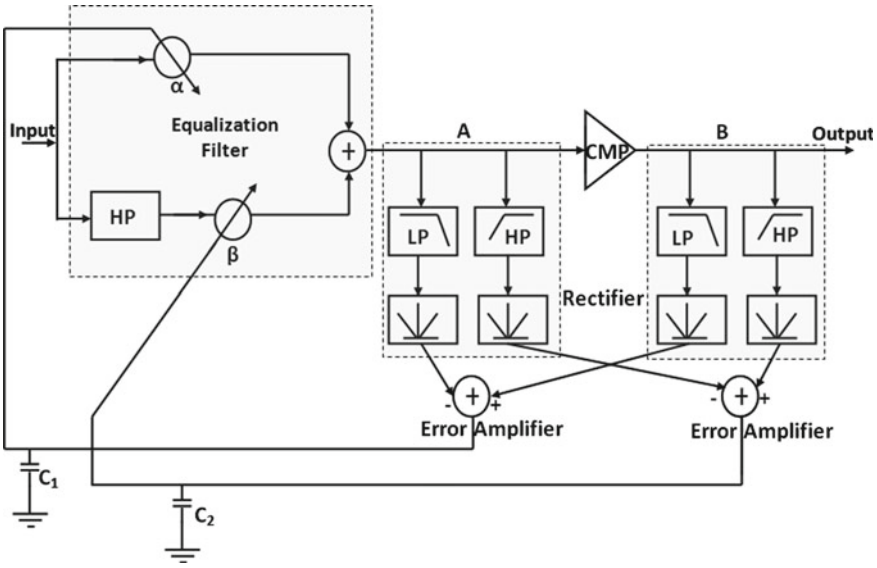

Fig. 7.6 Block diagram of the continuous time adaptive equalizer, with kind permission of [11]

Two pairs of high-pass filter and rectifier extract the high-frequency power of these signals, and an error amplifier calculates their difference. After being filtered by a loop capacitor C_2, this difference controls the high-frequency gain, which effectively adjusts the zero frequency of the equalizing filter. Two pairs of low-pass filter and rectifier extract the low-frequency power of the regulating comparator's input and output signals. Then the difference between these two quantities is amplified by the error amplifier.The difference signal is filtered by C_1 and the filtered signal controls the low-frequency gain of the equalizing filter (α).

Differential Source Degeneration Adjustable Equalizer: The equalizer filter can be implemented by the circuit shown in Fig. 7.7. This is a differential source degeneration equalizer. The operation of source degeneration was described in Sect. 7.3. The equalizer in Fig. 7.7 has two control inputs, the first control signal (V_z) controls the high-pass filter corner frequency and the second control signal (V_g) controls the low-frequency gain. The high cut-off frequency is controlled by varying the MOS capacitors M_5 and M_6. The additional passive resistor R_S is used to give a proper minimum low-frequency gain to prevent an undesired too small low-frequency gain when M_1 is off. Figure 7.8 indicates the frequency response of the adjustable equalizer with differential source degeneration as a function of low-frequency and high-frequency control signals.

The voltage gain transfer function of the differential source degeneration adjustable equalizer is given by:

$$A(\omega) = \frac{g_m R_L (J\omega R_s C_s + 1)}{g_m R_s + 1 + J\omega C_s R_s} \tag{7.10}$$

where

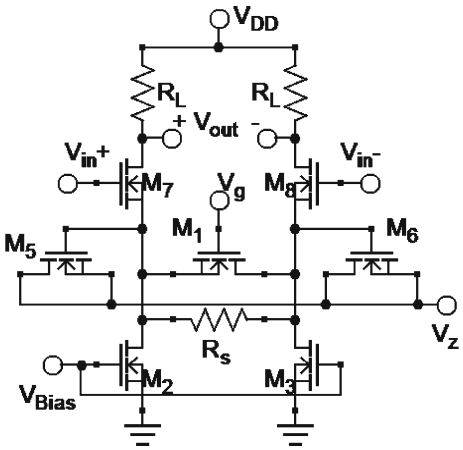

Fig. 7.7 Adjustable equalizer using differential source degeneration, [11]

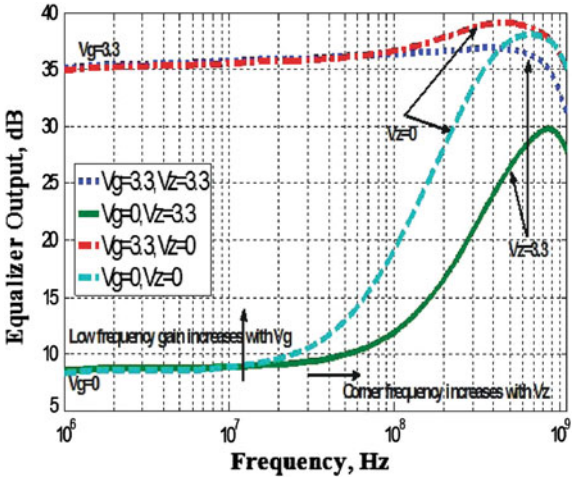

Fig. 7.8 Frequency response of differential source degeneration adjustable equalizer [12]

$$A(0) = \frac{R_L}{R_s}, A(\infty) = g_m R_L, \omega_z = \frac{1}{R_s C_s}, \omega_p = \frac{1 + g_m R_s}{C_s R_s} \tag{7.11}$$

To have enough high-frequency gain compared to the low-frequency gain, g_m should be much larger than $\frac{1}{R_s}$.

7.6.2 Discrete Time Adaptive Equalizer

Adaptive equalizers can be used in two modes:

Training Mode: To make equalizer suitable in the initial acquisition duration, a training signal is needed. In this mode of operation, the transmitter generates a data symbol sequence known to the receiver.

Decision Directed Mode: The receiver's decisions are used to generate the error signal. Decision directed equalizer adjustment is effective in tracking slow variations in the channel response. Once an agreed time has elapsed, the slicer output is used as a training signal and the actual data transmission begins. The forward and feedback filter coefficients are adjusted simultaneously using a least mean square (LMS) algorithm to minimize the MSE. The equalizer coefficient vector C in Fig. 7.9 is obtained according to the relation [13]:

$$c_j(k+1) = c_j(k) - \Delta G(k), k = 0, 1, 2, \ldots, N \tag{7.12}$$

with the gradient factor G(k) = −E(e(k)V^*)

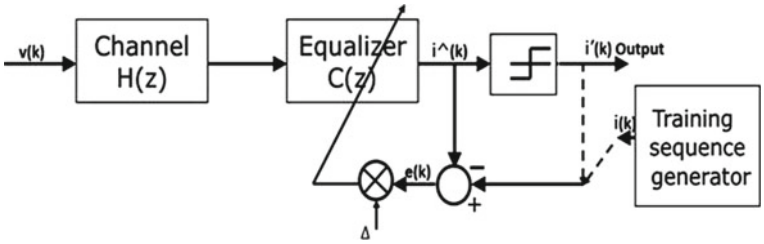

Fig. 7.9 Block diagram of an adaptive linear equalizer [5]

$$e(k) = I(k) - \hat{I}(k)(Trainingmode) \tag{7.13}$$

$$c_j(k+1) = c_j(k) + \Delta e'(k) v^*(k-j) \tag{7.14}$$

$$e'(k) = I'(k) - \hat{I}(k)(Decisionmode) \tag{7.15}$$

Δ indicates scale factors that control the rate of adjustment and $e(k) \cdot V^*(k-j)$ is an estimate of the cross correlation. The minimum MSE is reached when $G(k) = 0$. Following the training period, after which the equalizer coefficients have converged to their optimum values, the decisions at the output of the detector are reliable to be used to continue the coefficient adaptation process.

7.7 Continuous Time FIR Filter Implementation

The analog FIR filter is implemented using variable gain blocks and delay elements, Fig. 7.10. The passive delay line is easy to design but it requires a large CMOS area and suffers from limited design freedom because of characteristic impedance matching requirements, thus limiting the choice of usable on-chip inductors. For the above mentioned reasons, an active delay line version is better to use for integrated circuits [5].

The active delay line is more compact and has lower power consumption. Another important benefit of the active delay line is that it does not introduce DC voltage drops across each stage, and the overall bandwidth is not drastically degraded by additional delay stages. An example of an active delay line implementation is shown in Fig. 7.11. To integrate further and implement a varying time delay, the source follower as a load replaces the series combination of an on-chip inductor and resistor. The active delay-line cell is shown in Fig. 7.11 [14].

The overall voltage gain for a differential pair in the unit delay cell is [14]:

$$A(\omega) = g_{m3} \cdot (Z_{in} // C_L) \tag{7.16}$$

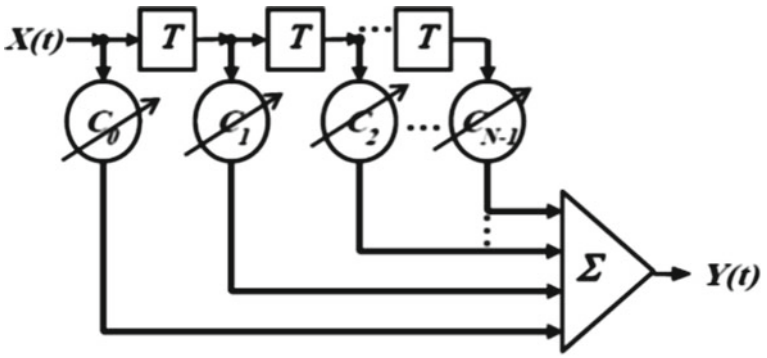

Fig. 7.10 Continuous time FIR filter implementation, [5]

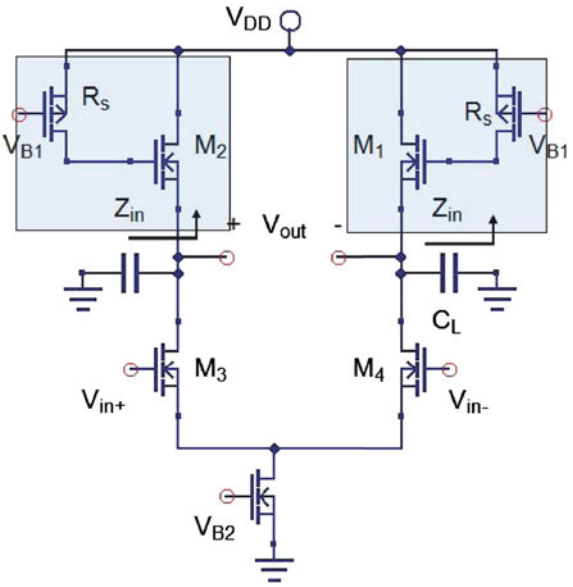

Fig. 7.11 Active delay line implementation using a MOSFET differential amplifier with active inductor loads, [14]

$$Z_{in} = (S \cdot R_s \cdot C_{gs2} + 1)//(S \cdot C_{gs2} + g_{m2}) \tag{7.17}$$

Then the overall voltage gain for a differential pair in the unit delay cell can be calculated by:

$$A(\omega) = \frac{g_{m3} \cdot (1 + J\omega R_s C_{gs2})}{g_{m2} + J\omega(C_{gs2} + C_L) - \omega^2 C_{gs2} C_L R_s} \tag{7.18}$$

where R_s is the turn-on resistance of M_1. There is a zero ω_z. By varying the value of R_s the zero location can be controlled.

$$\omega_z = \frac{1}{R_s C_{gs2}} \tag{7.19}$$

The Gilbert cell, shown in Fig. 7.12, is used as a gain stage for various tap gains. The tail current of a differential pair controls the tap gains ($C_0, C_1, \ldots, C_n$), while two control signals determine the polarity of the gain. The input signal is fed into M_1 and M_2. The upper differential pairs determine the sign of the gain depending on which pair is activated [15]. The Gilbert cell acts as a voltage-to-current converter so that the filter response can be combined from the tap at the summing node.

The simplest approach to implement the summing stage is to tie all the output currents from the multipliers together and then convert them into a voltage through a load resistor R_L. The bandwidth of the summing stage will be limited due to the large value of R_L to achieve the required gain. To improve the bandwidth, a low resistance is needed at the summing node, where the parasitic capacitance is expected to be large due to the contribution of all the multipliers. The bandwidth can be increased by decreasing R_L, but the gain will be decreased and the signal-to-noise ratio will be worse. To increase the bandwidth of the summing circuit and have enough gain, a transimpedance load is used, as shown in Fig. 7.13. Resistors R1 provide a feedback path for transistors M3 to implement a TIA. Resistors R2 bias the gate of transistors

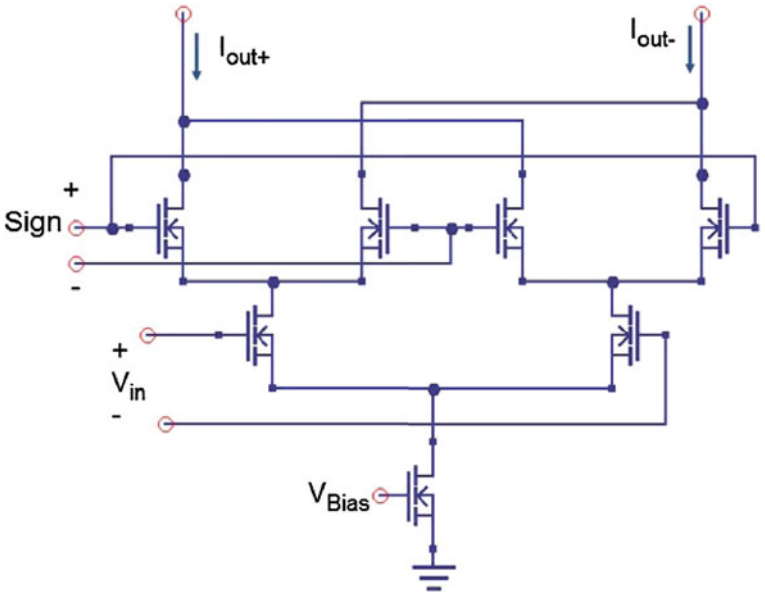

Fig. 7.12 Gilbert cell gain stage [16]

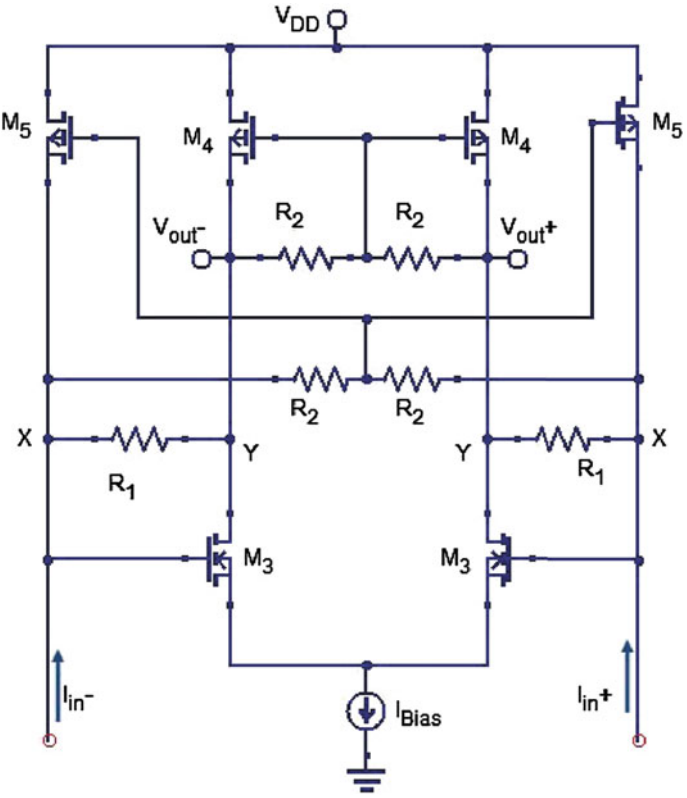

Fig. 7.13 Transimpedance amplifier for the implementation of the summing block, [16]

M_4 and M_5.The equivalent input resistance is given by $1/g_{m3}$ and the transimpedance gain is given by $-R_1$ for $R_2 \gg R_1 \gg 1/g_{m3}$ [16].

7.8 Discrete Time FIR Filter Implementation

Discrete time equalizers use FIR filters, see Fig. 7.14. The input stream x(k) propagates through a series of delay units. The signals taken between the delay units are multiplied by weighting factors (C_k). The outputs from the weights are summed to produce the transmitted pre-equalized output ($y(k)$). The number of taps depends on the length of the channel relative to the unit interval of the data. The equalizer vector coefficients $C(z)$ are:

$$C(z) = \sum_{k=0}^{N} c_{(k)} z^{-k} \tag{7.20}$$

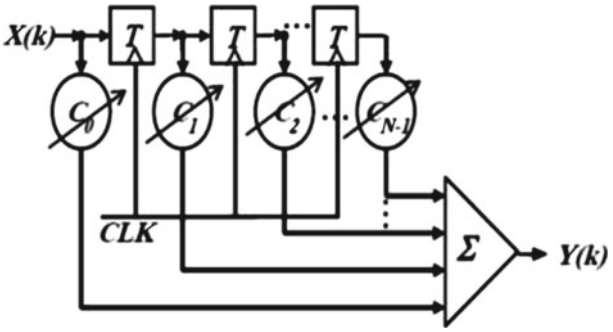

Fig. 7.14 Discrete time FIR filter implementation [5]

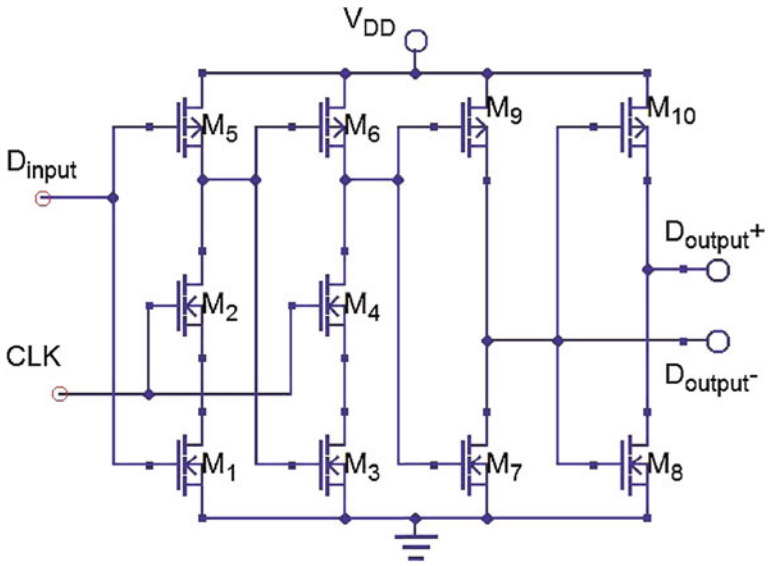

Fig. 7.15 Delay unit cell using a dynamic TSPC latch, [17]

The delay units can be implemented by a dynamic True Single-Phase Clocked (TSPC) latch as in Fig. 7.15 [17]. The output combiner circuit, shown in Fig. 7.16, makes the summation and gives the required weights for the delayed signals.

7.9 Nonlinear Equalization

The linear equalizer cannot distinguish between the signal and the noise. Both the high-frequency signal and the high-frequency noise is amplified. The data rate will be limited by the signal-to-noise ratio. Noise becomes a primary design consideration

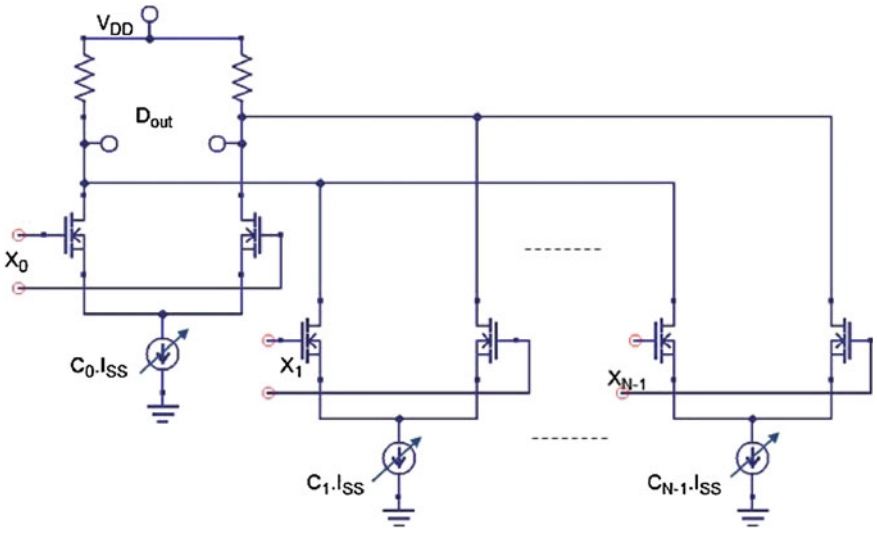

Fig. 7.16 The output combiner, [17]

with linear equalizers. An alternative to linear equalization is non-linear equalization. Non-linear equalization commonly is used in applications where the channel distortion is too severe. There are two effective non-linear equalization methods: Decision Feedback Equalization (DFE) and Maximum Likelihood Sequence Estimation (MLSE).

7.9.1 Decision Feedback Equalizer (DFE)

The DFE has a feedback from the filter output to the input as shown in Fig. 7.17. The feedback signal is summed to the input signal to provide input to a decision circuit (slicer), which decodes the analog input signal with the help of a comparator to a binary signal. The output from the slicer is used as input to the FIR filter. The basic idea is that if the values of the symbols already detected are known (past decisions are assumed to be correct), then the ISI contributed by these symbols can be canceled exactly. The error in current decisions will corrupt future decisions. There are coding methods to minimize this impact. The DFE corrects only for post-cursor ISI, see Fig. 7.18. Only the samples after h_0 can be equalized like samples $h_3, h_4, h_5, \ldots$ in Fig. 7.18. The DFE requires an input signal with open eye, so it is typically used with a feed forward linear equalizer on the front end of the DFE.

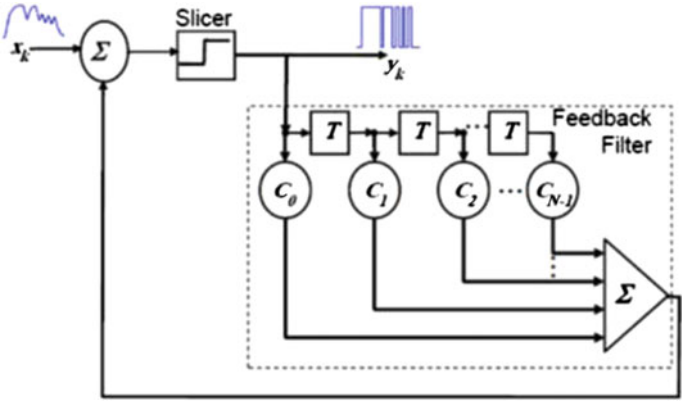

Fig. 7.17 Decision feedback equalizer(DFE) block diagram: **a** The received distorted pulse before equalization. **b** The output pulse after the DFE [5]

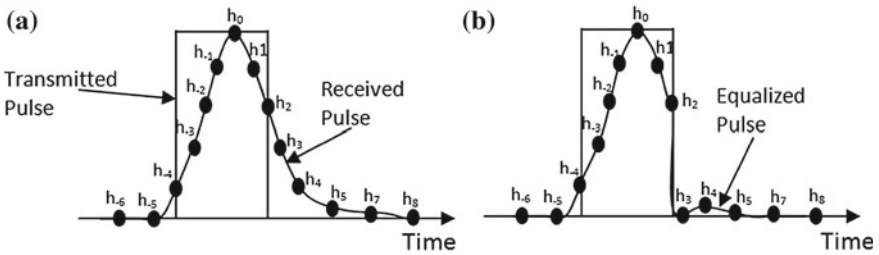

Fig. 7.18 The output pulse after the decision feedback equalizer

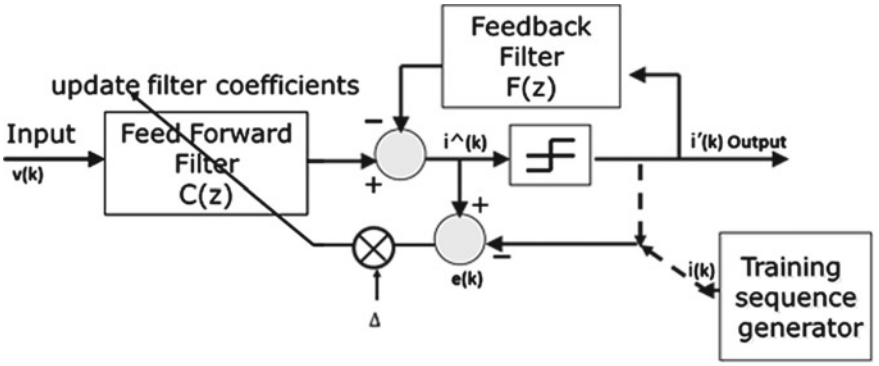

Fig. 7.19 Adaptive DFE block diagram, [5]

Adaptive DFE Equalization: The forward and feedback filter coefficients for the adaptive DFE (Fig. 7.19) are adjusted simultaneously to minimize the MSE using the steepest decent algorithm [13]:

$$c(k+1) = c(k) + \Delta E(e(k)V^*(k)) \tag{7.21}$$

$$e(k) = I(k) - \hat{I}(k)(Trainingmode) \tag{7.22}$$

$$c(k+1) = c(k) + \Delta e'(k)V^*(k) \tag{7.23}$$

$$e'(k) = I'(k) - \hat{I}(k)(Decisionmode) \tag{7.24}$$

7.9.2 Maximum Likelihood Sequence Estimator (MLSE)

In fact, it turns out that all the equalization methods discussed in the previous section are not optimal. Because of the fact that the effect of a symbol is spread to other symbols, it is intuitive that the optimal receiver should observe not only the segment of received signal concerning the desired symbol, but the whole received signal instead. Instead of deciding a transmitted symbol at a time, we can consider to decide the whole transmitted symbol sequence simultaneously from the received signal. In this way, we aim at minimizing the probability of choosing the wrong sequence of symbols instead of the average symbol error probability. With this sequence detection approach, we can employ the Maximum Likelihood principle to achieve our goal. The resulting "optimal" receiver is referred to as the Maximum Likelihood Sequence Estimator (MLSE) receiver. MLSE tests all possible data sequences (rather than decoding each received symbol by itself), and chooses the data sequence with the maximum probability as the output. The block diagram of an MLSE receiver is shown in Fig. 7.20. The MLSE receiver can be implemented with the Viterbi algorithm [18, 19]. The Viterbi algorithm can be simply described as an algorithm which finds the most likely path through a trellis, i.e. shortest path, given a set of observations. The trellis in this case represents a graph of a finite set of states from a Finite States Machine (FSM). Each node in this graph represents a state and each edge a possible transition between two states at consecutive discrete time intervals. An example of a trellis is shown below in Fig. 7.21.

MLSE usually has a large computational requirement. MLSE requires knowledge of the channel characteristics in order to compute the metrics for making decisions. MLSE also requires knowledge of the statistical distribution of the noise corrupting the signal. The MLSE for a channel with ISI has a computational complexity that grows exponentially with the length of the channel time dispersion. If the size of the symbol alphabet is M and the number of interfering symbols contributing to ISI is L, the Viterbi algorithm computes M^{L+1} metrics for each received symbol. In most channels of practical interest, such a large computational complexity is expensive to implement [13]. For time-invariant channel impulse response, the received symbols

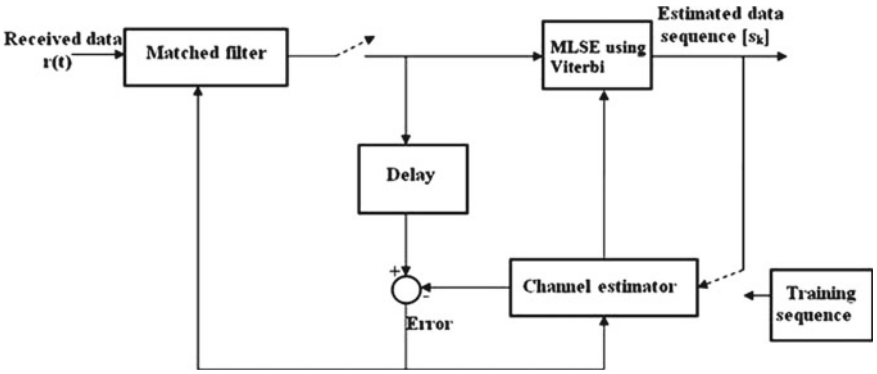

Fig. 7.20 Maximum Likelihood Sequence Estimator (MLSE) block diagram [20]

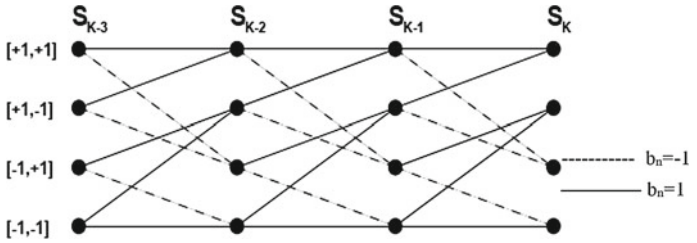

Fig. 7.21 Trellis diagram for L = 2, [21]

are described as:

$$r(k) = \sum_{j=0}^{L-1} h_j b_{(k-j)} + n_k \tag{7.25}$$

where r_k is the received symbol, b_k denotes the actual transmitted sequence, n_k is the Gaussian noise, and h_j is the coefficient of the estimated channel impulse response. The equalizer is responsible for reversing the effect of the channel on the transmitted symbols in order to produce the sequence of transmitted symbols with maximum confidence. To optimally estimate the transmitted sequence of length M in an optical communication system, the least mean square error function (7.26) must be minimized.

$$\sum_{k=1}^{M} \left| r(k) - \sum_{j=0}^{L-1} h_j s_{k-j} \right|^2 \tag{7.26}$$

Here, s = [s_1, …, s_M] is the most likely estimated transmitted sequence that will maximize the probability of P(s/r). The Viterbi MLSE equalizer is able to solve this problem exactly, with computational complexity linear in M and exponential in L [22]. The update relations of the states are depicted by the trellis diagram. The trellis

diagram for the case of L = 2 is given in Fig. 7.21. In a trellis, each node corresponds to a distinct state at a given time (S), and each edge represents a transition to some new state at the next instant of time. The trellis begins and ends at the known states S_0 and S_k. Its most important property is that to every possible state sequence S there is a unique path through the trellis, and vice versa.

References

1. D. Lee, J. Han, G. Han, S.M. Park, An 8.5-Gb/s fully integrated CMOS optoelectronic receiver using slope-detection adaptive equalizer. IEEE J. Soild-State Circuits **45**(12), 2861–2873 (2010)
2. W.-Z. Chen, S.-H. Huang, A 2.5 Gbps CMOS fully integrated optical receicer with lateral PIN detector, in *IEEE 2007 Custom Intergrated Circuits Conference (CICC)* (2007), pp. 293–296
3. C. Gimeno, C. Aldea, S. Celma, F. Aznar, A cost-effective 1.25-Gb/s CMOS receiver for 50-m large-core SI-POF links. IEEE Photonics Technol. Lett. **99** (2012)
4. M. Atef, R. Swoboda, H. Zimmermann, 1.25 Gbit/s over 50 m step-index plastic optical fiber using a fully integrated optical receiver with an integrated equalizer. J. Lightwave Technol. **30**(1), 118–122 (2012)
5. M. Atef, H. Zimmermann, *Optical Communication over Plastic Optical Fibers: Integrated Optical Receiver Technology* (Springer, Berlin, 2013)
6. C. Hermans, M. Steyaert, *Broadband Opto-Electrical Receivers in Standard CMOS* (Springer, Netherlands, 2007)
7. Y. Dong, K.W. Martin, A high-speed fully-integrated POF receiver with large-area photo detectors in 65 nm CMOS. IEEE J. Solid-State Circuits **47**(9), 2080–2092 (2012)
8. M. Atef, D. Abd-elrahman, 2.5 Gbit/s compact transimpedance amplifier using active inductor in 130 nm CMOS technology, in *21st International Conference on Mixed Design of Integrated Circuits & Systems (MIXDES)* (2014), pp. 103–107
9. M.S.F. Tavernier, High-speed optical receivers with integrated photodiode in nanoscale CMOS (Springer, New York, 2011)
10. M. Kiziroglou, A. Mukherjee, S. Vatti, A. Holmes, C. Papavassiliou, E. Yeatman, Self-assembly of three-dimensional au inductors on silicon. IET Microwaves, Antennas Propag. **4**, 1698–1703 (2010)
11. J.-S. Choi, M.-S. Hwang, D.-K. Jeong, A 0.18 um CMOS 3.5-Gb/s continuous-time adaptive cable equalizer using enhanced low-frequency gain control method. IEEE J. Solid-State Circuits **39**(3), 419–425 (2004)
12. M. Atef, R.Swoboda, H. Zimmermann, 1 Gbit/s transmission over step-index plastic optical fiber using an optical receiver with an integrated equalizer. Opt. Commun. **284**(21), 5153–5156 (2011)
13. J.G. Proakis, *Digital Communications*, 4th edn. (McGraw-Hill, New York, 2001)
14. M. Maeng, F. Bien, Y. Hur, H. Kim, 0.18 um CMOS equalization techniques for 10-Gb/s fiber optical communication. IEEE Trans. Microw. Theory Tech. **53**(11), 3509–3519 (2005)
15. H. Kim, F. Bien, Y. Hur, S. Chandramouli, J. Cha, E.Gebara, J. Laskar, A 0.25-um BiCMOS feed foward equalizer using active delay line for backplane communication, in *IEEE International Symposium on Circuits and Systems (ISCAS)* (2007), pp. 193–196
16. D. Hernandez-Garduno, J. Silva-Martinez, A CMOS 1 Gb/s 5-tap fractionally-spaced equalizer. IEEE J. Solid-State Circuits **43**(11), 2482–2491 (2008)
17. H. Wang, J. Lee, A 21-gb/s 87-mw transceiver with ffe/dfe/analog equalizer in 65-nm cmos technology. IEEE J. Solid-State Circuits **45**, 909–920 (2010)
18. J.K. Omura, On the Viterbi decoding algorithm. IEEE Trans. Inf. Theory **15**(1), 177–179 (1969)

19. G.D. Forney, Convolutional codes II. Maximum-likehood decoding. Inf. Control **25**(3), 222–266 (1974)
20. A. Idris, N. Abdullah, N.A. Hussein, D.M. Ali, Optimization of BER performance in the MIMO-OFDMA system for mobile WiMAX system using different equalization algorithm, in *Advanced Computer and Communication Engineering Technology: Proceedings of the 1st International Conference on Communication and Computer Engineering*, Chapter 25 (Springer International Publishing, Swizerland, 2015)
21. T. Wong, T. Lok, *Theory of Digital Communications*. Lecture Notes (University of Florida, 2004)
22. S. Benedetto, E. Biglieri, *Principles of Digital Transmission with Wireless Applications* (Kluwer Acad./Plenum Publ., 1999)

Chapter 8
Post Amplifiers

The output signal from the TIA is in the range of a few millivolts and more gain is needed to reach at least an amplitude of 200 mV required by the decision circuit. This additional gain will be introduced by the post amplifiers (PA) following the TIA. As stated in Chap. 6 the TIA specifications determine the performance of the optical receiver, such as the sensitivity and the dynamic range. The PA specifications have less impact, but insufficient PA bandwidth, gain, and sensitivity can degrade the optical receiver's overall performance. The effect of the PA specifications should be considered to achieve the required optical receiver performance (gain, bandwidth, sensitivity, dynamic range, …) at minimum or reasonable power consumption. There are two main types of post amplifiers (PA), the limiting amplifier (LA) and the automatic gain control (AGC) amplifier. The selection of LA or AGC amplifier depends on the required system performance, like linearity, gain and sensitivity. The DC offset should be minimized or completely canceled by using DC-offset canceling. The amplifier's bandwidth can be extended by using bandwidth enhancement techniques. All of these topics will be covered in this chapter.

8.1 Noise

The design of a high-speed TIA with high transimpedance gain in nanometer CMOS technologies is challenging, due to the low intrinsic gain, see Fig. 6.1. The output signal of the TIA is not strong enough and the effect of the PA's noise should be considered.

The input referred noise current for an optical receiver consisting of TIA noise in addition to PA noise is:

$$\overline{i^2_{n,total}} = \overline{i^2_{n,TIA}} BW_{TIA} + \frac{\overline{v^2_{n,PA}}}{R_{FB}} BW_{PA} \tag{8.1}$$

M. Atef and H. Zimmermann, *Optoelectronic Circuits in Nanometer CMOS Technology*, Springer Series in Advanced Microelectronics 54,
DOI 10.1007/978-3-319-27338-9_8

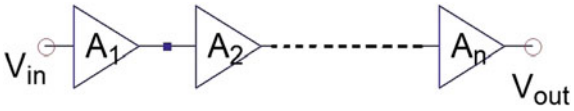

Fig. 8.1 A chain of cascaded amplifiers

To minimize the effect of PA noise on the optical receiver's sensitivity, the first stage of the PA should give the highest gain. The bandwidth of the first stage should be equal or a little bit higher as the signal data rate. The effect of the second PA stage on the receiver sensitivity can be neglected, so a lower gain and a higher bandwidth can be selected for the second and third PA stages. As the output buffer (50 Ω driver) has a large-area transistor to drive the load capacitance, the PA stages near to the output buffer should have higher biasing current than the PA stages near to the TIA. The load capacitance of each gain stage increases into the direction of the output buffer and so the biasing current of each stage is higher as it is closer to the output.

8.2 Cascaded Gain Stages

It is difficult to design a high-speed single-stage PA with high gain in nanometer CMOS technologies, due to the low intrinsic gain, see Fig. 6.1. A multistage PA is needed to deliver the required gain and bandwidth. Figure 8.1 shows a chain of cascaded amplifiers.

The voltage gain A_v is defined as the small-signal output voltage divided by the small-signal input voltage. The maximum gain for each stage must be small enough such that the amplified signal stays in the linear region, to have a good linearity without a gain compression or limiting effects. On the other hand, the higher the voltage gain the smaller the effect of the PA noise on the optical receiver's sensitivity. As the gain bandwidth product (GBW) is limited for a single stage to a value lower than the transit frequency f_T, so the higher the gain the lower the bandwidth.

$$GBW < f_T = \frac{g_m}{2\pi(C_{gs} + C_{gd})} \tag{8.2}$$

The gain bandwidth product of a single-stage amplifier cannot exceed f_T. By using cascaded amplifier stages it is possible to build an amplifier with a GBW that is larger than f_T. The total gain for a chain of N cascaded PAs is:

$$A_{total} = A_1 A_2 \ldots A_N \tag{8.3}$$

The first stage of the PA should contribute the highest gain A_1 to minimize the effect of PA noise on the optical receiver's sensitivity. The number of stages N depends on the required gain, bandwidth, and the power dissipation. As we will see in the next Sect. 8.3. Four to five amplifier stages are used for practical PA designs to achieve the required gain and bandwidth at a reasonable power consumption.

8.3 Bandwidth

In nanometer CMOS technology a chain of cascaded amplifiers is used to give the required gain. The total bandwidth of N cascaded stages is smaller than the smallest bandwidth of the stages. For simplicity consider N stages with equal bandwidth BW, considering each stage to behave as a first order low pass filter. The total bandwidth for N cascaded stages is calculated by [1]:

$$BW_{total} = BW\sqrt{\sqrt[N]{2} - 1} \tag{8.4}$$

In the first stage of the post amplifier the bandwidth should be higher than the target data rate; as the cascading of many stages will reduce the total amplifier bandwidth, (8.4). The last stage of the post amplifier will have a large input signal level, so the amplifier will work in the large-signal mode, not in the small-signal mode. The large input signal will switch the MOSFET ON and off, so the switching speed should be as high as the data rate. The definition of the small-signal bandwidth will not be valid here and the rise and fall time is the measuring factor, see (3.27).

The total gain bandwidth product GBW_{total} is larger than the GBW_{single} [1]:

$$GBW_{total} = A_N BW\sqrt{\sqrt[N]{2} - 1} \tag{8.5}$$

$$\frac{GBW_{total}}{GBW_{single}} = A_i^{(N-1)}\sqrt{\sqrt[N]{2} - 1} \tag{8.6}$$

$$r = \frac{GBW_{total}}{GBW_{single}} = A_{total}^{(1-1/N)}\sqrt{\sqrt[N]{2} - 1} \tag{8.7}$$

r is the ratio of the total gain bandwidth product to the single-stage bandwidth product. The r ratio in (8.7) is plotted in Fig. 8.2 as a function in number of stages N for different total gain A_{total}. As the number of stages N increases the ratio r will increase. For N larger than 4 the r increases slowly. For N larger than 6, r decreases. The larger the number of stages the larger the power consumption. The optimum number of stages to obtain a high r at a low power consumption is 4 or 5 stages.

The ratio r indicates how much the single stage gain bandwidth product can be extended by using a multistage amplifier. For $A_{total} = 30$ the multistage GBW is 6 times higher than the single stage GBW. The maximum value for the ratio r can be fitted by the formula:

$$r_{max} = 0.36 \cdot A_{total}{}^{0.8} \tag{8.8}$$

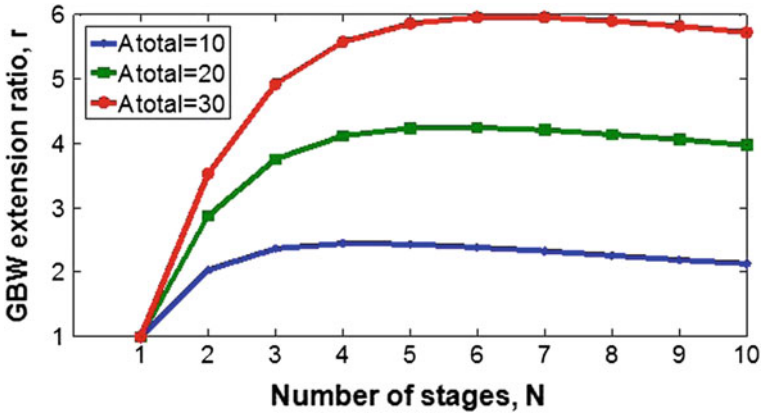

Fig. 8.2 The r ratio as a function in number of stages N for different total gain

8.4 Differential Post Amplifier

A differential amplifier has advantages over the single-ended amplifier. Figure 8.3 illustrate the block and circuit diagram of a differential amplifier. The differential amplifier has two out of phase outputs V_{out+} and V_{out-}. The differential amplifier's total output voltage is $V_{out} = V_{out+} - V_{out-}$. So the differential amplifier output swing is two times the single ended output swing.

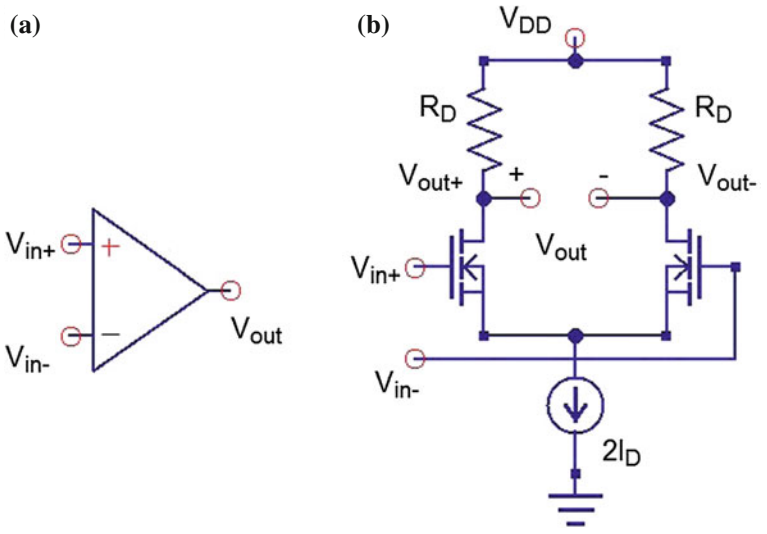

Fig. 8.3 **a** Block diagram and **b** circuitry of differential amplifier

A differential PA has improved immunity to power supply and substrate noise compared to a single-ended PA. The power supply and the substrate noise are coupled equally to both inputs of the differential PA and thus these noises are suppressed as a common-mode signal. Also the differential PA has a higher voltage swing compared to a single-ended PA. So, differential PAs are needed in noisy environments where the power supply and substrate noise are strong. A one-chip mixed-signal system where digital noise can be coupled to front-end low-noise analog circuits is an example where the differential structure is preferred. A differential output driver is also preferable from stability point of view compared to the single-ended output driver. The bonding wires of the power supply and ground pads can make a positive feedback path for single ended output driver. This reduce the stability of the single ended output and some kind of ringing and overshoot will happen [1]. As a result it is preferred to use a differential output driver with a differential pre-driver (PA).

8.5 Amplifier with Automatic Gain Control

The AGC reduces the gain of the amplifier to prevent the amplifier from being overloaded. Besides preventing the amplifier's overload, the AGC keeps the output signal amplitude constant to enable the next stage (decision circuit) to work correctly. The AGC amplifier preserves the signal linearity permitting analog signal post-processing (post equalization, …) to be performed on the amplified signal. To explain the function of the AGC, Fig. 8.4 shows the gain of the output signal of the AGC as a function of the input signal. The VGA amplifies the input signal with a constant gain at low input values less than V_{th1}. For input amplitude larger than V_{th1} and smaller than V_{th2} the AGC starts to work and the gain is reduced to have an almost constant output amplitude. For input signals larger than V_{th2} the AGC stops working and the gain will be fixed in order to prevent stability problems at low level

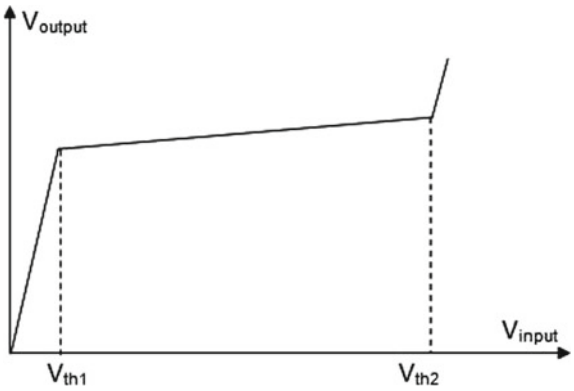

Fig. 8.4 The output voltage of an AGC as a function of the input signal

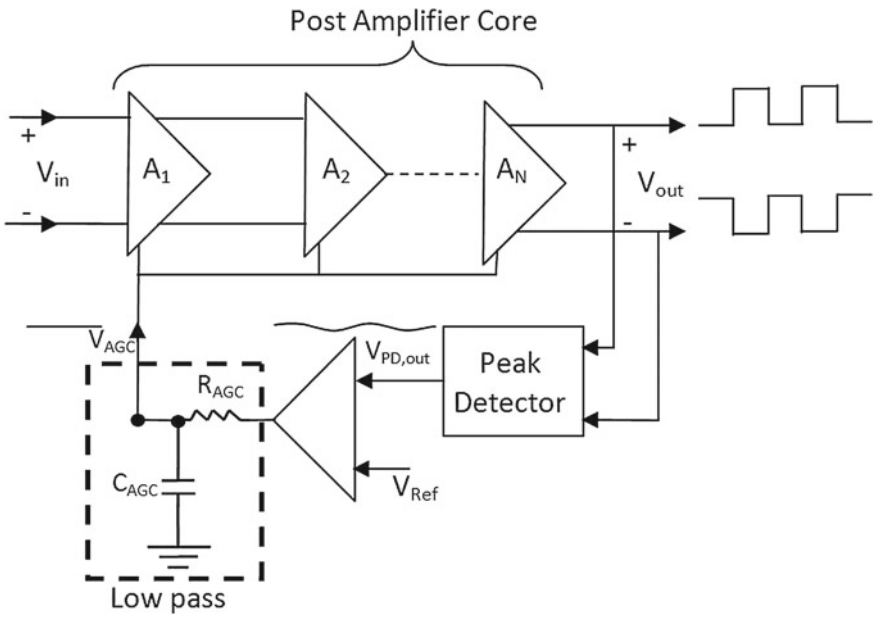

Fig. 8.5 Block diagram of an automatic gain control amplifier [2]

of gain until the amplifier saturates. The AGC's useful range of operation is between V_{th1} and V_{th2}.

For high linearity applications an AGC is a must. In other applications the non-linearity can be afforded, the simpler and limiting amplifier design with higher gain, bandwidth and simpler circuit is preferred.

The block diagram of an amplifier with automatic gain control is shown in Fig. 8.5. The peaks of the PA core output are detected by a peak detector (based on a full wave rectifier). The rectified output ($V_{PD,out}$) is fed to a comparator with reference voltage V_{Ref}. When the $V_{PD,out}$ is smaller than V_{Ref} the V_{AGC} is not enough to reduce the gain. As the $V_{PD,out}$ is greater than V_{Ref} then the V_{AGC} will start reducing the amplifier gain. The larger the PA output signal is the larger the $V_{PD,out}$ is and the V_{AGC} can reduce the PA gain more. The low-pass filter (R_{AGC} and C_{AGC}) is controlling the speed of the gain control loop. If there are fast changes in the PA output due to small overshoot or long patterns of ones or zeros, the filter should be slow enough to neglect these effects. On the other hand the low-pass filter should be fast enough to detect the change in the PA's output amplitude. If the filter is too fast this can affect the amplifier's stability, and if it is too slow it cannot correctly control the PA gain. The sensitivity and linearity are the main concern when we are controlling the PA stages gain. For high linearity reducing the PA first gain stage is important and the gain of the other PA stages can be controlled if needed. For high sensitivity requirements the gain of the PA first stage should be kept at maximum for the small amplitude input signal, and the gain of other PA stages should be reduced first.

Figure 8.6 shows a PA differential pair with variable source degeneration resistor R_{AGC}. The variable resistor R_{AGC} can be implemented with a MOSFET operating in the linear region in shunt with a fixed resistor R_s.

The gain of such a stage is calculated by:

$$Gain \approx \frac{g_m R_D}{1 + R_S // r_{AGC}} \tag{8.9}$$

The V_{AGC} controls the PMOS ($M_{P1,2}$) transistors working in the triode region. The V_{AGC} is set to Zero voltage for high gain operation as the PMOS resistor is has a low value. As the V_{AGC} going up the PMOS resistor is increasing reducing the stage's gain, see (8.9).

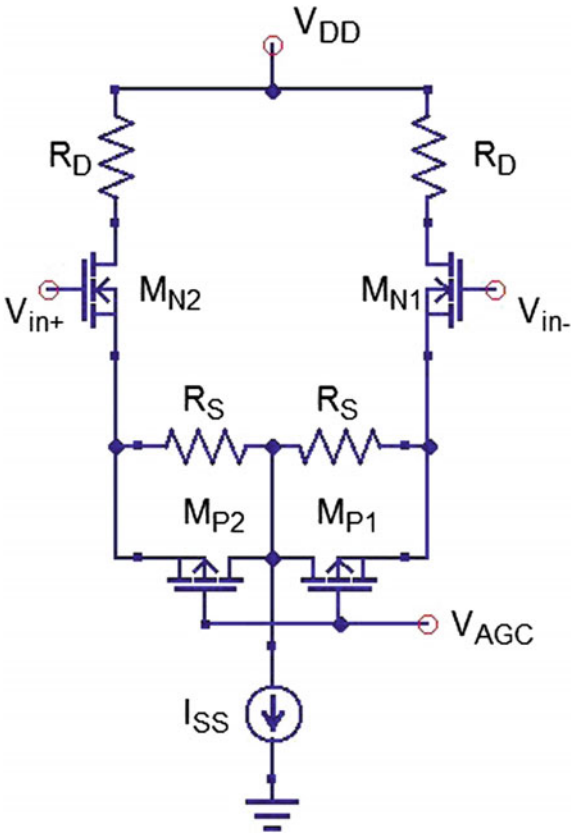

Fig. 8.6 Variable gain amplifier based on a variable source degeneration resistor

8.6 Limiting Amplifier

The limiting amplifier (LA) is easier to design than an AGC amplifier as it does not have an AGC. The LA has a fixed gain which can be considered as the maximum gain value of the AGC amplifier; so the LA performance (like sensitivity, bandwidth and power dissipation) will be better than the AGC amplifier's performance.

The high gain coming from the cascaded amplifier stages will limit the output signal amplitude especially for large amplitude input signals. For small input signals the amplifier works in the linear region. For large input signals nonlinear effects due to signal clipping may distort the output signal. The amplifier has a limited output swing: when the drain current through the MOSFET increases due to the large input signal on the gate, the output signal is clipped and a high nonlinearity occurs. The output amplitude is maximum when the constant current I_{SS} is completely directed to one of the two transistors of the differential pair. The transfer function of an LA for small and large input signals is shown in Fig. 8.7. For very small input signals the amplifier operates in the linear region, but for larger input signals, it crosses into the limiting regime where the output value is limited. As the LA switches between linear and limiting region the switching speed of the LA delay changes causing jitter in the output signal. A special concern should be given to reduce the output jitter coming from the limiting effect.

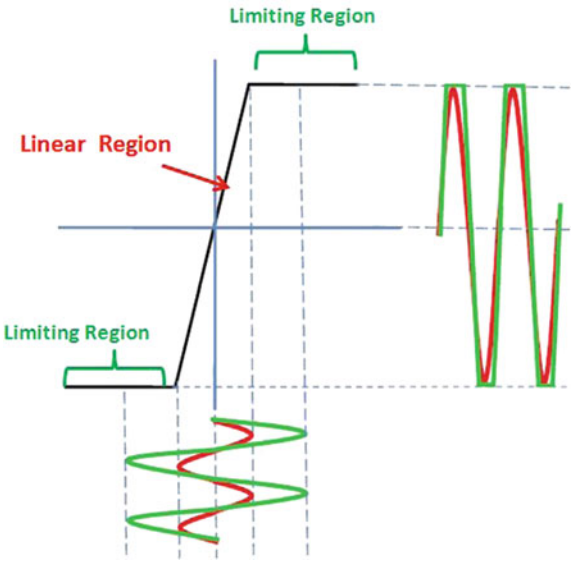

Fig. 8.7 The transfer function of an LA for small and large input signals

8.7 Offset Compensation

Due to mismatch and process varaiations there is a DC offset (from the DC output operating point) at the output of the PA. This output offset should be minimized or canceled to reduce its bad effect on the receiver's sensitivity. The threshold level is at its optimum value when the threshold is exactly centered in between the zero and one levels. The threshold level is not at its optimum value for the case of a decision-threshold offset. An offset voltage (V_{offset}) at the PA's output results in an error in the decision level, which will cause a higher BER, see Fig. 8.8. The $V_{PP,offset}$ is less than V_{PP} so the BER will increase for the same noise level. The power penalty of the offset can be calculated using (8.10) [2]. The offset power penalty (PP_{offset}) describes the reduction in optical sensitivity as a result of decision-threshold offset.

$$PP_{offset} = 10log\left(\frac{V_{out,PP} + 2V_{offset}}{V_{out,PP}}\right) dB \tag{8.10}$$

Another side effect of the offset is the pulse-width distortion. The random offset in the inverting and non-inverting signals due to process variations can randomly shift the crossover points in time causing pulse-width distortion and jittering, as shown in Fig. 8.8. The PW_{offset} is less than the PW which will increase the BER for the same noise level due to incorrect sampling instant.

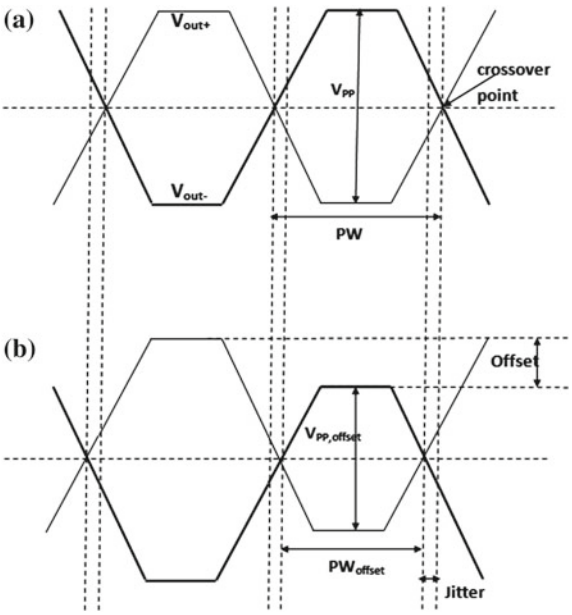

Fig. 8.8 **a** The original differential signal and **b** the differential signal with voltage offset

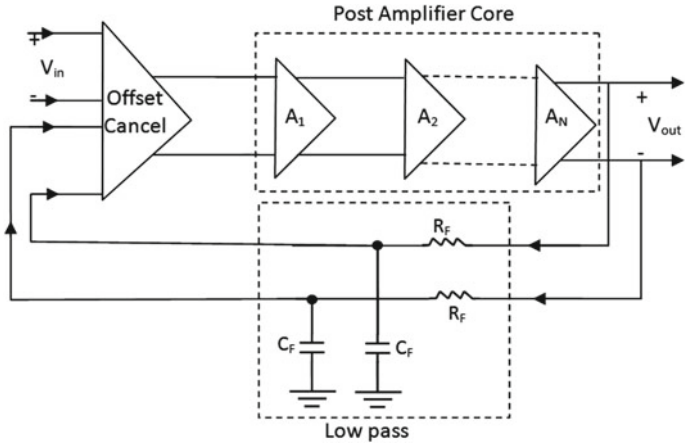

Fig. 8.9 The circuit diagram of post amplifier with offset cancellation stage [3]

This output offset should be minimized or canceled to prevent the amplifier from saturation at high offset values and to enhance the senstivity. An offset compensation circuitry is needed to reduce the offset voltage to the required value. Figure 8.9 shows the circuit diagram of a post amplifier with offset cancellation stage.

C_F and R_F are passive elements used in the low-pass filter to extract the output DC level of the PA. The error amplifier senses the DC component of the PA output and feeds back the error signal at the PA input to compensate for the offset. The error signal adjusts the output voltage until the output offset voltage reaches its minimum. The offset-compensation circuit introduces a low-cutoff frequency which causes baseline wander if the cutoff frequency is too high, see Sect. 2.9.

By neglecting the low cut-off frequency coming from the PA core, the offset canceling loop will determine the PA's low cut-off frequency. The time constant of the offset canceling loop is $\tau = C_F R_F$. The closed loop gain of the limiter is [3]:

$$\omega_C = \frac{A_{PA} A_F + 1}{2\pi C_F R_F} \tag{8.11}$$

where A_{PA} is the total gain of the PA, A_F is the gain of the error amplifier, τ denotes the time constant of the integrator. The low cut-off frequency A_{total}/τ depends on the integrator's time constant (τ). To minimize the chip area occupied by the passive elements, NMOS transistors operating in the triode region are employed as R_F and C_F (Fig. 8.10) is realized by PMOS transistors [3]. Therefore the integrator can be implemented with a low cut-off frequency with small integration capacitors saving the chip area.

The circuitry of the first gain cell and offset cancellation amplifier is shown in Fig. 8.11. The transistors M_3, M_4 in the feedback path act as transconductance cell which converts the PA output offset voltage to a current. The transconductance cell

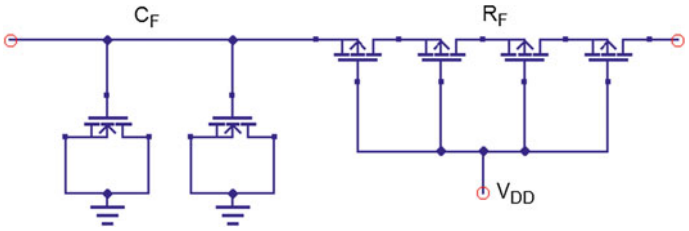

Fig. 8.10 Low pass filter implementation using NMOS and PMOS

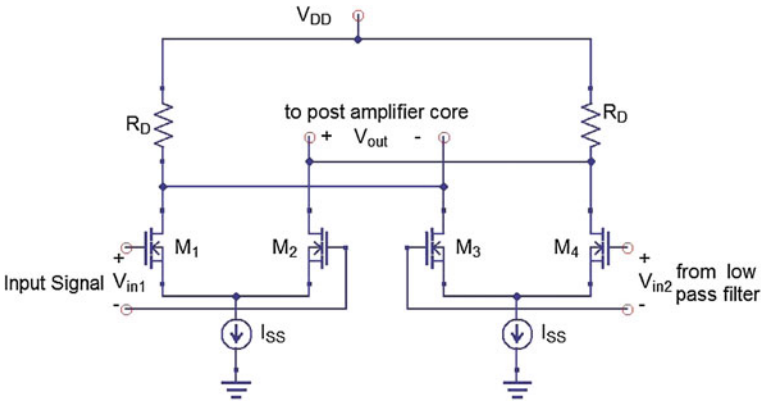

Fig. 8.11 The circuitry of the first gain cell and offset cancellation amplifier [3]

(M_3, M_4) is combined with the first gain stage of the PA core (M_1, M_2). The output DC current from the transconductance cell will be subtracted from the drain current of the PA's first gain stage. By this way the offset cancellation can be achieved [3].

8.8 Broad Band Amplifier Techniques

The gain-bandwidth product of a post amplifier cell is limited by the transit frequency f_t of the transistors. The amplifier bandwidth can be extended by using design techniques without sacrifying gain and power dissipation. The inductive peaking (Sect. 6.9.2), active inductive peaking (Sect. 6.9.3) and negative capacitance (Sect. 6.9.4) can be used to increase the TIA's bandwidth and can also be employed to extend the post amplifier's bandwidth. The reader can refer to Sect. 6.9 for these techniques. Capacitive degeneration utilizes capacitive elements to add an extra zero that compensates the dominant pole to extend the amplifier's bandwidth. The reduced low-frequency gain of the capacitive degeneration is the main shortcoming of this technique. Capacitive degeneration can be used as a post equalization technique; more details will be introduced about capacitive degeneration in the equalization

Chap. 7. Other techniques like Cherry-Hooper amplifiers and Interleaving Active Feedback will be presented in the following sections.

8.8.1 Cherry-Hooper Amplifiers

The Cherry-Hooper topology allows the amplifier gain and bandwidth to be tuned independently of each opposed to many other kinds of amplifiers like CS amplifiers. Figure 8.12 shows the circuit diagram of the Cherry-Hooper topology. It is composed of two stages, the first input stage converts (M_1, R_f) the input voltage signal to a current and the second stage (TIA (M_2, R_f)) converts the current coming from the first stage into the final output voltage V_{out}. Assuming $g_{m2} R_l < 1$, the Cherry-Hooper gain ($A = v_{out}/v_{in}$) can be approximated by:

$$A_S \approx \frac{g_{m1} R_f}{1 + SC_l/g_{m2}} \tag{8.12}$$

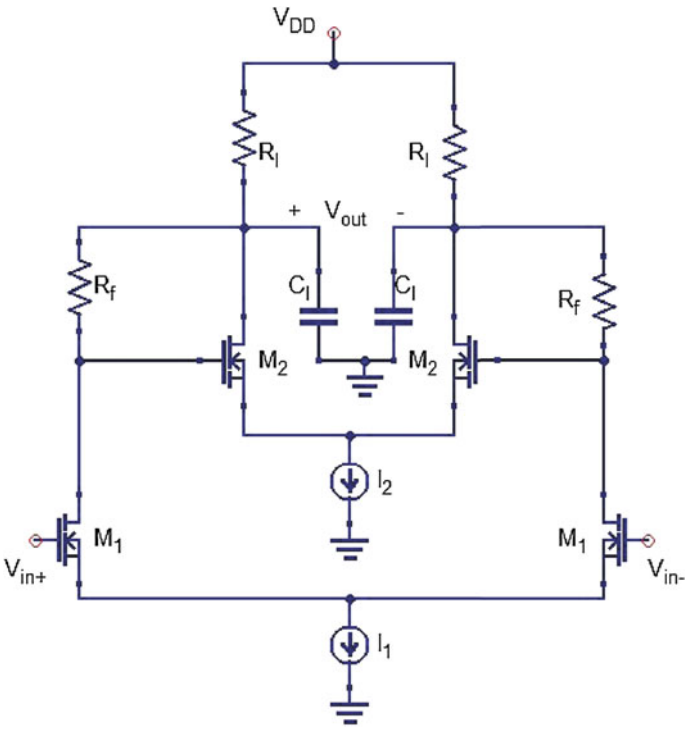

Fig. 8.12 Schematic of a differential Cherry-Hooper amplifier with resistive load [3]

The DC gain is equal to $g_{m1}\,R_f$ and the bandwidth can be approximated by g_{m2}/C_l. The gain and bandwidth of the Cherry-Hooper amplifier are independent, there is not a common parameter between them. The gain and bandwidth can be optimized independently without sacrificing one of them to increase the other one. This allowing the design of high-speed amplifiers at a reasonable gain. The drawback of using Cherry-Hooper topology appears at low voltage operation. There is a high voltage drop over R_l equal to $(I_1 + I_2)R_l$ and on R_f equal to $I_1 R_f$. The increase of the gain needs a high g_{m1} and R_f, so a high value of I_1 is required for high g_{m1}. To have a high bandwidth a high g_{m2} is needed leading to an increase in I_2. As a result for high speed and high gain requirements the voltage drop on R_l and R_f will be high. At low voltage operation M_1 and M_2 will enter the triode region instead of the saturation region required for correct amplifier operation.

8.8.2 *Interleaved Active Feedback*

By using the interleaving active feedback in the cascaded gain stages, the LA bandwidth can be extended while keeping a small area (inductor less topology). The circuit implementations of the gain and the feedback cells are realized by CMOS differential pairs with resistive loads as shown in Fig. 8.13. The transfer function of the third-order interleaved active feedback stage is given by [3]:

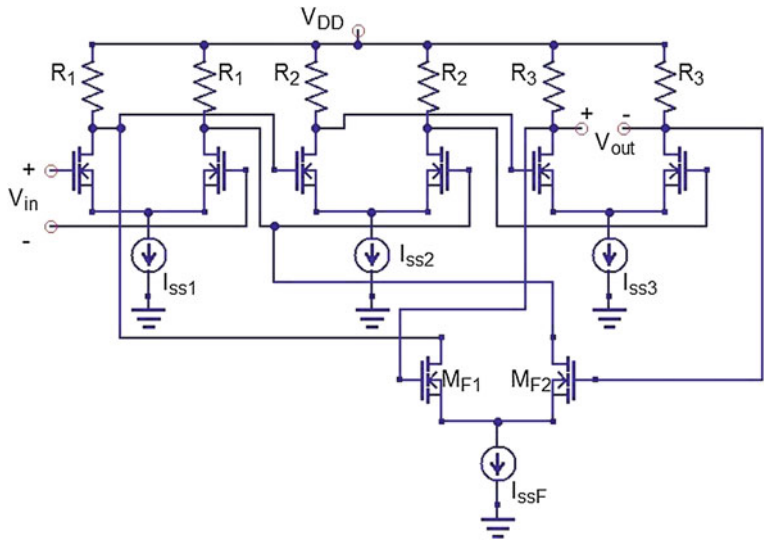

Fig. 8.13 Circuit schematic of third-order gain stages with interleaved active feedback [3]

$$A_{3rd}(S) = \frac{{A_0}^3}{(1 + S/\omega_0)^3 + {A_0}^2\beta} \tag{8.13}$$

where $\omega_0 = 1/\mathrm{RC}$, $A_0 = g_m\,\mathrm{R}$, $\beta = g_{mf}\,\mathrm{R}$, R and C are the resistive and capacitive loads of the differential pairs, respectively, and the transconductance of the cells are represented by g_m and g_{mf}.

The bandwidth increases with the feedback gain β at the cost of a gain reduction. The amount of gain peaking increases with β. For feedback gain equal to 0.3, the bandwidth of the third-order interleaved active feedback improves by a factor of 1.5 with 3.14 dB gain peaking [3]. Decreasing R or C, increasing g_{mf} can extend the amplifier bandwidth, but decreasing R or increasing g_{mf} will reduce the amplifier's gain. The excessive increase of g_{mf} will cause gain peaking and stability problems.

8.8.3 f_t Doubler

An f_T-doubler circuit was patented for the first time in [4]. The f_t doubler topology reduces the input capacitance of the differential stage from C_{gs} to $C_{gs}/2$. For the same gain the bandwidth will be doubled due to the two times reduction in the input capacitance. The f_t doubler topology and its input capacitance is shown in Fig. 8.14.

The biasing voltage V_{Bias} at the gates of M_2 and M_3 is equal to the common-mode voltage of the gates of M_1 and M_4 to cancel any DC offset.

As the gain bandwidth product of a single stage cannot exceed f_t ((8.2) where C_{gs} is larger than C_{gd}); the reduction of C_{gs} by a factor of two will approximately increase the gain bandwidth product to $2f_t$. This is why this topology is called f_t doubler.

As there is no free launch, the f_t doubler has double the power dissipation and double the load current, and double the output node capacitance. Doubling the load current will limit the low voltage supply operation. Increasing the output capacitance will decrease the output pole value and the space between the input and output poles will be decreased.

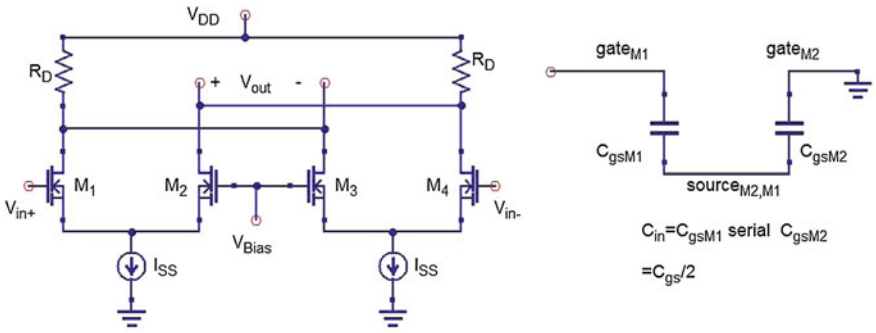

Fig. 8.14 The f_t doubler circuit topology and its input capacitance

References

1. B. Razavi, *Design of Integrated Circuits for Optical Communications* (McGraw-Hill, New York, 2003)
2. E. Säckinger, *Broadband Circuits for Optical Fiber Communication* (Wiley, New Jersey, 2005)
3. J.-C. Chien, H.-Y. Huang, L.-H. Lu, A 10-gb/s inductorless cmos limiting amplifier with third-order interleaving active feedback. IEEE J. Solid-State Circuits **42**(5), 1111–1120 (2007)
4. C.R. Battjes, Amplifier circuit. U.S. Patent 3,633,120, 16 Sept 1973

Chapter 9
Laser and Modulator Drivers

The main elements of an optical transmitter are the devices used for the electrical to optical conversion (laser, VCSEL, modulator …) and the circuit which drives the laser diode (laser driver or modulator driver). In this chapter the different types of the light sources used for optical communication and their characteristics will be discussed. The design considerations of different laser and modulator driver topologies will be presented.

9.1 LEDs, Laser Diodes, and VCSELs

Possible direct modulation light sources for optical communication are light emitting diodes (LEDs) for data rates up to 250 Mbps and resonant cavity light emitting diodes (RC-LEDs) for data rates up to 500 Mbps [1]. For higher data rates laser diodes are the preferred sources. There exist two main types of laser diodes: edge emitting laser diodes (LD) and vertical surface emitting diodes (VCSEL).

LEDs are large-area emitting components. Due to their large area they cannot be modulated very fast (250 Mb/s). Due to their wide emission angle (50°) the coupling loss to fibers is high. LEDs are very cheap and more simple and reliable than laser diodes. LEDs are mostly used in short-range optical communication over multi-mode fibers (MMF).

The resonant cavity LED (RC-LED) is different from the conventional LEDs due to the two additional semiconductor Bragg mirrors above and below the light generation layer. Due to the RC-LEDs special design they allow higher modulation speed (up to 500 Mb/s) than LEDs and a narrower spectral width (10 nm).

Conventional laser diodes (LD), like Fabry-Perot laser (FP) and Distributed-Feedback Laser (DFB), emit collimated light from the edge with a small emission angle which enhances the coupling efficiency to the fiber. FP lasers belong to the class of multi-mode lasers; as the FP lasers' spectrum of the emitted laser light has multiple peaks. The corresponding spectral line width of 2 nm is smaller than thatof

M. Atef and H. Zimmermann, *Optoelectronic Circuits in Nanometer CMOS Technology*, Springer Series in Advanced Microelectronics 54,
DOI 10.1007/978-3-319-27338-9_9

LED and RC-LED. As this 2 nm spectral line width is considered as large for the silica fiber transmission window; FP lasers are used primarily at the 1.3 μm wavelength where dispersion in an single-mode fiber (SMF) is low. The FP laser can be fast as a 10 Gb/s.

The distributed-feedback (DFB) laser has a very narrow line width (0.001 nm). It is an excellent continuous-wave source for external modulators. It can be modulated directly, but with a fairly narrow line width. DFB lasers are suitable for long-reach communication systems (1.55 μm). DFB laser diodes are very fast (up to 40 Gb/s). The laser diodes (FP and DFB) operate only if the current is larger than a threshold current (often more than 15 mA), which is changing strongly with the temperature, see Fig. 9.1. Also a high voltage drop of 1.5–3 V across the LDs is expected. For the operation in data communication an automatic power control (APC) is required, which makes the components more difficult. Because the light is emitted from the edge of the DFB lasers with a non-circular shape; the coupling to a fiber is not easy and on-wafer testing is not possible.

VCSELs radiate like LEDs from the upper surface (vertically) and the light is collimated almost like an edge emitting laser diode which leads to a better coupling efficiency and smaller launching NA to the fiber. There is a threshold current, but smaller than that of edge-emitting LD (typically a few milliamps) and VCSELs have a smaller voltage drop. The advantage of VCSELs over edge-emitting lasers (LD) is that they can be fabricated, tested, and packaged at lower costs than LDs. VCSELs are mostly used in data communication systems operating at short wavelengths (850 nm). Also VCSELs for long wavelengths (1.3 and 1.55 μm) are available. VCSEL is the fastest light source which can work up to DR 55 Gb/s. The VCSELs optical power is smaller than the LD's optical power and work at lower temperature; so a cooling system is required.

The important characteristics for LEDs, RC-LEDs, LDs, and VCSELs are illustrated in Table 9.1.

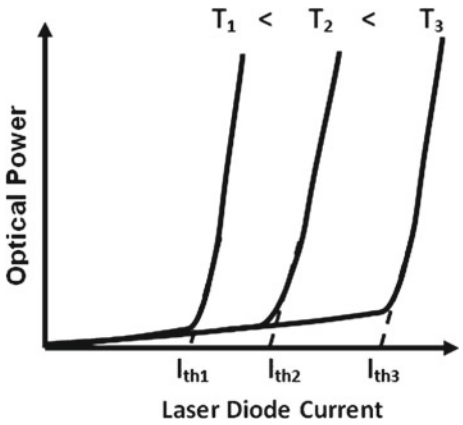

Fig. 9.1 Effect of temperature on the threshold current and the optical power of a laser diode

Table 9.1 Comparison of optical communication different light sources

	LED	LD	VCSEL	RC-LED
Main applications	Short-reach	Long-reach (1.33 and 1.55 μm)	Short-reach (850 nm)	Short-reach
Threshold current	No	More than 15 mA	Around 2 mA	No
Optical power (mW)	2	Up to 20	1	2
Data rate	250 Mb/s	40 Gb/s	55 Gb/s	500 Mb/s
Optical power sensitivity (dB/°C)	−0.05	−0.02	−0.1	−0.03
Wavelength sensitivity (nm/°C)	0.12	0.18	0.06	0.07
Spectral bandwidth	40 nm	2 nm@FP, 0.001 nm@DFB	1 nm	10 nm
Cost	Very cheap	Expensive	Cheap	Cheap
Operating temperature (°C)	85	85	50	85

9.1.1 Small-Signal Model

The small-signal model for light sources like LED, LD, or VCSEL is shown in Fig. 9.2. R_S is the diode's series resistance, C_D is the junction capacitance with a typical value around 2 pF. LDs have a small-signal series resistance of 4–7 Ω [2, 3]. Compared with edge-emitting lasers, the much smaller area and multilayer VCSELs tend to have a larger small-signal series resistance (around 45 Ω [2, 3]) and a smaller junction capacitance. For high frequency modeling the packaging parasitic capacitance C_P, bonding wire inductance L_B and the contact resistance R_C should be included in the model.

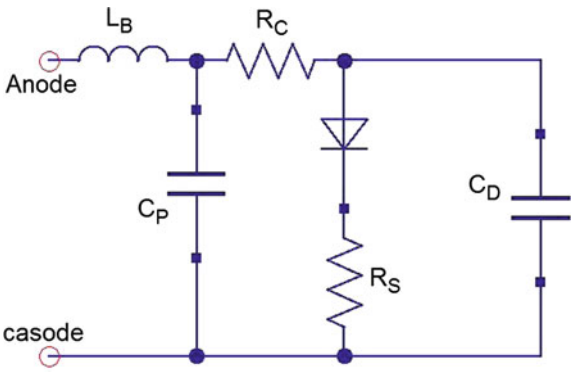

Fig. 9.2 Laser diode small-signal model

9.2 Laser and Modulator Driver

In optical data communication there exist two main principles to implement a light source. The first is a continuous-wave (CW) laser with external modulator and the second method is a direct modulated light source. Direct modulated light sources, using laser driver, offer a simple and low-cost solution for short-range data communication, where chromatic dispersion is not a dominant problem for short-length fiber. The laser driver modulates the current of a laser diode, whereas the modulator driver modulates the voltage across a modulator which in turn modulates the light intensity from a continuous-wave laser (see Fig. 9.3).

Turn on delay (ToD), relaxation oscillation (Sect. 9.3.4) and frequency chirping limit the switching speed of the laser diode. Because of these limitations external modulators are used for high speed (at 10 Gb/s and more) long distance optical communications [4]. The CW light source with the external modulator is commonly used in long range optical data communication in the 1500 nm or 1300 nm wavelength range. The advantage of this modulation technique is that the spectral purity of the emitted modulated light is high, since the laser source works in a constant bias point. The spectral purity is of large interest especially in long-reach applications, since chromatic dispersion limits the useful bandwidth and so the data rate of the link. Unfortunately, fast external modulators for the near infrared (IR) up to the ultraviolet (UV) wavelength range are still too expensive for low cost applications. Also external modulators have a high degree of non-linearity, and linearization techniques are necessary especially for analog application.

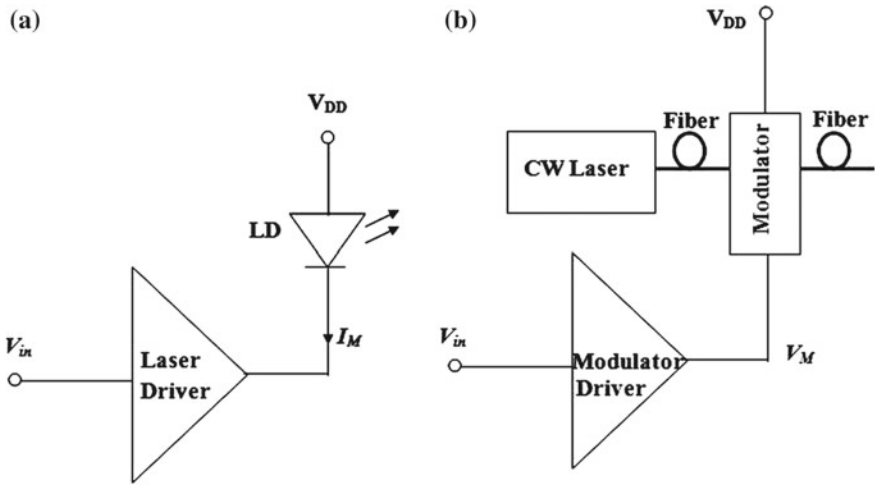

Fig. 9.3 Block diagram for **a** laser driver and **b** laser modulator driver

9.3 Laser Driver Specifications

9.3.1 Rise and Fall Times

The last stage of the laser driver will have a large input signal level, so the laser driver will work in the large-signal mode and not in the small-signal mode. The large input signal will switch the MOSFET (and the laser) ON and OFF, so the switching speed should be as high as the data rate. The definition of small-signal bandwidth will not be valid here and the rise and fall-times are the speed measuring factor, see Sect. 2.8. The maximum switching speed (data rate) can be calculated by:

$$DR = \frac{1}{t_r + t_f} \tag{9.1}$$

9.3.2 Modulation Current

The modulation current (I_{SS} in the laser driver shown in Fig. 9.4) passing thought the laser should be high (up to 100 mA) to generate a high output optical power required for long-reach communications. Such a high current passing through the laser driver increases the challenge for designing high-speed laser drivers especially at low voltage operation in nanometer CMOS technology. The high driver current needs a MOSFETs (M_1 and M_2) with a large gate width which therefore have a large input

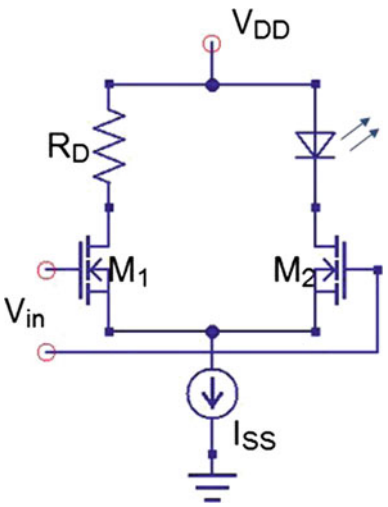

Fig. 9.4 Laser diode driver

and output capacitance. This large input/output capacitance reduces the switching speed. Also there is a large voltage drop across the laser diode (see Sect. 9.3.5) and the dummy load (R_D) resistor due to the high driving current (Fig. 9.4), leading to difficulties with low voltage power supplies.

9.3.3 *Extinction Ratio*

The optical power emitted by the laser diode is modulated by changing the current through the laser diode. The principle is shown in Fig. 9.5.

The conversion factor between modulation current and modulation output power in the laser operating mode is called slope efficiency S_{laser} (in units of W/A). The slop efficiency is the ratio of increase in light power intensity ΔP to the increase in injection (modulation) current ΔI, which is given by [5]:

$$S_{laser} = \frac{\Delta P}{\Delta I} \tag{9.2}$$

The minimum optical power is chosen to be larger than zero to improve the signal response of the laser diode, regarding turn on delay and overshoot. The ratio between

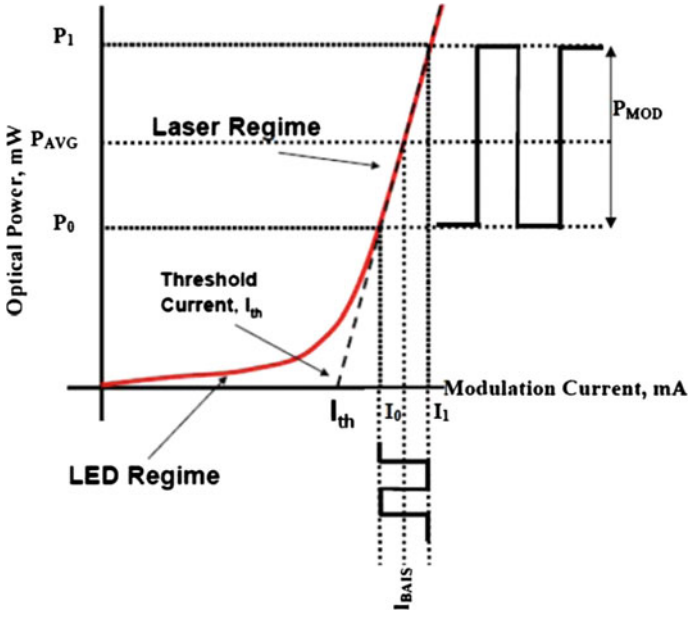

Fig. 9.5 Principle of laser diode modulation with electrical current

maximum optical power and minimum optical power is called extinction ratio (ER) [6, 7] and can practically range from 3 to 10.

$$ER = \frac{power(1)}{power(0)} \tag{9.3}$$

Another important parameter of a laser source is the average emitted laser power P_{AVG}.

$$P_{AVG} = \frac{power(1) + power(0)}{2} \tag{9.4}$$

The relation (9.5) calculates from extinction ratio and average optical power the modulation of the power. This is useful since the extinction ratio is usually known from the source and only average optical power can easily be measured [7].

$$P_{MOD} = 2P_{AVG}\frac{ER - 1}{ER + 1} \tag{9.5}$$

In the ideal case the extinction ratio would be infinite. In this case, the maximum modulation is reached for a given average power and relation (9.5) is reduced to P_{MOD} (ER $\rightarrow \infty$) $= 2P_{AVG}$. Due to the finite extinction ratio, the laser source has a power penalty compared to the ideal case with the same modulation swing. The power penalty in dB can be calculated from relation (9.6), and gives a power penalty of 1.76 dB for an extinction ratio of 5 [7].

$$ER_{power\,penalty} = 10Log\left(\frac{ER + 1}{ER - 1}\right) \tag{9.6}$$

9.3.4 Turn-on Delay (ToD)

As the laser diode turned on, it will produce coherent light only after a certain delay time called turn-on delay (ToD). Due to random spontaneous emission before the start of stimulated emission, a random turn-on delay jitter appears on the rising edge of an optical pulse. Another effect related to turning the laser diode on is the relaxation oscillation. Stimulated emission causes many carriers to recombine to produce more photons and increases the optical power. However, this high recombination rate reduces the carrier density and lowers the optical power. This change in carrier and photon densities causes a damped oscillation (relaxation oscillation). If the laser biasing current is kept close above the threshold current (I_{th}), then the turn on delay and the relaxation oscillation can be much reduced. On the other hand, biasing the laser diode close to the threshold current lowers the extinction ratio, see (9.3). This reduction in the ER is because the P(0) with biasing current near I_{th} will be higher than that P(0) without biasing, see Fig. 9.5.

9.3.5 Output Voltage (Compliance Voltage)

The laser diode has a large voltage drop of 1.5–3 V when it is ON. For blue laser diodes the voltage drop is higher (4 V) and much higher for green laser diodes (6 V). When the laser is OFF a small voltage drop (V_{th} around 0.5 V) which is required to pass the I_{th} through the laser diode. This large voltage drop variation due to laser diode ON and OFF makes the low-voltage driver design challenging. For a small laser voltage drop (OFF) the output voltage (compliance voltage) will be near V_{DD}, see (Fig. 9.4). As a result, V_{DS2} of the output MOSFET M_2 will be high and can exceed its breakdown voltage. The breakdown voltage is limited to the nominal supply voltage which is small in nanometer CMOS technology (around 1.1 V in 40 nm CMOS). For a large voltage drop across the laser diode (laser is ON) the output MOSFET's V_{DS2} will be small and the M_2 may be pushed into the linear region. As a result the laser driver produces large pulse-width distortions and the modulation current may be reduced below its desired value.

9.4 Laser Driver Circuit Design

9.4.1 Predriver

The capacitance of the MOSFETs at the output of the driver (Fig. 9.7) due to their large width can increase the laser driver's rise/fall times. Due to the small input voltage swing to the laser driver input, pre-drivers should be added in front of the output driver to enhance the laser driver's speed.

The pre-driver must be able to drive the large capacitive load of the output driver whereby the pre-driver must have a low input capacitance. The pre-driver also must give a sufficiently large voltage gain to ensure full switching of the output driver. The pre-drivers and the output driver should be jointly optimized for the highest switching speed of the laser driver. Due to the small input voltage swing to the laser driver input (around 100 mV for LVDS) which is not enough to perform a full switching to the LD modulation current (100 mA). The pre-drivers should amplify this small input signal to drive the next stage, see Fig. 9.6. The input voltage signal to the output driver (last stage) will be strong enough to drive the output MOSFETs having a large gate width. This tapered laser driver (Fig. 9.6) design can reduce the effect of the large size output MOSFETs on the laser driver speed.

9.4.2 Output Driver

The laser output driver is the last stage of the laser driver. The output driver MOSFETs have a large width to steer the LD and the dummy load with the tail current I_{SS} (up to 100 mA), see (Fig. 9.7). The output capacitance of the MOSFETs at the output

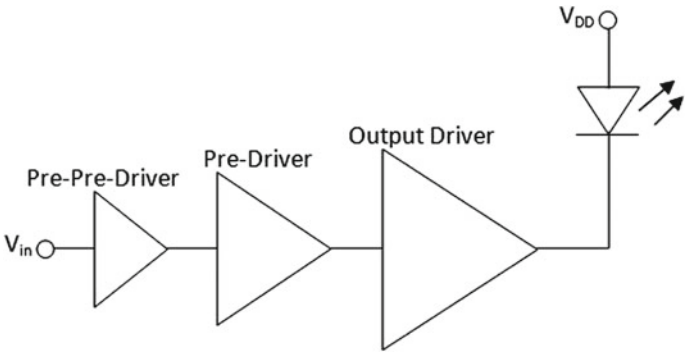

Fig. 9.6 Principle of laser diode modulation with electrical current

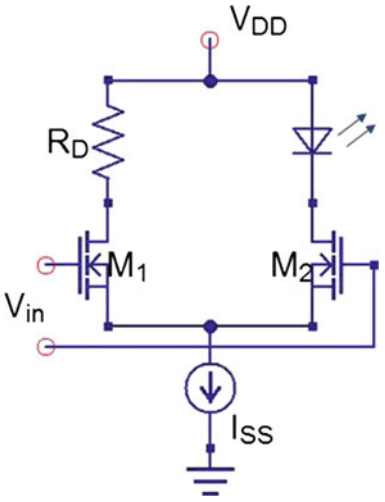

Fig. 9.7 Laser diode driver

of the driver due to their large width can increase the laser driver's rise/fall times. A differential driver is preferred to have all the differential topology advantages. A differential stage has improved immunity to power supply and substrate noise as well as a higher voltage swing compared to a single-ended stage. A differential laser driver is also preferable from stability point of view compared to single-ended output driver. The bonding wires of the power supply and ground pads can make a positive feedback path for single-ended output drivers. This reduces the stability of the single-ended output and some kind of ringing and overshot will happen, see Sect. 8.4.

The broadband techniques presented in Sect. 8.8 to extend the post-amplifier's bandwidth can be applied to the laser drivers to enhance its switching speed. The inductive peaking (Sect. 6.9.2), active inductive peaking (Sect. 6.9.3) and negative

capacitance (Sect. 6.9.4) used to increase the TIA's bandwidth can also be employed to increase the laser driver switching speed. The reader can refer to Sect. 6.9 for these techniques.

Another design alternative for the output driver stage is to use the differential cascode. The cascode will increase the performance through reducing the Miller capacitance ($C_{GD,M1}$) of the input MOSFET. Reducing the capacitance of the output driver MOSFET M_1 allows for easier pre-driver design. On the other hand, the cascode topology requires a higher supply voltage and consumes a higher power than the regular differential pair. The cascode M_C (CG transistor) which acts as a buffer will also contribute noise and unwanted jitter. The regular differential topology has a lower power consumption but the cascode will have a higher speed and higher voltage swing.

9.4.3 High Voltage Laser Driver

The output gain stage of the laser driver is a differential cascode. The M_C cascode increases the gain and reduces the Miller effect ($C_{dg,M1}$). The cascode is composed of a common-source stage (M_1) and a common-gate stag e(M_C). The cascode MOSFETs M_C biasing voltage is controlled via V_B (CG). The CG M_C presents a low input impedance to the drain of M_1 and provides a high impedance at the output node; so the cascode MOSFETs M_C are working as a buffer in this topology. To generate a higher output swing, a higher supply voltage is needed. The cascode of M_C should be implemented as a high voltage device (thick gate oxide) and M_1 as a low voltage device (with thin gate oxide). The use thin oxide MOSFET M_1 as input transistor is to increase the switching speed, while the thick oxide MOSFETs M_C will increase the gain and share part of the high voltage power supply. This will protect M1 from breakdown due to high V_{DS} coming from the high voltage power supply. By controlling the biasing voltage V_B and the size of the cascode M_C, a proper gain and switching speed can be obtained. The differential cascode is recommended to obtain high voltage swing and for the operation with high voltage laser diode (blue laser diodes need 4–5 V). It should be noted that due to the connection of sources and bulks of the MOSFETs in Fig. 9.8, a triple-well CMOS process is needed.

A high-voltage laser driver for GPON communications was designed in 40 nm CMOS technology [8]. The laser driver is fully differential and consists of four main blocks: a 50 Ω input stage, an inverter stage(pre-pre-driver), a pre-driver, and an output laser driver. As shown in Fig. 9.9, the input stage is a common-gate stage with 50 Ω input impedance ($M_{1,2}$, $R_{1,2}$ and $R_{3,4}$) to match the pattern generator output impedance. The output of the GC stage is connected to an inverter stage.

The inverter1 and inverter2 act as a pre-pre-drivers, which is implemented with larger transistors to drive a larger output load capacitance. The target is to boost the logic input signal for the output load of the pre-driver. The pre-driver is formed

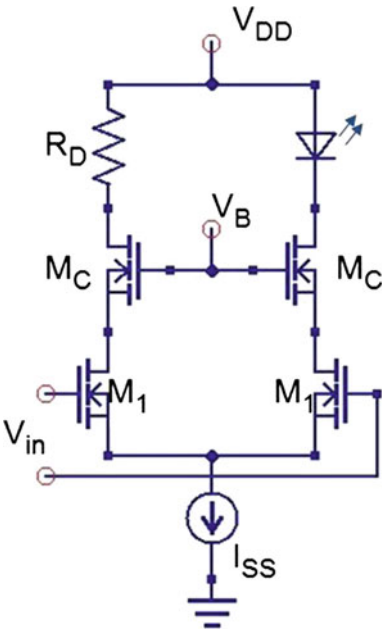

Fig. 9.8 Differential cascode output laser driver

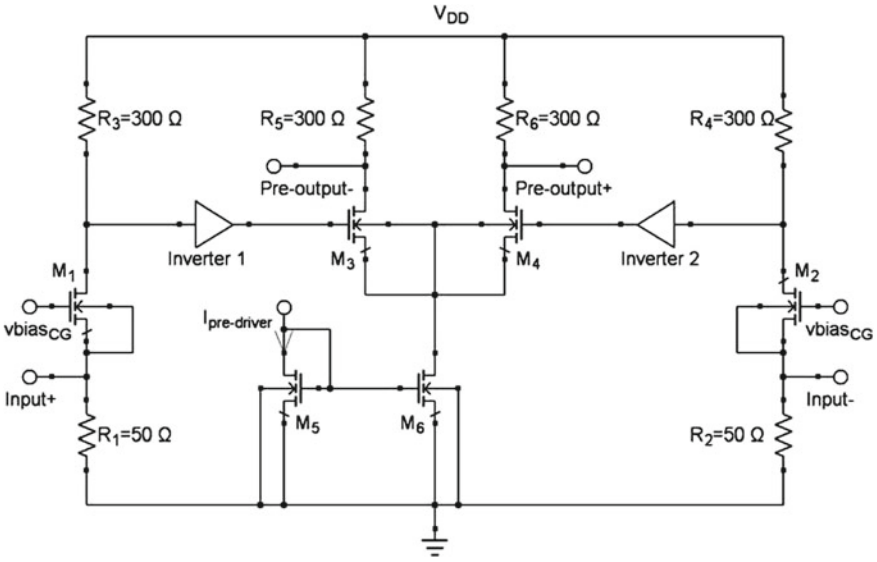

Fig. 9.9 The input and pre-driver stages

by a differential pair ($M_{3,4}$) with resistive loads ($R_{5,6}$). It is used for further signal amplification to drive the high capacitive load of the laser output driver input transistors. The pre-driver stage is optimized in order to keep its pole far away from the output driver pole. All the circuits in the input stage, pre-driver and pre-pre-drivers are biased with 1.1 V power supply and minimum channel length of 40 nm is used.

The laser output driver shown in Fig. 9.10 is designed for a 3.3 V power supply. For speed purposes only thin gate oxide transistors are used with 40 nm channel length for the input MOSFETs $M_{5,6}$. It has an open drain output at which the laser diode with a 5 Ω shunt resistance is connected.

The current through the laser output driver is switched between the two branches: laser diode and dummy resistive load. The laser diode maximum modulation current can be set by the cascoded current source ($M_{11,12,13}$) of the laser driver. Because of strong temperature dependence of the laser diode's optical power; a wide control range of the laser diode biasing current is needed. Laser diode currents up to about 100 mA can be processed by the laser driver. The laser diode current of the laser driver can be set by an external reference current. The high supply voltage operation of the laser driver is dangerous for the thin gate oxide transistors. All operating points

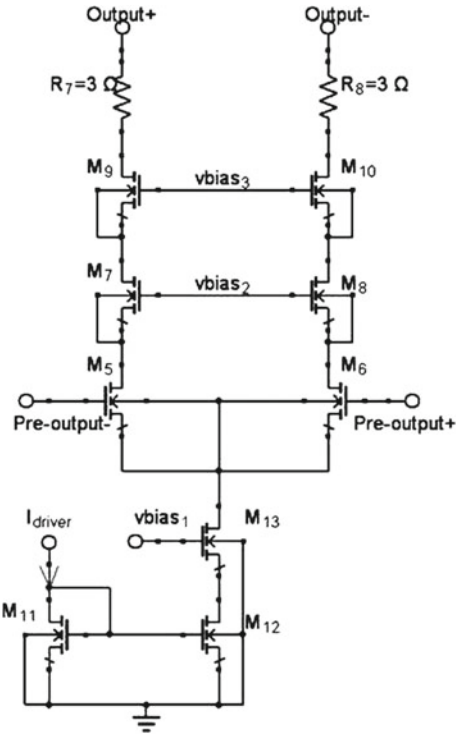

Fig. 9.10 The laser output driver

of the laser driver have to guarantee the voltage breakthrough limits of the transistors. Therefore a doubled cascode structure is implemented. This has two main advantages: the high supply voltage is divided by the cascodes and the Miller effect of the input transistors is reduced and the circuit is faster. Correct bias voltages of the cascodes guarantee the save operation of the laser driver. Monte Carlo simulations were done to ensure proper circuit functionality. Special focus was laid on the ESD protection. The output transistors of the output driver, i.e. the upper cascode transistors $M_{9,10}$, are implemented with a serial resistance. Each finger of the transistor has a resistance of 650 Ω, which is in sum approximately 3 Ω. Additionally silicon-controlled rectifiers (SCR) were added at all input, bias, and output pins to protect the laser driver from ESD actions. RC low-pass filters were applied at the gates of all the biasing transistors for further ESD protection. The circuit was fabricated in 40 nm standard CMOS technology. The area of the laser driver is 1 mm × 0.56 mm (0.56 mm^2). The chip consumes 10 mA and 90 mA from 1.1 V and 3.3 V, respectively. The laser driver test chip photo is shown in Fig. 9.11.

A $2^{31}-1$ non-return-to-zero (NRZ) PRBS was generated from a SYMPULS BPG-40G-128M pulse pattern generator with a data rate of 1.25 Gb/s. The eye diagram of the laser driver was measured using a Tektronix DS A8200 digital serial analyzer. The measured eye diagram is shown in Fig. 9.12.

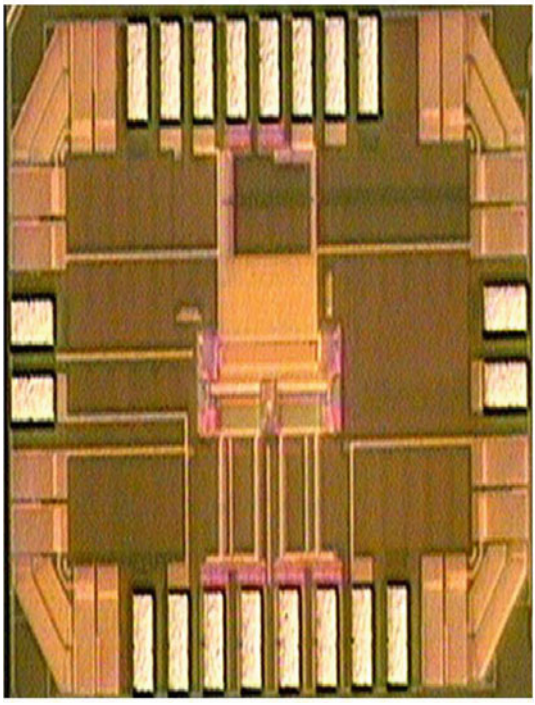

Fig. 9.11 The laser driver test chip photo

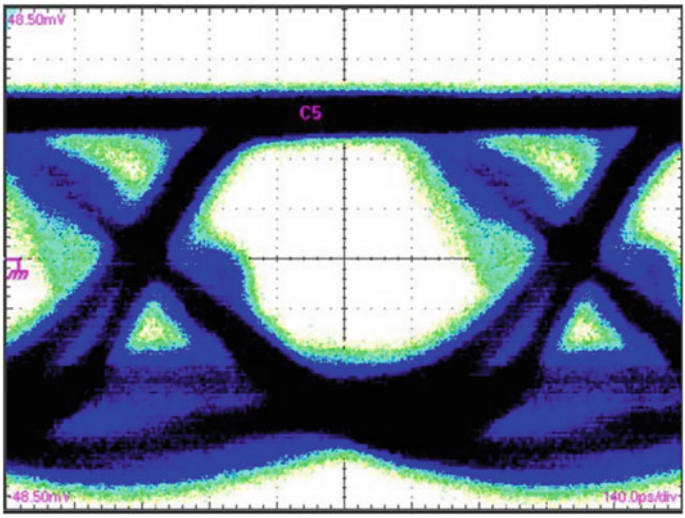

Fig. 9.12 Measured laser driver eye diagram at a $2^{31} - 1$ PRBS with a data rate of 1.25 Gb/s

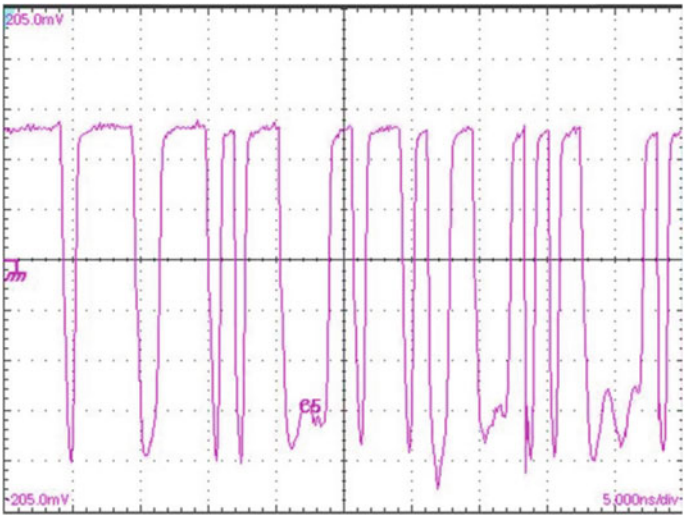

Fig. 9.13 Measured laser driver output signal bit stream at 10 Mbit/s used for rise and fall time measurements

The measured laser driver output signal bit stream at 10 Mbit/s used for rise and fall time measurements is shown in Fig. 9.13. The measured rise time and fall time are 385 and 485 ps, respectively [8]. The limiting part is the laser diode which can work up to a maximum data rate of 1.25 Gb/s.

9.5 Laser Automatic Power Control

Due to temperature variations and aging the threshold current changes and as a result the laser diode's optical power changes. For higher temperature the laser threshold current increases so the optical power decreases, see Fig. 9.1.

The laser's optical power should be constant during its operation. There is a need for the circuit to control the optical power and keep it constant. This circuit is called automatic power control (APC). Figure 9.14 shows the circuitry of a laser driver with an APC.

Most of the commercial LD packages have a monitoring PD beside the LD in the same package. A small fraction of the LD's optical power is coupled to the monitoring PD; the photo current (monitoring current) will give a measure of the LD's emitted optical power. The APC is composed of input resistor R_{APC} to convert the photocurrent coming from the monitoring PD to a voltage. The capacitor C_{APC} smoothes and gives an average voltage V_{Av} related to the sensed optical power by the monitoring PD. An operational amplifier will compare V_{Av} with V_{Ref} and give an output control signal V_C to control the biasing current of the laser diode I_B via tuning the gate voltage of M_B. If V_{Av} is less than V_{Ref} then V_C is low and the LD basing will be increased to increase its optical power. If V_{Av} is greater than V_{Ref} then V_C is low and the LD basing will be decreased to increase its optical power. The resulting

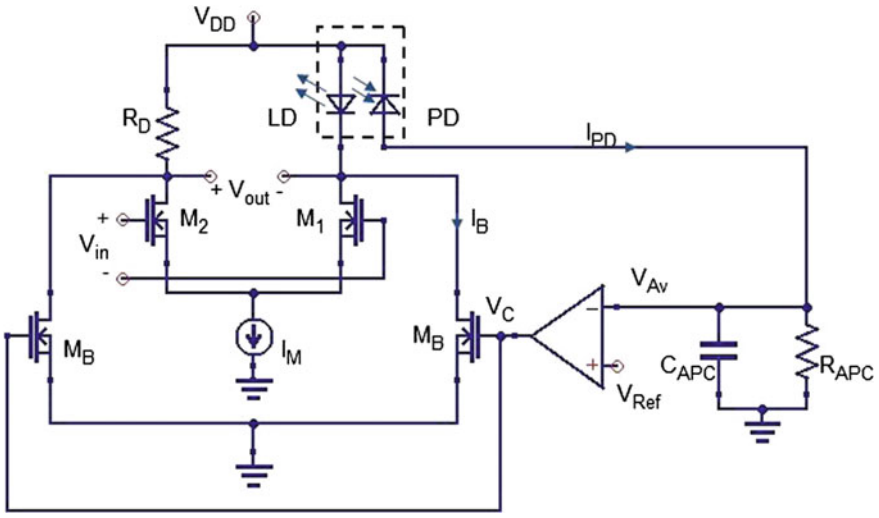

Fig. 9.14 Circuitry of a laser driver with an optical power control unit

control signal V_C adjusts the biasing current I_B such that V_{Av} approaches V_{Ref}. So, the optical power of the LD will remain constant regardless of changes in temperature and aging effects. The DC current source I_M specifies the modulation current.

9.6 Modulator Drivers

9.6.1 External Modulator

Turn on delay, relaxation oscillation and frequency chirping limit the switching speed of laser diodes. For these limitations, external modulators are used for high speed optical communications (at 10 Gb/s and more). The Mach-Zehnder device is a famous type of modulator for high speed optical communication. The propagation velocity of the light traveling through the Mach-Zehnder modulator is a function of the electric field inside it. This electric field is controlled by a external RF signal. As a result the traveling light will experience a certain delay depending on the external control signal. The Mach-Zehnder device has two different paths for light, see Fig. 9.15. In the first path the light travels without any external effects while in the second path the light experiences a certain phase shift due to an external modulating signal. The two paths are added and the net result will be low (ZERO) if the phase shift is 180^o and high (ONE) if there is not any phase shift. The amount of the phase shift is a function of the length of the Mach-Zehnder light path and of the peak value of the modulating signal [4] P-V characteristics of Mach-Zehnder light modulator is shown in Fig. 9.16.

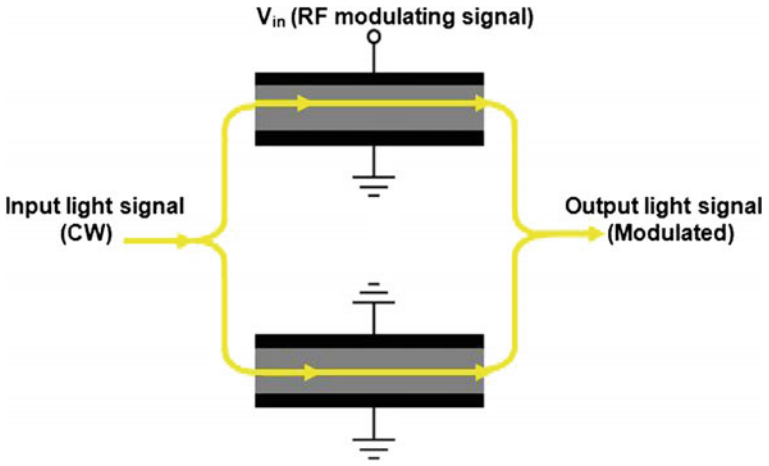

Fig. 9.15 Mach-Zehnder light modulator device

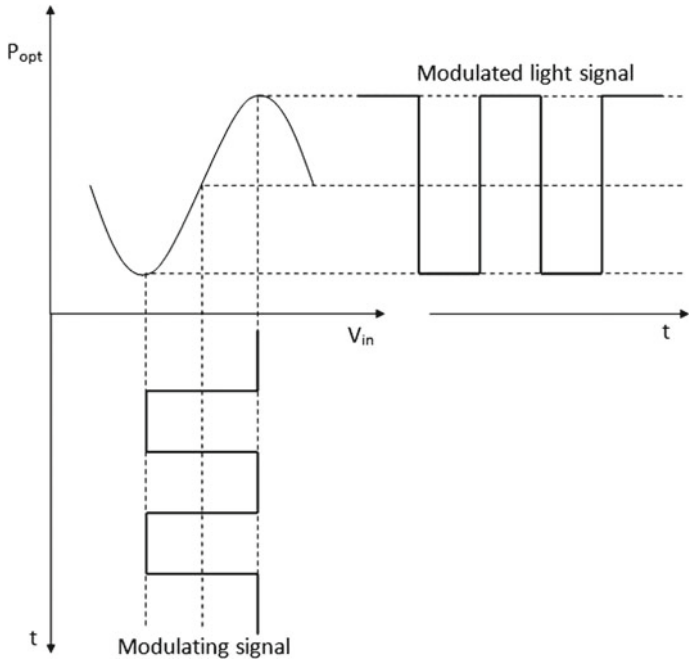

Fig. 9.16 P-V characteristics of Mach-Zehnder light modulator

9.6.2 Modulator Driver Circuitry

The 10 Gb/s laser modulator driver (LMD) fabricated in 180 nm CMOS technology and presented in [9] is illustrated in Fig. 9.17. The LMD needs to deliver around 2 V_{pp} across a 50 Ω external modulator (MZ). The large voltage swing must be delivered to the output stage demanding transistors with large widths, making the design of the pre-driver stage difficult. The pre-driver differential pair M_1 and M_2 is designed to amplify a 400 mV_{pp} single-ended input swing to 600 mV to drive the output stage M_3 and M_4. The design introduced two techniques to increase the bandwidth at the interface between the pre-driver and the output driver. The pre-driver employs an inductive peaking using 3 nH on-chip T-coil coupled inductors. Due to the low resistance value of $R_L = 40\,\Omega$, T-coils can increase the bandwidth by a factor of 2.2 even with the on-chip spiral coil. Inductors L_S create series peaking at the drains of the M_1 and M_2, partially canceling the effect of their drain junction capacitances. The output driver also employs negative capacitance (NIC) cancellation at nodes X and Y by the cross-coupled pair M_5 and M_6 and compensation capacitors C_C, see Sect. 6.9.4. The output driver tail current ($I_T = 90$ mA) is determined after iterations between the pre-driver and output driver to select the optimum value. The output differential

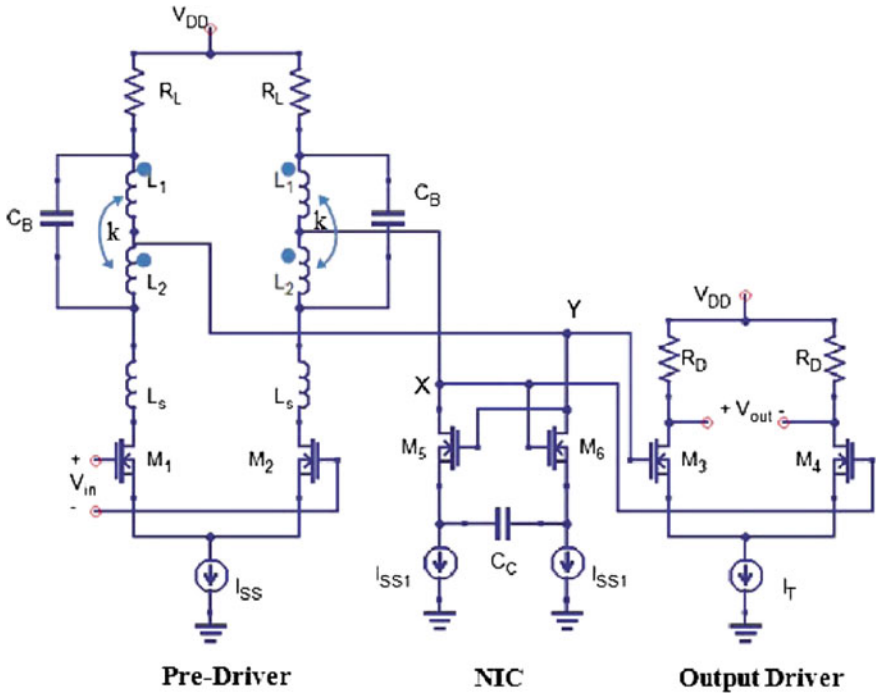

Fig. 9.17 Circuitry of a 10 Gb/s laser modulator driver (LMD) fabricated in 180 nm CMOS technology [9]

pair consisting of M_3 and M_4 having 600 μm widths. To match the 50 Ω off-chip load a 75 Ω on-chip resistor (R_D) termination is used; as the MOSFET output resistance (in parallel with the 75 Ω) will reduce the output stage's impedance to 50 Ω.

References

1. M. Atef, H. Zimmermann, *Optical Communication Over Plastic Optical Fibers: Integrated Optical Receiver Technology* (Springer, Berlin Heidelberg, 2013)
2. E. Säckinger, *Broadband Circuits for Optical Fiber Communication* (Wiley, New Jersey, 2005)
3. M.C.O' Farrell, Laser Driver Design in 0.18 m CMOS Technology. Master thesis, Queen's University, Kingston, Ontario, Canada, Sept 2010
4. B. Razavi, *Design of Integrated Circuits for Optical Communications* (McGraw-Hill, New York, 2003)
5. T. Numai, *Fundamentals of Semiconductor Lasers* (Springer, Berlin, 2004)
6. J.L. Miller, E. Friedman, *Optical Communications Rules of Thumb* (McGraw-Hill, 2002)
7. B. Chomycz, *Planning Fiber Optics Networks* (McGraw-Hill, 2008)
8. M. Hassan, H. Uhrmann, S. Schidl, H. Zimmermann, A 3 Gb/s high-voltage laser diode driver in 40 nm CMOS technology, in *Proceedings of the Microelectronic Systems Symposium* (MESS) (2014)
9. S. Galal, B. Razavi, 10-Gb/s limiting amplifier and laser/modulator driver in 0.18-um CMOS technology. IEEE J. Solid-State Circuits **38**(12), 2138–2146 (2003)

Chapter 10
Optoelectronic Circuits in Nanometer CMOS Technology

In this chapter three fully integrated optical receivers down to 40 nm CMOS are described. In addition two optical receivers with off-chip photodiode follow. Finally optical sensors are introduced to complete the application spectrum of optoelectronic nanometer CMOS circuits: one two-dimensional (2D) image sensor, two three-dimensional (3D) image sensors and a positron-emission tomography (PET) sensor for medical applications.

10.1 Fully Integrated Optical Receivers

Low cost and high speed optical receivers are required in modern telecommunication system. Low cost and high speed optical receivers can be extensively applied to short-distance optical communications, such as local area network, chip-to-chip and board-to-board interconnects. Fully integrated optical receivers with integrated silicon photodiodes provide some advantages over hybrid implementations, including low-cost as one chip will replace a multi-chip solution, reduced parasitic capacitance and no bond-wire inductance thanks to the integration of the photodiode. The fully integrated optical receiver is illustrated in Fig. 10.1c. The performance and cost improvement of using fully integrated optical receiver over the discrete solution was detailed in Chap. 1 Sect. 1.5.

M. Atef and H. Zimmermann, *Optoelectronic Circuits in Nanometer CMOS Technology*, Springer Series in Advanced Microelectronics 54,
DOI 10.1007/978-3-319-27338-9_10

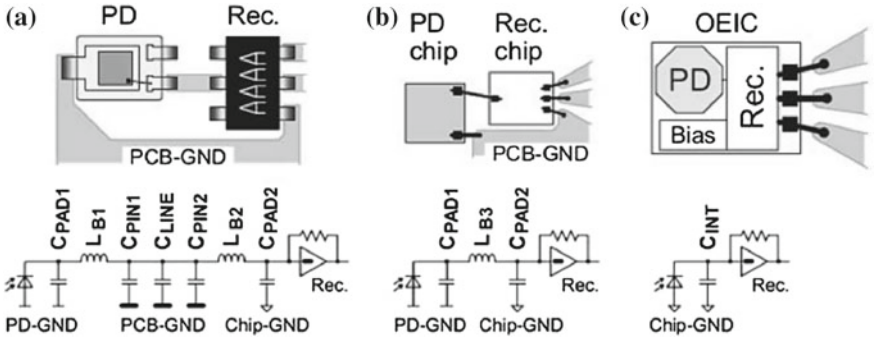

Fig. 10.1 Different ways of connecting the PD and the receiver: **a** packaged photodiode connected to packaged receiver, **b** photodiode die connected to receiver die via bond wires, **c** fully integrated optical receiver [1]

10.1.1 180 nm CMOS Fully Integrated Optical Receiver

The block diagram of the fully integrated optical receiver introduced in [2] is illustrated in Fig. 10.2. The fully integrated optical receiver fabricated in 180 nm CMOS is consisting of a integrated differential PD, a differential TIA, a Cherry Hooper (CH) NIC, an equalizer to enhance the intrinsic speed of the integrated PD, a limiting amplifier (LA) as post amplifier (PA) with offset-compensation followed by an f_T doubler as output buffer.

The differential photodiodes (like SML PD presented in Sect. 5.6) is integrated in this design for its high speed; due to the elimination of the carriers coming from slow carrier diffusion, see Sect. 5.6. The differential PD consists of 18 fingers, 9 fingers are active and 9 fingers are shielded with metal. To minimize the substrate resistance

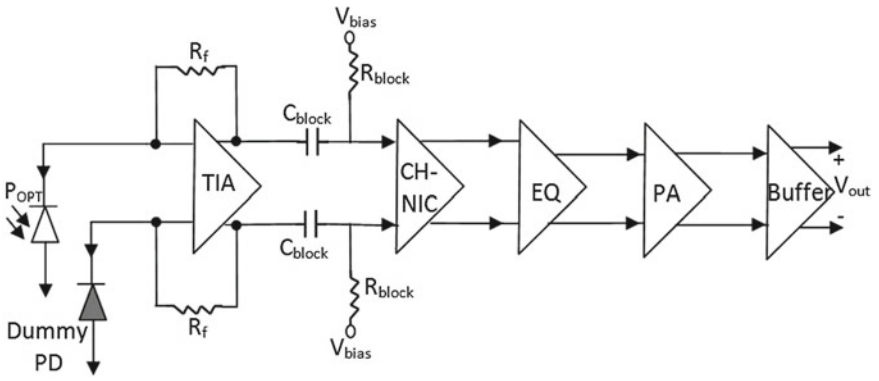

Fig. 10.2 Block diagram of the described 180 nm CMOS fully integrated optical receiver presented in [2]

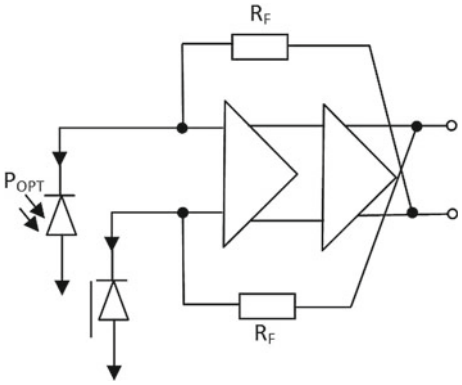

Fig. 10.3 TIA with two stages differential cross-coupled feedback common-source topology [2]

more substrate contacts are placed between the fingers, also every finger is cut into two parts to place extra substrate contacts, finally the PD is surrounded by a broad guard ring of substrate contracts. The junction capacitance of the active PD and the shielded one is equal to 159 fF with a reverse bias of 1.3 V.

The TIA is consisting of two cascaded stages of differential common-source topology (see Fig. 10.3). Cross-coupled feedback is applied to have a negative feedback for these two (even number) differential amplifying stages. The DC-levels at the outputs of the TIA are not constant and are a function of the optical power. A high-pass filter, C_{block} and R_{block}, is needed to set the DC-level at the input of the next stage. In this design, $C_{block} = 2$ pF and $R_{block} = 20\,\text{k}\Omega$, which gives a 3-dB frequency of 4 MHz.

The TIA is followed by an amplifier in Cherry-Hooper topology (CH) (Sect. 8.8) with negative impedance converter (NIC) (Sect. 6.9.4) to enhance bandwidth. The CH with NIC stage works as a subtractor to suppress the common-mode signal. The illuminated PD current consists of a fast drift and slow diffusion component, while the current from the dark junctions only consists of the diffusion component. After this differential CH with NIC stage, the difference signal is consisting only of the fast drift component.

The equalizer will compensate for the frequency roll-off of the integrated differential PD. The equalizer is a pseudo-differential stage, with second-order capacitive source degeneration (see Sect. 7.3 for capacitive source degeneration) to create zeros at the desired frequencies.

At 6 Gbit/s, an input voltage of 8 mV_{pp} was needed to achieve a BER of 10^{-12} [2]. The dynamic range is 37.5 dB at a data rate of 6 Gbit/s and with a BER of less than 10^{-12}. The total chip area, including bond pads and on-chip decoupling capacitors, is 1.5×3 mm^2. No additional masks or anti-reflective coatings have been used to enhance the PD's responsivity. The measured PD DC responsivity was 0.03 A/W. The current consumption of the complete front-end was 250 mA from a 1.8 V supply. The TIA and LA consume 40 mA each. 170 mA are dissipated in the output buffer to generate 500 mV_{pp} output signals. The optical measurements of the fully integrated

optical receiver were performed with a 850 nm laser source, electrically driven with a $2^7 - 1$ PRBS. For 10^{-12} BER and bit rates of 1.7 Gbit/s the sensitivity was −6 dBm [2].

10.1.2 65 nm CMOS Fully Integrated Optical Receiver

A fully integrated optical receiver fabricated in 65 nm standard CMOS technology was presented in [3]. The block diagram of the integrated optical receiver is shown in Fig. 10.4. The integrated PD is a NW/P-substrate PD fabricated in a standard 65 nm CMOS technology. In order to interface the large-core POF (125 μm to 1 mm), the PD is sized at 250 μm by 250 μm, consisting of 56 rectangular NW/P-substrate PD stripe cells. The integrated photodiode's capacitance is 14 pF at the target biasing voltage of 0.3 V; there is no external biasing for the PD. This high PD capacitance (14 pF) is requiring a very low input impedance TIA [3] to achieve the desired bandwidth. A hybrid current buffer based super-Gm TIA (SGM-TIA) has been proposed to cope with the large PD capacitance [4], see Sect. 6.9.1 for more details about the TIA. The illuminated PD and the dummy PD together with a differential super-Gm TIA convert the single-ended input light power into differential output voltages. A subtraction amplifier (S-AMP) amplifies the TIA output signals before the VGA. A two-stage continuous-time linear equalizer (CTLE) compensates for the limited bandwidth of the NW/P-sub PD, see Sect. 7.4 for the CTLE. An offset-cancellation network (OCN1) feeds the output voltages of the first MUX to the input of the SGM-TIA to allow cancellation of the DC offset. A LA is placed behind the CTLE to drive the current-mode output buffer (BUF). Another offset cancellation network (OCN2) is added to cancel the offsets inside the LA. High-speed current-mode two-to-one MUX and buffer amplifiers (AMPX/Y) are added into the POF receiver as auxiliary test circuits to allow bypassing the CTLE and the LA during characterization in the lab [3].

The active chip area is 0.24 mm^2 and the optical receiver consumes 46 mW from a 1 V DC power supply. The fully integrated optical receiver demonstrates a speed of 4 Gbps at a BER of $<10^{-12}$, at transmitted optical power of −3.2 dBm from a 2.5 Gb/s grade 670 nm VCSEL [3].

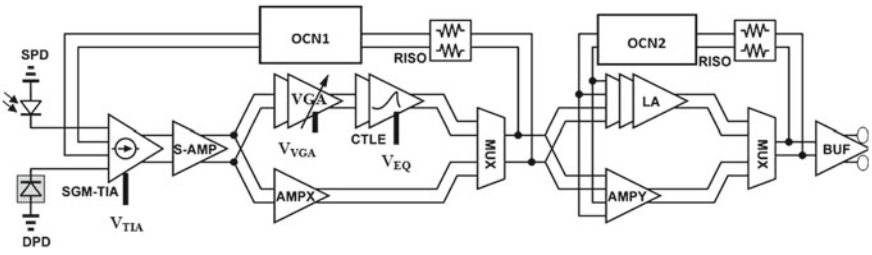

Fig. 10.4 Block diagram of a fully integrated POF receiver with 250 μm NW/P-sub integrated PDs fabricated in 65 nm CMOS [3]

10.1.3 40 nm CMOS Fully Integrated Optical Receiver

The presented fully integrated optical receiver described in [5] has been integrated in a 40 nm standard CMOS technology with 1.0 V supply voltage together with the N+/P-sub PD having an active area of 70 μm × 70 μm. The PD is biased with −5 V by using a negative substrate voltage. This is possible because all circuits are placed inside a deep N-well of simple N-well. The fully integrated optical receiver block diagram is shown in Fig. 10.5.

The optical receiver circuitry is shown in Fig. 10.6. The circuit operation is described in Sect. 5.6.

Figure 10.7 shows die photo of the proposed NC-TIA with the post amplifiers and the 50 Ω driver with integrated 70 μm × 70 μm N+/P-sub PD. The NC-TIA alone consumes 4.1 mW, and the total chip power dissipation is 12 mW. The output noise density measured by the spectrum analyzer is shown in Fig. 10.8. The measured input

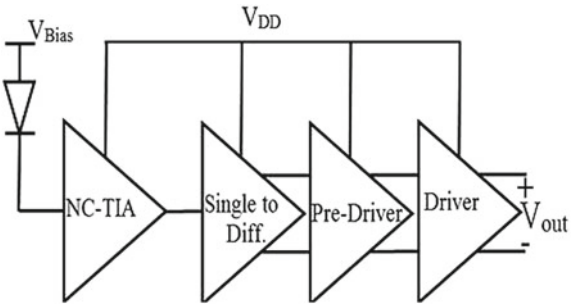

Fig. 10.5 Block diagram of fully integrated optical receiver with integrated 70 μm × 70 μm N+/Psub PD fabricated in 40 nm CMOS [5]

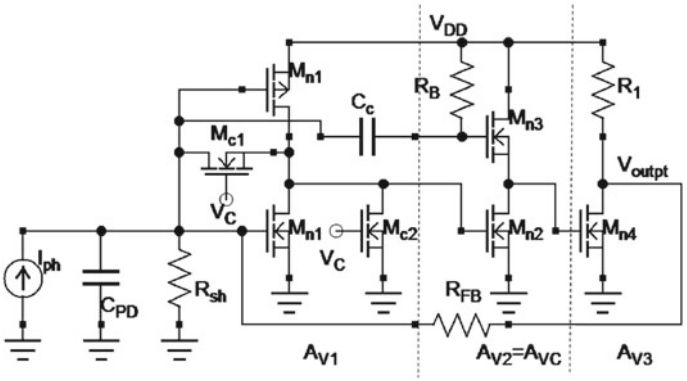

Fig. 10.6 Circuitry of the presented transimpedance amplifier using a noise cancellation technique (NC-TIA) [5, 6]

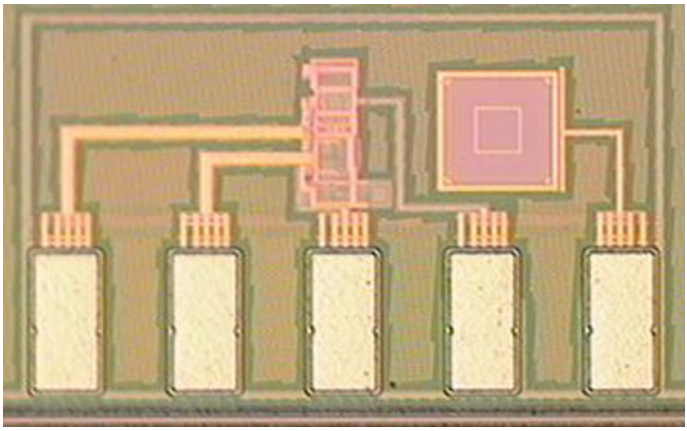

Fig. 10.7 Die photo of the fully integrated optical receiver fabricated in 40 nm CMOS with integrated 70 μm–70 μm N+/Psub PD [5]

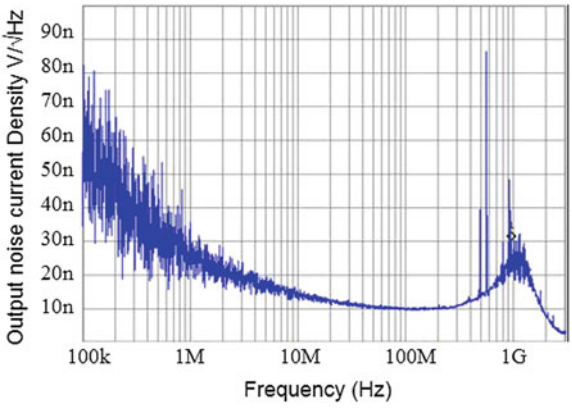

Fig. 10.8 Measured input referred noise current density [5]

referred noise current density is 7.2 pA/$\sqrt{Hz}$ and the integrated input referred noise current up to 1.5 GHz is 280 nA [5].

The optical receiver's frequency response was measured with a vector network analyzer (VNA). One of the optical receiver's differential output terminals was connected to the VNA and the other was 50 Ω terminated. A calibration was performed using a commercial 12 GHz photodiode to cancel the effect of the laser source and of the connected measurement cables on the receiver's frequency response. A flat band frequency response up to 12 GHz was obtained with the connected cables, the laser source, and the 12 GHz PD. Then the 12 GHz PD was replaced by the optical receiver under test. Figure 10.9 shows the measured frequency response for the optical receiver with an integrated PD using a 520 nm laser source. The measured −3 dB bandwidth is 1.5 GHz using the 520 nm laser source. The bandwidth of the optical

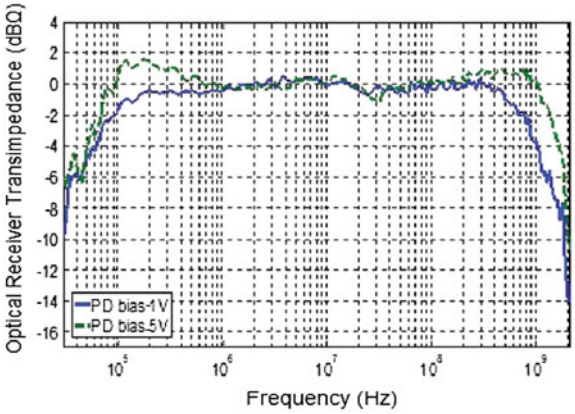

Fig. 10.9 Measured frequency response for the optical receiver with an integrated PD using 520 nm wavelength [5]

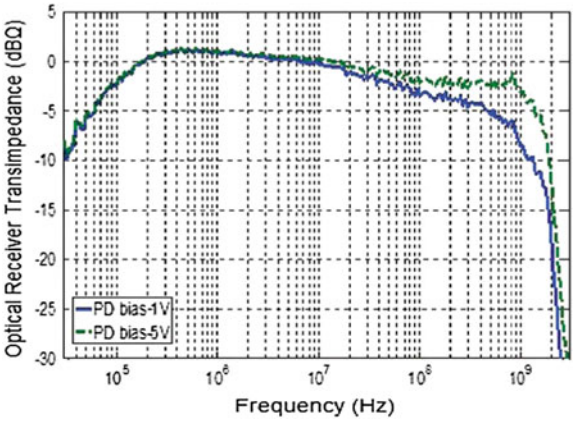

Fig. 10.10 Measured frequency response for the optical receiver with an integrated PD using 660 nm wavelength [5]

receiver with the integrated PD is 1.3 GHz for 660 nm due to carrier diffusion from the substrate of the N+/P-sub PD, see Fig. 10.10. The optical receiver shows a total transimpedance gain of 8 kΩ (78 dBΩ).

Sensitivity measurements were carried out at 660 nm wavelength. For these measurements a 660 nm VCSEL was modulated by a PRBS with a length of $2^7 - 1$. The BER was measured with a bit error rate test (BERT) set. Figure 10.11 shows the measured BER as a function of the average received optical power at a data rate of 2 Gbps and a PRBS of $2^7 - 1$. A sensitivity of −9.5 dBm is achieved at BER $= 10^{-12}$. Figure 10.12 shows an eye diagram of the differential output (the difference was built by the oscilloscope) at 2.5 Gbit/s with PRBS $= 2^7 - 1$ and with −9.5 dBm average received optical power.

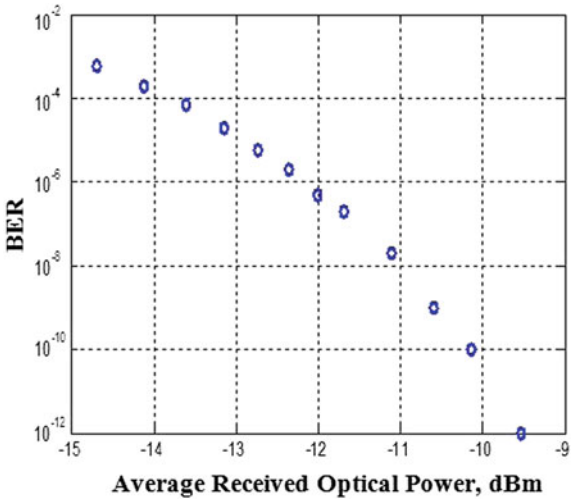

Fig. 10.11 Measured BER as a function of the average received optical power at a data rate of 2.5 Gbps and PRBS $= 2^7 - 1$ [5]

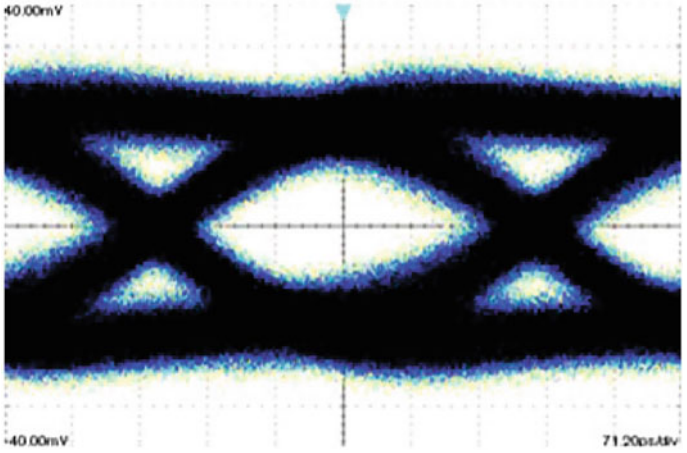

Fig. 10.12 Eye diagram of the differential output at 2.5 Gbit/s with PRBS $= 2^7 - 1$ and with -9.5 dBm average received optical power [5]

Table 10.1 compares the performance of the optical receiver based on the NC-TIA topology with integrated PD with other recently published Gigabit optical receivers with integrated PDs. It is clearly seen from Table I that the NC-TIA shows superiority in terms of Figure of Merit (FoM_1) which is defined by [7]:

Table 10.1 Comparison of nanometer CMOS Gigabit optical receivers having integrated PDs [5]

	[8]	[9]	[7]	[10]	[4]	[5]
Technology	180 nm CMOS	180 nm CMOS	130 nm CMOS	90 nm CMOS	65 nm CMOS	40 nm CMOS
Supply voltage (V)	1.8/3.3	1.8	1.2	1.2	1.2	1
Transimpedance gain (dBΩ)	110	80	105	92	78.5	78
Data rate Gb/s	2.5	5	4.5	3.125	3.125	2.5
Sensitivity@BER = 10^{-12} (dBm)	−4.5	−7.5	−3.4	−3.7	−3.8	−9.5
PD diameter (μm)	50	75	60	75	250	70
Power dissipation (mW)	138	183	74.16	46.3	50	12
F.O.M.1	8.3	34.1	55.5	192	3309	5673

$$FoM_1 = \frac{Gain\,(\text{dB}\Omega) \cdot Log\,|\text{BER}| \cdot Sensitivity\,(\text{dBm}) \cdot DR\,(\text{Gb/s}) \cdot PDDiameter^2\,(\mu\text{m}^2)}{Power\,Dissipation\,(\text{mW}) \cdot Technology^2\,(\text{nm}^2)} \tag{10.1}$$

10.2 Infrared Optical Receivers with External Photodiode

GaAs, InGaAs, InP, and Ge and related compound materials are interesting for the realization of high-speed and high-responsivity photodiodes. III–V photodetectors allow much higher data rates and responsivity than Si photodetectors. For IR optical receiver the III–V photodiode is a must (see Chap. 4) as Si cannot detect wavelengths larger than 1.1 μm. High speed III–V photodetectors can be combined with silicon CMOS integrated circuits in hybrid OEICs to have the benefits of high integration and low cost of Si circuits and the high performance of III–V photodetectors.

10.2.1 Infrared Optical Receiver in 90 nm CMOS with External Photodiode

The integrated optical receiver introduced in [11] has been fabricated in a 90 nm standard CMOS technology with 1.2 V supply voltage for the TIA and 1.6 V for the output stages. Block diagram of complete optical receiver is shown in (Fig. 10.13). The optical receiver uses a multistage inverter based TIA to increase the gain-bandwidth product (Fig. 10.14), see Sect. 6.5.5 for more details. The optical receiver is optimized for an off-chip Ge waveguide PD [12]. The PD has a bandwidth being sufficient for a data rate of 40 Gb/s and a responsivity of 0.8 A/W for 1.55 μm light.

The total optical receiver's input capacitance including PD capacitance, ESD protection and input bondpad is 300 fF. Finally, the output buffer amplifier provides enough driving capability for the 50 Ω input measurement devices. Figure 10.15 shows the circuit diagram for the post amplifiers and the 50 Ω driver. The post amplifiers and the 50 Ω driver consume 30 mA from the 1.6 V supply.

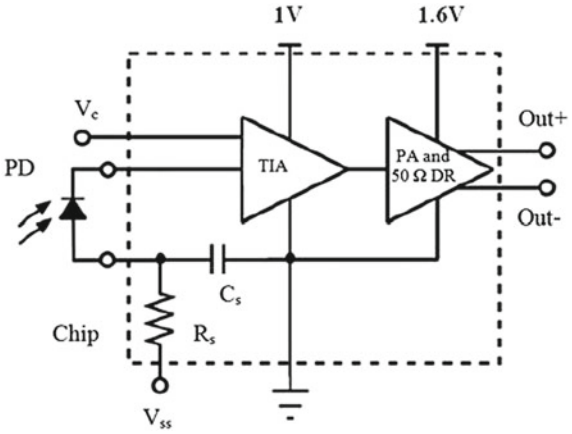

Fig. 10.13 Block diagram of complete optical receiver [11]

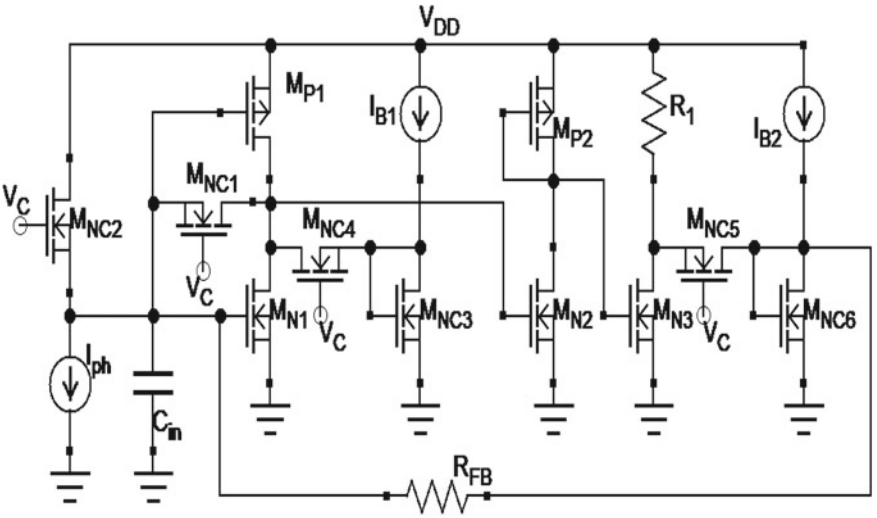

Fig. 10.14 Circuitry of multi-stage inverter based TIA with compression and stability compensation stages [11]

The multistage-inverter based TIA alone consumes 12 mW and the total chip power dissipation is 60 mW. The block diagram of complete optical receiver is shown in Fig. 10.13. The total chip area is 0.46 mm^2 and the TIA occupies only 0.0036 mm^2. The die photo is shown in Fig. 10.16.

For the frequency response measurements a VNA was used to drive a 1.55 μm external laser modulator. A single mode fiber with a core diameter of a 9 μm was coupled to the proposed optical receiver. The measured frequency response of the optical receiver is shown in Fig. 10.17.

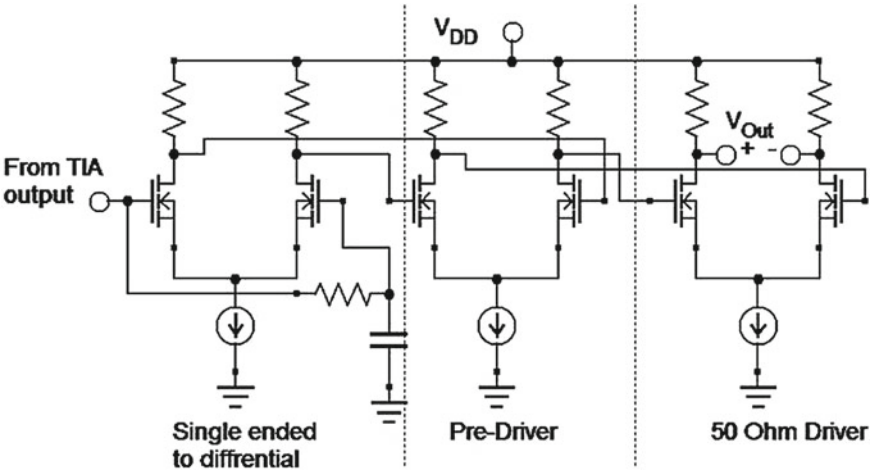

Fig. 10.15 Circuit diagram of the post amplifiers and the 50 Ω driver [11]

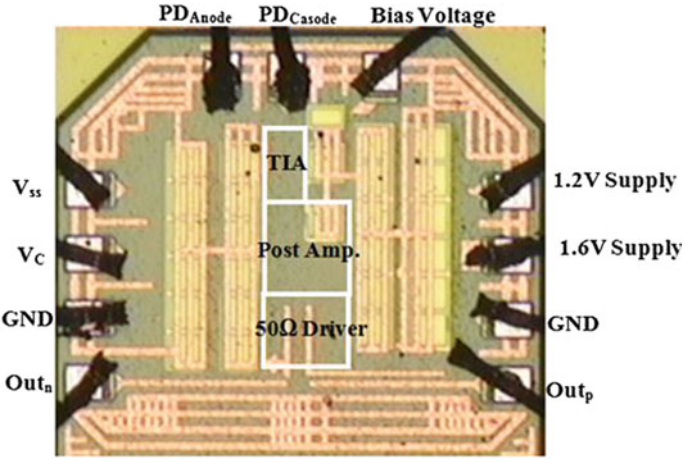

Fig. 10.16 Die photo of the optical receiver fabricated in 90 nm CMOS [11]

The optical receiver's measured bandwidth is equal to 5 GHz. The overall transimpedance is 72 dBΩ. The TIA alone has a simulated bandwidth of 6 GHz and a transimpedance value of 63 dBΩ. The measured average input noise current is 2 μA. For the BER measurements the 1.55 μm external laser modulator was modulated by a PRBS signal of length $2^{15} - 1$. The received data stream from the integrated optical receiver was compared with the original transmitted signal by bit-error counting using a BERT. The measured BER in dependence on the received optical power at different data rates (8 and 10 Gbit/s) is illustrated in Fig. 10.18.

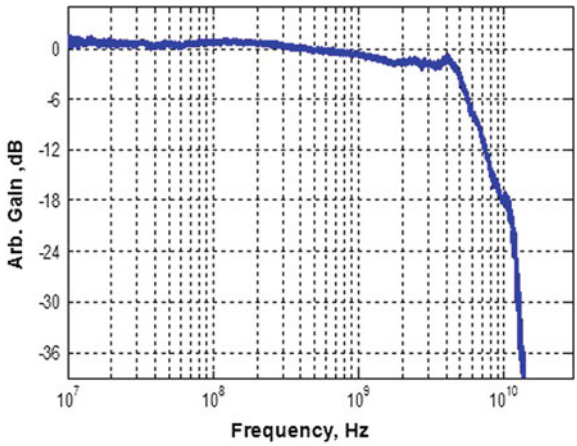

Fig. 10.17 The measured frequency response of the integrated optical receiver in 90 nm CMOS [11]

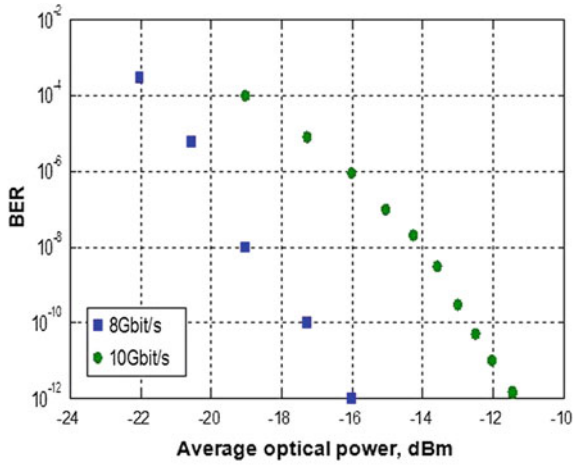

Fig. 10.18 The measured BER of the integrated optical receiver in 90 nm CMOS [11]

A sensitivity of −18.3 dBm is obtained for the presented optical receiver for BER $= 10^{-9}$ at a data rate of 8 Gbit/s. A sensitivity of −13.3 dBm is obtained for BER $= 10^{-9}$ at a data rate of 10 Gbit/s. Figure 10.19 shows eye diagrams at PRBS = $2^{15} - 1$ and data rates of 8 and 10 Gbit/s with an average received optical power of −18.3 dBm and −13.3 dBm, respectively. To study the effect of the compression technique on the dynamic range, the optical power was increased to obtain up to 500 μA photocurrent and the output eye diagrams are displayed in Fig. 10.19.

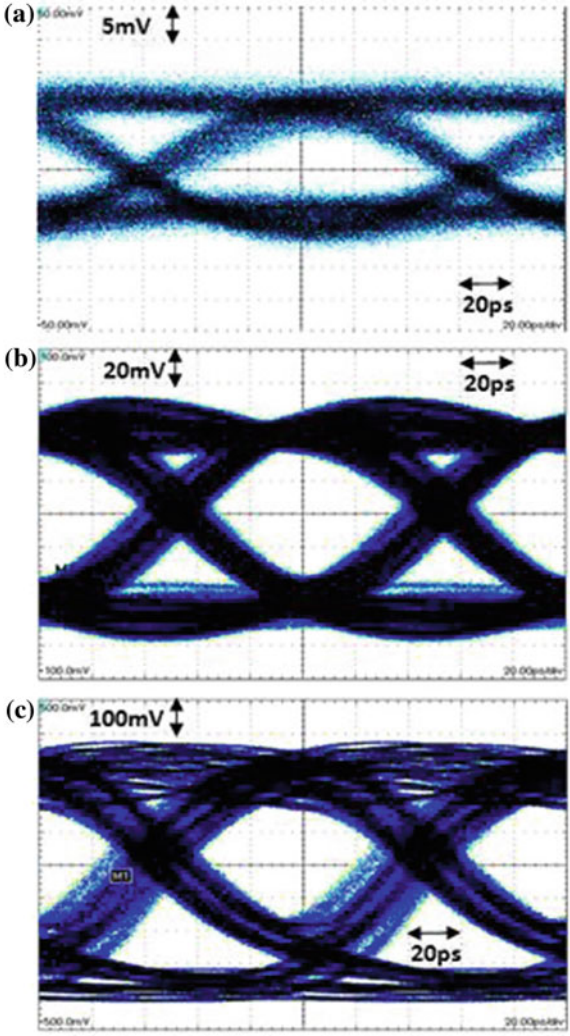

Fig. 10.19 Eye diagrams at PRBS $= 2^{15} - 1$ and data rate of **a** 8 Gbit/s with an average received optical power of −18.3 dBm, **b** 10 Gbit/s with an average received optical power of −13.3 dBm, **c** 10 Gbit/s with an average received optical power of −2 dBm [11]

It is clear from Fig. 10.19 (a, b, c) that the photocurrent current can go from 12 μA (optical power −18.3 dBm) to its maximum of 500 μA (optical power −2 dBm) without any stability problems. Figure 10.20 shows the peak to peak output voltages for different generated photocurrents at 8 Gbit/s. The gain is constant and equal to 4.1 kΩ up to 160 μA photocurrent. When the photocurrent increases, the gain decreases and reaches 1.5 kΩ at 500 μA.

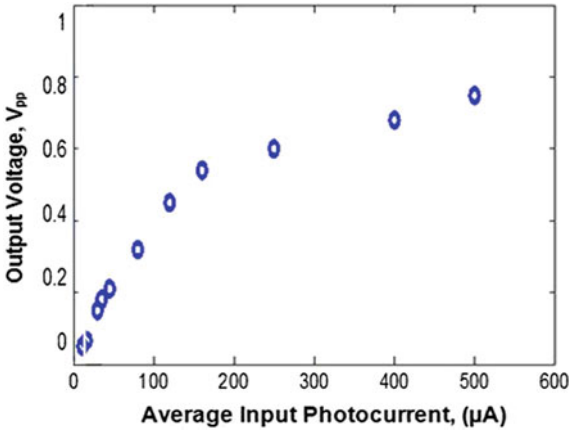

Fig. 10.20 The peak to peak output voltages for different generated photocurrents at 8 Gbit/s [11]

10.2.2 Infrared Optical Receivers in 40 nm CMOS with External Photodiode

The ICDF TIA (Fig. 10.21) was investigated in 40 nm CMOS technology and the values of the TIA elements were optimized to reach a bandwidth of 8 GHz at the largest possible transimpedance gain to minimize the noise. The transistor's sizing, resistor's values, and biasing currents were swept to reach the desired specifications. The reader can see Sect. 6.4.3 for ICDF-TIA details. A post amplifier and a 50 Ω driver were integrated to have enough gain and to interface to the measurement equipment. The gain stage amplifies the signal from the transimpedance stage and provides the signal isolation and buffering between the transimpedance stage and output buffer to avoid that the transimpedance stage directly connects to a large input capacitance of the output driver and thus results in reduction of its bandwidth. Finally, the output buffer amplifier provides enough driving capability for the 50 Ω input measurement devices. Figure 10.22 shows the circuit diagram for the post amplifiers and the 50 Ω driver.

The differential output stage is better than a single-ended one with respect to common-mode rejection and power-supply noise. The ICDF-TIA single-ended output is fed to a differential post amplifier. The second input for the differential post amplifier is biased through a low-pass filter coming from the TIA output. The next stage after the first post amplifier is a pre-driver stage to increase the gain and make the interface to the 50 Ω driver. The last stage in the optical receiver is a 50 Ω differential output driver to make the interface between chip and the measurement setup. A reference current of 100 μA is generated with an on chip reference current circuit. The 100 μA reference currents are mirrored to bias the amplifiers with fixed ratio. The single-ended-to-differential converter's, the pre-driver's and the 50 Ω driver's biasing currents are 2.2 mA, 2.2 mA, and 8.4 mA, respectively. The post amplifiers have a designed gain of 5.6 dB and a designed bandwidth of 20 GHz each. The output driver has a unity gain. The TIA's bandwidth is affected by the post ampli-

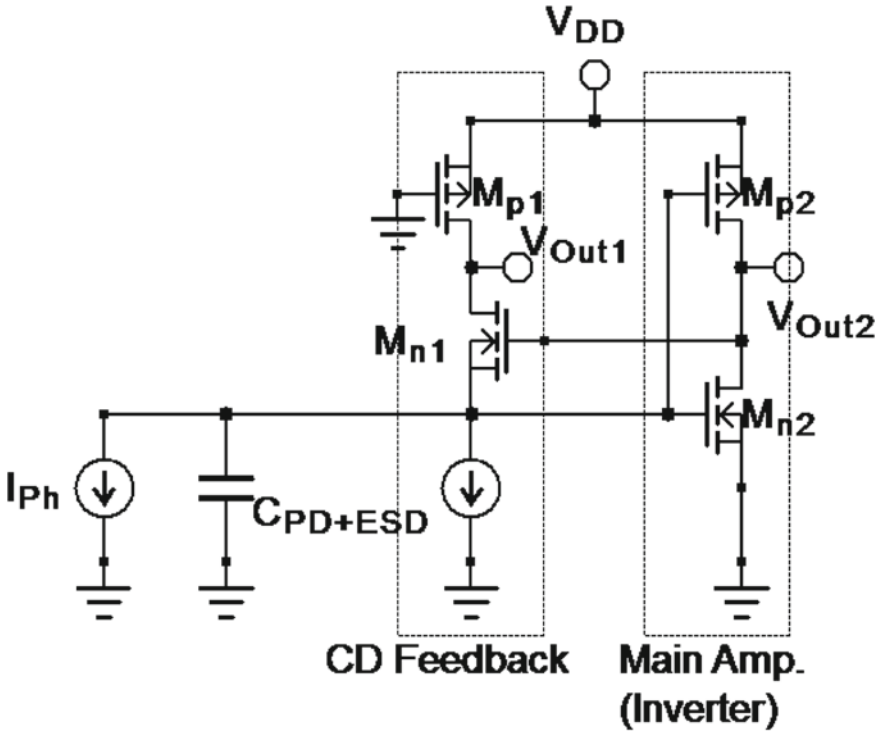

Fig. 10.21 Inverter based common-drain feedback TIA [13]

fier's loading which reduces the actual bandwidth of the TIA. Part of the simulation results of the ICDF-TIA was introduced in [14].

The optical receiver has been integrated in a 40 nm standard CMOS technology with 1.1 V supply voltage [14]. The optical receiver is optimized for an off-chip commercial InGaAs PIN-PD. The PD has a bandwidth being sufficient for a data rate of 10Gb/s and a responsivity of 1.1A/W for 1.55 μm light. The total optical receiver's input capacitance including PD capacitance, ESD protection and input PAD is 450fF. Finally, the output buffer amplifier provides enough driving capability for the 50 Ω input measurement devices. The TIA alone consumes 2.03 mW, and the total chip power dissipation is 17 mW. The TIA alone occupies 0.0002 mm^2 whereas the complete optical receiver including the TIA, post amplifiers and the output driver occupies a chip area of 0.16 mm^2. The die photo is shown in Fig. 10.23.

The measured input referred noise current density is 22 $pA/\sqrt{Hz}$ and the integrated input referred noise current is 1.85 μA up to 7 GHz. For BER measurements a PRBS signal of length $2^{15} - 1$ was used to drive a 1.55 μm laser modulator. The received data stream from the integrated optical receiver was compared with the original transmitted signal by bit-error counting using a BERT. The measured BER in dependence on the received optical power at different data rates (8 and 10 Gbit/s)

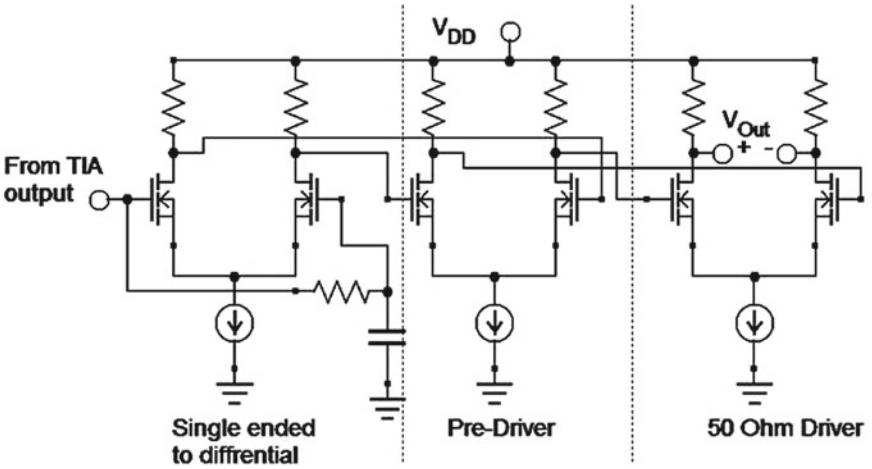

Fig. 10.22 Circuit diagram of the post amplifiers and the 50 Ω driver [14]

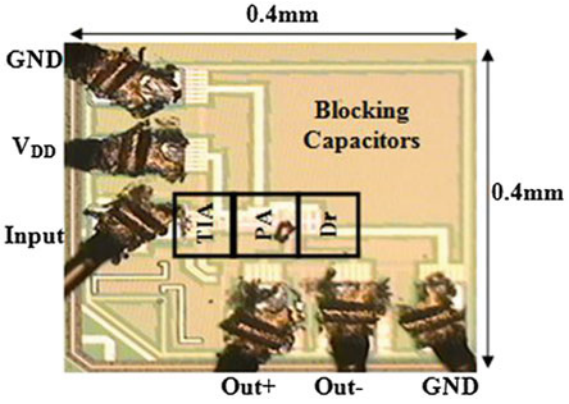

Fig. 10.23 Die photo of the optical receiver. The test chip is bonded to the PCB and PD by bonding wires [14]

is illustrated in Fig. 10.24. A sensitivity of −17.7 dBm is obtained for the presented optical receiver for BER = 10^{-12} at a data rate of 8 Gbit/s. A sensitivity of −16.2 dBm is obtained for BER = 10^{-9} at a data rate of 10 Gbit/s. Figure 10.25 shows eye diagrams at PRBS = $2^{15} - 1$ and data rates of 8 Gbit/s and 10 Gbit/s with an average received optical power of −17 dBm and −16 dBm, respectively.

Table 10.2 compares the measured performance of the ICDF-TIA along with other recently published 10 Gb/s TIAs with external PD. It is clearly seen that the presented work shows superiority in terms of Figure of Merit (FoM_2) which is defined by:

$$FoM_2 = \frac{Gain\,(\Omega) \cdot Bandwidth\,(\mathrm{GHz}) \cdot C_{in}\,(\mathrm{pF})}{Power\,Dissipation\,(\mathrm{mW}) \cdot Input\,Noise\,Current\,(\mu\mathrm{A})} \tag{10.2}$$

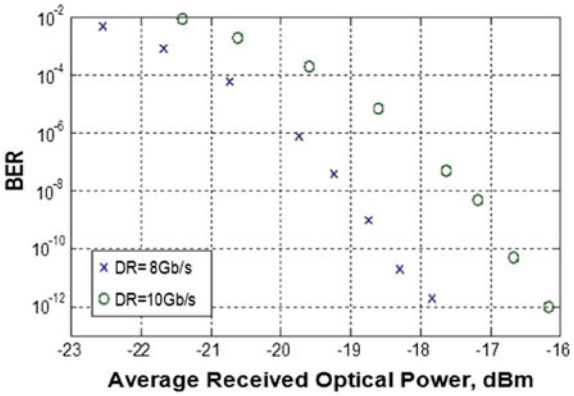

Fig. 10.24 Measured BER in dependence on the received average optical power at different data rates [14]

where C_{in}(pF) is the input capacitance in pF. As the technology is scaled the transit frequency f_t is increased. To include the effect of technology scaling FoM_3 is introduced:

$$FoM_3 = \frac{Gain\,(\Omega) \cdot Bandwidth\,(\text{GHz}) \cdot C_{in}\,(\text{pF})}{Power Dissipation\,(\text{mW}) \cdot Input Noise Current\,(\mu\text{A}) \cdot f_t\,(\text{GHz})} \tag{10.3}$$

A new technique to design an inductorless TIA was introduced in [15]. This technique uses N similar TIAs in parallel configuration to boost the overall bandwidth. The current sensitivity was 22.5 μA (−15 dBm) for BER 10^{-12} at PRBS = $2^{31} - 1$. The TIA in [15] shows the worst FoM due to the large power consumption needed for the N parallel TIAs to reach the required bandwidth. A three-stage inverter based TIA with 3.9 kΩ shunt feedback resistor was introduced in [16]. The small PD capacitance of 40 fF and the high transit frequency of 40 nm CMOS technology enable to design a 7 GHz TIA with a low power consumption of 3.95 mW and a high sensitivity of −15 dBm for BER = 10^{-12} at PRBS = $2^{31} - 1$. The ICDF optical receiver presented in [14] has the best FoM_2 due to the low-noise feature of the inverter stage, the higher bandwidth and the lower power consumption achieved by the CD feedback. The optical receiver in [11] based on multi-stage inverter TIA has the best FoM_3 due to the multistage inverter high gain. This gives the designer a hint that the multistage inverter based TIA can give also the best FoM_2 if fabricated in 40 nm CMOS technology instead of the 90 nm. The multistage technique can give a high performance in nanometer CMOS technology as the MOSFET intrinsic gain is decreasing and the transit frequency is increasing with technology scaling. As the transit frequency is 20 times higher than the required bandwidth the multistage will give a better performance than the single-stage TIA. For small-bandwidth TIAs, the multi-stage amplifier is the best design choice. It gives more gain without decreasing the stability; as there is enough space in the frequency domain to put the poles apart to ensure the stability, see Sect. 6.2.

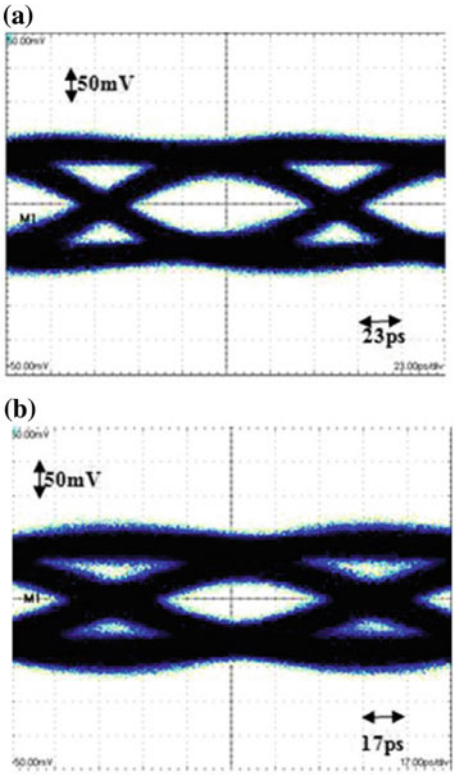

Fig. 10.25 Eye diagram of the optical receiver with PRBS $2^{15} - 1$ and data rate of **a** 8 Gbit/s and −17 dBm average optical power, **b** 10 Gbit/s, and −16 dBm average optical power [14]

10.3 Optical Sensors

10.3.1 2D Image Sensors

For the mobile-phone market, low-power and small area implementations are necessary. Deep-sub-μm and nanometer CMOS image sensors can fulfill these requirements. Serial signal-processing topologies, in which programmable-gain amplifiers (PGAs) and an analog-digital converter (ADC) can be shared by column-level correlating double sampling (CDS) circuits [18–22], reduce chip area and power consumption. A 10 bit self-differential offset-cancelled pipeline SAR-ADC was implemented in a 2D image sensor IC in 0.13 μm 2P4M CMOS technology [23]. It operates with the reference voltage of the ADC's half full-scale voltage, leading to a reduction of 50 % capacitor digital-analog converter (CDAC) area and of 80 % switching power. Proof-of-concept circuits were implemented in a 1.4 Mpixel CMOS image sensor consuming 51 mW (at 1.5 V analog-digital, 1.8 V I/O, and 2.5 V analog supply) at a frame rate of 17 frames per second and a read noise of 187.5 μV_{rms} at an analog gain

Table 10.2 Comparison of nanometer CMOS Gigabit TIAs with external PDs [14]

	[15]	[17]	[11]	[16]	[14]
Technology	130 nm CMOS	180 nm CMOS	90 nm CMOS	40 nm CMOS	40 nm CMOS
Supply voltage (V)	2	1.8	1.2	0.95	1.1
Transimpedance gain (dBΩ)	62	56	63.5	71.8	47
Bandwidth (GHz)	6	8.27	6	7	8
Sensitivity@BER = 10^{-12} (dBm)	−15	−15	−16	−15	−16.2
PD + ESD capacitance (pF)	0.25	0.3	0.3	0.04	0.45
Power dissipation (mW)	98	35	12	3.95	2.03
F.O.M.2	12.52	31.7	112.5	161.26	218.92
F.O.M.3	0.203	0.713	1.26	0.806	1.09

of 8.1. The sensor core contained a 1720×832 pixel array with a pitch of 2.2 µm. The total test chip area was $9.0 \times 7.2\,\text{mm}^2$ [23].

An 8 Mpixel CMOS image sensor (CIS) was introduced in [24]. In order to fulfill the requirements of low chip area and increased functionality of cellphone cameras a stacked CMOS image sensor consisting of a conventional back-illuminated image sensor in 90 nm 1P4M CIS technology and a 65 nm standard 1P7M digital logic chip was presented. Both chips were connected by through silicon vias (TSVs). The CIS chip had to be thinned. The use of the 65 nm technology enabled 2400k gates, which were mainly used for image-processing circuits. Stacking of the two chips resulted in 70 % of the chip area of conventional 1/4-inch image sensor chips with an equivalent of only 500 k gates [24]. A high-dynamic range (HDR) high-definition (HD) mode was implemented. HD (1080×720) videos can be taken by the stacked image sensor with 30 frames per second. The number of effective pixels was given with 3280×2464. Each pixel had a size of $1.12 \times 1.12\,\mu\text{m}^2$. These pixel saturated at 5000 electrons. The sensitivity of green pixels was given with 4800 e–/lx/s. The RMS random noise was specified with 2.2 electrons (e–). A conversion gain of 63.2 µV/e– and a dynamic range of 60 dB plus 24 dB in HDR mode were reported. The power consumption was below 200 mW [24].

10.3.2 3D Image Sensors

Stereo vision is one of the principles being used for 3D cameras. Such a stereo-vision based camera chip was realized in 0.11 µm [25]. A very compact design implementing the right-eye and the left-eye sensor on one sensor chip. Figure 10.26 depicts the single-lens principle and the block diagram of this sensor IC. Each half of the lens represents one eye of the 3D vision system. The 3D image sensor chip contains the photodiode array, 12 bit ADCs, and drivers [25].

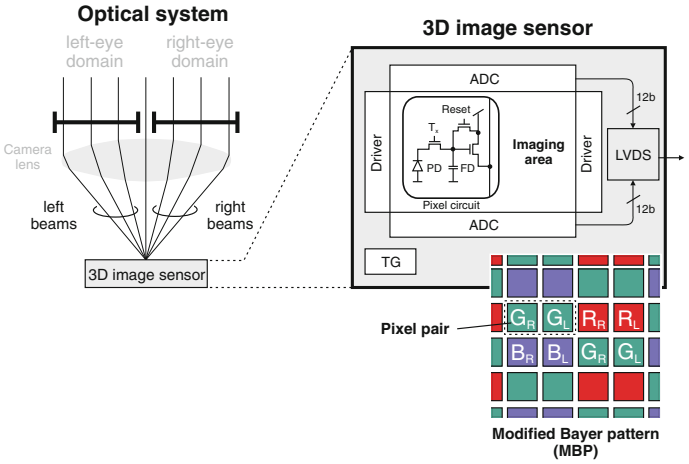

Fig. 10.26 Single-lens stereo vision system

The photodiodes for each eye are placed alternately in the matrix based on a modified Bayer pattern color filter as shown in Fig. 10.27. A so-called digital micro lens (DML) [26, 27] directs the light beams into the photodiodes. One micro lens serves for each right and left eye pair of photodiodes. 2.1 million pixels were implemented on a chip with an area of $8.45 \times 7.00\,\mathrm{mm}^2$. Each pixel includes a conventional 3-

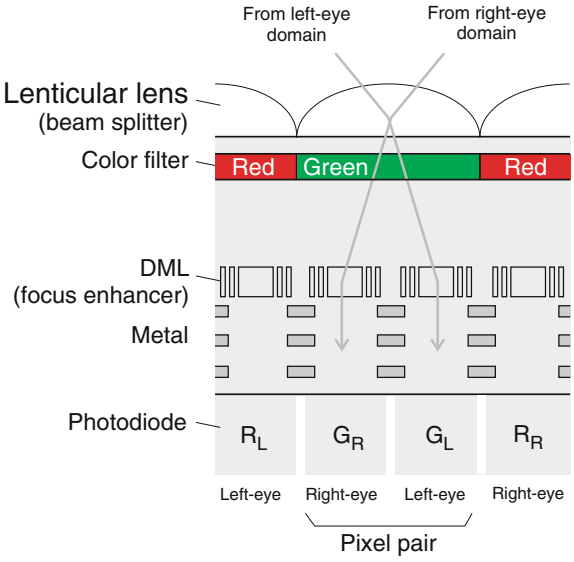

Fig. 10.27 On-chip single-lens optics

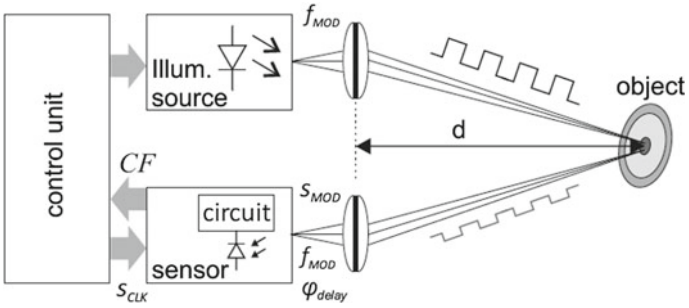

Fig. 10.28 Basic principle of an optical time-of-flight sensor

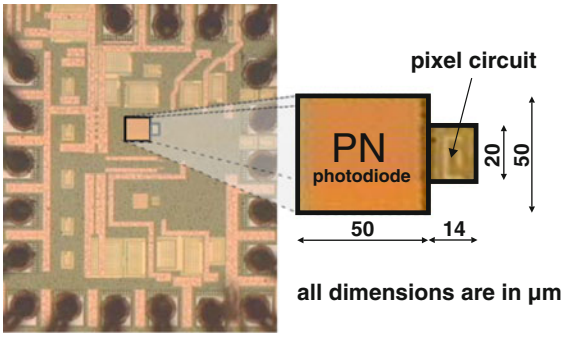

Fig. 10.29 Die photo of a distance sensor pixel in 90 nm CMOS [28]

transistor circuit and occupies an area of $2.75 \times 2.75\,\mu m^2$. 60 frames per second were achieved.

Figure 10.28 shows the basic setup of an optical time-of-flight (TOF) 3D-sensor. A continuous-wave signal is emitted by a modulated light source. At the object a very small portion of the emitted optical power is reflected and detected in the TOF 3D-sensor IC. The phase delay of this detected signal is determined by in-pixel correlators. At least four different phase shifts have to be used in order to be able to eliminate the reflectivity of the object.

A test chip for a TOF 3D-sensor IC was designed and fabricated in 90 nm CMOS. The chip photo of this test chip is shown in Fig. 10.29.

The dimension of one pixel is $64 \times 50\,\mu m^2$, whereby the pixel circuit occupies only an area of $14 \times 20\,\mu m^2$ [28]. Due to the 90 nm technology, this small circuit area and a record high TOF pixel fill factor of 90 % were possible. Figure 10.30 depicts the circuit diagram of a bridge correlator circuit used in a time-of-flight (TOF) pixel in 90 nm CMOS. With the help of the bridge circuit it is possible to subtract the charges stemming from background light immediately in the same clock period and to achieve record high immunity against background light. In [29], 180 klx ambient light were possible with the measured distance changing less than 3 cm. The photodiode node is regulated in its potential to avoid fast saturation of the pixel.

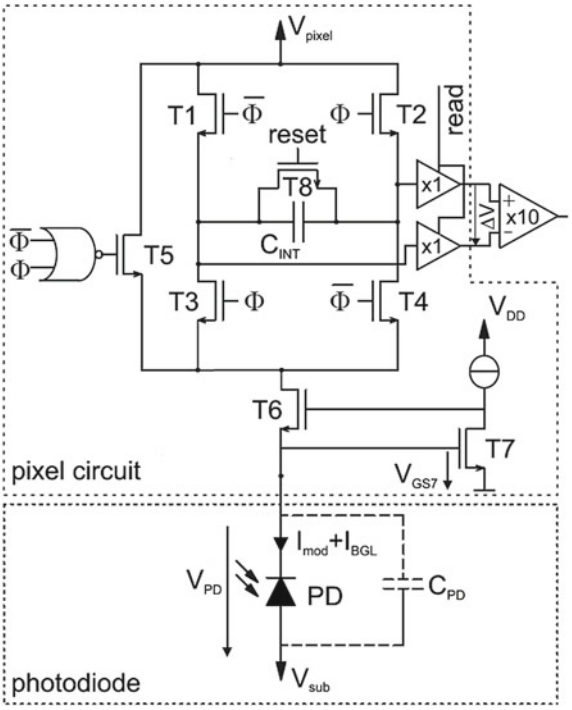

Fig. 10.30 Bridge correlator circuit

The electrical power consumption of this pixel was 2 µW. The standard deviation of the measured distance of 1.2 m was 2 cm covering the range from 0.2 to 3.2 m at 120 klx background light. The integration capacitor of 200 fF was realized as MIM capacitor and stacked with the pixel circuit facilitating the small pixel area. An optimized multipixel design could use the area of $14 \times 50\,\mu m^2$ (compare Fig. 10.29) for two correlator circuits of adjacent pixels in order to obtain an equivalent pixel pitch of $57 \times 50\,\mu m^2$ justifying the optical fill factor of about 90 %.

10.3.3 Medical Sensors

One medical sensor working with an optical principle is a positron-emission tomography (PET) sensor. PET is a nuclear imaging technique providing functional 3D images of the body e.g. in clinical oncology and brain-function analysis. A ring of scintillator crystals absorbs gamma quants and emits photons typically in the blue spectral range. These photons can be detected by single-photon avalanche diodes (SPADs). A sensor IC containing 92,000 SPADs and signal processing circuits was introduced in [30]. This rather complex integrated PET sensor was realized in a 1P4M CMOS Imaging Sensor (CIS) 0.13 µm technology. Such a CIS technology

may implement an anti-reflection coating. SPADs with a diameter of 16.27 μm and a P+/N-well structure similar to [31] with a photon detection probability (PDP) of 28 % were implemented in MiniSiPMs with 12 × 15 SPADs. 4 MiniSiPMs (Mini Silicon Photomultipliers) formed one pixel. The four independent digital SiPMs (called MiniSiPMs) share a data-managing circuit (DAMAC), which is responsible for photon counting, time stamping and data delivery. All SPAD cathodes are connected to a common bias line. The SPAD anodes are individually fed into a quenching transistor. The SPAD output is digitized by a Schmitt trigger inverter. A compression circuit combines the 180 SPAD outputs into a single channel which provides one short pulse (250ps) per photon count. Two time-to-digital converters (TDCs) in ping-pong mode and two counters are present in each DAMAC. A discriminator with two configurable thresholds distinguishes a gamma event from noise (dark counts and spurious photons). The energy resolution of 12.9 % was reported for 511 keV together with a LYSO scintillator. A detection time resolution including time-to-digital converter (TDC) of 253 ps (FWHM) was reported. The total chip area was $9.85 \times 5.45\,\text{mm}^2$.

References

1. M. Fortsch, Monolithically Integrated Optical Receivers for Low-Cost Data Communication and Optical Storage Systems. Ph.D. Dissertation, Vienna University of Technology (2007)
2. C. Hermans, M. Steyaert, *Broadband Opto-Electrical Receivers in Standard CMOS* (Springer, Netherlands, 2007)
3. Y. Dong, K.W. Martin, A high-speed fully-integrated POF receiver with large-area photo detectors in 65 nm CMOS. IEEE J. Solid-State Circuits **47**(9), 2080–2092 (2012)
4. Y. Dong, K. Martin, *A Monolithic 3.125 gbps Fiber Optic Receiver Front-End for POF Applications in 65 nm CMOS* (Sept 2011), pp. 1–4
5. M. Atef, H. Zimmermann, Optical receiver using noise cancelling with an integrated photodiode in 40nm cmos technology. IEEE Trans. Circuits and Systems I (TCAS I) (2013)
6. M. Atef, H. Zimmermann, 2.5 Gbit/s transimpedance amplifier using noise cancelling for optical receivers, in *IEEE International Symposium on Circuits and Systems (ISCAS)* (Seoul, Korea (South), 2012), pp. 1740–1743
7. F. Tavernier, M. Steyaert, High-speed optical receivers with integrated photodiode in 130 nm cmos. IEEE J. Solid-State Circuits **44**, 2856–2867 (2009)
8. W.-Z. Chen, S.-H. Huang, A 2.5 Gbps CMOS fully integrated optical receicer with lateral PIN detector, in *IEEE 2007 Custom Intergrated Circuits Conference (CICC)* (2007), pp. 293–296
9. T.-C. Kao, F. Musa, A. Carusone, A 5-gbit/s cmos optical receiver with integrated spatially modulated light detector and equalization. IEEE Trans. Circuits Syst. I: Regul. Pap. **57**, 2844–2857 (2010)
10. A. Rousson, T.C. Carusone, A multi-lane optical receiver with integrated photodiodes in 90 nm standard cmos, in *Optical Fiber Communication Conference*, p. JTh2A.21 (Optical Society of America, 2012)
11. M. Atef, F. Aznar, S. Schid, A. Polzer, W. Gaberl, H. Zimmermann, 8 Gbit/s inductorless transimpedance amplifier in 90 nm CMOS technology. Analog Integr. Circuits Sig. Proces. **79**(1), 27–36 (2014)
12. L. Vivien, A. Polzer, D. Marris-Morini, J. Osmond, J.M. Hartmann, P. Crozat, E. Cassan, C. Kopp, H. Zimmermann, J.M. Fdli, Zero-bias 40 Gbit/s germanium waveguide photodetector on silicon. Opt. Express **20**(2), 1096–1101 (2012)
13. M. Atef, H. Zimmermann, 10 Gbit/s 2 mW inductorless transimpedance amplifier, in *IEEE International Symposium on Circuits and Systems (ISCAS)* (Seoul, Korea (South), 2012), pp. 1728–1731

14. M. Atef, H. Zimmermann, Low-power 10 Gb/s inductorless inverter based common-drain active feedback transimpedance amplifier in 40 nm CMOS. Analog Integr. Circuits Sig. Process. **76**(3), 367–376 (2013)
15. O. Momeni, H. Hashemi, E. Afshari, A 10-Gb/s inductorless transimpedance amplifier, in *IEEE Transactions on Circuits and Systems II (TCAS II)*, vol. 57, no. 12 (2010), pp. 926–930
16. F.Y. Liu, D. Patil, J. Lexau, P. Amberg, M. Dayringer, J. Gainsley, H.F. Moghadam, X. Zheng, J.E. Cunningham, A.V. Krishnamoorthy, E. Alon, R. Ho, 10-Gbps, 5.3-mW optical transmitter and receiver circuits in 40-nm CMOS. IEEE J. Solid-State Circuits **PP**(99), 1 (2012)
17. H. Chen, C. Chen, W. Yang, J. Chiang, Inductorless CMOS receiver front-end circuits for 10-Gb/s optical communications. Tamkang J. Sci. Eng. **12**(4), 449–458 (2009)
18. I. Takayangi et al., High-resolution CMOS video image sensors. Proc. IEEE **99**, 1–13 (2012)
19. R. Johansson et al., A 1/13-Inch 30 fps VGA SoC CMOS image sensor with shared reset and transfer-gate pixel control, in *IEEE International Solid-State Circuits Conference* (2011), pp. 414–415
20. J. Moholt et al., A 2Mpixel 1/4 Inch CMOS image sensor with enhanced pixel architecture for camera phones and PC cameras, in *IEEE International Solid-State Circuits Conference* (2008), pp. 58–59
21. K.B. Cho et al., A 1/2.5 Inch 8.1Mpixel CMOS image sensor for digital cameras, in *IEEE International Solid-State Circuits Conference* (2007), pp. 508–509
22. M.J. Loinaz et al., A 200-mW, 3.3 V CMOS color camera IC producing 352 × 288 24-b video at 30 frames/s. IEEE J. Solid-State Circuits **33**(12), 2092–2103 (1998)
23. J. Deguchi, F. Tachibana, M. Morimoto, M. Chiba, T. Miyaba, H. Tanaka, K. Takenaka, K.A. S. Funayama, K. Sugiura, R. Okamoto, S. Kousai, A 187.5 μV_{rms}-read-noise 51 mW 1.4 Mpixel CMOS image sensor with PMOSCAP column CDS and 10b self-differential offset-cancelled pipeline SAR-ADC, in *IEEE International Solid-State Circuits Conference* (2013), pp. 494–495
24. S. Sukegawa, T. Umebayashi, T. Nakajima, H. Kawanobe, K. Koseki, I. Hirota, T. Haruta, M. Kasai, K. Fukumoto, T. Wakano, K. Inoue, H. Takahashi, T. Nagano, Y. Nitta, T. Hirayama, N. Fukushima, A 1/4-inch 8Mpixel back-illuminated stacked CMOS image sensor, in *IEEE International Solid-State Circuits Conference* (2013), pp. 484–485
25. S. Koyama, K. Onozawa, K. Tanaka, Y. Kato, A 3D Vision 2.1 Mpixel image sensor for single-lens camera systems, in *IEEE International Solid-State Circuits Conference* (2013), pp. 492–493
26. K. Toshikiyo et al., A MOS image sensor with microlenses built by sub-wavelength patterning, in *IEEE International Solid-State Circuits Conference* (San Francisco, CA, 2007), pp. 514–515
27. K. Onozawa et al., A MOS image sensor with a digital microlens. IEEE Trans. Electron Devices, 986–991 (2008)
28. M. Davidovic, G. Zach, K. Schneider-Hornstein, H. Zimmermann, Range finding sensor in 90 nm CMOS with bridge correlator based background light suppression, in *ESSCIRC* (Seville, 2010), pp. 298–301
29. M. Davidovic, G. Zach, K. Schneider-Hornstein, H. Zimmermann, TOF range finding sensor in 90 nm CMOS capable of suppressing 180klx ambient light. in *IEEE Sensors Conference* (Kona, HI, 2010), pp. 2413–2416
30. L. Braga, L. Gasparini, L. Grant, R. Henderson, N. Massari, M. Perenzoni, D. Stoppa, R. Walker, An 8×16-pixel 92kSPAD time-resolved sensor with on-pixel 64 ps 12b TDC and 100 MS/s real-time energy histogramming in 0.13 μm CIS technology for PET/MRI applications, in *IEEE International Solid-State Circuits Conference (ISSCC 2013)* (San Francisco, USA, 2013), pp. 486–487
31. J. Richardson, L. Grant, R. Henderson, Low dark count single-photon avalanche diode structure compatible with standard nanometer scale CMOS technology. IEEE Photonics Technol. Lett. **21**, 1020–1022 (2009)

Index

M. Atef and H. Zimmermann, *Optoelectronic Circuits in Nanometer CMOS Technology*, Springer Series in Advanced Microelectronics 54,
DOI 10.1007/978-3-319-27338-9