

**AN INTEGRATE-AND-DUMP RECEIVER
FOR FIBER OPTIC NETWORKS**

ANDREW E. STEVENS

Submitted in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY
1995

©1995

Andrew E. Stevens

All Rights Reserved

ABSTRACT

AN INTEGRATE-AND-DUMP RECEIVER FOR FIBER OPTIC NETWORKS

by

ANDREW E. STEVENS

The receiver in an optical communication system is a key component in determining the overall system performance. This thesis will present a new optical receiver which mathematically integrates the arriving bits with respect to time in order to improve receiver sensitivity. This so-called integrate-and-dump circuit is a classical result from communication theory, but has not previously been adapted to modern-day integrated circuit technology.

This thesis will explore the theoretical basis for the integrate-and-dump circuit and will explain the advantages and disadvantages of this method. A comparison of the performance of several receiver noise filtering methods is made which shows that the integrate-and-dump scheme has the best realizable performance. A new architecture is proposed which uses parallel signal processing in order to relax the requirements on the integrators. In addition, an new biasing method allows connection of the photodetector to the integrator without ac coupling capacitors, thereby improving low-bitrate performance. A balanced symmetrical circuit topology and differential output allows cancellation of circuit and external noise for improved noise rejection.

A test chip for the integrate-and-dump receiver was implemented in a 1.2 μm CMOS technology. The chip achieves a sensitivity of -49.4 dBm at 10 Mb/s with a bit error rate of 10^{-9} . In addition, the chip can operate at bit rates exceeding 100 Mbits/s. The total active circuit area is approximately 1 mm \times 1 mm, and the total static power dissipation is approximately 290 mW.

CONTENTS

1	INTRODUCTION	1
1.1	OPTICAL NETWORKS	1
1.2	BROADCAST-AND-SELECT TOPOLOGY	2
1.3	REQUIREMENTS FOR WDMA	5
1.4	IMPLEMENTATION ISSUES	6
1.5	CHOICE OF TECHNOLOGIES	8
1.6	THESIS OUTLINE	11
2	THEORY OF RECEIVERS	13
2.1	OPTICAL RECEIVER BASICS	14
2.1.1	Photodetector	15
2.1.2	Preamplifier	17
2.1.3	Noise-Shaping Filter	21
2.1.4	Timing Recovery and Decision Circuit	25
2.2	OPTIMAL NOISE FILTERING	30
2.3	BIT ERROR RATE	33
2.3.1	Quantum Noise Limit	33
2.3.2	BER and Gaussian noise	34
2.4	COMPARISON OF NOISE FILTERS	37
2.4.1	Matched Filter Performance	39
2.4.2	2nd-Order Filter Performance	40
2.4.3	Raised-Cosine-Output Filter Performance	42
2.4.4	Integrate-and-Dump Performance	44
2.4.5	Comparison of Sensitivities	46
3	INTEGRATE-AND-DUMP RECEIVER TOPOLOGY	51
3.1	INTEGRATE-AND-DUMP TOPOLOGY	52
3.2	DETECTOR BIASING TOPOLOGY	57
3.3	PREAMPLIFIER DESIGN	60
3.4	BIAS CANCELLATION CIRCUIT	64
3.5	COMPLETE RECEIVER FRONT-END	66
3.6	GAIN CONTROL	66
3.7	CMOS SWITCH DESIGN	69

3.8	COMPLETE CIRCUIT	71
3.9	CLOCK GENERATOR	71
3.10	PREDICTED PERFORMANCE	75
	3.10.1 Noise Response of Integrate-and-Dump Circuit	75
	3.10.2 Sensitivity	81
4	IMPLEMENTATION AND TEST RESULTS	85
4.1	CHIP LAYOUT	85
4.2	TEST SETUP	86
4.3	PERFORMANCE	90
	4.3.1 Amplifier offset	90
	4.3.2 Common mode noise	93
	4.3.3 Sensitivity	94
	4.3.4 Dynamic Range	98
4.4	IMPROVEMENT OF DESIGN	99
5	SUMMARY AND FUTURE WORK	103
5.1	THESIS SUMMARY	103
5.2	COMPARISON WITH PUBLISHED RESULTS	105
5.3	AREAS FOR FURTHER STUDY	105
	Bibliography	107
A	A HIGH-SLEW INTEGRATOR FOR SWITCHED-CAPACITOR CIRCUITS	113
A.1	INTRODUCTION	113
A.2	DESCRIPTION OF BOOSTED INTEGRATOR	115
A.3	CIRCUIT DESIGN	118
A.4	TEST RESULTS	120
A.5	CONCLUSION	124
A.6	REFERENCES	124

ACKNOWLEDGMENTS

I am very indebted to my advisor, Prof. Ed Yang, for his encouragement and devotion to my completion of this work. I also owe a great deal to Dr. Paul Green and Dr. Frank Tong of the IBM T.J. Watson Research Center for their unfailing support of my Ph.D. thesis. And I am very thankful to the IEEE Solid-State Circuits Council for their generous fellowship for 1993–94.

I am grateful to Prof. John Khoury, Dr. Ken Suyama, and Prof. Jordan Spencer for agreeing to serve on my defense committee.

I am also thankful to: Peicheng Ju, Seema Varma, Homer Wang, Yoshi Horio, and all the members of the “mad” VLSI lab; Lourdes de La Paz and Judy Nicholson in the EE office; Chung-Sheng Li, Young Kwark, Frank Canora, Bardia Pezeshki, Dennis Rogers, Karen Liu, Thomas Schrans, Jeff Kravitz, Eric Hall, and Rajiv Ramaswami at IBM; all my friends at the Postcrypt Coffeehouse.

The appendix of this thesis represents work which I completed at Columbia during 1991-92 under a previous thesis advisor. This work was presented at the VLSI Circuits Symposium in Kyoto, Japan, in 1993 and was published in the *IEEE Journal of Solid-State Circuits* in 1994. I would like to thank Gerry Miller and Paul Ferguson at Analog Devices, and David Vallancourt at AT&T Bell Labs for their assistance in this work.

This thesis was phototypeset using Leslie Lamport’s \LaTeX macro package, Rusty Wright’s *ucthesis* style, and Donald Knuth’s \TeX formatter. The drawings were done with William Chia-Wei Cheng’s *tgif* program, and the plots were made with David Harrison’s *xgraph* and Steven Wolfram’s *Mathematica*. The figures were integrated into the text with Trevor Darrell’s *psfig* macros and Sebastian Rahtz’s *rotating* macros.

Finally, thanks to my parents, sister, and Theresa Adams for their unfailing support through this wild adventure.

CHAPTER 1

INTRODUCTION

The rapidly-expanding field of fiber optic communications has created a large demand for low-cost components and circuits necessary for implementation. Originally, application of optical transmission to long-haul telephone traffic allowed the high cost of the transmitters and receivers to be shared by the tens of thousands of simultaneous users, thus presenting a low cost per user. In contrast, in computer networking or fiber-to-the-home applications, each computer or customer must have a dedicated optical transceiver in order to talk directly to the network. Since the potential marketplace for services such as high-speed computer networking, digital video, or interactive television numbers in the billions of nodes, the drive to lower the cost per node becomes paramount.

This thesis will describe an attempt to redesign one piece of the puzzle, that of the optical receiver. By utilizing some classical results of communication theory, and combining them with today's low-cost CMOS technology, this thesis will present a new integrated circuit design which is intended to compete with prior (primarily gallium arsenide) designs. The design is particularly well-suited for use in high-speed computer networks where all of the components must be located on one card inside each computer, or perhaps within one CMOS chip on each motherboard.

1.1 OPTICAL NETWORKS

Most of today's existing computer networks are either based on copper cable (such as Ethernet) or point-to-point optical links as a replacement for copper (such as Fiber Distributed Data Interface, or FDDI). By comparison, an *all-optical network* does not require

any intermediate optical-to-electronic or electronic-to-optical conversions (e.g., active hubs or repeaters) between the transmitting and receiving ends. This allows the transmitters and receivers to utilize the unique properties of fiber which are not available on copper or point-to-point optical links.

One of the key advantages of fiber is the ability to send multiple colors of light simultaneously through the same fiber. This technique, called *wavelength division multiplexing* (WDM) is illustrated in Fig. 1.1. In a simple non-WDM system, a single transmitter sends light down a fiber to a single receiver. In a WDM system, there are multiple transmitters and receivers at each end of the fiber. Each transmitter is assumed to operate at a distinct wavelength or range of wavelengths which are independent of the other transmitters. Likewise, each receiver is presumed to only be sensitive to a particular wavelength or range of wavelengths. A typical wavelength spectrum is shown in Fig. 1.1(c).

In a non-WDM system, the total throughput is generally determined by the maximum speed of the electronics inside the transmitter and receiver at each end. In a WDM system, the throughput is determined by the speed of the electronics multiplied by the total number of unique wavelengths utilized in the system. This total number of wavelengths is determined by a number of factors, including the attenuation characteristics of the fiber, the monochromaticity of the transmitters, the crosstalk between channels, and the wavelength selectivity of the receivers. However, an upper bound on the total throughput for WDM systems has been estimated at 25,000 GHz [1].

1.2 BROADCAST-AND-SELECT TOPOLOGY

An example of an all-optical network utilizing WDM is shown in Fig. 1.2. This type of architecture has been proposed for the interconnection of mainframe computers and workstations and has been named *wavelength division multiple access* (WDMA) to reflect the multiplicity of different connections possible on a single fiber. In this “broadcast-and-select” topology, each node (computer) on the network is assigned a unique optical wavelength on which to broadcast data to all of the other nodes in the network. In addition, each node has a wavelength-tunable optical receiver in order to receive transmissions from any of the nodes on the network. In order for any two nodes to communicate, each node must first set up the connection by tuning its receiver to the other node’s transmit wavelength. This setup is accomplished through some specified protocol, perhaps via a dedicated signaling

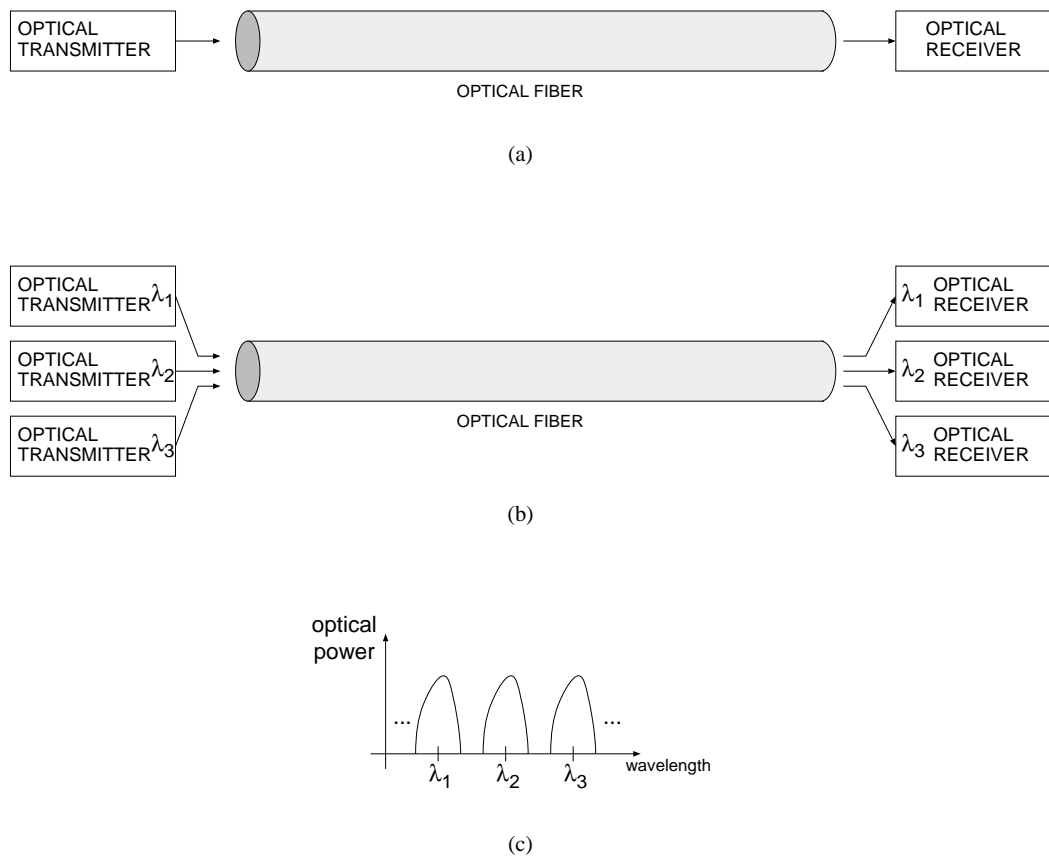


Figure 1.1. Wavelength division multiplexing (WDM). (a) In a single wavelength system, one transmitter and one receiver communicate over one fiber. (b) In a WDM system, multiple transmitters at different wavelengths communicate with multiple receivers, with each receiver tuned to only receive a specific wavelength. (c) Wavelength power spectrum of a WDM system.

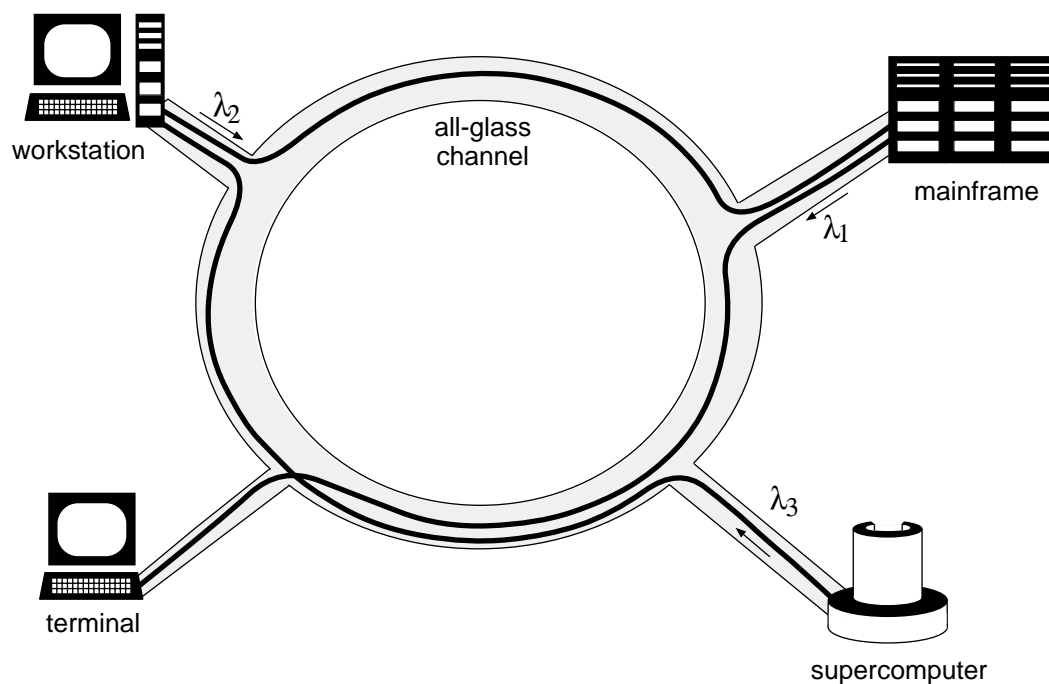


Figure 1.2. Broadcast-and-select optical network.

channel at a fixed wavelength.

An all-fiber channel is used to connect the nodes and typically consists of an $N \times N$ *star coupler* which is essentially an N -way optical power splitter. The star coupler may be passive (whereby the optical power of a given transmitter is split equally among the N nodes) or active (containing erbium-doped fiber amplifiers (EDFAs) to cancel the splitting loss).

The main advantage of this type of network is that the high bandwidth requirements are spread into the wavelength domain instead of the time or frequency domain. Because each network node transmits on its own dedicated wavelength, the total usable bandwidth of the network is equal to the sum of the bandwidths of the individual nodes. For example, in a 32-node broadcast-and-select network with a maximum per-node transmission rate of 1 Gbit/s, the total maximum network throughput is 32 Gbit/s. An equivalent time-division multiplexing (TDM) or frequency-division multiplexing (FDM) network at 32 Gbit/s would require electronic components to operate at rates approaching 32 GHz, which would be prohibitively difficult and expensive. Instead, WDMA networks achieve high capacity by

exploiting the high bandwidth properties of the fiber in the wavelength domain, without placing undue speed requirements on the system electronics. Another advantage of WDMA is that each wavelength may operate at a bitrate or protocol which is totally independent of the other wavelengths in the system. Thus, protocol conversion or bit stuffing is not necessary (as in TDM networks).

1.3 REQUIREMENTS FOR WDMA

Several parameters will determine the overall performance of a broadcast-and-select network, including

- the total size of the network (maximum distance and number of nodes)
- the selectivity of the network (spacing of nodes in the wavelength domain)
- the routing capabilities of the network (circuit vs. packet switching)

The total size of the network is limited by several factors, including optical loss due to the fiber and splicing (cumulatively referred to as the *link budget*), splitting loss in the star coupler, transmitter power, and receiver sensitivity. In the case of a local area network, the link budget is dominated by fiber splices which can typically cause approximately 0.5 dB loss per splice. In larger (metropolitan and wide area) networks, the actual loss in the fiber becomes more significant (at about 0.2 dB/km).

The star coupler is the major source of loss in the broadcast-and-select topology. In a 32-node network, the transmitted light must be split equally 32 ways, thus resulting in 15 dB loss from the transmitter to each individual receiver. A simple way to overcome this loss would be to use a high-power laser in the transmitter. However, high-power lasers are costly, hot, and perhaps most seriously, are unsafe for the commercial environment due to their potential for human eye damage. Most lasers used in commercial products are limited in power to approximately 1 mW (=0 dBm).

Given these circumstances, it becomes apparent that a high-quality tunable optical receiver becomes a major objective of any WDMA network. Assuming a -20 dBm (=0.01 mW) optical signal entering the tunable receiver (assuming a 0 dBm transmitter, a 5 dB link budget, and a 15 dB splitting loss), the receiver must optically select the desired wavelength, convert that wavelength to an electronic current, and then amplify the current

into a digital logic signal. These operations must be performed within an acceptable noise margin, typically one bit error every 10^9 bits.

The wavelength selectivity of the receiver will determine its ability to reject crosstalk from adjacent channels. This directly determines the channel spacing in the wavelength domain. Typical numbers are 1 nm spacing around a center wavelength of 1550 nm. Given that the optical spectral “window” of usable wavelengths for a single-mode fiber is approximately 1500 to 1600 nm, this places an upper bound of approximately 100 on the total possible number of channels in a system.

As already pointed out, the large loss associated with the star coupler could be reduced by incorporating EDFAs with the star coupler. EDFAs allow amplification of an optical signal directly in the optical domain without using electronic amplifiers, and more importantly, without any optical/electronic conversions. However, as expected, the cost of an active star coupler is a trade-off with the cost of highly-sensitive receivers. In addition, the optical bandwidth of an EDFA is currently only 30 nm, thus limiting the total number of nodes. However, the broadening of EDFA bandwidth is a topic of current research [2].

The applicability of a broadcast-and-select network to circuit or packet switching will be determined by the speed with which the receiver can switch channels, i.e. the *tuning time*. In circuit switching, a connection between two nodes is held for the length of the conversation, and other nodes requesting service are locked out (similar to a telephone busy signal). In packet switching, data is split up into small packets which are sent one at a time, thus allowing time sharing and multiple virtual connections between multiple nodes (as in Ethernet). Packet switching is more desirable for computer networking due to its flexibility and resource sharing, particularly when multiple networks are connected together with routers. In a broadcast-and-select network, the ability to perform packet switching requires that the receiver have a short tuning time, typically in the hundreds of nanoseconds range.

1.4 IMPLEMENTATION ISSUES

A possible implementation of the broadcast-and-select network in Fig. 1.2 is shown in Fig. 1.3. In this design, the transmitter operates at a single fixed wavelength, designated λ_n . In the simplest case, this is achieved by measuring and handpicking a laser diode of the desired wavelength. A more sophisticated (and costlier) design might use a tunable laser

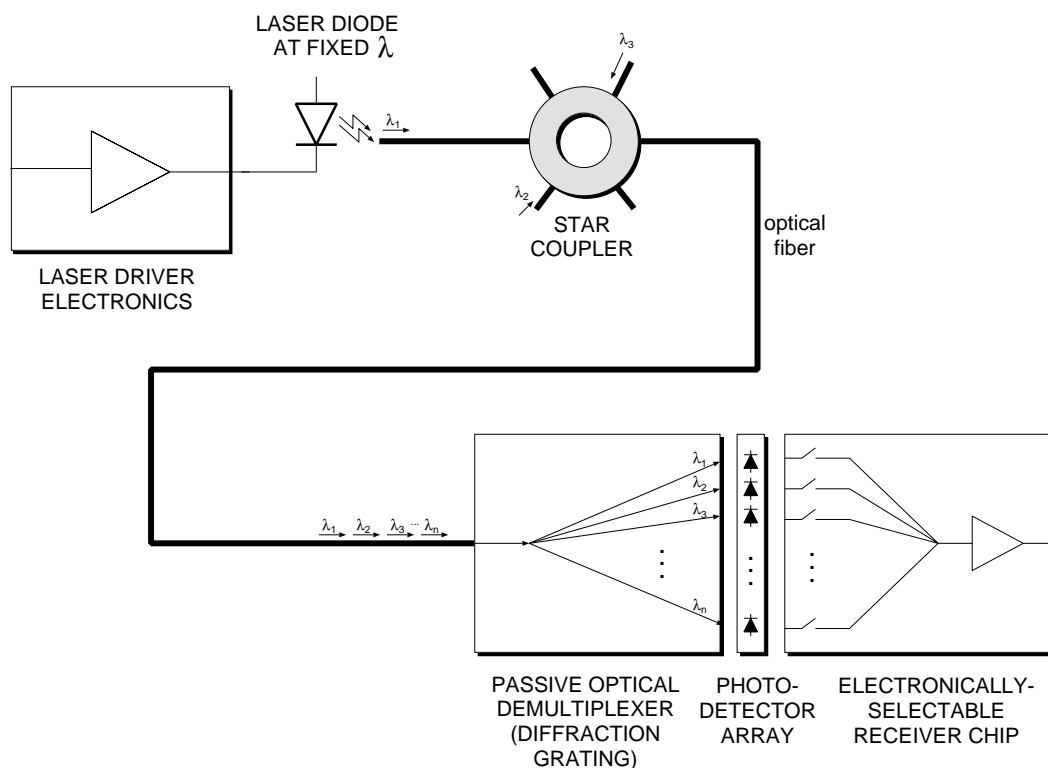


Figure 1.3. Implementation of broadcast-and-select network.

for more flexibility.

The receiver in this topology consists of three distinct components: (a) a passive optical demultiplexer which takes a single multiwavelength fiber and splits the different wavelengths with respect to space; (b) a photodetector array which is spaced in accordance with the splitting capability of the optical demultiplexer; and (c) an electronically-selectable integrated-circuit receiver which can amplify any of the photocurrents simply by closing the appropriate selection switch in order to choose the proper photodetector.

The optical demultiplexer can be implemented by a diffraction grating. The grating consists of a sawtooth surface which reflects light at an angle which is dependent on wavelength. The resulting reflected light will vary in space with respect to wavelength of the incoming light. By properly choosing the spacing and shape of the teeth in the grating, the output light can be coupled directly into an array of output optical fibers (in the case of a bulk grating) or planar optical waveguides (in the case of a planar grating).

The photodetector array consists of a strip of semiconductor containing an array of photosensitive structures such as PIN diodes or MSM (metal-semiconductor-metal) photodetectors. The material and structure will be determined by the desired optical wavelength and fabrication technology. The spacing and diameter of the photodetectors will depend on the physical dimensions and coupling between the grating and the detector strip.

The receiver chip consists of an array of switches, preamplifiers, postamplifiers, and clock recovery circuitry. In practice, these functions may actually be split among a series of chips and then integrated in a hybrid package.

1.5 CHOICE OF TECHNOLOGIES

Fig. 1.4 shows a possible physical arrangement for the elements in the tunable receiver [3]. In this design, an optical fiber containing multiple wavelengths is connected to the edge of a planar grating chip. The grating splits the different wavelengths with respect to space and couples them to output waveguides which run to the edge of the device. The light is then reflected down by a 45° mirror onto an array of photodetectors. The photodetectors are then connected to an adjacent electronic chip which may contain an array of preamplifiers or perhaps only a few amplifiers selected by switches (as shown) in order to reduce power consumption. The electronic chip may also contain postamplification and clock recovery circuitry. The choice of technology for each element will have a large effect on the overall design. In this case, planar technologies are used throughout, thereby allowing a stacked packaging arrangement which is very compact and robust.

The diffraction grating may be implemented in a silica on silicon process which has standard etching steps similar to those used for integrated circuits. In such a technology, three layers of SiO_2 are deposited onto a Si carrier substrate. The center layer of SiO_2 has a higher refractive index than the first and third layers, thus forming a propagation medium for light coupled to the edge of the device. Optical structures such as waveguides and gratings may then be etched into the silica using standard semiconductor techniques.

An advantage of this method is that the silica waveguides will have similar properties to fiber with respect to index matching, etc. However, depending on the wavelength of light used, the grating could be implemented directly in semiconductor. For example, indium phosphide conducts light at 1550 nm and has been used to build integrated grating and photodetector chips [4]. Theoretically, the required electronic circuitry could also be

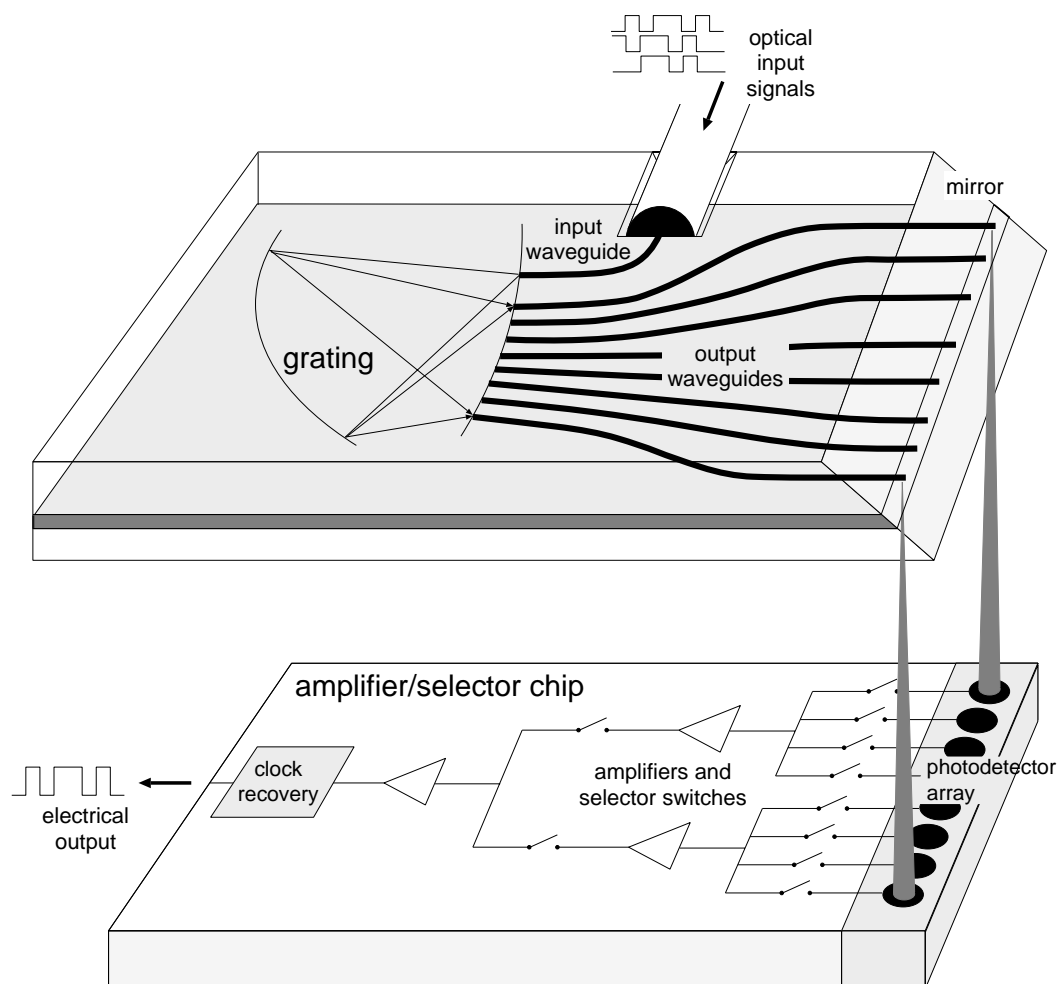


Figure 1.4. Physical design of tunable receiver.

TABLE 1.1. Photodetector materials and their useful wavelengths.

Si	850 nm
GaAs	850 nm
InGaAs	1300/1550 nm
InP	1300/1550 nm
Ge	1300/1550 nm

integrated into such a structure to create a complete tunable receiver on a single substrate. However, InP technology is not quite mature enough for this to be possible today. Gallium arsenide technology could theoretically be used to build a complete grating, detector, and preamplifier chip on a single substrate [5]. However, because GaAs only conducts and detects light at around 850 nm, such a device would be of limited usefulness due to the poorer propagation performance of the 850 nm wavelength down a standard fiber.

The choice of technology for the photodetector array depends largely on the optical wavelength of interest. Different materials will be sensitive to different wavelengths as shown in Table 1.1. The cheapest materials are Si and GaAs, which are both highly sensitive at a “short” wavelength of 850 nm. In addition, these photodetectors are advantageous because they may be integrated onto the same substrate as electronic preamplifiers. However, the disadvantage to short-wavelength transmission is that the fiber is more lossy (~ 2 db/km) than at “long” wavelengths such as 1300 nm or 1550 nm (~ 0.2 db/km), thus limiting the overall size of the network. Conversely, the improved transmission at long wavelengths requires detectors made from InGaAs, InP, or Ge, thereby precluding (as of yet) the integration of the detector on the same die as the preamplifier. Thus, the packaging and the connections between the photodetector and the preamplifier become a major issue in long wavelength designs.

The technology chosen for the preamplifier and related electronic circuitry is generally driven by the desired bitrate. Silicon (CMOS and BJT) is the least expensive and provides speeds into the gigabit-per-second regime. Gallium arsenide is more expensive, but operates faster than silicon by approximately an order of magnitude. More advanced technologies utilizing more novel devices (e.g. HBTs) are a current topic of research [6, 7, 8] and are still considered too expensive for use in a real system.

Given the switched architecture of the preamplifier chip as shown in Fig. 1.4, a natural

choice for the technology becomes CMOS due to the availability of high-quality switches. In addition, the insulated-gate structure of the MOSFET provides an excellent lownoise interface to the photodetector. The availability of complementary devices also allows a balanced symmetrical design which is generally more difficult in bipolar or gallium arsenide designs. Finally, CMOS is a mature lowcost technology, and today's submicron processes allow circuit operation at gigahertz speeds [9, 33]. Thus, CMOS seems a logical choice for implementation into a tunable receiver for optical WDMA networks.

1.6 THESIS OUTLINE

Given the choice of CMOS for the receiver circuitry, the design of optimized receiver topology becomes of high interest. Certainly, satisfactory CMOS receivers have been designed in the past [10, 11, 12, 13]. However, this research will introduce a new integrate-and-dump receiver which takes advantage of the many of the circuit techniques which are unique to CMOS technology and provides a more flexible design with higher performance.

Chapter 2 of this thesis describes the theory behind optical receiver design, including descriptions of the photodetector, preamplifier, noise filter, and timing recovery circuitry. Three types of preamplifiers are described, including the high impedance, transimpedance, and integrate-and-dump designs. The signal-to-noise ratio and bit error rate performance of binary receivers are also analyzed. The noise performance of four types of filtering schemes are also compared against each other and with the quantum noise limit for optical receivers.

Chapter 3 details the overall design of an integrate-and-dump receiver chip in CMOS. A method for using multiple preamplifiers in parallel is introduced in order to relax the performance requirements and simplify the circuit design of each individual preamplifier. In addition, a new technique for biasing the photodetector with dual opposite-type (n-type and p-type) preamplifiers is introduced. A complete noise analysis of the integrate-and-dump design is also included.

Chapter 4 details the test results of a prototype integrate-and-dump receiver chip in a 1.2 μm CMOS technology. The chip contains four parallel integrate-and-dump preamplifiers, a four-phase nonoverlapping complementary clock generator, and output buffers. Bonding pads are provided for connecting a photodetector adjacent to the chip. Measurement of the circuit noise performance were made in the lab with a bit error rate tester.

Chapter 5 discusses some of the lessons learned from this prototype, and potential areas of further research.

CHAPTER 2

THEORY OF RECEIVERS

Since most modern-day computers operate with electronic signals, it is obviously necessary to convert the optical signals of a fiber optic network into electrical signals usable by a computer. The function of the optical receiver is to perform this transformation. On the optical side, the ones and zeros are typically represented by the presence or absence of light, also known as on-off keying or OOK. On the electrical side, the receiver must output bits at standard logic levels which are usable by the computer bus.

The following are some of the requirements on the optical side of the receiver:

- The receiver should be able to reliably receive bits at very low incident light levels, thus allowing for loss in the optical fiber due to distance, splices, etc. This is referred to as the receiver *sensitivity*.
- The receiver should be able to receive bits over a large range of bright and dim incident light levels, i.e. have a large *dynamic range*.
- The receiver should not be adversely affected by the quality of the OOK modulation, i.e. the residual optical power which gets transmitted during the supposed “off” state. This “off” power depends on the *extinction ratio* (ratio of “off” and “on” optical power) of the transmitter.

The following are the requirements for the electrical side of the receiver:

- The receiver should output a digital bitstream at standard logic levels such as TTL, CMOS, or ECL.

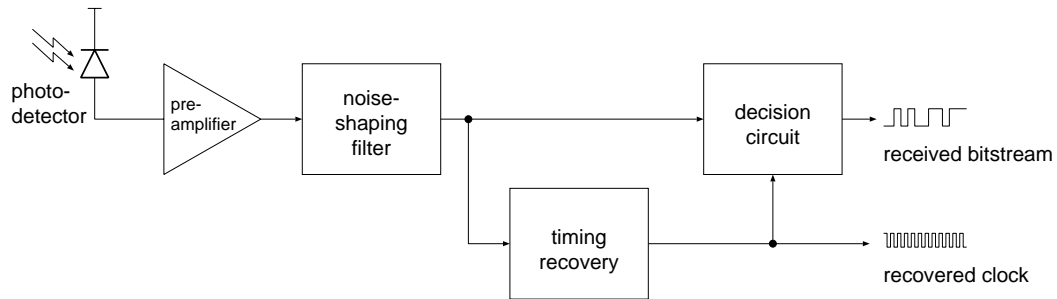


Figure 2.1. Optical receiver block diagram for on-off keying (OOK).

- For synchronization with the computer clock, the receiver should also output a recovered clock which is synchronized with the received bitstream.
- The digital bitstream should be accurate to some specified noise level or *bit error rate*. Digital communication systems are typically accurate to one bit in 10^9 or one bit in 10^{12} .

This chapter will discuss the optical receiver design given the above requirements.

2.1 OPTICAL RECEIVER BASICS

A block diagram of an optical receiver is shown in Fig. 2.1. Starting at the left of the diagram, a photodetector is used to convert the OOK-modulated incident light into an electrical current. This current is amplified and converted to a voltage by the preamplifier. The preamplifier output is then filtered by a noise-shaping filter in order to achieve the best signal-to-noise ratio (SNR). The output of the noise-shaping filter is simultaneously fed to a timing recovery circuit and a decision circuit (usually a flip flop or comparator). The timing recovery circuit extracts a clock from the incoming bitstream and uses it to trigger the decision circuit at the optimum point in the bit time. Both the timing recovery and decision circuits output voltages at standard logic levels.

The following sections will describe the operation of each of the elements of the receiver mentioned above.

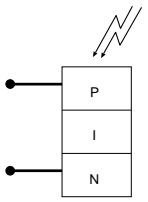

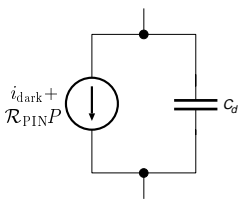
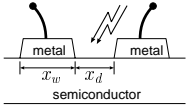


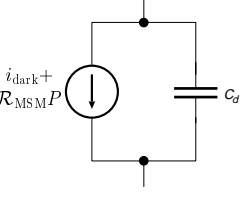
	STRUCTURE	CIRCUIT SYMBOL	CIRCUIT MODEL
P-I-N PHOTODIODE:			
M-S-M PHOTODETECTOR:	<p>cross-section:</p>  <p>top view:</p> 		

Figure 2.2. Summary of two major types of photodetectors, and their equivalent circuits.

2.1.1 Photodetector

Photodetector operation is based on the generation of photocarriers by low-level injection into a semiconductor. An electric field across the semiconductor region causes the generated carriers to drift to the detector electrodes, thereby causing a signal current. Common types of photodetectors include the PIN diode (the “I” stands for *intrinsic*) and the metal-semiconductor-metal (MSM) photodetector. These are summarized in Fig. 2.2. A third type of photodetector, the avalanche photodiode (APD), will not be considered here due to the high bias voltage necessary for its operation.

The PIN diode is a vertical structure consisting of a top layer of p-type semiconductor, a middle layer of intrinsic (undoped) semiconductor, and a bottom layer of n-type material. The PIN diode is normally operated under reverse bias. When a reverse bias of a few volts is applied, an electric field is formed across the junction depletion region. Light shined on this region generates electron-hole pairs and thus causes a photocurrent to flow through the diode. The purpose of the intrinsic layer is to increase the size of the depletion region, and thus increase the efficiency of the device.

In the MSM photodetector, an electric field is set up in the semiconductor by biasing

metal electrodes at a fixed potential. When light is shined on the biased region, the generated electron-hole pairs drift to the electrodes, thus creating a photocurrent. The MSM detector is a symmetrical device, and thus its circuit symbol does not differentiate between the two terminals.

Photodetectors are characterized by their ability to convert optical power (measured in watts) into electrical current (measured in amperes). The responsivity \mathcal{R} of the photodetector is the measure of this property in units of A/W. In the ideal case, each photon of light generates one electron-hole pair, giving

$$\mathcal{R}_{\text{ideal}} = \frac{q}{h\nu} \quad (2.1)$$

where q is the electron charge, and $h\nu = hc/\lambda$ is the wavelength-dependent energy of a single incident photon. In practice, the responsivity of a given semiconductor will also have a dependence on the bandgap, and different materials such as Si, GaAs, and InGaAs will be sensitive to different parts of the optical spectrum.

In a PIN device, some photons are lost due to surface reflection, absorption in the p-type region, and other non-idealities. These effects are taken into account by defining the quantum efficiency η of the device to be the ratio between the actual absorbed light and the total incident light. Thus, in reality, the responsivity of the PIN is

$$\mathcal{R}_{\text{PIN}} = \eta \frac{q}{h\nu} \quad (2.2)$$

For typical values of $\eta=0.75$ and $\lambda=1.5 \mu\text{m}$, $\mathcal{R}_{\text{PIN}}=0.91 \text{ A/W}$.

In an MSM photodetector, the presence of the metal electrodes will lower the overall responsivity of the device due to the light which is blocked by the electrode area. This shadowing factor will depend on the width x_w of the electrode and the distance x_d between adjacent electrodes. The transmitted light will be proportional to $x_d/(x_d + x_w)$, giving

$$\mathcal{R}_{\text{MSM}} = \eta \frac{q}{h\nu} \frac{x_d}{x_d + x_w} \quad (2.3)$$

For typical values of $\eta=0.75$, $\lambda=1.5 \mu\text{m}$, and $x_d=x_w=2 \mu\text{m}$, $\mathcal{R}_{\text{MSM}}=0.46 \text{ A/W}$. Thus, the responsivity of the MSM device will always be worse by approximately a factor of two versus the PIN.

Each type of photodetector is also characterized by a capacitance C_d which determines the maximum operating frequency of the device. In the case of the PIN diode, C_d is determined by the capacitance of the PIN junction, which is a parallel-plate capacitance

and depends on the device area and the thickness of the depletion region. In the MSM case, C_d is determined by the electrode spacing, electrode width, and device area. The MSM detector will have approximately half of the capacitance of a PIN of similar dimensions [14] due to the lateral structure of the MSM and the absence of any parallel plate geometry.

Each photodetector will also have a *dark current* (i_{dark}) which is defined as the leakage current through the biased device with zero incident light. The dark current is typically on the order of nanoamperes and will determine the minimum amount of light which is resolvable by the detector. It occurs due to thermal electron-hole pair generation within the semiconductor, and is considered the primary noise source in the detector.

The circuit model for either type of detector consists of two elements: (a) a current source with a constant portion (i_{dark}) and a dependent portion which is proportional to the responsivity times the incident light level $\mathcal{R} \times P$; and (b) the parallel detector capacitance C_d .

The choice between PIN and MSM photodetectors is often a tradeoff between electrical characteristics (responsivity and capacitance) and physical characteristics (technology and packaging). The PIN device is a vertical structure and thus requires a back contact or special processing in order to contact the cathode of the device. The MSM is a planar structure and may be easily integrated with the preamplifier into a single monolithic *optical-electronic integrated circuit* (OEIC) [15]. However, the choice of optical wavelength and amplifier technology may preclude this degree of integration. For example, GaAs OEICs may only be used at wavelengths of around 850 nm because GaAs photodetectors have poor responsivity at longer wavelengths (e.g. 1500 nm).

2.1.2 Preamplifier

The purpose of the preamplifier is to provide a low-noise interface in order to receive the small detector photocurrent. Ideally, the preamplifier is a high-quality current-in/voltage-out amplifier with high bandwidth to pass the desired bit rate. Unfortunately, most solid-state FET amplifiers are of the voltage-in/voltage-out or voltage-in/current-out variety. There are several methods of overcoming this problem.

In the simplest case, the photocurrent is converted into a voltage and then amplified by a voltage amplifier. This may be accomplished by using a load resistor R_L (Fig. 2.3). The resistor also serves to bias the photodetector (assuming that there is negligible IR drop

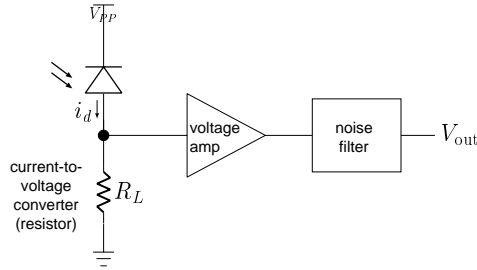


Figure 2.3. Optical preamplifier using a current-to-voltage converter (high-impedance topology).

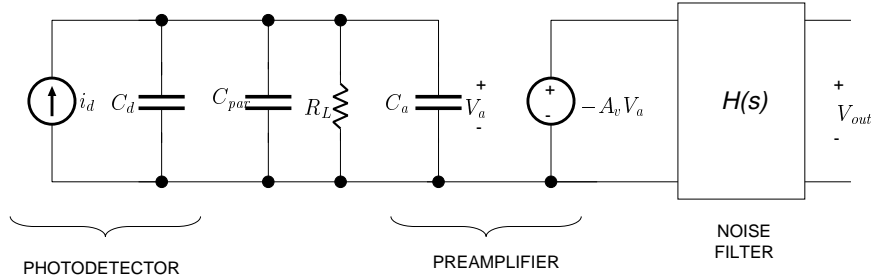


Figure 2.4. Equivalent circuit for high-impedance optical receiver design.

across R_L). This type of optical receiver is called the *high-impedance* design for reasons which will soon become clear. The amplifier is presumed to be ideal, with all frequency dependence lumped into the noise filter. The equivalent circuit for this design is shown in Fig. 2.4, with amplifier input capacitance C_a and gain $-A_v$ (we will assume an FET amplifier with infinite input resistance). Parasitic interconnect capacitance between the detector and amplifier is represented by C_{par} .

The resulting transfer function is

$$\frac{V_{out}(s)}{I_d(s)} = \frac{-A_v H(s)/C_T}{s + \frac{1}{R_L C_T}} \quad (2.4)$$

where the total input capacitance $C_T = C_d + C_{par} + C_a$. Note the high-impedance node at the input, creating a pole at $1/R_L C_T$. Because it is desirable to make R_L as large as possible in order to reduce the effects of thermal noise (this will be discussed later), this high impedance will cause a pole to occur at a low frequency such that the circuit will

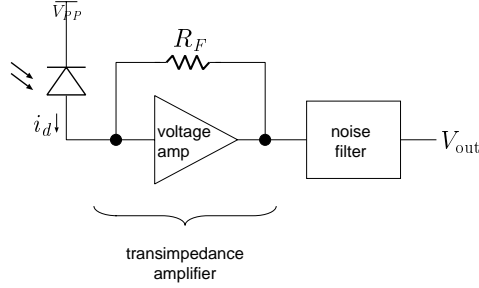


Figure 2.5. Optical preamplifier using a transimpedance amplifier.

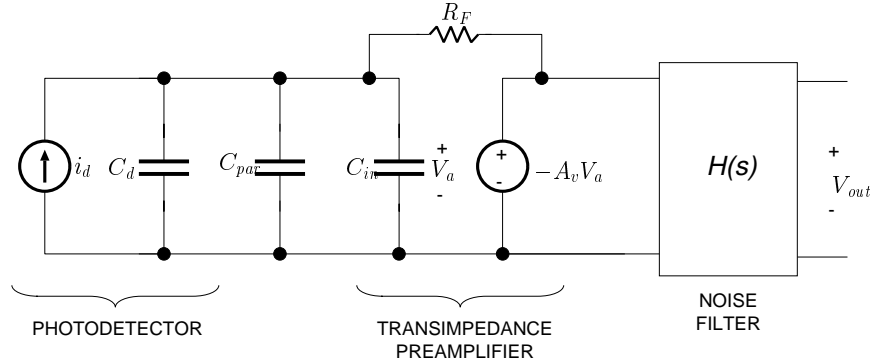


Figure 2.6. Circuit model for the transimpedance topology.

behave as an integrator at frequencies above $1/R_L C_T$. While this integrating behavior may be accounted for by adding an appropriate canceling zero into the noise filter $H(s)$, a more serious problem is that the integration can cause amplifier saturation after long strings of ones or zeros in the incoming bit sequence.

An alternative to the high-impedance design is the *transimpedance* design of Fig. 2.5. In this topology, a voltage-in/voltage-out amplifier is converted to current-in/voltage-out by using a resistor R_F in feedback. The amplifier input is assumed to be biased at ground in order to bias the photodiode. The equivalent circuit is shown in Fig. 2.6. The resulting transfer function will be

$$\frac{V_{out}(s)}{I_d(s)} = \frac{-A_v H(s)/C_T}{s + \frac{A_v + 1}{R_F C_T}} \quad (2.5)$$

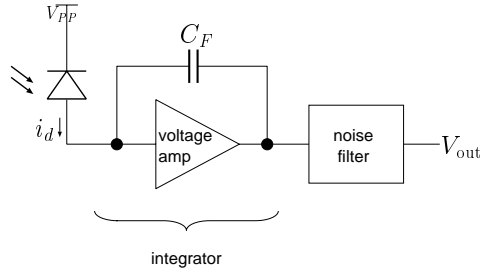


Figure 2.7. Optical preamplifier with noiseless feedback.

It is apparent from (2.5) that the pole due to the input node has been pushed out by a factor of $A_v + 1$, thereby eliminating the integrating behavior of the circuit. Thus, the transimpedance design has several distinct advantages over the high-impedance design: (a) R_F may now be made large without causing the amplifier to behave like an integrator; (b) no pole-canceling zero is necessary in the noise filter. In actual designs, the parasitic capacitance of the feedback resistor R_F may cause the transimpedance design to have a lower frequency cutoff than the above ideal analysis, thus limiting the bitrate of the receiver. An alternate solution involves lowering the value of R_F to reduce the parasitic capacitance, thus trading off noise performance for bandwidth.

A third class of preamplifier modifies the transimpedance design and uses a noiseless element as the feedback element. This is accomplished by using a capacitor in feedback instead of a resistor (Fig. 2.7). The resulting transfer function is

$$\frac{V_{out}(s)}{I_d(s)} = \frac{-A_v H(s)}{C_T + (A_v + 1)C_F} \frac{1}{s} \quad (2.6)$$

The highest gain is achieved in the limit of $C_F \rightarrow 0$,

$$\frac{V_{out}(s)}{I_d(s)} = \frac{-A_v H(s)}{C_T} \frac{1}{s} \quad (2.7)$$

In this design, the preamplifier is intentionally designed to have an integrating response. However, instead of canceling the pole at the origin with a corresponding zero (as in the high-impedance design), a switch may be added around the amplifier in order to reset (dump) the integrator to zero initial conditions. This reset switch is closed briefly after every received bit and is then reopened to receive each subsequent bit. In addition, the size

of C_F (or its omission) may be used to easily control the gain of the receiver in order to optimize the dynamic range for large or small inputs.

The obvious advantage to this *integrate-and-dump* topology is the absence of any noise-generating resistor. However, an added drawback is the necessity to generate a signal to operate the reset switch.

2.1.3 Noise-Shaping Filter

The purpose of the noise-shaping filter is to reduce the wideband circuit noise without significantly altering the signal waveform. Typically, the noise shaper is a lowpass filter which simply bandlimits the noise while at the same time removing the high-frequency components of the bit waveform. Thus, transmitted bits which are rectangular will have their edges rounded by the filter, and in the extreme case, will become sine-like in appearance. The bit rounding may also cause *intersymbol interference* when a given bit is smeared into the timeslot of an adjacent bit.

In order to choose the proper noise shaping filter, it is first necessary to determine the noise response of the preamplifier. The equivalent noise model for an FET preamplifier is shown in Fig. 2.8. In this model, the resistor R_L represents *either* the load resistor (for the high-impedance design) or the feedback resistor (for the transimpedance design). For the integrate-and-dump design, $R_L \rightarrow \infty$. Noise in the receiver comes from several sources:

- shot noise due to the dark current in the photodetector $\overline{i_{\text{dark}}^2}$
- thermal noise in the bias resistor $\overline{i_L^2}$
- input-referred thermal noise in the amplifier input device $\overline{v_a^2}$
- shot noise due to leakage current in the amplifier input device $\overline{i_a^2}$

where the notation $\overline{a^2}$ refers to the integrated mean-square noise of parameter a . We will neglect the shot noise current due to the signal photocurrent in the photodetector, which is presumed small compared to the signal photocurrent itself.

We will assume that the noise of the whole receiver is determined solely by the noise of the input circuit, including the detector, bias resistor, and the input transistor of the preamplifier. This will be true if the preamplifier has high gain. We also model the preamplifier as an ideal gain $-A_v$, and the noise-filter frequency response as $H(s)$. Note that

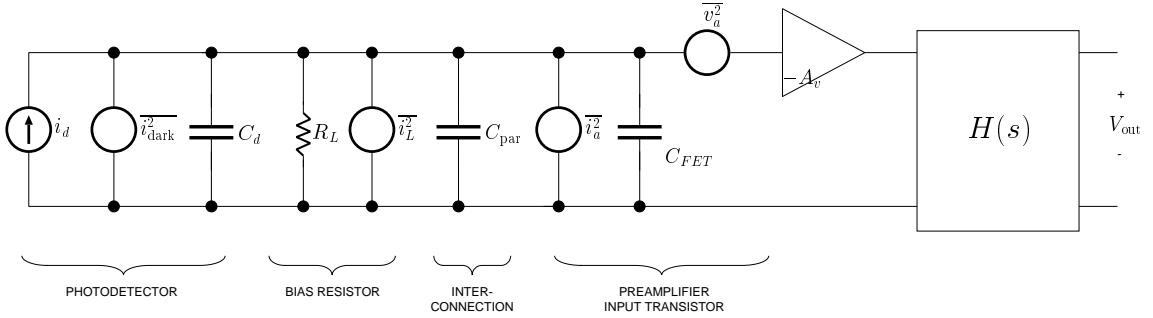


Figure 2.8. Noise sources for the optical receiver front-end.

$H(s)$ is *not* the response of the whole receiver, which we will define as

$$Z_T(s) = \frac{V_{\text{out}}(s)}{I_d(s)} = \frac{-A_v H(s)/C_T}{s + \frac{1}{R_L C_T}} \quad (2.8)$$

where $Z_T(s)$ is the *transimpedance* of the circuit in units of ohms, and $C_T = C_D + C_{\text{par}} + C_{\text{FET}}$.

In order to calculate the total noise, we must first determine the effect of each individual noise source at the output of the receiver. The noise sources may be grouped into two types: *parallel* noise sources ($\overline{i_{\text{dark}}^2}$, $\overline{i_L^2}$, and $\overline{i_a^2}$) which shunt the input current source, and *series* noise sources ($\overline{v_a^2}$) which are in series with the input. The output noise will depend on the power spectral densities of the parallel and series noise sources, and their respective transfer functions to the output. Thus, the noise at the output will be

$$\overline{n_{\text{parallel}}^2} = \int_0^\infty \frac{di_{\text{parallel}}^2}{df} |Z_{\text{parallel}}(s)|^2 df \quad (2.9)$$

$$\overline{n_{\text{series}}^2} = \int_0^\infty \frac{dv_{\text{series}}^2}{df} |H_{\text{series}}(s)|^2 df \quad (2.10)$$

where $\frac{di_{\text{parallel}}^2}{df}$ and $\frac{dv_{\text{series}}^2}{df}$ are the conventional *single-sided* power spectral densities (hence the integration limits from 0 to ∞), and $Z_{\text{parallel}}(s)$ and $H_{\text{series}}(s)$ are the transfer functions for the parallel and series noise sources, respectively. For Fig. 2.8, these transfer functions are

$$Z_{\text{parallel}}(s) = \frac{-A_v H(s)/C_T}{s + \frac{1}{R_L C_T}} \quad (2.11)$$

$$= Z_T(s) \quad (2.12)$$

$$H_{\text{series}}(s) = -A_v H(s) \quad (2.13)$$

$$= Z_T(s) \left(sC_T + \frac{1}{R_L} \right) \quad (2.14)$$

In the case of the parallel noise sources $\overline{i_{\text{dark}}^2}$, $\overline{i_L^2}$, and $\overline{i_a^2}$, the noise power spectral density for each source will be

$$\frac{\overline{di_{\text{dark}}^2}}{df} = 2qi_{\text{dark}} \quad (2.15)$$

$$\frac{\overline{di_a^2}}{df} = 2qi_{gs} \quad (2.16)$$

$$\frac{\overline{di_L^2}}{df} = \frac{4k\Theta}{R_L} \quad (2.17)$$

where q is the electron charge, k is Boltzmann's constant, Θ is temperature (not to be confused with T for bit period), and i_{gs} is the gate leakage current of the preamplifier input transistor. Thus, the total parallel noise spectrum at the output will be

$$\frac{\overline{dn_{\text{out,parallel}}^2}}{df} = \left[2qi_{\text{dark}} + 2qi_{gs} + \frac{4k\Theta}{R_L} \right] |Z_{\text{parallel}}(f)|^2 \quad (2.18)$$

$$= \left[2qi_{\text{dark}} + 2qi_{gs} + \frac{4k\Theta}{R_L} \right] |Z_T(f)|^2 \quad (2.19)$$

Likewise, for the series noise, the noise power spectral density at the input will be [56]

$$\frac{\overline{dv_a^2}}{df} = \frac{4k\Theta\Gamma}{g_m} \quad (2.20)$$

where Γ is the thermal noise constant for FETs ($\simeq 0.7$ for MOSFETs or $\simeq 1.1$ for MESFETs), and g_m is the transconductance of the preamplifier input transistor. The total output series noise spectrum will be

$$\frac{\overline{dn_{\text{out,series}}^2}}{df} = \frac{4k\Theta\Gamma}{g_m} |H_{\text{series}}(s)|^2 \quad (2.21)$$

$$= \frac{4k\Theta\Gamma}{g_m} |Z_T(s)|^2 \left((2\pi f C_T)^2 + \left(\frac{1}{R_L} \right)^2 \right) \quad (2.22)$$

Combining (2.19) and (2.22), and assuming that the series and parallel noises are independent, the total output noise spectrum will be

$$\frac{\overline{dn_{\text{out}}^2}}{df} = \frac{\overline{dn_{\text{out,parallel}}^2}}{df} + \frac{\overline{dn_{\text{out,series}}^2}}{df} \quad (2.23)$$

$$= \left[2qi_{\text{dark}} + 2qi_{gs} + \frac{4k\Theta}{R_L} \left(1 + \frac{\Gamma}{g_m R_L} \right) + \frac{4k\Theta\Gamma(2\pi C_T)^2}{g_m} f^2 \right] |Z_T(s)|^2 \quad (2.24)$$

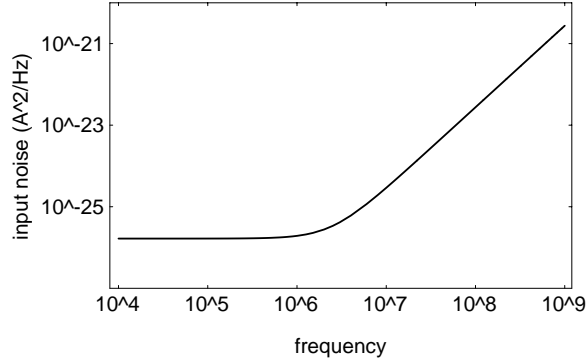


Figure 2.9. Input-referred noise spectrum. Typical values are assumed: $i_{\text{dark}}=10^{-12}$ A, $i_{gs}=10^{-12}$ A, $R_L=1$ M Ω , $g_m=0.006$ S, $C_T=6$ pF.

We may also determine the *input-referred noise spectrum* by dividing the output noise spectrum by the magnitude-square of the transfer function, and thus

$$\frac{\overline{dn_{\text{in}}^2}}{df} = 2qi_{\text{dark}} + 2qi_{gs} + \frac{4k\Theta}{R_L} \left(1 + \frac{\Gamma}{g_m R_L}\right) + \frac{4k\Theta\Gamma(2\pi C_T)^2}{g_m} f^2 \quad (2.25)$$

This input noise spectrum is plotted in Fig. 2.9. There is a characteristic *noise corner* where the spectrum assumes an f^2 dependence at

$$f_{\text{noise corner}} = \frac{\sqrt{g_m R_L / \Gamma + 1}}{(2\pi)^2 R_L C_T} \quad (2.26)$$

where the dark current and gate leakage are assumed small.

By inspection of (2.25), we may make the following observations in order to minimize the receiver noise:

- dark current and leakage current should be minimized
- load/feedback resistance R_L should be maximized
- preamplifier transconductance g_m should be maximized
- input capacitance C_T should be minimized
- We may optimize the last term in (2.25) by using the implicit relationship between g_m and C_{FET} for any FET. For example, for a MOSFET amplifier with input transistors operating in the saturation region, $g_m = \frac{W}{L} \mu C_{ox} (V_{GS} - V_t)$ and $C_{FET} \approx \frac{2}{3} W L C_{ox}$, where

W is the gate width, L is the gate length, μ is the channel mobility, C_{ox} is the gate oxide area capacitance, V_{GS} is the gate-source voltage, and V_t is the threshold voltage. Substituting into the last term of (2.24) gives

$$\frac{4k\Theta\Gamma\left(2\pi\left(C_D + C_{par} + \frac{2}{3}WLC_{ox}\right)\right)^2}{\frac{W}{L}\mu C_{ox}(V_{GS} - V_t)} f^2 \quad (2.27)$$

This term may be minimized (by differentiating with respect to W and setting to zero), yielding the well-known result that the input capacitance of the preamplifier should be matched to the detector [16, 17]

$$C_{FET,opt} = C_D + C_{par} \quad (2.28)$$

We may now also determine the variance of the output noise by integrating (2.24) over all positive frequencies

$$\overline{n_{out}^2} = \int_0^\infty \frac{dn_{out}^2}{df} df \quad (2.29)$$

$$\begin{aligned} &= \left[2qi_{\text{dark}} + 2qi_{gs} + \frac{4k\Theta}{R_L} \left(1 + \frac{\Gamma}{g_m R_L} \right) \right] \int_0^\infty |Z_T(s)|^2 df \\ &\quad + \frac{4k\Theta\Gamma(2\pi C_T)^2}{g_m} \int_0^\infty |Z_T(s)|^2 f^2 df \end{aligned} \quad (2.30)$$

2.1.4 Timing Recovery and Decision Circuit

The recovery of a clock from the received bitstream is necessary in order to synchronize the bits to the local equipment. In addition, in the case of the integrate-and-dump receiver, the recovered clock is used in the receiver itself to operate the reset switch. There are two common methods used to recover the clock: *open-loop synchronizers* and *closed-loop synchronizers*.

In the open-loop case, a spectral line at the bitrate is extracted directly from the incoming bitstream. For example, in certain cases such as return-to-zero (RZ) line coding, the frequency spectrum of the bitstream will have a component at the bitrate, and a linear bandpass filter is used to isolate and amplify that component. In other cases, such as non-return-to-zero (NRZ) line coding, there is no spectral component at the bitrate. In these cases, the bitstream is first “filtered” with a nonlinear function such as square law or absolute value in order to create a spectral line at the bitrate. Then, a bandpass filter is used as in the RZ case (Fig. 2.10a).

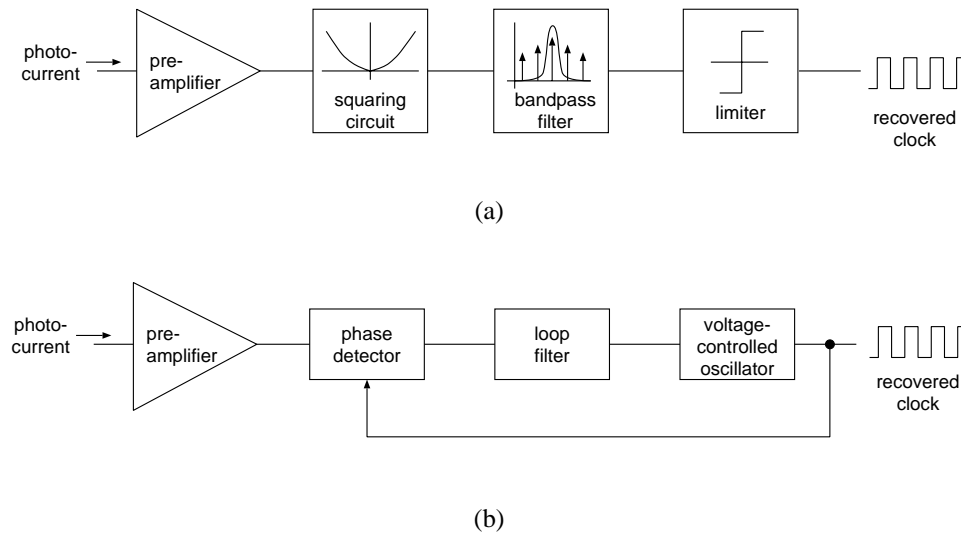


Figure 2.10. Two methods of clock recovery. (a) In an open-loop synchronizer, the preamplifier output is squared and then bandpass-filtered to recover the spectral component at the bitrate. (b) In a closed-loop synchronizer, a phase-locked loop is used to recover the clock.

The main drawback to open-loop synchronizers is that the clock recovery will be highly dependent on the signal-to-noise ratio of the received bitstream. For the smallest received signals, the recovered clock will contain a significant jitter component due to the noise in the bitstream. This will seriously degrade performance at low signal-to-noise levels, i.e. low incident light levels.

In closed-loop synchronizers, a local variable-frequency oscillator is locked onto the received bitstream by using a phase-locked loop (PLL). This is achieved through traditional PLL methods [18] by using a phase detector, loop filter, and voltage-controlled oscillator (VCO) as shown in Fig. 2.10b. The PLL has a long enough time constant such that it remains locked to the incoming bitstream even with the occasional absence of bit transitions (as in the case of long strings of “ones” or “zeros”). However, in the case of the integrate-and-dump receiver topology, a problem arises: since the recovered clock is necessary to operate the receiver (in order to close the reset switch) and receive the bitstream, how is the bitstream recovered in order to extract the clock? In this case, an early-late gate synchronizer loop may be used [19] as shown in Fig. 2.11.

In the early-late gate synchronizer, each bit is divided into two equal pieces: an early

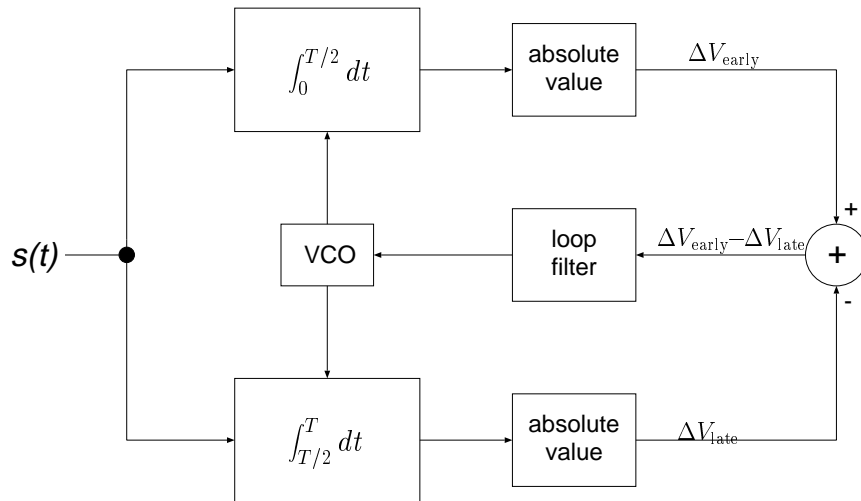


Figure 2.11. Early-late gate data synchronizer [19].

gate (first half of each bit) and a late gate (second half of each bit). Then, two integrators are used to measure each bit during the two gates. The magnitudes of the integrations are then subtracted and used to control a VCO. This is illustrated in the timing diagram of Fig. 2.12, which shows the circuit operation for three cases: a correctly timed bit, an early bit, and a late bit. For the correctly timed case, the bit starts exactly at the beginning of the early gate, and ends exactly at the end of the late gate. Assuming the bit is a “one,” the results of the early and late integrations will be identical, i.e. $\Delta V_{\text{early}} = \Delta V_{\text{late}}$. Because the difference is zero, there will be no correction sent to the VCO.

In the case of an early bit, the early and late gates will not line up on the bit boundaries, and the integration results will not be equal. In this case, the VCO control voltage is adjusted in proportion to the difference in the integrations. Similarly, for a late bit, the integration results will be different and the VCO frequency will be adjusted by a proportional (but opposite) amount. Assuming that the gains are set properly within the loop, this synchronizer will eventually lock the VCO onto the proper clock frequency. This type of loop has been analyzed in detail in [18].

In the case of the integrate-and-dump receiver, the preamplifier may be enclosed within the clock recovery loop by simply taking samples of the output at the appropriate moment with sample-and-hold circuits (Fig. 2.13). The integrator output is sampled three times: at

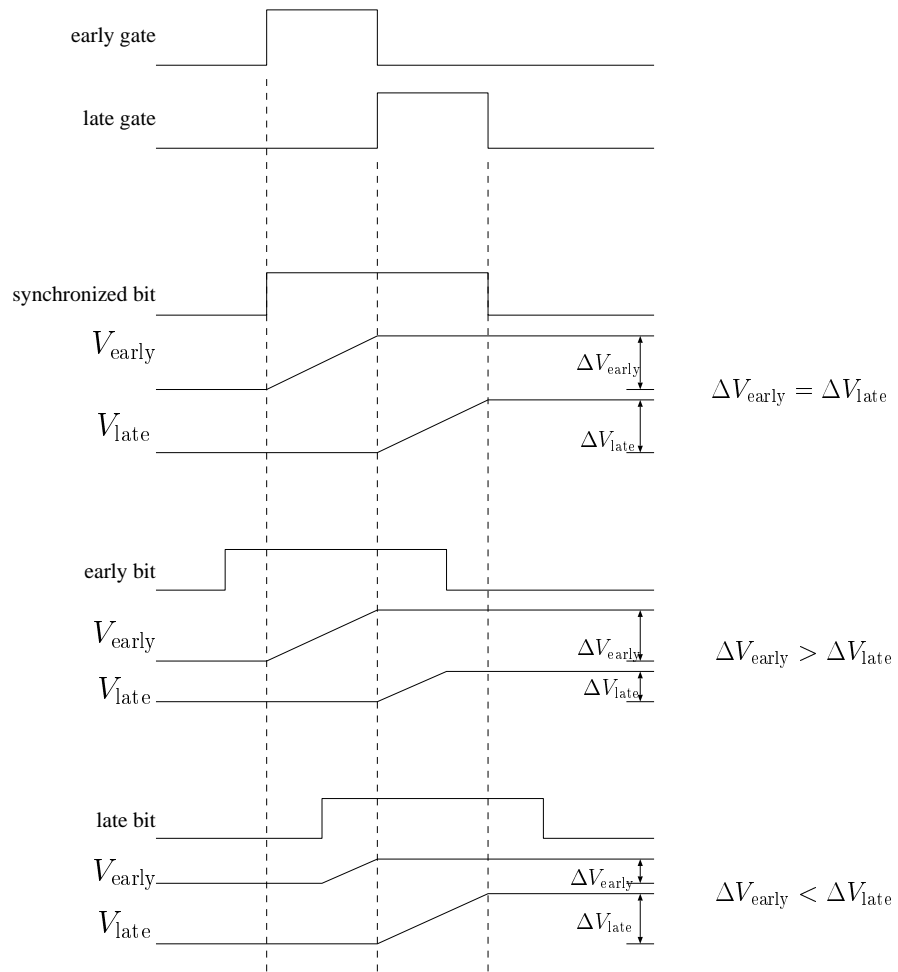
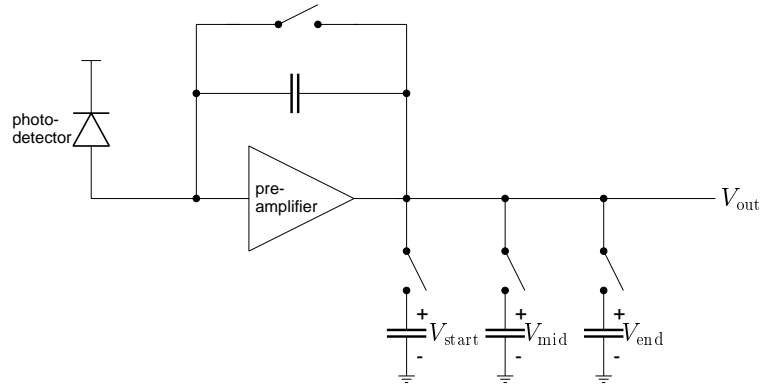
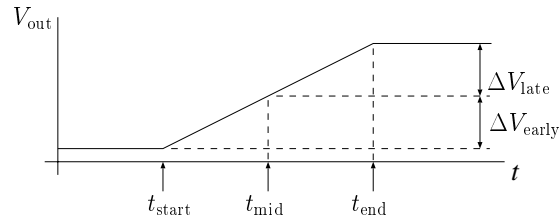


Figure 2.12. Timing diagram for early-late gate synchronizer.



(a)



(b)

Figure 2.13. Implementation of early-late gate synchronizer with integrating front-end. (a) The preamplifier output is sampled three times: at the start, middle, and end of the bit period. (b) Then, the VCO control signals are formed by generating $\Delta V_{\text{early}} = V_{\text{mid}} - V_{\text{start}}$ and $\Delta V_{\text{late}} = V_{\text{end}} - V_{\text{mid}}$.

the beginning of the integration (V_{start}), at the exact center of the integration time (V_{mid}), and at the end of the integration (V_{end}). The early gate and late gate integrations may then be formed by generating the following differences (not shown in the diagram)

$$\Delta V_{\text{early}} = V_{\text{mid}} - V_{\text{start}} \quad (2.31)$$

$$\Delta V_{\text{late}} = V_{\text{end}} - V_{\text{mid}} \quad (2.32)$$

This type of arithmetic is easily done using switched-capacitor circuitry [20].

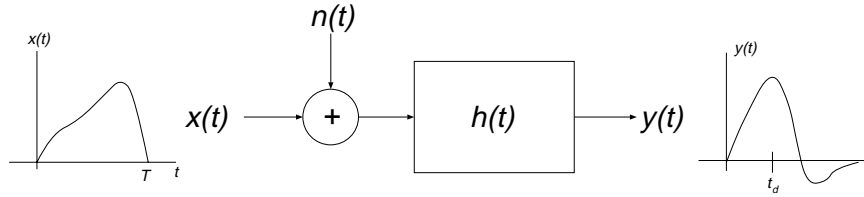


Figure 2.14. Linear system with additive noise. The output of the filter $h(t)$ is sampled at t_d in order to decide the presence or absence of the input pulse $x(t)$.

2.2 OPTIMAL NOISE FILTERING

Signal-to-noise ratio (SNR) is defined as the ratio of the peak output signal power to the rms output noise power [21, 22]. In the case of binary digital transmission, we want to filter each bit in order to maximize the signal component and minimize the noise component. In addition, we want to sample the filter output at the moment when it is at its maximum. Fig. 2.14 shows a simple linear system with input $x(t)$, output $y(t)$, additive noise $n(t)$, and filter impulse response $h(t)$. $x(t)$ is assumed to be a single digital bit of arbitrary shape and finite length. By inspection, we know that

$$y(t) = h(t) * x(t) + h(t) * n(t) \quad (2.33)$$

$$= y_s(t) + y_n(t) \quad (2.34)$$

where $y_s(t) \equiv h(t) * x(t)$ is the signal component of $y(t)$, and $y_n(t) \equiv h(t) * n(t)$ is the noise component of $y(t)$. Our goal is to determine the best $h(t)$ in order to maximize the SNR.

The SNR at the output of this system is defined as

$$\text{SNR} \equiv \frac{[y_s(t)]_{\max}^2}{y_n^2} \quad (2.35)$$

where the term on top refers to the square of the peak value of $y_s(t)$, and the bottom term refers to the time-average value of the square of $y_n(t)$. Note that the squares are necessary because SNR is always defined as a ratio of *powers* and not of signal *magnitudes*.

Let us define t_d as the instant in time when the signal component $y_s(t)$ is at its peak. Thus,

$$[y_s(t)]_{\max} = y_s(t_d) \quad (2.36)$$

$$= h(t) * x(t)|_{t=t_d} \quad (2.37)$$

$$= \mathcal{F}^{-1}[H(f)X(f)]|_{t=t_d} \quad (2.38)$$

$$= \int_{-\infty}^{\infty} H(f)X(f)e^{j2\pi ft_d} df \quad (2.39)$$

where $H(f)$ is the Fourier transform of $h(t)$, $X(f)$ is the Fourier transform of $x(t)$, and \mathcal{F}^{-1} denotes the inverse Fourier transform defined as $X(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft} dt$ and $x(t) = \int_{-\infty}^{\infty} X(f)e^{j2\pi ft} df$.

By applying Parseval's Theorem, we may express $\overline{y_n^2}$ in terms of the power spectral density of $y_n(t)$, denoted $S_{Y_N}(f)$

$$\overline{y_n^2} = \int_{-\infty}^{\infty} |S_{Y_N}(f)|^2 df \quad (2.40)$$

$$= \int_{-\infty}^{\infty} S_N(f) |H(f)|^2 df \quad (2.41)$$

where $S_N(f)$ is the power spectral density of $n(t)$.

Substituting (2.39) and (2.41) into (2.35), we get a frequency-domain representation for the SNR:

$$\text{SNR} = \frac{\left| \int_{-\infty}^{\infty} H(f)X(f)e^{j2\pi ft_d} df \right|^2}{\int_{-\infty}^{\infty} |H(f)|^2 S_N(f) df} \quad (2.42)$$

In order to maximize the SNR, we may apply the Schwarz inequality, which states

$$\frac{\left| \int_{-\infty}^{\infty} V(f)W(f) df \right|^2}{\int_{-\infty}^{\infty} |V(f)|^2 df} \leq \int_{-\infty}^{\infty} |W(f)|^2 df \quad (2.43)$$

where $V(f)$ and $W(f)$ are arbitrary functions. By comparing (2.42) with (2.43), we can identify

$$V(f) = H(f)\sqrt{S_N(f)} \quad (2.44)$$

$$W(f) = X(f)e^{j2\pi ft_d} / \sqrt{S_N(f)} \quad (2.45)$$

Note that because the power spectral density is defined as real and positive [23], the square root $\sqrt{S_N(f)}$ is real. Since (2.42) has the same form as the left side of (2.43), we can maximize the left side (2.43) and thus maximize the SNR by replacing the less-than-or-equal sign of (2.43) with an equal sign. Making this replacement and solving (2.43) gives

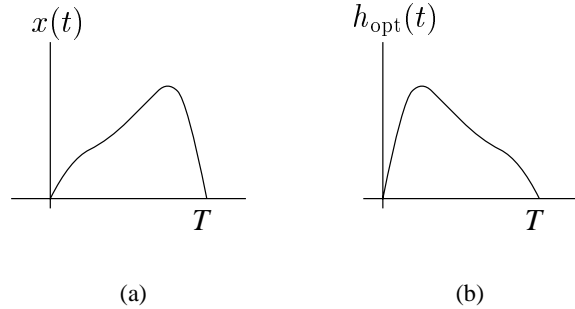


Figure 2.15. The *matched filter* receiver has an impulse response $h_{\text{opt}}(t)$ which is a time-reversed copy of the received input pulse $x(t)$. White noise is assumed.

$V(f) = KW^*(f)$, and thus substituting (2.44) and (2.45) gives

$$H(f)\sqrt{S_N(f)} = K \left[X(f)e^{j2\pi ft_d} / \sqrt{S_N(f)} \right]^* \quad (2.46)$$

where K is an arbitrary constant. Solving (2.46) for $H(f)$ gives an expression for the optimum filter response $H_{\text{opt}}(f)$ for maximum SNR

$$H_{\text{opt}}(f) = \frac{KX^*(f)e^{-j2\pi ft_d}}{S_N(f)} \quad (2.47)$$

and the maximum SNR is

$$\text{SNR}_{\text{opt}} = \int_{-\infty}^{\infty} \frac{|X(f)|^2}{S_N(f)} df \quad (2.48)$$

In the special case of white gaussian noise, the noise spectrum $S_N(f) = N_o$ is constant. Then, converting (2.47) to time domain gives

$$h_{\text{opt}}(t) = K'x(t_d - t) \quad (2.49)$$

where $K' = K/N_o$. Qualitatively, (2.49) tells us that $h_{\text{opt}}(t)$ is a time-reversed copy of $x(t)$ pivoted about the sampling time [50, 51, 61]. This is shown in Fig. 2.15, which shows an arbitrary input pulse and its associated matched filter impulse response.

The solution for $H_{\text{opt}}(f)$ in 2.47 provides the theoretical best-case filtering for a given pulse shape and noise power spectrum. However, it gives no insight into the implementation of such a filter, or even its realizability. For example, $H_{\text{opt}}(f)$ may turn out to be noncausal or infinite at certain points. In general, an approximation to the matched response used instead.

2.3 BIT ERROR RATE

In this section, we will examine the predicted noise performance for several different types of noise filtering. We will use standard probabilistic models to determine the sensitivity of a receiver at a given *bit error rate* (BER). Bit error rate is expressed as a ratio of errors to nonerrors, e.g. a BER of 10^{-9} corresponds to one error in every 10^9 bits. The characterization of noise in terms of BER is useful because BER may be measured directly in the laboratory by using a bit error rate tester.

2.3.1 Quantum Noise Limit

We will first determine the best-case receiver sensitivity based on quantum statistics. Although this best case can never actually be implemented, it does provide a baseline limit with which to check the results of our later calculations.

Planck's equation gives the amount of energy in a single photon of wavelength λ

$$E_{\text{photon}} = \frac{hc}{\lambda} \quad (2.50)$$

Given the randomness of events, it is impossible to ascertain the exact arrival times of the photons at the photodetector, or the time of their conversion into electron-hole pairs. Thus, for a given *average arrival rate* of photons ℓ , the probability of N photons arriving during interval T is given by the Poisson distribution [23]

$$\text{Prob}(N) = \frac{(\ell T)^N e^{-\ell T}}{N!} \quad (2.51)$$

We may compute the average rate of photon arrival from the incident power P (in units of joules/second, or watts) and the photon energy (in units of joules)

$$\ell = \frac{P}{E_{\text{photon}}} \quad (2.52)$$

$$= \frac{P\lambda}{hc} \text{ photons/second} \quad (2.53)$$

We will now make the following assumptions about the transmission of light down the fiber:

- The received light is modulated in an ideal OOK format, and thus no photons are present during a “zero” bit

- The receiver is an ideal noiseless “photon detector,” and outputs a “one” upon reception of one or more photons during a given bit period T . If no photons are received during time T , then the receiver outputs a “zero.”
- There is an equal probability of a “one” or a “zero” being transmitted.

Give the above constraints, the total probability of error depends on the probability $Prob(0|1)$ that a “zero” is detected given that a “one” was transmitted, and the probability $Prob(1|0)$ that a “one” is detected given that a “zero” was transmitted.

$$\begin{aligned}
 Prob(\text{bit error}) &= Prob(0)Prob(0|1) + Prob(1)Prob(1|0) & (2.54) \\
 &= \frac{1}{2} \times (\text{probability of a photon received for a transmitted “zero”}) + \\
 &\quad \frac{1}{2} \times (\text{probability of no photons received for a transmitted “one”})
 \end{aligned}$$

We will assume that photons cannot be accidentally detected in a noiseless photon detector, and thus the first term on the righthand side of (2.54) is zero. However, due to the Poisson distribution of photon arrivals during a “one” bit, there is a finite possibility that no photons will arrive during the bit time T . Thus, substituting (2.51) into (2.54), and solving for a BER of 10^{-9} gives

$$10^{-9} = \frac{1}{2}(0) + \frac{1}{2} \frac{(\ell T)^0 e^{-\ell T}}{0!} \quad (2.55)$$

Letting $\ell = P\lambda/hc$ and solving for P gives

$$P = \frac{-hc}{\lambda T} \ln(2 \times 10^{-9}) \quad (2.56)$$

Stated another way, (2.55) shows that, on average, $\ell T = 20$ photons per “one” bit must be transmitted in order to achieve a BER of 10^{-9} .

Receiver sensitivity is generally expressed in terms of *average power* \bar{P} , where $\bar{P} = (P_{\text{high}} + P_{\text{low}})/2$. The expression derived in (2.56) is for the power in a “one” bit, i.e. $P = P_{\text{high}}$. Thus, for comparison purposes, the quantum limit on receiver sensitivity (assuming $P_{\text{low}} = 0$) is

$$\left(\bar{P}\right)_{\text{quantum-limit}} = \frac{-hc}{2\lambda T} \ln(2 \times 10^{-9}) \quad (2.57)$$

2.3.2 BER and Gaussian noise

Since it is impossible to build an ideal photon detector as a receiver, we will now examine the performance of real receivers using noisy amplifiers. Thus, we must translate

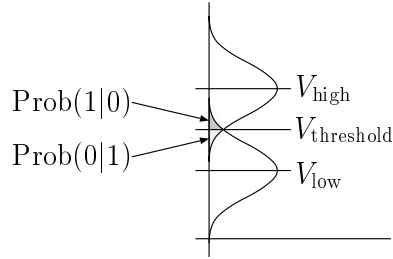


Figure 2.16. Determination of bit error rate for a gaussian noise process. The errors are indicated by the shaded regions.

the signal-to-noise ratios derived previously into bit error rates. This is accomplished by assuming gaussian statistics as shown in Fig. 2.16. We assume that the high and low voltages at the output of the receiver are random variables and follow the standard distribution with probability density function

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad (2.58)$$

We want to calculate the two shaded areas in Fig. 2.16 which correspond to $Prob(0|1)$ and $Prob(1|0)$. These two quantities may be calculated by integrating under the two curves:

$$Prob(0|1) = \frac{1}{\sqrt{2\pi}} \int_{Q_1}^{\infty} e^{-x^2/2} dx \quad (2.59)$$

$$Prob(1|0) = \frac{1}{\sqrt{2\pi}} \int_{Q_2}^{\infty} e^{-x^2/2} dx \quad (2.60)$$

where $Q_1 = (V_{\text{threshold}} - V_{\text{low}})/\sigma_{\text{noise}}$ and $Q_2 = (V_{\text{high}} - V_{\text{threshold}})/\sigma_{\text{noise}}$ are integration limits which are normalized by the mean and variance of the noise. We will assume that the noise variance σ_{noise} is the same for both “ones” and “zeros.” We will also assume that the distribution of “ones” and “zeros” is equal, and thus the optimum threshold level is $(V_{\text{high}} + V_{\text{low}})/2$. In this case, we can define $Q = Q_1 = Q_2 = \frac{V_{\text{high}} - V_{\text{low}}}{2\sigma_{\text{noise}}}$. We also have already calculated $\sigma_{\text{noise}}^2 = \overline{n_{\text{out}}^2}$ in (2.30). Thus, the total number of errors will be

$$\text{BER} = Prob(1)Prob(0|1) + Prob(0)Prob(1|0) \quad (2.61)$$

$$= \frac{1}{2}Prob(0|1) + \frac{1}{2}Prob(1|0) \quad (2.62)$$

$$= \frac{1}{\sqrt{2\pi}} \int_Q^{\infty} e^{-x^2/2} dx \quad (2.63)$$

The bit error rate as a function of the signal-to-noise ratio Q is plotted in Fig. 2.17. For a typical BER of 10^{-9} , $Q=6$.

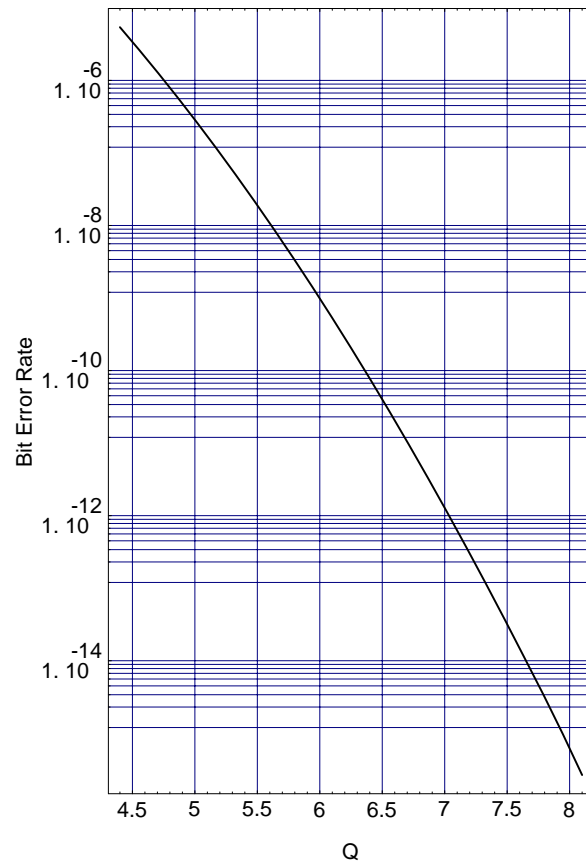


Figure 2.17. Plot of bit error rate vs. Q .

V_{high} and V_{low} are related to the input power by the responsivity of the detector and the output voltage of the receiver

$$V_{\text{high}} = \mathcal{R}_{\text{det}} P_{\text{high}} [y_s(t)]_{\text{max}} \quad (2.64)$$

$$V_{\text{low}} = \mathcal{R}_{\text{det}} P_{\text{low}} [y_s(t)]_{\text{max}} \quad (2.65)$$

where $[y_s(t)]_{\text{max}}$ is the maximum output voltage of the receiver during a given bit period, and P_{high} and P_{low} are the optical powers for a high and low bit, respectively. We may define the average incident input power \bar{P} as the average of the powers for high and low inputs

$$\bar{P} = \frac{P_{\text{high}} + P_{\text{low}}}{2} \quad (2.66)$$

We may also define the *extinction ratio* r as the ratio between the low and high power levels

$$r = \frac{P_{\text{low}}}{P_{\text{high}}} \quad (2.67)$$

Combining (2.66), (2.67), (2.64) and (2.65) yields an expression for the average incident input power level necessary to obtain a given bit error rate:

$$\bar{P} = \left(\frac{1+r}{1-r} \right) \left(\frac{1}{\mathcal{R}_{\text{det}}} \right) Q \left(\frac{\sqrt{n_{\text{out}}^2}}{[y_s(t)]_{\text{max}}} \right) \quad (2.68)$$

When describing receiver sensitivity, it is common to factor out the detector-dependent parameters in order to compare receivers independently of the type of photodetector used [16, 1, 14]. Thus, for the PIN diode case, using $\mathcal{R}_{\text{det}} = (\eta q)/(h\nu)$, we define the *receiver sensitivity* $\eta\bar{P}$ as

$$\eta\bar{P} = \left(\frac{1+r}{1-r} \right) \left(\frac{h\nu}{q} \right) Q \left(\frac{\sqrt{n_{\text{out}}^2}}{[y_s(t)]_{\text{max}}} \right) \quad (2.69)$$

In the ideal case where no power is transmitted for a “zero” bit ($r=0$), the best-case sensitivity is

$$\eta\bar{P} = \left(\frac{h\nu}{q} \right) Q \left(\frac{\sqrt{n_{\text{out}}^2}}{[y_s(t)]_{\text{max}}} \right) \quad (2.70)$$

2.4 COMPARISON OF NOISE FILTERS

Given the expression (2.70) for the receiver sensitivity, we may now compare the performances of several types of transfer functions for the receiver. Four different types of transfer functions will be considered:

1. The matched filter as derived in Section 2.2
2. A simple second-order transfer function which simulates cases where the amplifier frequency response is used as the noise filter
3. A transfer function which generates a raised-cosine frequency spectrum at the output, as described by Personick [24]
4. An integrate-and-dump preamplifier

The choice of these four cases may be rationalized for several reasons. Although generally unrealizable, the matched filter case will provide an upper limit on the sensitivity of a given optical receiver with a given input noise spectrum. On the other hand, the second-order response case attempts to simulate the performance of many realized systems in the world today. The third choice, that of a raised-cosine output spectrum, has been used in most theoretical sensitivity analyses in the literature to date [24, 25, 16, 26, 27, 28]. The fourth choice, that of the integrate-and-dump response, will be shown to have the highest level of performance, except for the matched filter case.

We must examine the sensitivity performance for the four cases under identical conditions. For the input noise spectrum, we will assume a standard preamplifier front-end similar to Fig. 2.8, and a noise spectrum

$$S_N(f) = 2k\Theta \left(\frac{1}{R_L} + \frac{\Gamma}{g_m R_L^2} + \frac{\Gamma(2\pi C_T)^2}{g_m} f^2 \right) \quad (2.71)$$

This is a *double-sided* version of (2.25), and thus has half of the magnitude. In addition, we neglect detector dark current and gate leakage current, which are presumed to be small compared to the other noise terms.

For the input pulse shape, we will assume an ideal square pulse $x(t)$ of unity height and width T , as shown in Fig. 2.18. This corresponds to a non-return-to-zero (NRZ) input signal for a “one” bit. This pulse is a reasonable assumption for optical networks where the fiber dispersion is negligible. This pulse has Fourier transform

$$X(f) = \frac{\sin \pi f T}{\pi f} \quad (2.72)$$

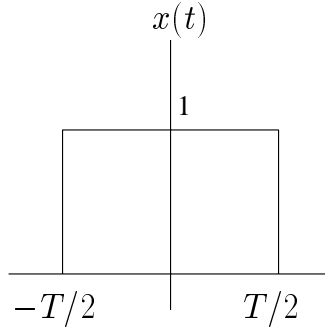


Figure 2.18. Ideal input pulse for receiver sensitivity analysis.

2.4.1 Matched Filter Performance

Given the input pulse spectrum $X(f)$ and the noise power spectrum $S_N(f)$, we may substitute into (2.47) to get the matched filter response

$$Z_{\text{matched}}(f) = \frac{\sin \pi f T}{\pi f} \frac{1}{2k\Theta \left(\frac{1}{R_L} + \frac{\Gamma}{g_m R_L^2} + \frac{\Gamma(2\pi C_T)^2}{g_m} f^2 \right)} e^{-j2\pi f t_d} \quad (2.73)$$

We will make no attempt to normalize the DC value of $Z_{\text{matched}}(f)$, as it will cancel when computing the SNR.

In order to proceed with the determination of the signal portion of the SNR, we must first find the time t_d at which the output signal is at its maximum. This is easily done by observing the symmetry of the signals in the time domain (Fig. 2.19). In the time domain, $z_{\text{matched}}(t)$ consists of the convolution of the three parts of (2.73), which consists of a square pulse, a cusp, and a delta function [49]. Because the square pulse and cusp are both positive, even, and maximum at $t=0$, their convolution will be maximum at $t=0$ and thus $z_{\text{matched}}(t)$ will be maximum at $t=0$. We may therefore set $t_d=0$ and

$$[y_{s,\text{matched}}(t)]_{\text{max}} = y_{s,\text{matched}}(0) \quad (2.74)$$

$$= \int_{-\infty}^{\infty} X(f) Z_{\text{matched}}(f) df \quad (2.75)$$

The variance of the output noise of the matched filter is obtained by simply integrating over the output noise spectrum

$$\overline{n_{\text{matched}}^2} = \int_{-\infty}^{\infty} S_N(f) |Z_{\text{matched}}(f)|^2 df \quad (2.76)$$

$$z_{\text{matched}}(t) = \mathcal{F}^{-1} \left[\frac{\sin \pi f T}{\pi f} \right] * \mathcal{F}^{-1} \left[\frac{1}{2k\Theta \left(\frac{1}{R_L} + \frac{\Gamma}{g_m R_L^2} + \frac{\Gamma(2\pi C_T)^2}{g_m} f^2 \right)} \right] * \mathcal{F}^{-1} [e^{-j2\pi f t_d}]$$

Figure 2.19. Time domain representation of $z_{\text{matched}}(t)$.

Substituting (2.71), (2.72), (2.75) and (2.76) into (2.70) gives the sensitivity of the matched filter

$$\left(\eta \bar{P} \right)_{\text{matched}} = \left(\frac{h\nu}{q} \right) Q \left(\int_{-\infty}^{\infty} \left(\frac{\sin \pi f T}{\pi f} \right)^2 \frac{1}{2k\Theta \left(\frac{1}{R_L} + \frac{\Gamma}{g_m R_L^2} + \frac{\Gamma(2\pi C_T)^2}{g_m} f^2 \right)} df \right)^{-1/2} \quad (2.77)$$

2.4.2 2nd-Order Filter Performance

The use of a simple second-order noise filter response simulates the case of a high-impedance or transimpedance front-end where the preamplifier cutoff is used for the noise filter. Thus, we will assume the response in (2.4) and (2.5) with

$$H(s) = \frac{\omega_n^2(s + p_e)}{s^2 + 2\zeta\omega_n s + \omega_n^2} \quad (2.78)$$

where ζ is the damping factor, ω_n is the natural frequency, and p_e is a pole-canceling zero. For the high-impedance case, $p_e=1/(R_L C_T)$, and for the transimpedance case, $p_e=(A_v + 1)/(R_F C_T)$. Thus, the pole due to the resistance and capacitance of the receiver input circuit is assumed to be exactly canceled by the noise shaping filter, and thus p_e provides the ideal best-case equalization in the noise filter. Letting $s=j2\pi f$, the receiver transfer function for the second-order case is

$$Z_{\text{2nd-order}}(f) = \frac{f_n^2}{-f^2 + f_n^2 + j2\zeta f_n f} \quad (2.79)$$

where $f_n=\omega_n/2\pi$. We will now make several assumptions about this response:

- We will assume that the natural frequency of the response is a fixed factor of the bitrate. We would ideally like to make f_n as low as possible, limited by the smearing of the bit into the adjacent bit time (which results in intersymbol interference). We will set f_n to be 0.7 times the bit rate, or $f_n=0.7/T$, as a tradeoff between SNR and intersymbol interference. This corresponds to 40% excess bandwidth [29, 11].
- We will assume that $\zeta=1$, i.e. that the filter is critically damped.

Applying these assumptions to (2.79) gives

$$Z_{2\text{nd-order}}(f) = \left(\frac{0.7/T}{jf + 0.7/T} \right)^2 \quad (2.80)$$

We must now find the optimum sampling time t_d for the second-order case. This is easily done directly in the time domain. The step response of $Z_{2\text{nd-order}}(f)$ is

$$z_{2\text{nd-order,step}}(t) = \mathcal{L}^{-1} \left[\frac{1}{s} \frac{\omega_n^2}{(s + \omega_n)^2} \right] \quad (2.81)$$

$$= 1 - e^{-((0.7)2\pi/T)t} (1 + ((0.7)2\pi/T)t) \quad (2.82)$$

where \mathcal{L}^{-1} is the inverse Laplace transform. For a finite pulse of length T , $z_{2\text{nd-order,step}}(t)$ will be maximum at $t=T$ (Fig. 2.20). Thus,

$$[y_{2\text{nd-order}}(t)]_{\text{max}} = z_{2\text{nd-order,step}}(T) \quad (2.83)$$

$$= 1 - e^{-(0.7)2\pi} (1 + (0.7)2\pi) \quad (2.84)$$

$$= 0.934 \quad (2.85)$$

The variance of the output noise for the second-order case is simply the output noise power spectrum integrated over frequency

$$\overline{n_{2\text{nd-order}}^2} = \int_{-\infty}^{\infty} S_N(f) |Z_{2\text{nd-order}}(f)|^2 df \quad (2.86)$$

Substituting into (2.70) gives the sensitivity of the second-order filter

$$\begin{aligned} (\eta\overline{P})_{2\text{nd-order}} &= \\ & \left(\frac{h\nu}{q} \right) Q \frac{1}{0.934} \sqrt{\int_{-\infty}^{\infty} \left(2k\Theta \left(\frac{1}{R_L} + \frac{\Gamma}{g_m R_L^2} + \frac{\Gamma(2\pi C_T)^2}{g_m} f^2 \right) \right) \left| \left(\frac{0.7/T}{jf + 0.7/T} \right)^2 \right|^2 df} \end{aligned} \quad (2.87)$$

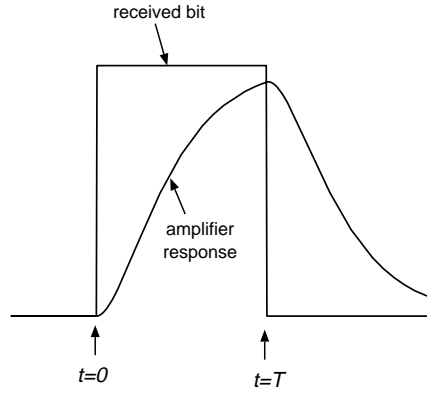


Figure 2.20. For a second-order response, the maximum height of the filtered bit will be at the end of the bit period T .

2.4.3 Raised-Cosine-Output Filter Performance

A common choice for the receiver transfer function for theoretical analysis was originated by Personick [24], where a sinc-like *output* pulse shape is assumed. The advantage of such an output pulse lies in the values of the sinc function at the discrete periodic sampling points $0, T, 2T, 3T, \dots$

$$\frac{\sin \pi B n T}{\pi B n T} = \begin{cases} 1 & n = 0 \\ 0 & n = \pm 1, \pm 2, \pm 3, \dots \end{cases} \quad (2.88)$$

where B is the bitrate, T is the bit period, and n is an integer. By using such a function for the output pulse, we are guaranteed that there will be no intersymbol interference, assuming that the sampling occurs exactly at multiples of T . Unfortunately, a sinc function has a Fourier transform which is a “brickwall” filter, which is unrealizable. However, a convenient class of sinc-like functions is known as the *raised cosine* family. The raised cosine $C(f)$ consists of a flat top with sloping sides which are cosine shaped. This is illustrated in Fig. 2.21 and is mathematically expressed as follows

$$C(f) = \begin{cases} 1 & 0 < |f| < \frac{1-\beta}{2T} \\ \frac{1}{2} \left[1 - \sin \left(\frac{\pi |f| T}{\beta} - \frac{\pi}{2\beta} \right) \right] & \frac{1-\beta}{2T} < |f| < \frac{1+\beta}{2T} \\ 0 & \text{otherwise} \end{cases} \quad (2.89)$$

where $0 < \beta < 1$ is a parameter which indicates the width of the flat top region, and T is the bit period. In the limit of $\beta \rightarrow 1$, $C(f)$ is a single cycle of a cosine wave, and in the

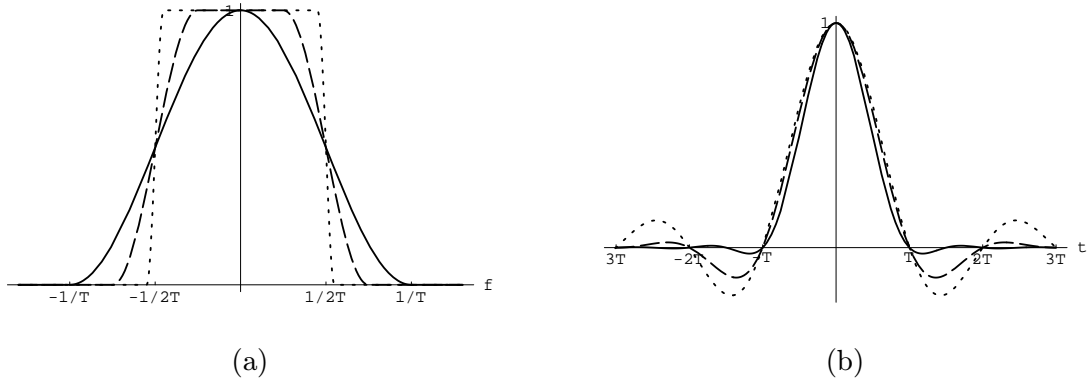


Figure 2.21. (a) Frequency spectrum of the raised-cosine response. (b) Inverse Fourier transform of the raised-cosine spectrum. The above plots are for $\beta=1$ (solid line), $\beta=0.5$ (dashed line) and $\beta=0.1$ (dotted line).

limit of $\beta \rightarrow 0$, $C(f)$ approaches a brickwall response. If we inverse-Fourier-transform the raised-cosine spectrum, we get the sinc-like response $c(t)$

$$c(t) = \frac{\sin\left(\frac{\pi t}{T}\right) \cos\left(\frac{\pi \beta t}{T}\right)}{\frac{\pi t}{T} \left[1 - \left(\frac{2\beta t}{T}\right)^2\right]} \quad (2.90)$$

Given the desired raised-cosine spectrum $C(f)$ of the output signal, and the known square-pulse input signal $X(f)$, the resulting transfer function which implements this system is

$$Z_{\text{raised-cos}}(f) = \frac{C(f)}{X(f)} \quad (2.91)$$

It is interesting to note that the use of the raised-cosine output response by Personick seems purely motivated by the ease of computation, given that such a filter is nonrealizable (because it is noncausal) and does not provide optimum noise filtering (as does the matched filter). However, several other authors [26, 27, 28] have continued along the lines of Personick, and thus the raised cosine is included here for completeness.

The integrated output noise for the raised-cosine-output response is

$$\overline{n_{\text{raised-cos}}^2} = \int_{-\infty}^{\infty} S_N(f) \left| \frac{C(f)}{X(f)} \right|^2 df \quad (2.92)$$

Substituting into (2.70) gives, assuming $\beta=1$

$$\begin{aligned}
\left(\eta\bar{P}\right)_{\text{raised-cos}} = & \\
\left(\frac{h\nu}{q}\right) Q & \sqrt{\frac{\int_{-1/T}^{1/T} \left(2k\Theta\left(\frac{1}{R_L} + \frac{\Gamma}{g_m R_L^2} + \frac{\Gamma(2\pi C_T)^2}{g_m} f^2\right)\right) \left|\frac{\frac{1}{2}(1 + \cos \pi f T)(\pi f)}{\sin \pi f T}\right|^2 df}{\int_{-1/T}^{1/T} \frac{1}{2}(1 + \cos \pi f T) df}}
\end{aligned} \tag{2.93}$$

2.4.4 Integrate-and-Dump Performance

For an integrate-and-dump receiver, we will assume the transfer function of (2.7) with a second-order preamplifier response

$$Z_{\text{i-d}}(s) = \frac{1}{C_T s} \frac{\omega_n^2}{(s^2 + 2\zeta\omega_n s + \omega_n^2)} \tag{2.94}$$

Similar to the prior case of a second-order filter, we will assume that $\zeta=1$, i.e. that the amplifier is critically damped. We also know automatically that because the filter is an integrator, its time response will be at its maximum at the end of the bit time, and therefore $t_d=T$. Thus, we can obtain the output signal for the integrate-and-dump circuit by finding the step response of (2.94) and evaluating at $t=T$. The step response will be [30]

$$z_{\text{i-d,step}}(t) = \mathcal{L}^{-1} \left[\frac{1}{C_T s^2} \frac{\omega_n^2}{(s + \omega_n)^2} \right] \tag{2.95}$$

$$= \frac{1}{C_T} \left(t - \frac{2}{\omega_n} + \frac{2}{\omega_n} e^{-\omega_n t} \left(1 + \frac{\omega_n t}{2} \right) \right) \tag{2.96}$$

We will also assume that the filter cutoff ω_n is a fixed ratio of the bitrate. However, because the integrate-and-dump receiver inherently rejects all intersymbol interference (due to the integrator resetting between bits), the optimum choice for ω_n will be different than the second-order case. In this case, the choice will be a tradeoff between the increased noise rejection of a low filter bandwidth, and the increased signal output of a high filter bandwidth. We define the parameter γ to be the ratio between the filter bandwidth and the bitrate, thus

$$\gamma = \frac{\omega_n}{2\pi/T} \tag{2.97}$$

Fig. 2.22 illustrates the effect of different values of γ . For infinite bandwidth, the filter is an ideal integrator, and the response to a pulse is an ideal ramp. For high γ (and thus high

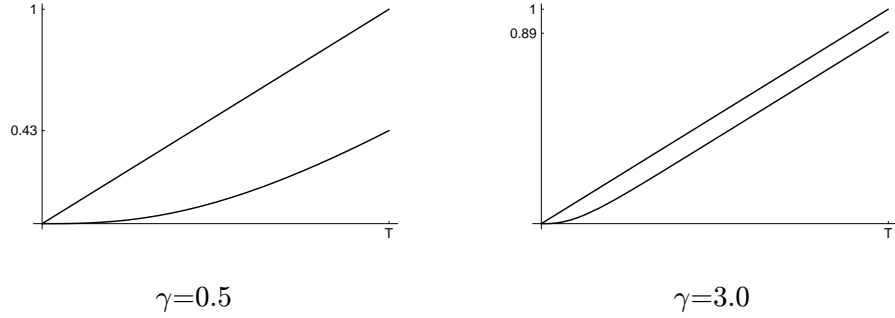


Figure 2.22. Pulse response for integrate-and-dump filter for different amplifier bandwidths, compared with the ideal ramp response. The higher the bandwidth, the closer the output is to an ideal ramp.

filter bandwidth), the output at time T approaches that of the ideal ramp. For a low γ , the output only reaches a fraction of the ideal ramp value.

Substituting (2.97) into (2.96) gives an expression for the filter output at $t=T$

$$[y_{i-d}(t)]_{\max} = z_{i-d, \text{step}}(T) \quad (2.98)$$

$$= \frac{T}{C_T} \left(1 - \frac{1}{\gamma\pi} + \frac{e^{-\gamma 2\pi}}{\gamma\pi} (1 + \gamma\pi) \right) \quad (2.99)$$

The output noise variance for the integrate-and-dump case is

$$\overline{n_{i-d}^2} = \int_{-\infty}^{\infty} S_N(f) |Z_{i-d}(f)|^2 df \quad (2.100)$$

Substituting into (2.70) gives the sensitivity of the integrate-and-dump filter in terms of γ

$$\begin{aligned} (\eta\overline{P})_{i-d} &= \left(\frac{h\nu}{q} \right) Q \frac{C_T}{T \left(1 - \frac{1}{\gamma\pi} + \frac{e^{-\gamma 2\pi}}{\gamma\pi} (1 + \gamma\pi) \right)} \\ &\quad \times \sqrt{\int_{-\infty}^{\infty} \frac{2k\Theta\Gamma(2\pi C_T)^2}{g_m} f^2 \left| \frac{1}{C_T j 2\pi f} \left(\frac{\gamma/T}{jf + \gamma/T} \right) \right|^2 df} \quad (2.101) \end{aligned}$$

where we have assumed that $R_L \rightarrow \infty$ for the integrate-and-dump configuration. This expression may be optimized graphically in terms of γ as shown in Fig. 2.23 which plots (2.101) at a fixed bitrate in terms of γ . The optimum point for best sensitivity is at $\gamma=0.9$, meaning that the best tradeoff between maximum signal output and noise filtering occurs when the amplifier bandwidth is 0.9 times the bitrate.

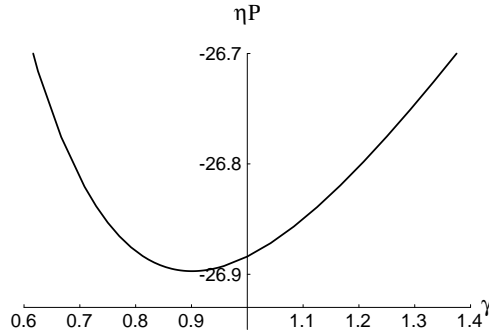


Figure 2.23. Optimization of the amplifier bandwidth with respect to the bitrate for the integrate-and-dump filter. The best sensitivity is achieved when the bandwidth is 0.9 times the bitrate.

Thus, the optimum expression for the sensitivity (2.101) may be simplified and rewritten as

$$(\eta\bar{P})_{i-d} = \left(\frac{h\nu}{q}\right) Q \frac{C_T}{0.651T} \sqrt{\int_{-\infty}^{\infty} \frac{2k\Theta\Gamma(2\pi C_T)^2}{g_m} f^2 \left| \frac{1}{C_T j 2\pi f} \left(\frac{0.9/T}{jf + 0.9/T} \right) \right|^2 df} \quad (2.102)$$

2.4.5 Comparison of Sensitivities

Fig. 2.24 shows plots of the filter transfer functions for the four cases of the matched filter, the second-order filter, the raised cosine output, and the integrate-and-dump filter. The parameters used are as shown in the figure caption, which provide a noise corner of approximately 1 MHz. In addition, the bitrate is presumed to be 1 Gb/s.

There are few similarities among the four responses. In the case of the matched filter, there is a second-order pole at the noise corner, and zeros at multiples of the bitrate. Even an approximation of this noncausal filter would be difficult to implement because the integrating behavior near the bitrate. This would necessitate some kind of resetting scheme (“filter-and-dump”) in order to avoid saturating the preamplifier.

The second-order and raised-cosine-output cases are simply flat filters with rolloffs near the bitrate. The rolloff of the raised-cosine-output case would be particularly difficult to implement due its sharpness.

The integrate-and-dump filter has a first-order rolloff up to the bitrate, and a third-order rolloff at frequencies beyond the bitrate. This response is very simple to implement,

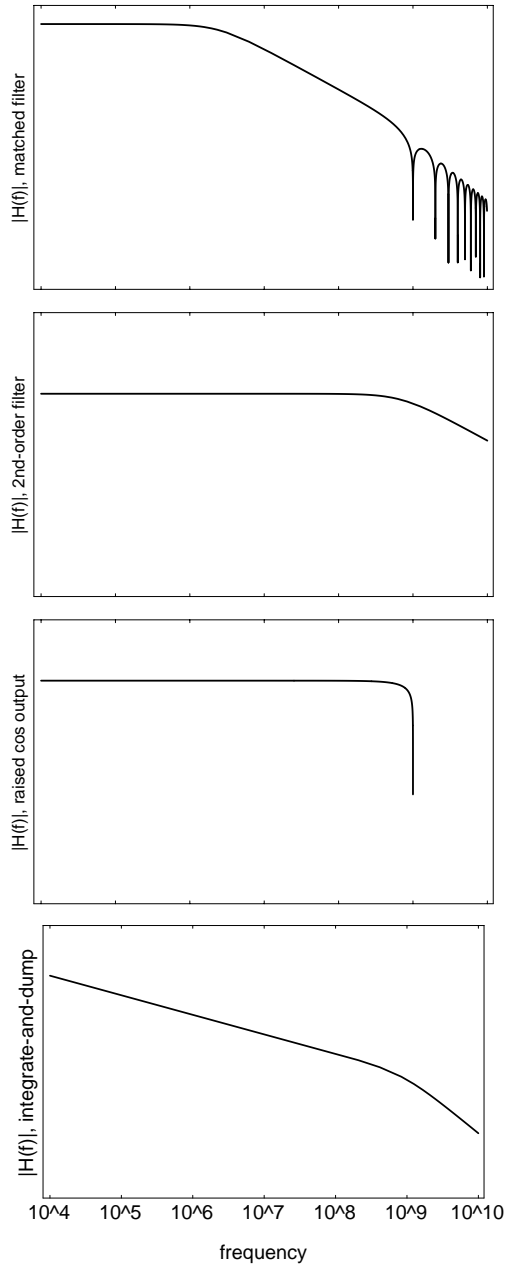


Figure 2.24. Comparison of transfer functions for the matched filter, the second-order filter, the raised cosine output, and the integrate and dump filter. $C_T=6$ pF, $R_L=1$ M Ω , $g_m=0.006$ U, bitrate=1 Gb/s.

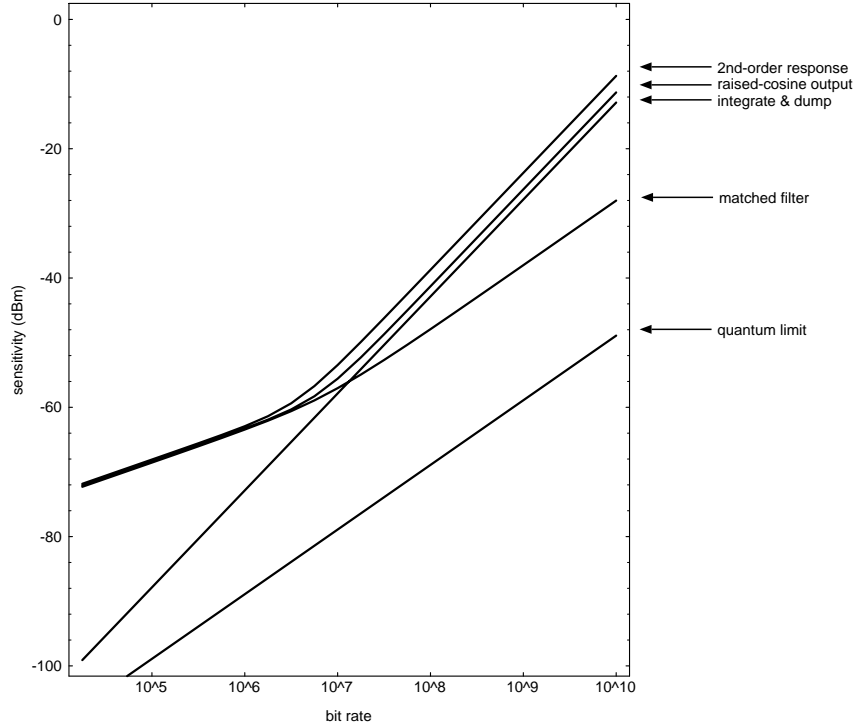


Figure 2.25. Comparison of sensitivities for 10^{-9} BER for the 2nd order, raised cosine, integrate and dump, and matched filter receivers. $C_T=6$ pF, $\Theta=300$ K, $R_L=1$ M Ω , $\Gamma=0.7$, $g_m=0.006$ U, $\lambda=1550$ nm, $Q=6$.

as will be shown in the succeeding chapters of this thesis.

Given expressions for the sensitivity for the four different filtering schemes, we may now plot (2.77), (2.87), (2.93), and (2.101) on common axes in order to compare the effectiveness of each scheme. However, in order to compute numerical results, we must select values for several key parameters. We will arbitrarily set the transconductance of the amplifier input transistor to be $g_m=6$ mU, the total input node capacitance to be $C_T=6$ pF, and the load or feedback resistor to be $R_L=10$ M Ω . These numbers provide a reasonable estimate for a typical high-performance receiver. They are also the same numbers used in [16]. We will also assume room temperature conditions, and a bit error rate of 10^{-9} , i.e. $Q=6$. In addition, we will assume that the optical wavelength of interest is 1550 nm.

Fig. 2.25 shows the sensitivities for the four types of filtering. In addition, a fifth line shows the fundamental quantum limit for optical receivers. Several observations are

TABLE 2.1. Receiver sensitivities at 1 Gb/s

Second-order filter	-23.7 dBm
Raised-cosine output	-26.3
Integrate-and-dump filter	-27.9
Matched filter	-38.0
Quantum limit	-58.9

TABLE 2.2. Receiver sensitivities at 100 kb/s

Second-order filter	-68.1 dBm
Raised-cosine output	-68.3
Integrate-and-dump filter	-87.9
Matched filter	-68.5
Quantum limit	-98.9

apparent from this plot:

- At bitrates above the noise corner, the matched filter achieves significantly better sensitivity performance than the other filtering schemes.
- Table 2.1 shows the sensitivities for the five cases at 1 Gb/s. Of the two realizable cases, the integrate-and-dump filter shows a 4.2 dB improvement over the simple second-order filter.
- At bitrates below the noise corner, the integrate-and-dump scheme is clearly the best performer. However this result is somewhat deceiving because all the other filtering schemes assume a bias resistor of $1\text{ M}\Omega$, whereas the integrate-and-dump filter assumes the absence of any bias resistor at all.
- At bitrates below the noise corner, the advantage of the matched filter becomes small. Table 2.2 shows the receiver sensitivities at 100 kb/s. The matched filter only shows an improvement of 0.4 dB over the second-order filter.
- At a certain low bitrate, the integrate-and-dump case will “bottom out” as the levels of dark current and leakage current become non-negligible. This is not depicted in Fig. 2.25, and the noise level of the integrate-and-dump filter will never be lower than the quantum limit.

To summarize, the matched filter is the best receiver but is unrealizable. A simple second-order lowpass filter with a bias resistor at the input (high impedance or transimpedance receiver designs) and ideal zero cancellation provides a baseline performance level. A second-order amplifier with capacitive feedback (integrate-and-dump) provides approximately 4.2 dB improvement over the simple second-order case. The raised-cosine-output filter is useful for comparison purposes and provides performance slightly better than the second order case, but is unrealizable. None of the designs come remotely close to achieving the quantum noise limit.

CHAPTER 3

INTEGRATE-AND-DUMP RECEIVER TOPOLOGY

This chapter will discuss the implementation of an integrate-and-dump optical receiver at the transistor level. There are many practical issues which affect the circuit implementation of an optical preamplifier:

- the desired bitrate(s)
- the choice of a single-ended or differential design
- the biasing conditions for the photodetector
- the desired dynamic range
- the application to circuit or packet switching (continuous or burst mode)

The higher the bitrate, the faster the integrated circuit technology necessary to implement that bitrate. In the case of high-speed terrestrial point-to-point links (such as those used in modern telephone networks), the drive is to achieve the highest bitrate in order to squeeze the most number of telephone conversations onto a single fiber. In this case, sophisticated transmitters and receivers implemented in esoteric heterojunction technologies are the area of current interest [6, 7, 8] because the high cost of such receivers may be split among the many end-users.

On the other hand, in the cases of optical networking and fiber-to-the-home, the cost of the transmitter/receiver becomes paramount because the cost is borne entirely by the

individual user. CMOS technology becomes particularly attractive for these applications due to its low cost, simplicity, and high level of integration. In addition, today's submicron CMOS technologies can achieve subnanosecond switching speeds which can provide bitrates exceeding 1 GHz. This rate provides enough bandwidth for many useful applications such as HIPPI, Fiber Channel, SONET, or HDTV.

Given CMOS as the chosen technology, several advantages become apparent. The availability of high-quality switches in CMOS makes the technology attractive for applications where a multiplicity of photodetectors are connected to a single preamplifier, as in the broadcast-and-select optical network topology described in Chapter 1. In addition, switches allow the easy implementation of time-variant and integrate-and-dump circuits for improved processing of binary signals. CMOS also has other advantages such as the availability of complementary devices, the ease of prototyping, and the overall maturity of the technology. We will assume a 1.2 μm double-metal double-poly p-well technology available from Orbit Semiconductor [31].

The design of an integrate-and-dump receiver in CMOS will rely heavily on existing knowledge learned in design of switched-capacitor circuits and systems [20]. Interestingly enough, up until now, these techniques largely have yet to be applied to the field of optical receivers.

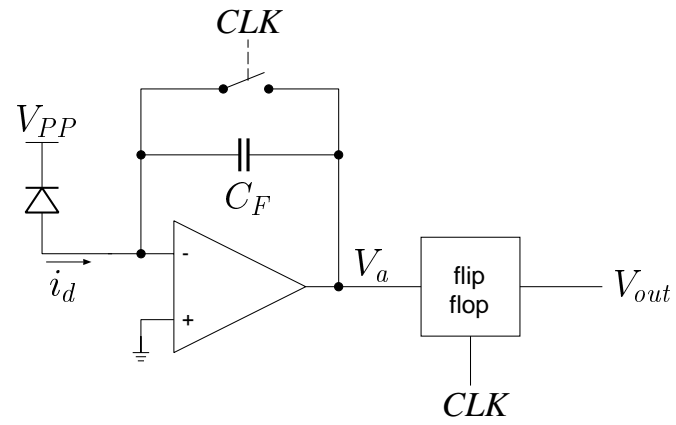
3.1 INTEGRATE-AND-DUMP TOPOLOGY

The “traditional” integrate-and-dump circuit is shown in Fig. 3.1. In this simple implementation, an op amp is used with capacitive feedback to create an integrator with a response

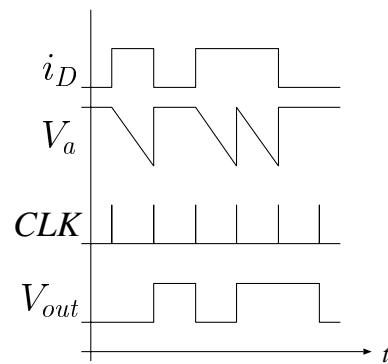
$$V_{out} = -\frac{1}{C_F} \int_0^T i_d dt \quad (3.1)$$

Thus, as shown in Fig. 3.1b, the response to a step of current i_d is a negative ramp at the amplifier output V_a . At the end of each bit, a short clock pulse CLK is used to reset the integrator for reception of the next bit. A flip flop at the output of the amplifier is also clocked by CLK and is used to sample V_a at the end of each bit to determine a “one” or a “zero.”

The major drawback to this approach is in the implementation of the short clock pulse. This clock signal must have an extremely fast rise and fall time in order to quickly reset the integrator and then receive the next bit. Ideally, such a clock consists of impulse-like spikes



(a)



(b)

Figure 3.1. (a) A simple integrate-and-dump circuit. (b) Circuit timing.

containing very high frequency components. Such a clock is extremely difficult to achieve for several reasons. First, the clock spikes will cause substrate noise coupling and transients throughout the chip which will degrade sensitivity. Secondly, a clock with finite width will reduce the time of integration due to time spent in the reset (dump) phase [32]. Finally, the necessity of generating such fast clocks will reduce the maximum bitrate achievable for a monolithic receiver. For example, in a given IC technology, the maximum receiver bitrate will be some fraction of the technology cutoff frequency f_T . However, the generation of a fast spike clock on-chip which contains high-frequency components (at perhaps ten times the bitrate) means that a given technology may only support one tenth the bitrate over receiver architectures which do not use integrate and dump. Stated another way, the technology for a given bitrate would need be ten times faster due to the necessity of generating a high-speed spike clock.

Fig. 3.2 shows a parallel circuit topology which avoids many of the problems with the simple implementation just discussed. In this architecture, three preamplifiers are used in parallel in order to avoid using a fast spike clock. The amplifiers are controlled by using a three-phase clock as shown in Fig. 3.3. By splitting the work of the integrate-and-dump circuit into three parallel circuits, the requirements on each individual circuit may be relaxed.

In this topology, the integrate-and-dump operation is divided into *integrate*, *readout*, and *dump* phases. At any given instant, each of the preamplifiers is performing one of these three operations as determined by the position of the switches. As depicted in Fig. 3.2, clock ϕ_1 is active and thus preamplifier 1 is connected to the photodetector and is integrating; preamplifier 2 is shorted output to input, and thus is dumping; and preamplifier 3 is connected to the decision circuit, and thus is reading out. During the subsequent ϕ_2 and ϕ_3 clock cycles, each preamplifier performs the other operations in sequence, as indicated by the letters *I*, *R*, and *D* in the timing diagram. The clear advantage to this method is the absence of any spike clock. Switching within the circuit only occurs at the bit rate, and thus does not require a faster technology for implementation. An obvious disadvantage is the necessity for three times as much circuitry. In addition, noise performance will be reduced due to the added thermal noise and parasitic capacitance due to the switches at the input [15]. However, given the sensitivity advantage of the integrate-and-dump receiver scheme, this might be a worthwhile tradeoff.

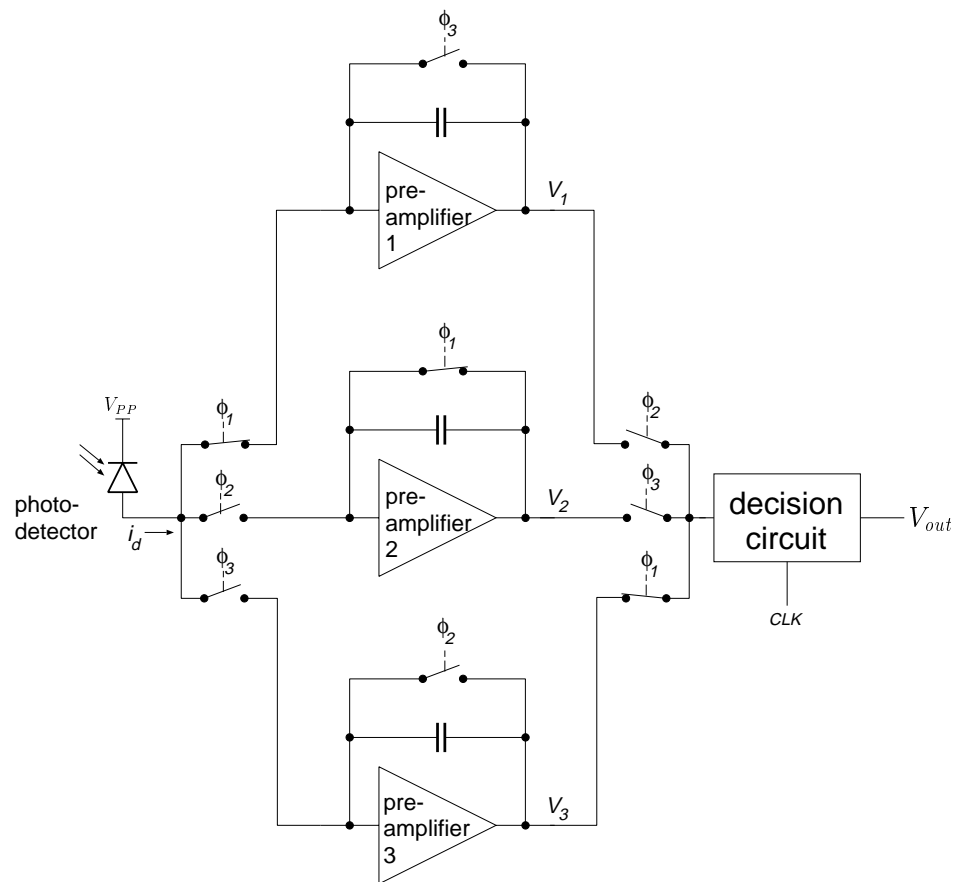


Figure 3.2. Parallel integrate-and-dump receiver.

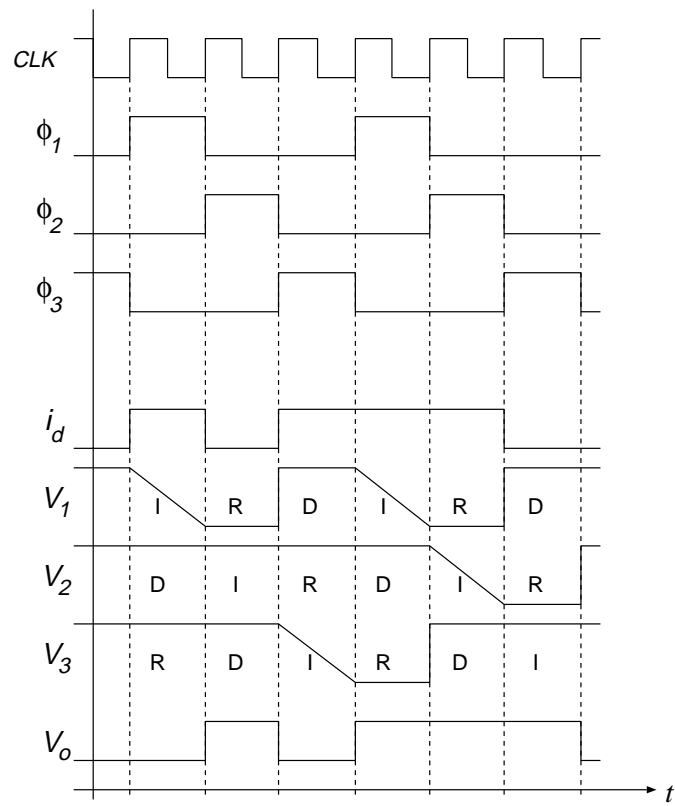


Figure 3.3. Circuit timing diagram for parallel integrate-and-dump receiver. I = integrate cycle, R = readout cycle, D = dump cycle.

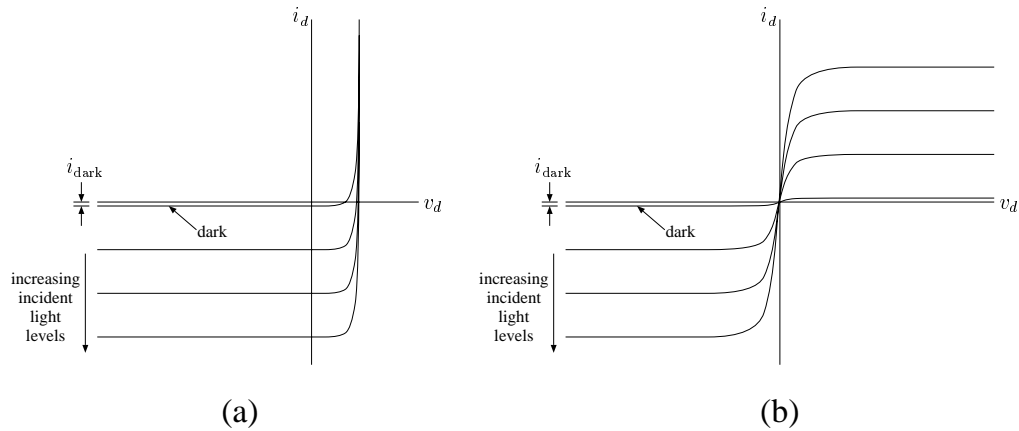
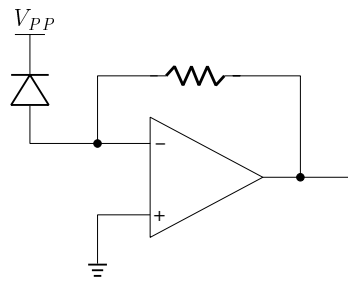


Figure 3.4. Typical IV characteristic for (a) PIN diode photodetector and (b) MSM photodetector.

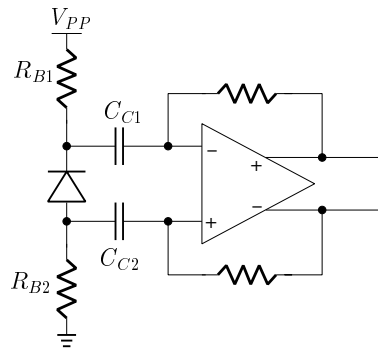
3.2 DETECTOR BIASING TOPOLOGY

The photodetector in an optical receiver must be biased at the correct voltage level in order for it to operate as a light-sensitive dependent current source. Typical I-V characteristics for PIN and MSM photodetectors are shown in Fig. 3.4. The PIN diode behaves like an ordinary diode while under forward bias, but as dependent current source while under reverse bias. The amount of reverse current varies from a low leakage level (i_{dark}) to an amount proportional to the incident light power ($i_d = \mathcal{R}_{\text{PIN}}P$). It is clear from the drawing that even a few volts of reverse bias on the PIN diode is enough to place it into the photosensitive regime. The MSM detector has no forward bias region, and depends only on the magnitude of the biasing and not the sign. Similar to the PIN, the MSM has a dark current i_{dark} and requires only a few volts of bias in order to work as a light-dependent current source.

Fig. 3.5 shows two common methods of biasing the photodetector for use in an optical receiver [33, 12, 34, 13, 15]. In the first method, an operational amplifier is used to set up a virtual ground at the anode of the photodiode. The cathode of the detector is connected directly to the supply voltage V_{PP} , thus placing a bias of V_{PP} across the detector. In an actual design, a single-ended amplifier is usually used instead of an op amp with differential inputs, and the amplifier input bias voltage is set up with appropriate bias networks.



(a)



(b)

Figure 3.5. Traditional methods of biasing the photodetector. (a) In a single-ended design, the detector is biased between the supply V_{PP} and the virtual ground at the amplifier negative input. (b) In a differential design, the detector is biased with large resistors and is ac-coupled to the amplifier through large capacitors.

The main drawback to this biasing method is that it uses single-ended circuitry, and thus lacks the symmetrical noise cancellation properties of balanced circuits. A fully-differential biasing topology is shown in Fig. 3.5b. This method uses bias resistors R_{B1} and R_{B2} to place a bias voltage of V_{PP} across the photodetector. R_{B1} and R_{B2} are made large enough that their thermal noise is negligible, but small enough that the detector dark current does not cause an appreciable IR drop. The detector photocurrent is then passed to the amplifier through coupling capacitors C_{C1} and C_{C2} .

The drawback to ac coupling the photodetector to the amplifier is that the circuit will now have a low frequency cutoff determined by the size of C_{C1} and C_{C2} . This cutoff is particularly cumbersome because it is dependent on the pattern of the received bits. For example, if a long pattern of “ones” or “zeros” is received, the resulting dc-like signal will cause C_{C1} and C_{C2} to discharge and subsequently cause eye closure due to dc wander [58, 59, 60]. The low-frequency cutoff may be made arbitrarily small by using large values of C_{C1} and C_{C2} , however this requirement is at odds with most IC technologies where large capacitors take up large amounts of die area. In addition, in the case of a burst-mode receiver for packet-switched applications, the charging/discharging of C_{C1} and C_{C2} can cause significant delays in the overall acquisition time of the receiver. Thus, designs with ac coupling will only operate at a limited range of bitrates [15].

Fig. 3.6 shows a new method for dc biasing the photodetector in the receiver front-end. In this case, two amplifiers are used to set up the bias voltage for the detector. The top amplifier in Fig. 3.6 has its positive input tied to $V_{\text{bias}+}$, while the bottom amplifier has its positive input tied to $V_{\text{bias}-}$. Assuming high-gain amplifiers, the resulting bias voltage across the photodetector will be $V_{\text{bias}+} - V_{\text{bias}-}$. The outputs of the two amplifiers are combined to form a single balanced output signal. This may be accomplished by using a differential difference amplifier [35] or with switched-capacitor techniques (to be described later). In addition, although conceptually drawn as having differential inputs, the amplifiers in Fig. 3.6 may be implemented as single-ended amplifiers whose outputs are combined into a balanced signal. This idea will be shown in the next section.

The advantage of this biasing method is that it allows a defined bias voltage determined by $V_{\text{bias}+} - V_{\text{bias}-}$ while maintaining symmetry in the circuit. In addition, there are no coupling capacitors to limit low-bitrate operation. A previous method [36] used an intentional offset voltage to create a dc bias at the input of a differential amplifier, thereby eliminating the need for coupling capacitors, but also introducing asymmetry into the circuit. The

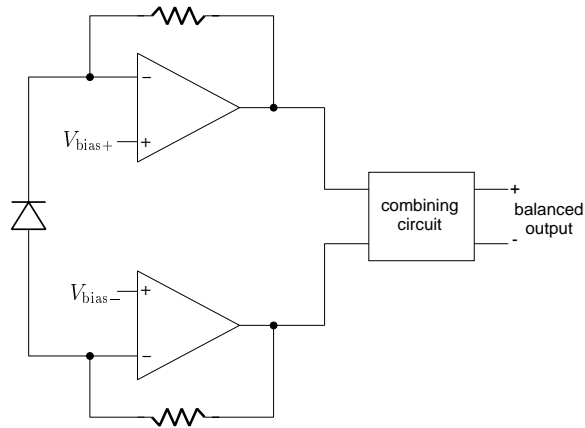


Figure 3.6. New method of biasing the photodetector. Two amplifiers with separate dc bias points are used to set up a voltage across the detector

method presented here does not intentionally imbalance the circuit in any way.

3.3 PREAMPLIFIER DESIGN

Most ordinary optical preamplifier designs are optimized with respect to bitrate, frequency response, transimpedance, and input pole location. However, in the case of an integrate-and-dump design, many of these parameters do not apply. For example, by its very nature, the integrate-and-dump circuit does not have a load/feedback resistor to determine the transimpedance. In addition, because we are *intentionally* designing an integrating amplifier, the input pole location is automatically placed at the origin (as we shall see).

Fig. 3.7 shows a simple preamplifier design using a pair of common-source amplifiers. The amplifiers are identical except that the top one is constructed with a p-channel transistor, and the bottom one uses an n-channel transistor. Note that the detector bias method of Fig. 3.6 is implemented by using these single-ended common-source stages instead of differential amplifiers. This is advantageous for several reasons. First, it eases the implementation immensely due to the simple design of common-source stages over differential amplifiers. In addition, there is no need to generate the external bias voltages V_{bias+} and V_{bias-} because appropriate voltages will already be present at the inputs of the two opposite-type common-source stages. Finally, it is straightforward to combine the outputs of the two amplifiers

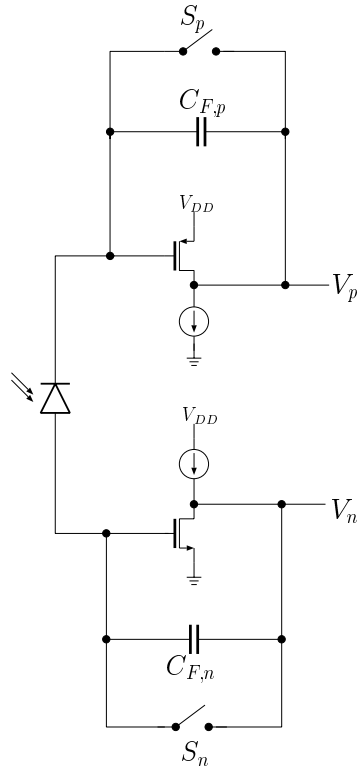


Figure 3.7. Dual integrating preamplifier design.

into a single differential signal by using a simple switched-capacitor circuit. The disadvantage of using common-source inputs is that the circuit is not truly differential throughout, but rather only after the preamplifier/combining stage. However, the symmetry of the configuration still allows first-order noise cancellation.

Biasing for the preamplifiers is done dynamically by closing their respective reset switches. This places both common-source stages into the MOS saturation region because $V_{gs}=V_{ds}$ for both transistors, and therefore we are guaranteed that $|V_{ds}| > |V_{gs} - V_T|$ for both cases. When the switches are reopened, the gate bias voltage is presumed to remain present on the capacitances present (parasitic or otherwise) at the MOSFET gates. This idea is similar to that of DRAMs or current copier circuits [37] where capacitor leakage is presumed to have a much longer time constant than the system clock. This type of biasing is especially convenient because the integrator is reset after each bit received. Thus, for each bit received and each reset cycle, the preamplifiers are rebiasing as well.

A small signal circuit model for the preamplifier is shown in Fig. 3.8. In this drawing, the photodetector has been modeled by two identical current sources i_d in series, with the common terminal between them connected to ground. In addition, the detector capacitance C_d has been split into two identical series capacitors with magnitude $2C_d$. The purpose of this detector model is to allow separate and independent calculations of the output voltages V_p and V_n . Thus, each preamplifier sees its own “photodetector” with photocurrent i_d and detector capacitance $2C_d$.

The parameters available to the designer for the preamplifier include the gate length L , gate width W , and drain current I_D of the common-source input transistors. We will presume that L is always set to the minimum feature size of the technology in order to maximize transconductance and amplifier gain. W is usually chosen to match the amplifier input capacitance with the detector as in (2.28). This leaves I_D , and by extension, g_m , as the only parameter which is adjustable to set the gain and the bandwidth of the preamplifiers.

The resulting transfer functions for the small signal circuit are

$$\frac{V_n}{i_d} = \frac{g_{m,n}r_{o,n} - sr_{o,n}C_{B,n}}{s(sr_{o,n}C_{T,n}C_{B,n} + ((g_{m,n}r_{o,n} + 1)C_{B,n} + C_{T,n}))} \quad (3.2)$$

$$\frac{V_p}{i_d} = \frac{g_{m,p}r_{o,p} - sr_{o,p}C_{B,p}}{s(sr_{o,p}C_{T,p}C_{B,p} + ((g_{m,p}r_{o,p} + 1)C_{B,p} + C_{T,p}))} \quad (3.3)$$

where $C_{T,p}$ and $C_{T,n}$ represent the total capacitance at the input of each preamplifier ($2C_d + C_{par} + c_{gs}$) and $C_{B,p}$ and $C_{B,n}$ represent the total feedback capacitance around each preamplifier ($c_{gd} + C_F$). By assuming high gain in the preamplifiers ($g_m r_o \gg 1$), the corresponding pole and zero locations for the n-channel and p-channel amplifiers will be

$$p_n = -\left(\frac{g_{m,n}}{C_{T,n}} + \frac{1}{C_{B,n}r_{o,n}}\right) \quad z_n = \frac{g_{m,n}}{C_{B,n}} \quad (3.4)$$

$$p_p = -\left(\frac{g_{m,p}}{C_{T,p}} + \frac{1}{C_{B,p}r_{o,p}}\right) \quad z_p = \frac{g_{m,p}}{C_{B,p}} \quad (3.5)$$

In the case of a receiver with the highest possible sensitivity, we assume $C_F=0$, and thus $z_n=g_{m,n}/c_{gd,n}$ and $z_p=g_{m,p}/c_{gd,p}$. In MOSFETs operating in the saturation region, the gate-drain capacitance is always smaller than the gate-source capacitance $c_{gd} < c_{gs}$, and thus these zeros will be neglected because they are above the cutoff frequency $\omega_T=g_m/c_{gs}$ of the devices. Then, assuming $s \ll j\omega_T$ and $g_m r_o \gg 1$, the transfer functions become

$$\frac{V_n}{i_d} = \frac{g_{m,n}}{C_{T,n}c_{gd,n}} \frac{1}{s\left(s + \left(\frac{g_{m,n}}{C_{T,n}} + \frac{1}{r_{o,n}c_{gd,n}}\right)\right)} \quad (3.6)$$

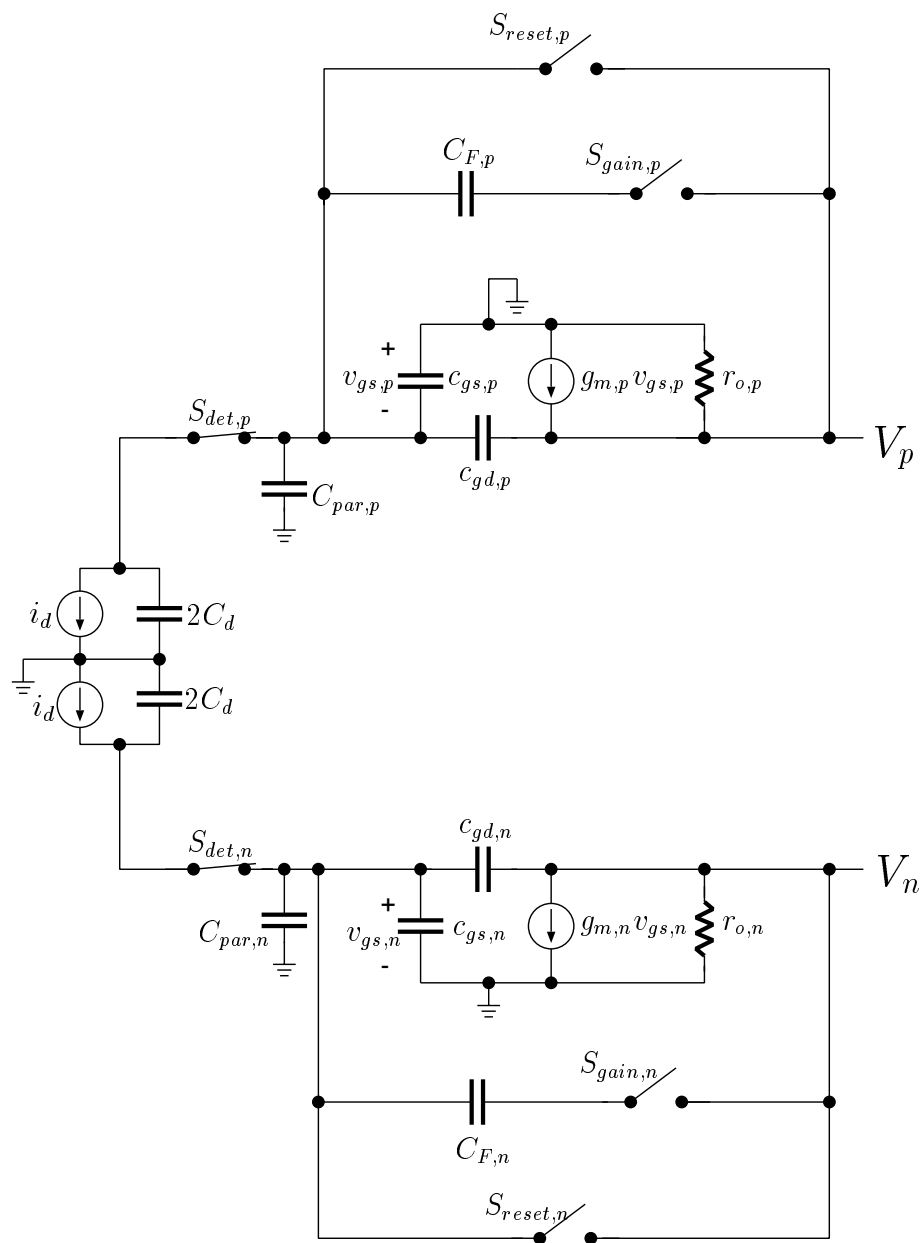


Figure 3.8. Preamplifier circuit model.

$$\frac{V_p}{i_d} = \frac{g_{m,p}}{C_{T,p}c_{gd,p}} \frac{1}{s \left(s + \left(\frac{g_{m,p}}{C_{T,p}} + \frac{1}{r_{o,p}c_{gd,p}} \right) \right)} \quad (3.7)$$

These responses have an integrating pole at the origin and a second pole at $(g_m/C_T) + (1/r_{o,p}c_{gd})$. This second pole will limit the maximum bitrate of the amplifiers, and is at a frequency below (but approaching) the f_T of the input transistors. However, as we shall later see, the maximum bitrate of the receiver will be ultimately determined by the maximum speed of the switches and clock generator and not the preamplifiers.

3.4 BIAS CANCELLATION CIRCUIT

In order to generate a balanced signal from the two single-ended preamplifiers, it is necessary to somehow combine the outputs V_p and V_n from Fig. 3.7. A bias cancellation circuit is shown in Fig. 3.9. Switches $S1$ – $S8$ are used to sample and then subtract the initial and final voltages at the output of the two preamplifiers.

Initially (before a bit arrival), the switches $S1$ – $S2$ and $S7$ – $S8$ are closed as shown in Fig. 3.9a. This operation places sampling capacitors at the outputs of the amplifiers and charges the capacitors to the amplifier output voltages $V_{p,init}$ and $V_{n,init}$. The right sides of the capacitors are tied to a constant voltage V_{ref} which provides a dc reference voltage for the output bitstream. V_{ref} is typically placed halfway between the supply rails to prevent clipping.

After the storage of the initial voltage, all of the switches are then opened and the preamplifiers are allowed to integrate the input photocurrent for a given bit. At the end of the bit, switches $S3$ – $S6$ are closed, as in Fig. 3.9b. This operation places the initial stored voltages in series with the amplifier outputs, thereby subtracting the initial values from the final values. Thus, the output differential voltage depends only on the integration result independent of any dc bias conditions. If $V_{ref}=0$, then

$$V_{NET} = \Delta V_p - \Delta V_n \quad (3.8)$$

$$= V_{p,final} - V_{p,init} - (V_{n,final} - V_{n,init}) \quad (3.9)$$

This type of bias cancellation is similar to the offset voltage cancellation common in MOS operational amplifier design [38]. In addition, the sampling operation serves to implement *correlated double sampling*, a technique which reduces the effect of $1/f$ noise in the circuit (this will be discussed later).

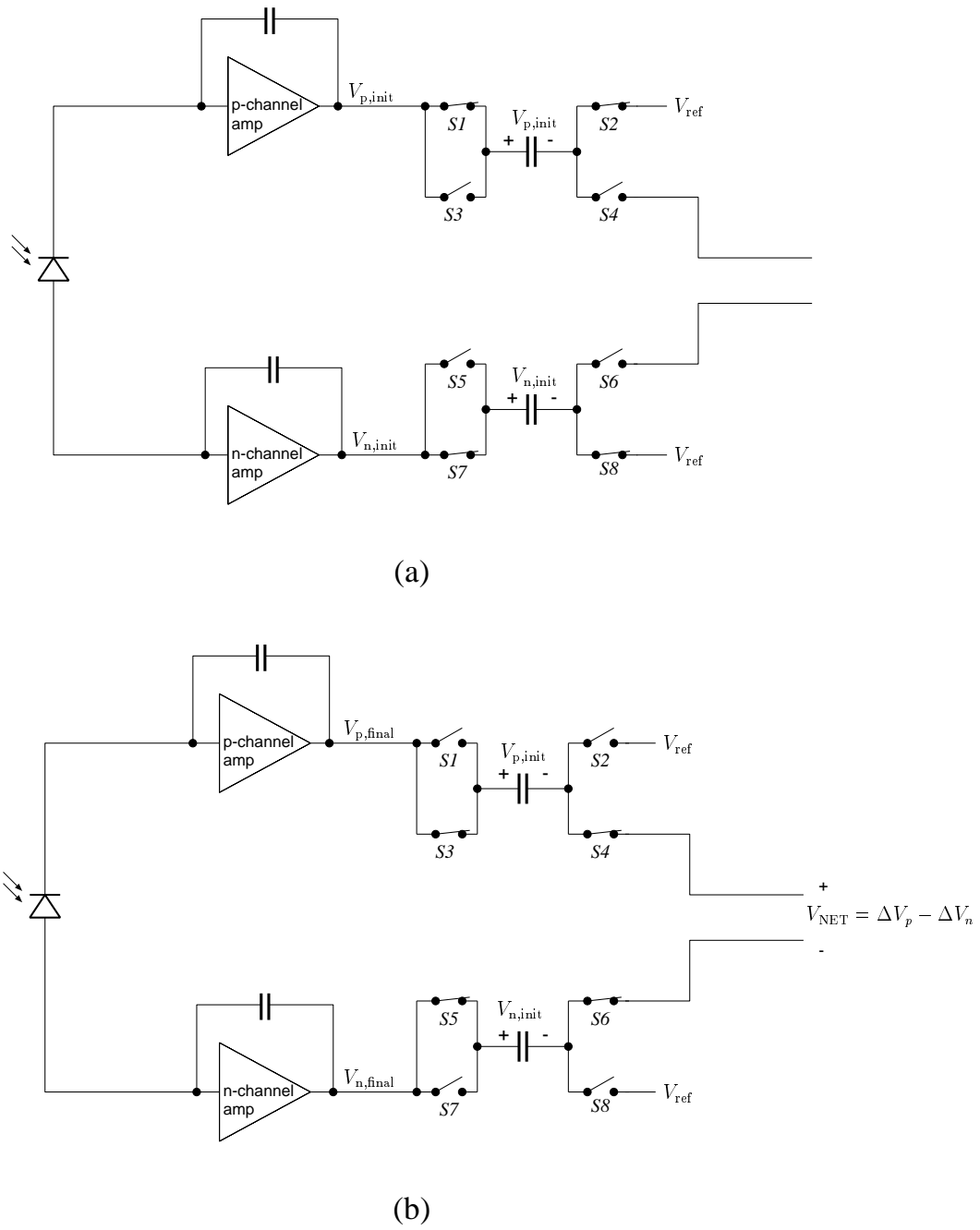


Figure 3.9. Bias cancellation circuit. (a) Before the integration, the initial values of V_p and V_n are stored on the capacitors by closing switches $S1$ – $S2$ and $S7$ – $S8$. (b) After the integration, the capacitors are placed in series with the amplifier outputs by closing switches $S3$ – $S6$. The resulting output V_{NET} will be the sum of the net changes at the outputs of the amplifiers.

3.5 COMPLETE RECEIVER FRONT-END

Fig. 3.10 shows a complete parallel-processing optical receiver front-end. The circuit uses four parallel sets of dual integrating preamplifiers which are connected through switches. The switches are controlled by a four-phase nonoverlapping clock generator. Each preamplifier pair has its own bias cancellation circuit. At any given moment, each of the four sets of preamplifiers is performing one of four operations: *integrate*, *dump*, *readout*, or *bias store*. The fourth clock phase is necessary (vs. Fig. 3.2) in order to allow time for the initial voltage storage in the bias cancellation circuit. It is the parallel architecture which allows each preamplifier pair to be designed with minimal complexity.

Each of the four bias cancellation circuits is isolated from the other outputs by a pair of buffers. This prevents any charge sharing between the bias storage capacitors and any parasitic capacitances introduced by the switches. These buffers are simple p-channel source followers.

In addition, a pair of fast high-power buffers are used to drive off the chip. These buffers are placed after the output switches for the four parallel preamplifier circuits. These buffers (Fig. 3.11) are simple n-channel source followers with the source and bulk terminals shorted for optimal gain. The buffers are capable of driving approximately 350 mV_{p-p} into a 50 Ω cable through a dc block.

3.6 GAIN CONTROL

In order to obtain a large dynamic range in an optical receiver, it is necessary to implement some kind of gain control in order to optimize the preamplifier for high or low incident light levels. For the integrate-and-dump design, this may be accomplished by altering the size of the integration capacitor. This is shown in Fig. 3.12 where the amplifier feedback capacitance may be adjusted by opening or closing the switches *S1* or *S2*. From (3.2) and (3.3), we know that the amplifier gain is inversely proportional to the total feedback capacitance $C_B = c_{gd} + C_F$. Thus, the highest gain (and most sensitive performance) is achieved when $C_F=0$ and there is no external feedback capacitance.

In order to *reduce* the gain and subsequently extend the dynamic range to higher incident light levels, an external feedback capacitance may be added by closing *S1* or *S2*. If we assume that $C_{F1}=5c_{gd}$ and $C_{F2}=5C_{F1}$, this allows a 31:1 variation on the value of C_B ,

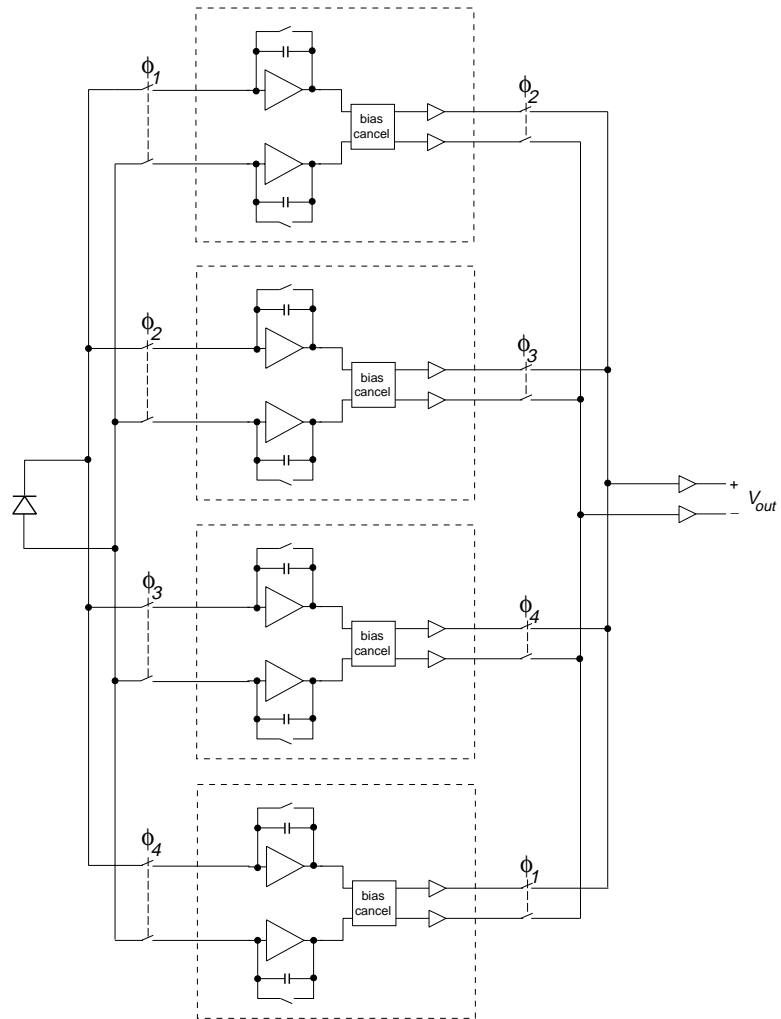


Figure 3.10. Complete circuit. At any given moment, each of the four sets of preamplifiers is performing one of four operations: *integrate*, *dump*, *readout*, or *bias store*.

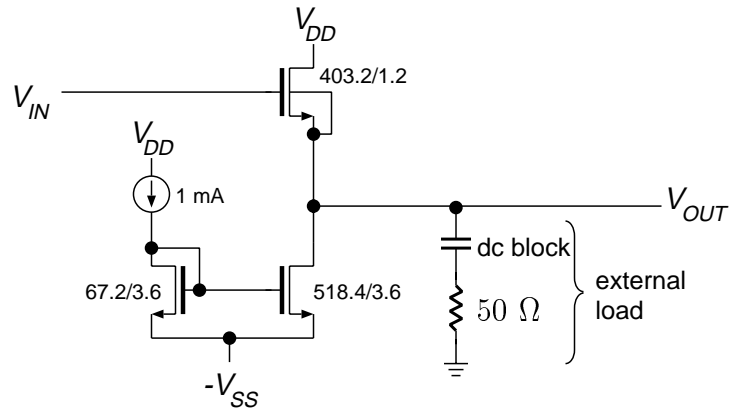


Figure 3.11. Output buffer used to drive off the chip. The buffer is capable of driving approximately 350 mV_{p-p} into a 50Ω cable through a dc block.

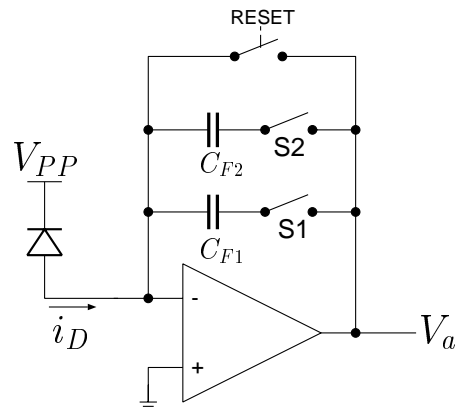


Figure 3.12. The gain of the preamplifiers may be adjusted by adding feedback capacitance around the amplifier. This is done by operating switches $S1$ or $S2$.

which corresponds to a 15 dB extension of the dynamic range.

3.7 CMOS SWITCH DESIGN

There are two issues to be addressed in the switches in the integrate-and-dump circuit

- the switch “on” resistance
- capacitive coupling from the gate voltage, or “clock feedthrough.”

The switch “on” resistance has two effects: (a) thermal noise and (b) the RC time constant which is formed by the switch and the node capacitance. For the circuit of Fig. 3.8, we will assume that only the switches $S_{det,n}$ and $S_{det,p}$ will have significant thermal noise because they are at the input of the preamplifier and thus will see significant gain.

The effect of the RC time constant may be seen as follows. For a given clock rate, we would like the switching transient to settle within a reasonable portion of the bit, typically within the first third of the bit time. Assuming that the transient lasts approximately three time constants, this means that RC should be about one tenth of the bit period. Thus, to achieve 500 Mb/s, the time constant should be $2 \text{ ns} \div 10$, or 0.2 ns. For a worst-case node capacitance of 2 pF, this corresponds to a switch resistance of 100 Ω .

The resistance of a CMOS transistor in the linear region is given by

$$R_{sw} = \frac{1}{\mu C_{ox} \frac{W}{L} (v_{gs} - V_T)} \quad (3.10)$$

The only variable available to the designer to control the switch resistance is the gate width W , since L is always assumed at the minimum gate length and v_{gs} is at the maximum value limited by the supply. However, excessively increasing W also increases the parasitic capacitances due the source and drain areas, and also increases the drive requirements due to the larger gate capacitance. Thus, a finite amount of switch resistance is unavoidable.

The problem of clock feedthrough is shown in Fig. 3.13. Every MOS switch has a parasitic capacitance c_{ovl} due to overlap of the gate with the drain and source areas. When a voltage transition is made on the gate as in Fig. 3.13a, voltage divider is formed between c_{ovl} and C_H , and a residual voltage is left on V_C . This residual voltage is called “clock feedthrough.” A similar overlap capacitance exists between the gate and drain of the switch, but is omitted from the drawing because of the low impedance seen at the drain terminal due to V_S .

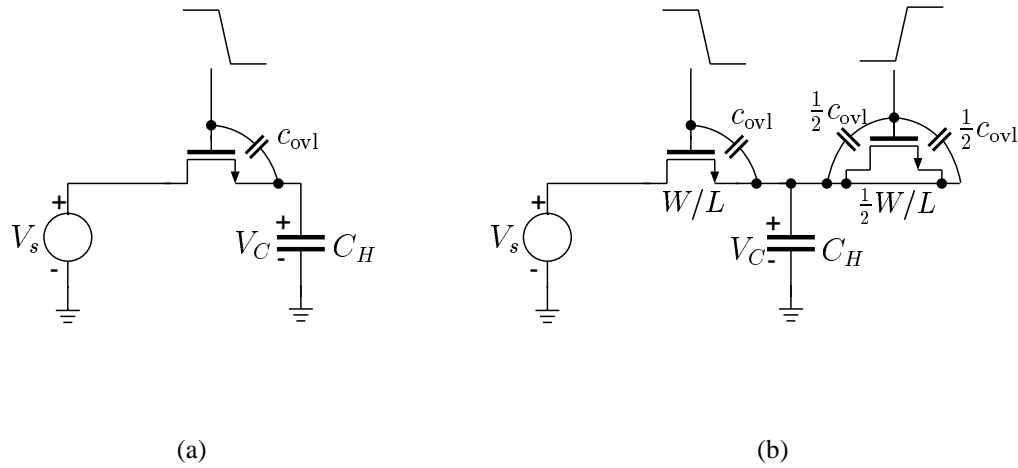


Figure 3.13. (a) Clock feedthrough due to capacitive coupling will corrupt the node voltage V_C . (b) By adding a half-size dummy transistor with a complementary clock, the capacitive coupling may be canceled to first order and V_C remains unchanged.

A simple way to mitigate the feedthrough effect would be to increase the value of C_H compared to c_{ovl} . However, as just described, this would have the effect of increasing the RC time constant of the switch, and thereby slow down the circuit. Another method to reduce the feedthrough is shown in Fig. 3.13b. In this circuit, a dummy transistor with shorted drain and source is connected to C_H . The dummy transistor is half the size of the switch transistor, but the sum of its gate-source and gate-drain overlap capacitors is the same as the gate-source overlap in the switch. By applying a complementary waveform to the gate of the dummy transistor, the feedthrough due to the switch clock will be canceled to first order [39].

This simplistic analysis neglects the effects of the channel charge injection and the nonlinear gate and drain area capacitances in the MOSFETs. However, due to the binary nature of the received waveform in the receiver, we are less concerned with the accuracy of the switch errors than their reproducibility. Thus, as long as the switching errors do not vary significantly from bit to bit, they may only cause a constant offset which may then be canceled in the decision circuit at the output of the receiver.

3.8 COMPLETE CIRCUIT

Fig. 3.14 shows a circuit schematic for a dual integrate-and-dump preamplifier. Four of these circuits are combined in parallel to create the complete receiver.

$M3$ and $M17$ are the common source transistors for the p-channel and n-channel preamplifiers respectively. $M7$ and $M21$ are the reset switches for the integrators. $M1$ and $M15$ are used to connect the photodetector to the preamplifiers. $M8$ and $M22$ are used to control the integrator gain by adding capacitors in feedback.

The bias cancellation circuit is formed by $M9$, $M11$, and $M14$ on the p-channel side, and $M23$, $M25$, and $M28$ on the n-channel side. $M29$ and $M32$ are source follower output buffers which prevent coupling between the outputs of the four parallel stages. Output selection is determined by switches $M34$ and $M35$.

For the n-channel amplifier and its associated bias cancellation circuit, all of the switches are simple n-channel transistors (as opposed to CMOS transmission gates). This is because the voltages which are switched on the n-channel side are all close to the negative supply rail, and n-channel transmission gates are good at passing low voltages. Similarly, all of the switches on the p-channel side are p-channel transistors because the voltages will be near the positive supply rail and p-channel transmission gates are good at passing high voltages. All of the switches for the preamplifiers and bias cancellation circuits use dummy transistors to reduce the clock feedthrough.

3.9 CLOCK GENERATOR

In order to operate the four parallel receivers, it is necessary to generate a four-phase nonoverlapping clock. This type of clocking has previously been used for switched-capacitor circuits and microprocessors [20, 40]. The principle for generating n such clocks is shown in Fig. 3.15. In this circuit, each clock phase cannot go high until the previous phase has gone low. For example, ϕ_1 is unable to go high unless ϕ_n goes low first, and ϕ_2 is unable to go high unless ϕ_1 goes low first. The clocks are activated by sequentially asserting enable signals $\phi_{1E}, \phi_{2E}, \dots, \phi_{nE}$. These enable signals may be generated with a simple shift register or Johnson counter. The amount of overlapping between the clocks is determined by the gate delay and transition time of the NOR gates.

Fig. 3.16a shows the complete logic diagram of a four-phase nonoverlapping clock gen-

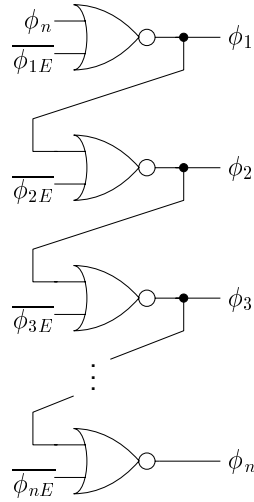


Figure 3.15. Method of generating n nonoverlapping clocks. [40]

erator. This circuit uses four NOR gates at the output similar to Fig. 3.15. However, rather than using feedback from the adjacent NOR gate to ensure nonoverlap, this circuit uses an inverter-delayed copy of the input of the adjacent NOR gate. This allows more careful control of the nonoverlap time by adjusting the delay of the inverters. For example, the overlap time could be increased by adjusting the W/L ratio of the inverter or by using a chain of three inverters instead of one. The four NAND gates on the left of the circuit form a 4-bit shift register which enables the NOR gates. The NANDs are clocked with a complementary clock at half the bit rate with a 50% duty cycle. The resulting timing diagram is shown in Fig. 3.16b. The nonoverlap time is determined by the inverter delay t_{inv} .

This clock generator performs well at high speeds because the feedback loop in the sequential part of the circuit (the NAND gates) contains only one gate delay. In addition, each NAND is only loaded by two other NANDs. By making the gate widths of the NANDs large, the loading by the inverters and NOR gates may be neglected and the fall time of the NAND output is [41]

$$\begin{aligned}
 t_{\text{fall}} &\approx 4 \frac{C_{\text{load}}}{\mu_n C_{ox} \frac{W_n}{L_n} V_{DD}} \\
 &\approx 4 \frac{2(W_n L_n C_{ox} + W_p L_p C_{ox})}{\mu_n C_{ox} \frac{W_n}{L_n} V_{DD}}
 \end{aligned} \tag{3.11}$$

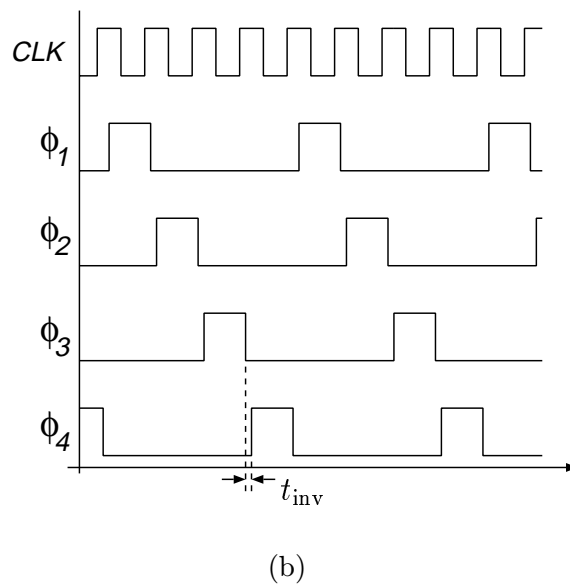
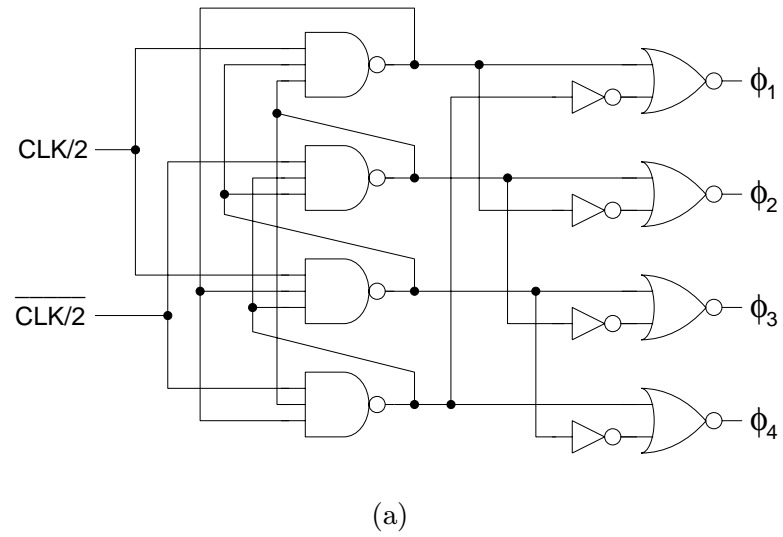


Figure 3.16. (a) Logic diagram of nonoverlapping four-phase clock generator. (b) Timing diagram.

Similarly, the rise time of the NAND is

$$t_{\text{rise}} \approx 4 \frac{2(W_n L_n C_{ox} + W_p L_p C_{ox})}{\mu_p C_{ox} \frac{W_n}{L_n} V_{DD}} \quad (3.12)$$

and thus the minimum bit period is

$$\begin{aligned} T_{\text{min}} &= t_{\text{fall}} + t_{\text{rise}} \\ &= \frac{8L^2}{V_{DD}} \left(\frac{1 + \frac{W_p}{W_n}}{\mu_n} + \frac{1 + \frac{W_n}{W_p}}{\mu_p} \right) \end{aligned} \quad (3.13)$$

where it is assumed that $L=L_n=L_p$. For a 1.2 μm CMOS technology with $V_{DD}=5$, $\mu_n=0.043$, $\mu_p=0.012$, and $W_p/W_n=2$, the minimum output period for the shift register is approximately 500 ps.

The rest of the circuit (NORs and inverters) consists of two levels of combinational logic in order to generate the nonoverlap time. Thus, the clock generator cannot be operated at its minimum bit period of 500 ps due to the necessity of having a small but finite nonoverlap time between the phases. Fig. 3.17 shows a simulation of the clock generator at 250 MHz. With an inverter-chain buffer at the outputs of the NOR gates, a nonoverlap time of 600 ps and an “on” time of 3.4 ns may be achieved. The satisfaction of the nonoverlap condition on the clock generator ultimately limits the maximum clock rate of the integrate-and-dump circuit.

3.10 PREDICTED PERFORMANCE

3.10.1 Noise Response of Integrate-and-Dump Circuit

Since the integrate-and-dump filter is a time-variant circuit, certain aspects of ordinary linear noise theory may not be directly applied [52, 53, 54, 55]. For example, in Section 2.4.4, we assumed that the noise for the integrate-and-dump case was only due to the amplifier noise at the sampling instant. In reality, because the integrate-and-dump receiver is a clocked system with four clock cycles involved in each bit measurement, there will be other noise sources which must be accounted for. We will describe this more realistic noise performance in a method similar to the analysis by White for CCD sense amplifiers [42].

Because the integrate-and-dump circuit is a sampled-data system, we must keep track of the noise at each sampling point. We will use the circuit of Fig. 3.18 as an example. The circuit contains a voltage amplifier with input capacitance C_a . Capacitor C_d models

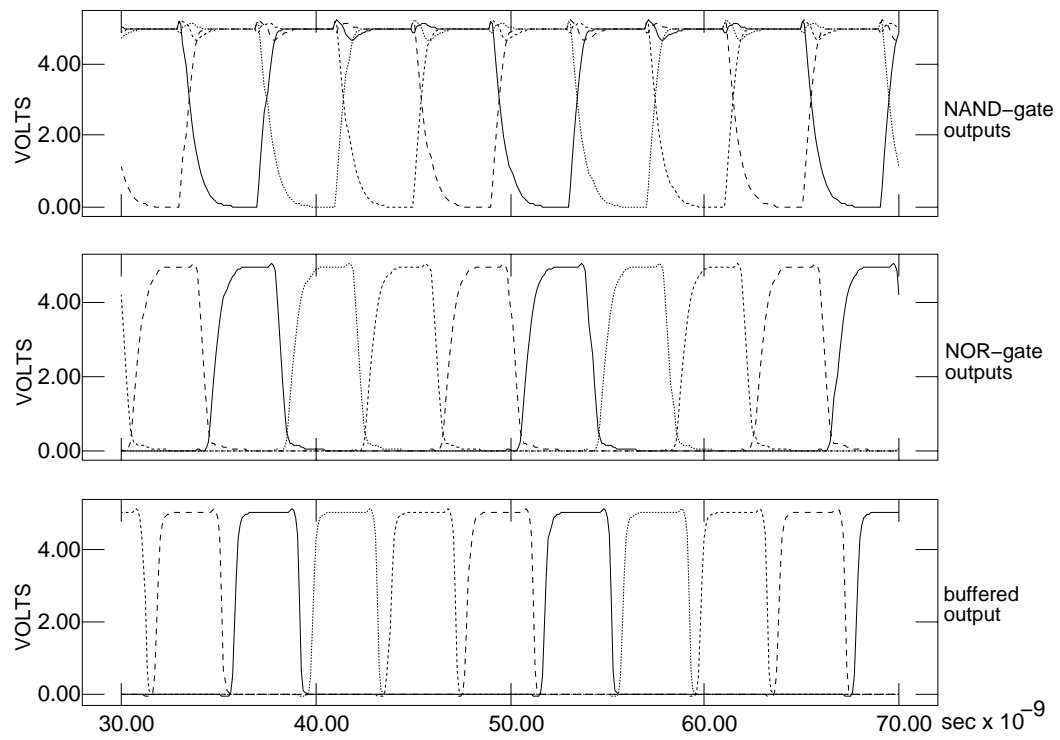


Figure 3.17. Simulation of clock generator at 250 Mb/s. Top trace: outputs of cross-coupled NAND shift register. Center trace: nonoverlapping outputs of NOR gates. Bottom trace: buffered outputs.

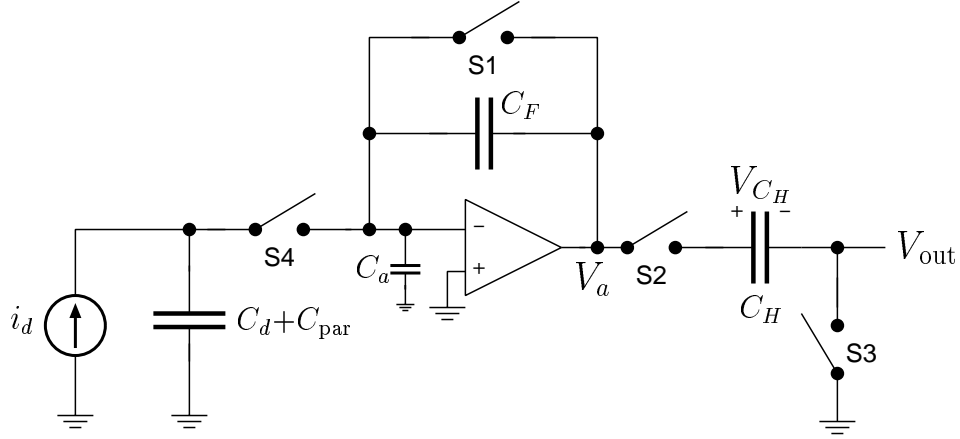


Figure 3.18. Integrate-and-dump circuit model for noise determination.

the detector capacitance, C_{par} models the parasitic capacitance due to wiring, and C_F is in feedback around the amplifier. The four switches $S1$, $S2$, $S3$, and $S4$ are closed and opened in sequence as shown in Fig. 3.19.

Initially, switch $S1$ is closed to reset the integrator, thereby setting the amplifier output $V_a=0$ (neglecting amplifier offset voltage). When $S1$ is reopened, V_a will be affected in three ways: (a) clock feedthrough from $S1$ will cause a residual charge to be left on C_F , thereby leaving a dc offset component $\Delta V_{ft,S1}$ at V_a ; (b) thermal noise due to the resistance of switch $S1$ will be sampled and held on C_F , leaving an offset component $\Delta V_{n,S1}$ at V_a ; and (c) thermal noise due to the amplifier will be sampled and held on C_F , leaving $\Delta V_{n,amp,1}$ (the subscript indicates “noise due to the amplifier in cycle l ”) at V_a . Thus, at the end of the reset cycle, $V_a = \Delta V_{ft,S1} + \Delta V_{n,S1} + \Delta V_{n,amp,1}$.

During the second timing cycle, the switches $S2$ and $S3$ are closed, thereby placing the sampling capacitor C_H at the output of the amplifier. Thus, the amplifier output voltage V_a is placed on C_H . In addition, the voltage held on C_H will contain a component $\Delta V_{n,amp,2}$ due to the instantaneous value of the thermal noise of the amplifier at the end of cycle 2. When $S2$ and $S3$ are opened, the voltage held on C_H will be $V_{C_H} = \Delta V_{ft,S1} + \Delta V_{n,S1} + \Delta V_{n,amp,1} + \Delta V_{n,amp,2}$. The effects of the thermal noise and feedthrough of switches $S2$ and $S3$ are presumed to be small compared to other circuit noises and will be neglected.

During the third timing cycle, the switch $S4$ is closed, thereby connecting the photodetector (modeled as a current source and a capacitance) to the amplifier input. Feedthrough

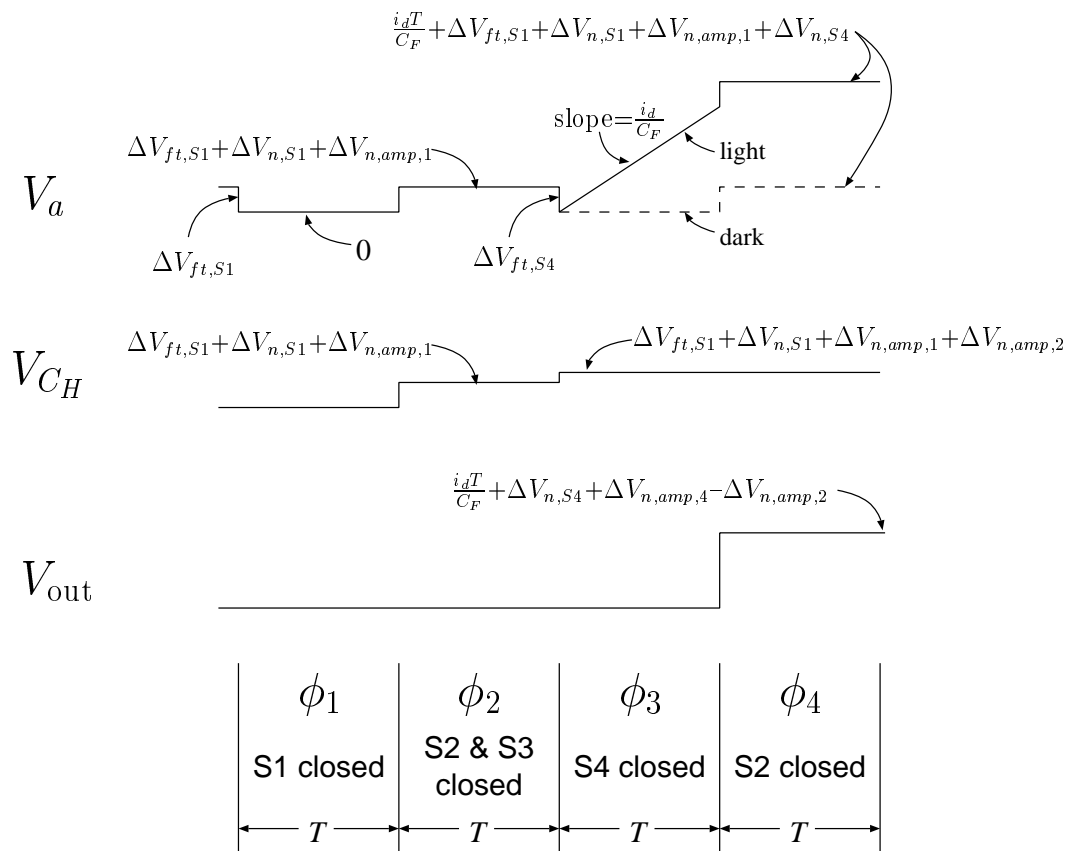


Figure 3.19. Timing diagram of integrate-and-dump circuit showing noise sampling and clock feedthroughs.

from $S4$ will cause a change of $\Delta V_{ft,S4}$ on V_a . However, this voltage will be canceled when $S4$ turns off. Depending on the presence or absence of the current i_d (corresponding to a “one” or “zero” bit), V_a will ramp up or stay constant. At the end of the cycle, the integration result will be stored on C_F , along with additive sampled noise voltage $\Delta V_{n,S4}$ due to the resistance of $S4$. Thus, at the end of the third timing cycle, $V_a = i_d T / C_F + \Delta V_{ft,S1} + \Delta V_{n,S1} + \Delta V_{n,amp,1} + \Delta V_{n,S4}$, where i_d is the photodetector current and T is the bit length (in seconds).

During the fourth timing cycle, switch $S2$ is closed and C_H is placed in series with the amplifier output. A final noise from the amplifier $\Delta V_{n,amp,4}$ will be added to V_a due to the sampling of the amplifier output by the subsequent (unshown) data recovery circuitry. This sample will be assumed to occur at the *end* of the fourth timing cycle. Thus, the net output voltage will be

$$V_{out} = V_a - V_{C_H} \quad (3.14)$$

$$= \frac{i_d}{C_F} T + \Delta V_{n,amp,4} + \Delta V_{ft,S1} + \Delta V_{n,S1} + \Delta V_{n,amp,1} + \Delta V_{n,S4} - (\Delta V_{ft,S1} + \Delta V_{n,S1} + \Delta V_{n,amp,1} + \Delta V_{n,amp,2}) \quad (3.15)$$

$$= \frac{i_d}{C_F} T + \Delta V_{n,S4} + \Delta V_{n,amp,4} - \Delta V_{n,amp,2} \quad (3.16)$$

Thus, V_{out} will consist of a signal component $i_d T / C_F$, a thermal noise component $\Delta V_{n,S4}$ due to $S4$, and the difference between two time-separated samples of the amplifier noise $\Delta V_{n,amp,4} - \Delta V_{n,amp,2}$. Depending on the autocorrelation of the amplifier thermal noise, the difference $\Delta V_{n,amp,4} - \Delta V_{n,amp,2}$ will partially cancel the effect of the amplifier thermal noise. This noise reduction operation is called *correlated double sampling* (CDS) [42]. The effect may be seen by examining the noise transfer function from input to output. If we assume a transfer function $H(s)$ for the system, then the transfer function for the time-delayed difference of two input samples will be

$$H_{diff}(s) = H(s) - e^{-s\tau} H(s) \quad (3.17)$$

$$= H(s) (1 - e^{-s\tau}) \quad (3.18)$$

where τ is the delay time between the samples. In order to find the output noise spectral density $S_{n,out}(f)$, we set $s = j2\pi f$ and multiply the input noise spectrum $S_{n,in}(f)$ by the magnitude-square of the transfer function

$$S_{n,out}(f) = S_{n,in}(f) |H_{diff}(j2\pi f)|^2 \quad (3.19)$$

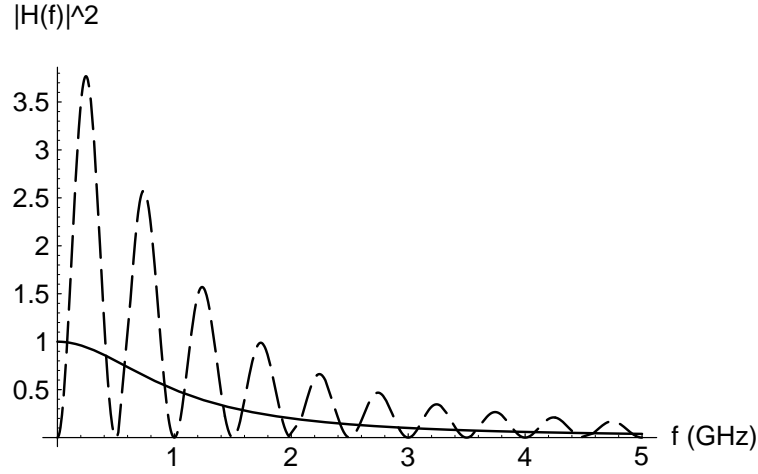


Figure 3.20. Plot of a typical transfer function (*solid line*) and the effect of applying correlated double sampling (*dashed line*).

$$= S_{n,in}(f) |H(j2\pi f)|^2 |1 - e^{-j2\pi f\tau}|^2 \quad (3.20)$$

$$= S_{n,in}(f) |H(j2\pi f)|^2 (4 \sin^2 \pi f\tau) \quad (3.21)$$

The effect of CDS on the output noise spectrum is shown in Fig. 3.20, where the magnitude-square of the transfer function is modified by multiplication by a $4 \sin^2$ function. In cases where the two samples are taken very close together (such as oversampled switched-capacitor filters), the added zero at the origin will effectively suppress much of the inband noise. In the case of the integrate-and-dump receiver, the sampling rate and the filter cutoff are roughly the same (i.e. there is little correlation between the two subtracted samples), and the net effect is to approximately double the total integrated output noise [43]. However, the zero at dc does eliminate the effect of any low-frequency noise components such as $1/f$ noise. This allows us to neglect MOSFET $1/f$ noise throughout this analysis.

This analysis also neglects several additional imperfections in the circuit, including amplifier offset voltage and channel charge due to the MOS switches. In the case of the offset voltage, it is straightforward to show that the offset voltage will be stored and then canceled by the output sampling capacitor C_H . The channel charge will have a more visible but also inconsequential effect. The channel charge distribution is difficult to model but has been shown to depend consistently on the terminal voltages of the switch and the transition time of the gate [44]. We will make the assumption that the switch-off conditions for each of

the switches $S1$ – $S4$ are identical and repeatable for each received bit from the photodetector. In this case, the net effect of the channel charge is to leave a small unpredictable residual dc voltage $\Delta V_{\text{channel}}$ at the output V_{out} . However, if $\Delta V_{\text{channel}}$ is constant for every bit, then it is easily canceled out in later processing, particularly in the data recovery threshold circuit.

We can examine each switch individually and see if it meets the above assumption. In the case of $S1$, both switch terminals will have be at the same dc voltage when the switch is turned off. The same is also true for $S3$ and $S4$. If the control voltages for these switches have a consistent transition time from bit to bit, then the channel charge effect will be constant. In the case of $S2$, the switch terminal voltages will vary from bit to bit, but because $S2$ is not part of the preamplifier input circuit, we will assume that its channel charge is negligible or may be canceled by standard methods such as with dummy transistors [39].

To summarize, the noise of the switched integrate-and-dump circuit as found in (3.16) consists of one sample of the thermal noise of the input switch $S4$, and two samples of the amplifier thermal noise. Fig. 3.21 illustrates the effect of the thermal noise of $S4$ by showing the output noise power spectrum of an integrate-and-dump amplifier with different values of resistance of $S4$. Because this resistance is a parallel noise source, it gets added in quadrature with the amplifier thermal noise and has a white output spectrum. For resistance values below 100 Ω , the net increase in the flat portion of the output noise spectrum is less than 1 dB.

3.10.2 Sensitivity

We may now attempt to predict the sensitivity of the parallel integrate-and-dump design with all of the implementation-specific noise effects added. These effects include

- the extinction ratio of the received bitstream
- the combination of signal and noise from the dual (p-channel and n-channel) preamplifiers
- the added thermal noise due to the switch at amplifier input
- the double sampling of the amplifier thermal noise

In addition, although the analysis of Section 2.4.4 assumes that the amplifier has an optimal cutoff at 0.9 times the bitrate, we will assume that the amplifier has a fixed band-

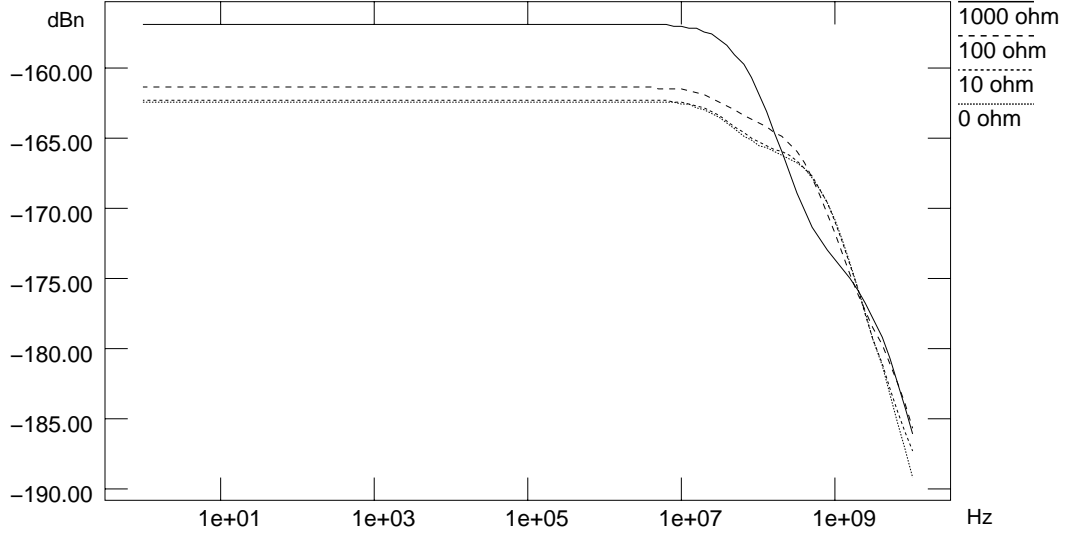


Figure 3.21. The effect of thermal noise of the input switch for the integrate-and-dump receiver. For values of $R_{\text{switch}} < 100 \Omega$, the increase in the flat portion of the output noise spectrum is less than 1 dB. Assumed parameters: $C_d + C_{\text{par}} = 2$ pF, $g_m = 11.4$ mS, $C_a = 7$ pF, $C_F = 0$.

width which is determined by the circuit design. Proceeding along the lines of Section 2.4.4, we may write transfer functions for the n-channel and p-channel preamplifiers from (3.6) and (3.7)

$$H_n(s) = \frac{g_{m,n}}{C_{T,n} C_{B,n}} \frac{1}{s \left(s + \frac{1}{\tau_n} \right)} \quad (3.22)$$

$$H_p(s) = \frac{g_{m,p}}{C_{T,p} C_{B,p}} \frac{1}{s \left(s + \frac{1}{\tau_p} \right)} \quad (3.23)$$

where $1/\tau_n = (g_{m,n}/C_{T,n} + 1/r_{o,n}C_{B,n})$ and $1/\tau_p = (g_{m,p}/C_{T,p} + 1/r_{o,p}C_{B,p})$ are the high frequency 3-dB bandwidths of the two preamplifiers. These transfer functions correspond to the following step responses [30]

$$y_{n,step}(t) = \frac{g_{m,n}\tau_n}{C_{T,n}C_{B,n}} \left(t - \tau_n + \tau_n e^{-t/\tau_n} \right) \quad (3.24)$$

$$y_{p,step}(t) = \frac{g_{m,p}\tau_p}{C_{T,p}C_{B,p}} \left(t - \tau_p + \tau_p e^{-t/\tau_p} \right) \quad (3.25)$$

The maximum signal for the integrate-and-dump amplifier is the sum of the two integrator outputs at $t=T$

$$[y_s(t)]_{\max} = y_{n,step}(T) + y_{p,step}(T) \quad (3.26)$$

$$= \frac{g_{m,n}\tau_n}{C_{T,n}C_{B,n}} (T - \tau_n + \tau_n e^{-T/\tau_n}) + \frac{g_{m,p}\tau_p}{C_{T,p}C_{B,p}} (T - \tau_p + \tau_p e^{-T/\tau_p}) \quad (3.27)$$

Applying the results of Section 3.10.1, the power spectral density $S_N(f)$ of input noise is

$$\begin{aligned} S_N(f) &= 2(\text{thermal noise due to preamplifier}) \\ &+ (\text{thermal noise due to preamplifier input switch}) \end{aligned} \quad (3.28)$$

Thus, at the input of the n and p amplifiers, the noise spectrums $S_{N,n}(f)$ and $S_{N,p}(f)$ will be

$$S_{N,n}(f) = 2k\Theta(2\pi C_{T,n})^2 f^2 \left(\frac{2\Gamma}{g_{m,n}} + R_{sw,n} \right) \quad (3.29)$$

$$S_{N,p}(f) = 2k\Theta(2\pi C_{T,p})^2 f^2 \left(\frac{2\Gamma}{g_{m,p}} + R_{sw,p} \right) \quad (3.30)$$

where $R_{sw,n}$ and $R_{sw,p}$ are the resistances of the switches at the inputs of the n and p amplifiers, respectively.

The total output noise is then

$$\overline{n_{\text{out},n}^2} = \int_{-\infty}^{\infty} 2k\Theta(2\pi C_{T,n})^2 \left(\frac{2\Gamma}{g_{m,n}} + R_{sw,n} \right) f^2 \left| \frac{g_{m,n}}{C_{T,n}C_{B,n}} \frac{1}{j2\pi f (j2\pi f + \frac{1}{\tau_n})} \right|^2 df \quad (3.31)$$

$$\overline{n_{\text{out},p}^2} = \int_{-\infty}^{\infty} 2k\Theta(2\pi C_{T,p})^2 \left(\frac{2\Gamma}{g_{m,p}} + R_{sw,p} \right) f^2 \left| \frac{g_{m,p}}{C_{T,p}C_{B,p}} \frac{1}{j2\pi f (j2\pi f + \frac{1}{\tau_p})} \right|^2 df \quad (3.32)$$

The receiver sensitivity as a function of the bit period T is

$$\left(\eta \overline{P} \right)_{\text{circuit}} = \left(\frac{h\nu}{q} \right) \left(\frac{1+r}{1-r} \right) Q \frac{\sqrt{\overline{n_{\text{out},n}^2} + \overline{n_{\text{out},p}^2}}}{\frac{g_{m,n}\tau_n}{C_{T,n}C_{B,n}} (T - \tau_n + \tau_n e^{-T/\tau_n}) + \frac{g_{m,p}\tau_p}{C_{T,p}C_{B,p}} (T - \tau_p + \tau_p e^{-T/\tau_p})} \quad (3.33)$$

A plot of (3.33) is shown in Fig. 3.22 as a function of the bitrate $1/T$. This curve differs from the straight line in Fig. 2.25 because of the effect of the frequency response of the amplifiers. In the ideal noise analysis of Chapter 2, the amplifier bandwidth was assumed

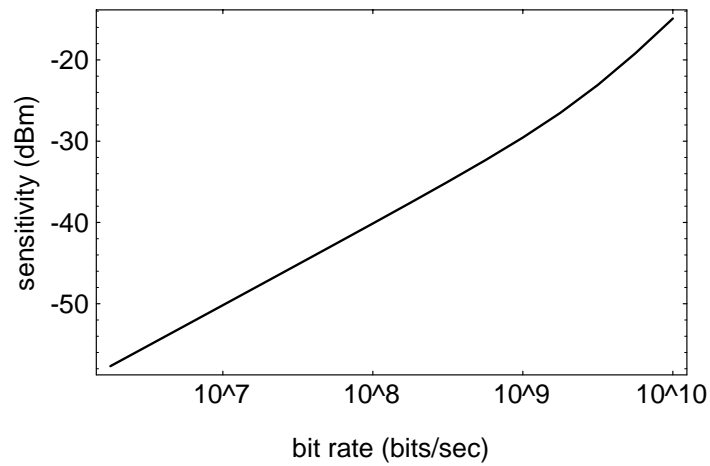


Figure 3.22. Predicted sensitivity performance of the integrate-and-dump optical receiver for 10^{-9} BER. Assumed: $\Theta=300$ K, $\Gamma=0.7$, $C_{T,n}=C_{T,p}=2$ pf, $g_{m,n}=11.2$ m Ω , $g_{m,p}=7.7$ m Ω , $c_{gd,n}=117$ fF, $c_{gd,p}=131$ fF, $r_{o,n}=7.5$ k Ω , $r_{o,p}=1.4$ k Ω , $\lambda=1550$ nm, $r=0.2$, $Q=6$, $R_{sw,n}=68$ Ω , $R_{sw,p}=327$ Ω .

to track the bitrate by a factor of 0.9. In this real analysis, the amplifier bandwidth is fixed at around 1.5 GHz, and thus the sensitivity shows a bend at that frequency as seen in Fig. 3.33.

CHAPTER 4

IMPLEMENTATION AND TEST RESULTS

In order to fully investigate the characteristics of the integrate-and-dump topology, a test chip was implemented in a 1.2 μm p-well double-poly double-metal CMOS technology available from Orbit Semiconductor, Inc [31]. The Orbit foundry service is similar to the more well-known MOSIS foundry [45] but runs weekly, provides 8-week turnaround time, and is slightly more costly.

4.1 CHIP LAYOUT

The test chip layout was done with the MAGIC design tool written at University of California, Berkeley [46], and simulations were done with the HSPICE circuit simulator [47] on an IBM RS/6000 workstation. The particulars of the chip are listed in Table 4.1. Because of the p-well technology, the n-type substrate was grounded and a single -5 V supply was used.

A photograph of the circuit is shown in Fig. 4.1. The complete test circuit occupies approximately 25% of the total die area. The circuit consists of four parallel dual-preamplifiers (top), a four-phase non-overlapping complementary clock generator (lower left), and an output buffer (lower right). In addition, the two octagonal bonding pads at the top are for the attachment of a photodetector which is situated adjacent to the chip. All biasing currents are controlled by external variable resistors. The clock generator uses an external reference clock at one half the bitrate and a 50% duty cycle. Thus, no clock recovery circuit or decision circuit is included on the test chip.

The reference clock input is buffered at the pad by an inverter chain. In addition,

TABLE 4.1. Test chip specifications.

Technology	1.2 μm CMOS p-well double-poly
Area	2.25 mm \times 2.25 mm
Power supplies	0, -5V
Static power dissipation (not including test structures)	290 mW

all analog input pads for the bias currents contain reverse-biased diodes for electrostatic discharge protection.

4.2 TEST SETUP

Fig. 4.2 shows the complete packaging including the receiver chip, photodiode, and decoupling capacitors all silver-epoxied into the well of a TriQuint MLC 132/84 ceramic package. This package was chosen for its good high-speed performance which includes 50Ω impedances on all of the signal leads and subnanosecond transition times. The purpose of having the decoupling capacitors directly in the well is to reduce lead inductance and therefore improve power supply noise rejection. In addition, all dc bias voltages are decoupled directly in the well.

An Epitaxx ETX 75 CER-F PIN photodiode was also epoxied into the well and bonded to the test chip for measurement purposes. The photodiode consists of a $75\ \mu\text{m}$ diameter InGaAs PIN diode on a ceramic submount. The specifications of this device are detector capacitance $C_D=0.36\ \text{pF}$, responsivity $\mathcal{R}=0.95\ \text{A/W}$, and dark current $i_{\text{dark}}=0.02\ \text{nA}$.

The packaged chip was tested using the equipment shown in Fig. 4.3. Bit patterns were generated on a pattern generator and were used to modulate a laser at a wavelength of $1550\ \text{nm}$. The extinction ratio of the laser was 0.2. The amount of optical power was controlled by feeding the laser light through a programmable optical attenuator. Light was coupled to the photodiode by positioning a bare fiber approximately $50\ \mu\text{m}$ above the photodiode.

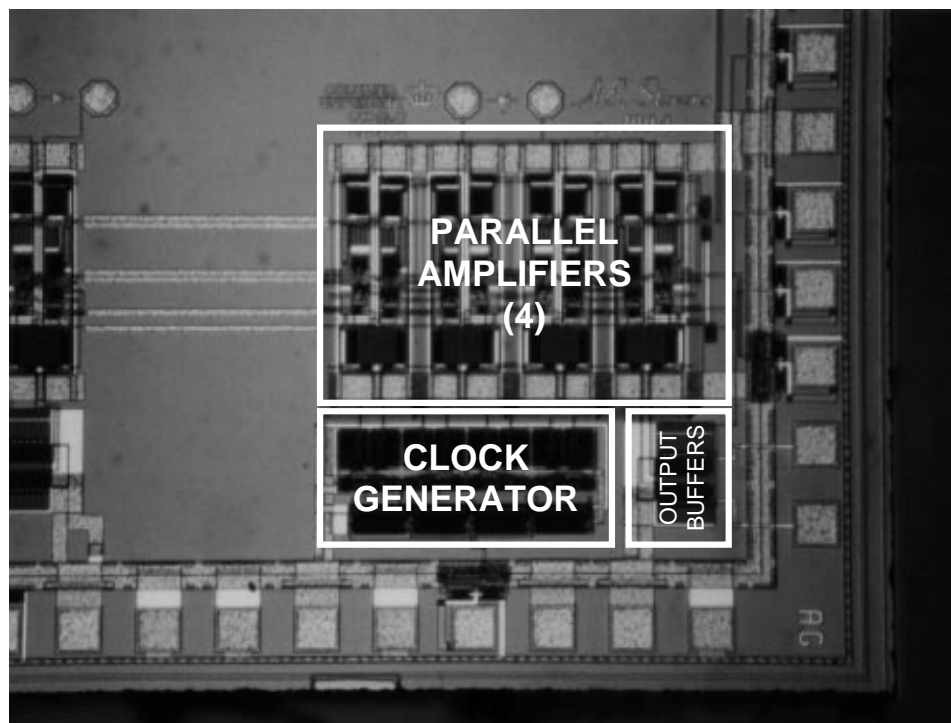


Figure 4.1. Die photograph. Approximate size of active circuit area (not including pads): 1 mm × 1 mm.

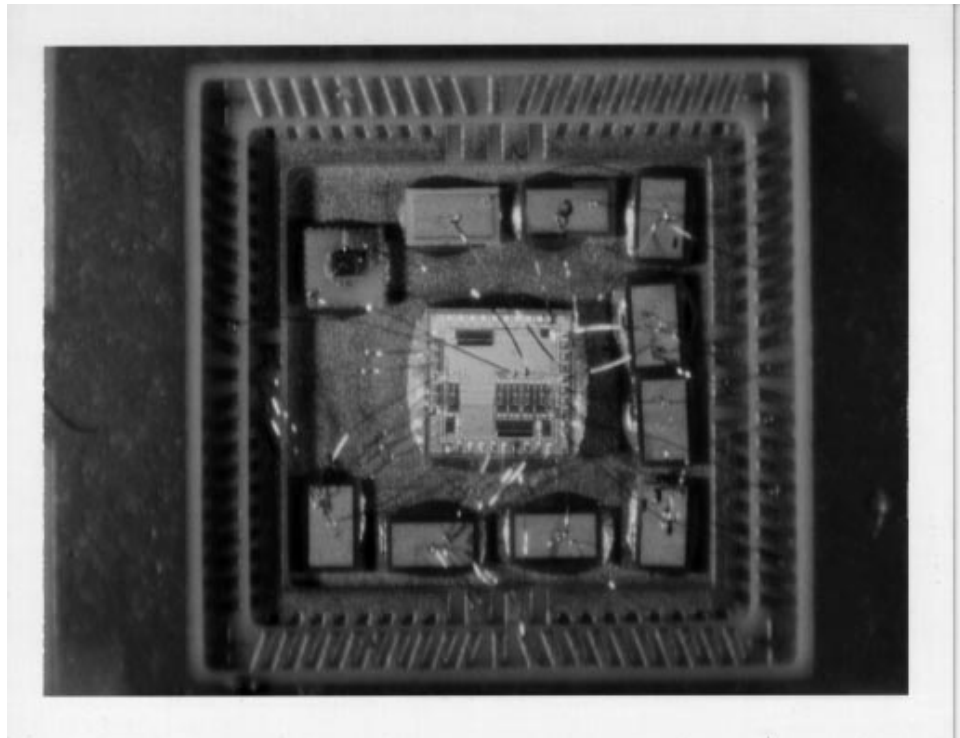


Figure 4.2. Photograph of packaged chip in TriQuint MLC 132/84 ceramic flatpack. Decoupling chip capacitors surround the die. The photodetector is the square in the upper left corner.

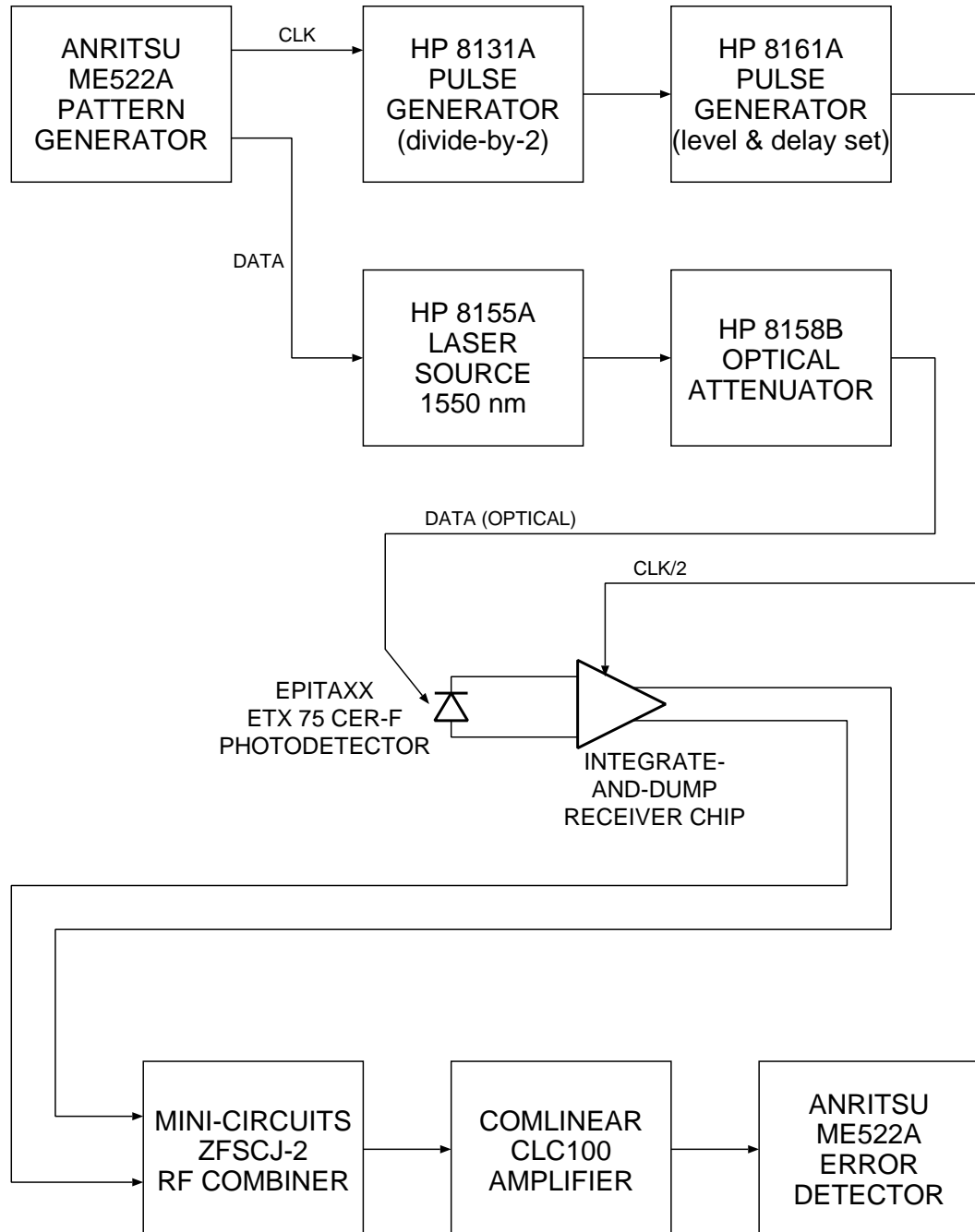


Figure 4.3. Equipment setup for laboratory testing of the integrate-and-dump receiver.

The clock output from the bit pattern generator was used to clock the integrate-and-dump chip. However, because the chip requires an input clock at one-half the bit rate, a triggered pulse generator was used to divide the bit clock by two. This signal was then used to trigger a second pulse generator which was used as a programmable delay and also generated an output voltage at CMOS logic levels (0 and $-5V$).

The differential output of the receiver chip was combined into a single-ended signal by a passive RF combiner. The resulting signal was amplified by high-bandwidth amplifiers and then fed to a bit error rate tester.

4.3 PERFORMANCE

4.3.1 Amplifier offset

Fig. 4.4 shows an oscilloscope photograph of the output of the receiver for an alternating (101010) bit pattern. It is apparent from the photograph that every other “one” is at a different height. Similarly, every other “zero” has a slight offset. This variation is caused by mismatch among the four parallel amplifiers in the receiver. The magnitude of this problem may be shown by sending a random bit sequence through the receiver and triggering the oscilloscope at the bit rate. This produces an eye pattern as shown in Fig. 4.5.

Ordinarily, in a continuous-time receiver, the thickness of the “one” and “zero” lines in the eye pattern correspond to the noise in the receiver. For this integrate-and-dump design, Fig. 4.5 shows that the eye thickness is dominated by the mismatch between the amplifiers. This effect may be more clearly seen by triggering the oscilloscope on every fourth bit, as seen in Fig. 4.6. This photo shows four distinct alternating eyes with different threshold voltages corresponding to the four amplifiers in the receiver.

The eyes in Fig. 4.6 more closely resemble an ordinary eye pattern, with some important differences. Most notable is the fact that each logic level within a given eye actually consists of two distinct voltage levels. This effect is due to intersymbol interference (ISI) and can be attributed to the dynamic biasing technique for the preamplifiers and photodetector. As was shown in Fig. 3.10, the photodetector is connected to four preamplifiers through switches. As the photodetector is switched from amplifier to amplifier, it will be “rebiased” each time to the input voltage of each given amplifier.

During the reception of a “one” bit, the voltage across the detector will change slightly

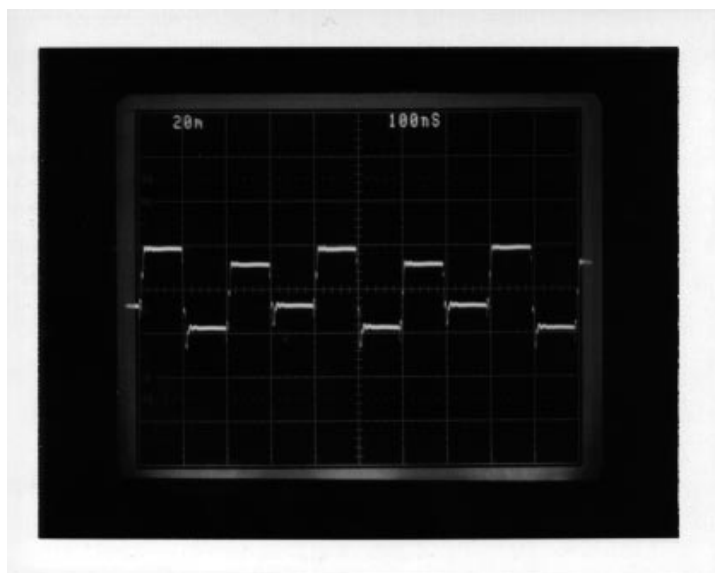


Figure 4.4. Output of optical receiver with an alternating (101010) bit pattern at 10 Mbps.

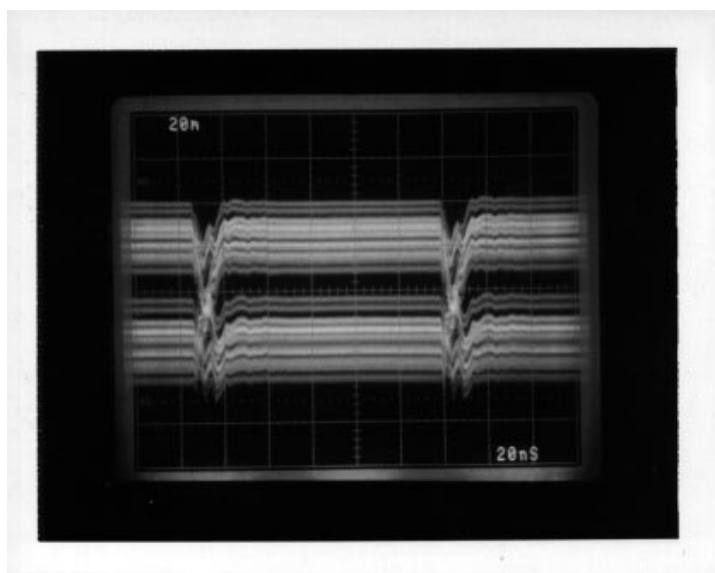


Figure 4.5. Eye pattern at 10 Mbps for a 2^7-1 pseudo-random bit sequence.

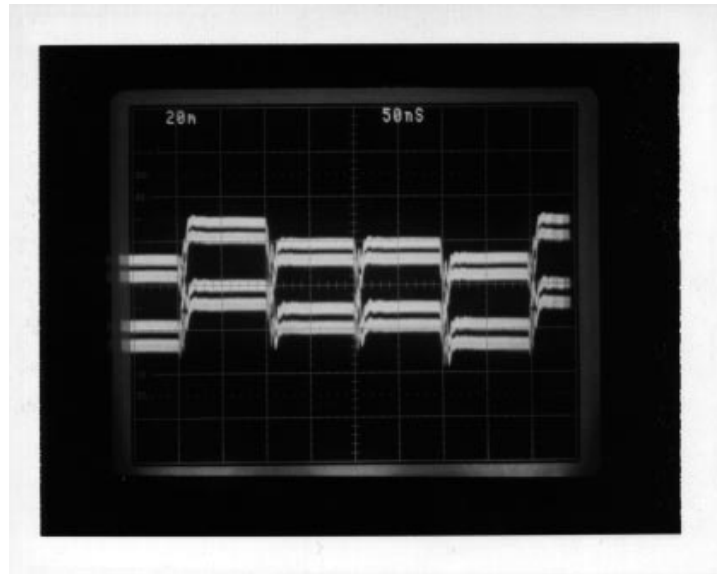


Figure 4.6. Eye pattern at 10 Mbps for a 2^7-1 PRBS with the oscilloscope triggered on every fourth bit. The mismatch between the four amplifiers is apparent.

as the input node capacitance is charged by the photocurrent. When the switches are then operated to connect the photodetector to the next successive amplifier, the voltage across the detector is slightly different (on the order of millivolts) than the dc bias voltage at the input of the next amplifier. Thus, at the instant that the switches close, a charge sharing will take place between the detector and the amplifier in order to equalize the voltage between the two. This will cause a net voltage change at the output of the amplifier. This charge sharing will not occur on “zero” bits because the detector voltage will not change in the absence of any photocurrent. The net result is that there will be a small error offset voltage for each bit which follows a “one” bit, but not a “zero” bit. Hence, for any given bit, the output level will vary depending on the value of the previous bit.

This charge-sharing effect is not ISI in the usual sense. In most optical receivers, the ISI is due to bandlimiting effects and smearing of bits. In this case, the ISI is due to the mechanics of the circuit design, and is not related to the bandwidth at all.

Another peculiarity of the eye in Fig. 4.6 is the transients which are present at the bit transitions. This effect is more readily apparent when the chip is operated at higher bitrates. Figs. 4.7 and 4.8 show the receiver eye diagrams at bitrates of 40 and 100 Mb/s.

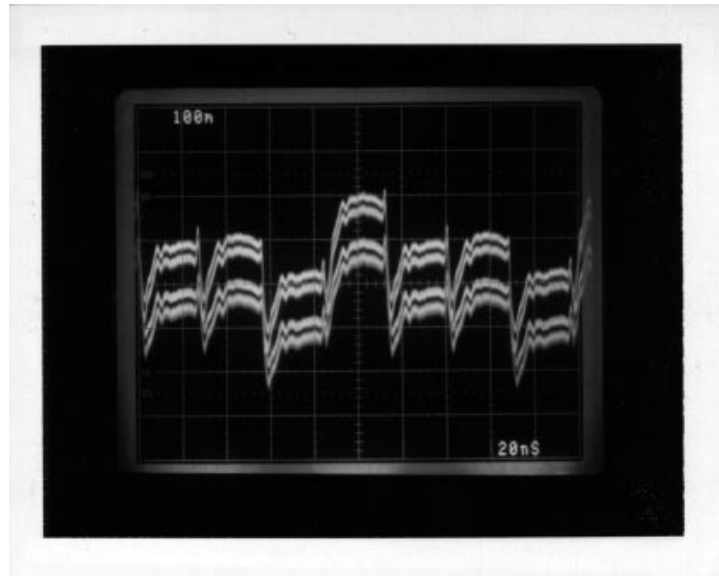


Figure 4.7. Eye pattern at 40 Mbps for a 2^7-1 PRBS with the oscilloscope triggered on every fourth bit.

The transients at the bit transitions become more apparent as they occupy a larger portion of the total bit time.

These transients are due to the switched-capacitor nature of the receiver design and are the expected behavior for this kind of circuit. They occur due to clock feedthroughs and switching transients as the amplifiers are reset and switched. However, because the circuit is a sampled-data system, the shape of the waveforms is unimportant as long as the final value is settled to appropriate accuracy. Thus, as long as we can sample the waveform at the appropriate time with a decision circuit or flip flop, the transients may be ignored. This is a common result in switched-capacitor circuit design [20].

Some of the ripple in Figs. 4.7 and 4.8 is due to transmission line reflections and power supply decoupling noise in the test setup. These effects would be reduced in a more monolithic design, particularly if a decision circuit was integrated directly on the chip.

4.3.2 Common mode noise

Fig. 4.9 further illustrates the effect of common mode noise and its suppression by the balanced circuit design. The two traces in Fig. 4.9a show the differential outputs of the



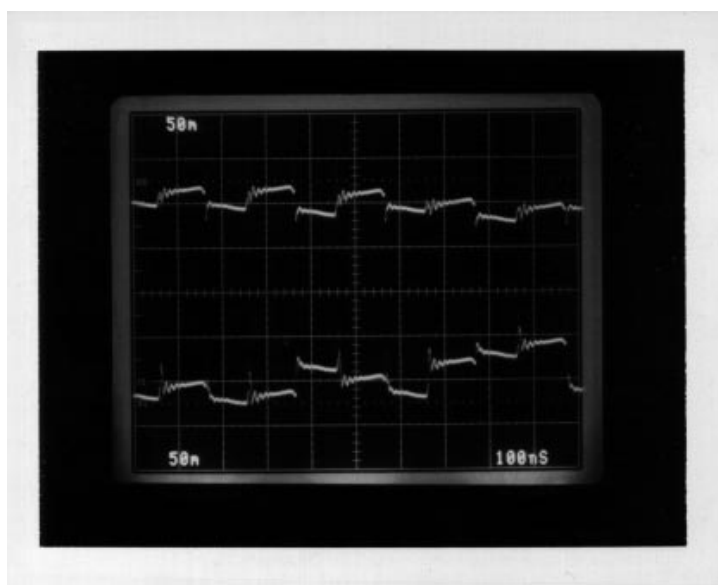
Figure 4.8. Eye pattern at 100 Mbps for a 2^7-1 PRBS with the oscilloscope triggered on every fourth bit.

receiver. These waveforms were generated by a pseudo-random bit sequence at 10 Mb/s and are unrecognizable as binary signals. The waveforms are dominated by noise and loading effects in the output buffers.

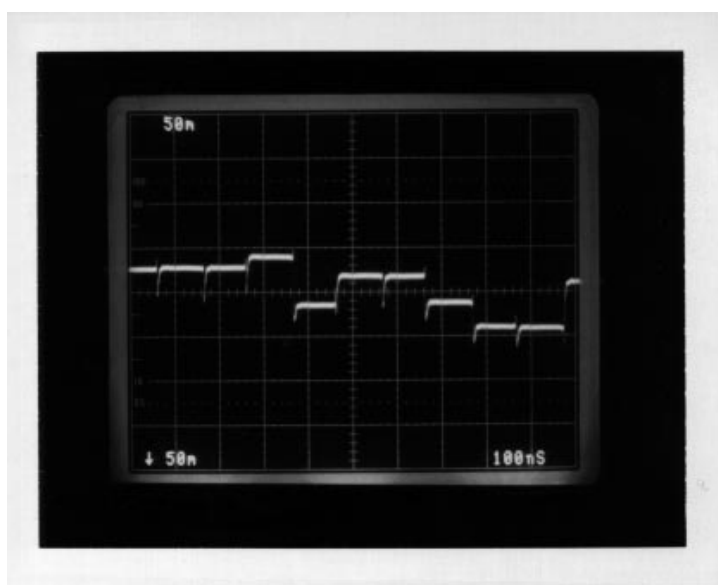
Fig. 4.9b shows the result of subtracting the two differential waveforms. The loading effects are canceled and recognizable bits appear (albeit with a differing threshold at every fourth bit as discussed previously). This simple example shows the huge advantage gained by using a differential output from the chip vs. a single-ended output.

4.3.3 Sensitivity

Given the problem of different threshold voltages among the four parallel amplifiers shown in Fig. 4.6, the determination of the sensitivity of the receiver becomes more difficult than in an ordinary receiver. However, we can get a close approximation of the sensitivity by clocking the chip at four times the bitrate, or stated another way, by transmitting the bits at one fourth the chip clock. This is illustrated in Fig. 4.10, which shows an input bitstream at one fourth the clock rate, and the resulting output bitstream. Each output “bit” actually consists of four consecutive measurements, with each measurement by one of



(a)



(b)

Figure 4.9. (a) Oscilloscope photographs showing the differential outputs of the chip. (b) When the two traces are subtracted, the transients are canceled and usable bits appear.

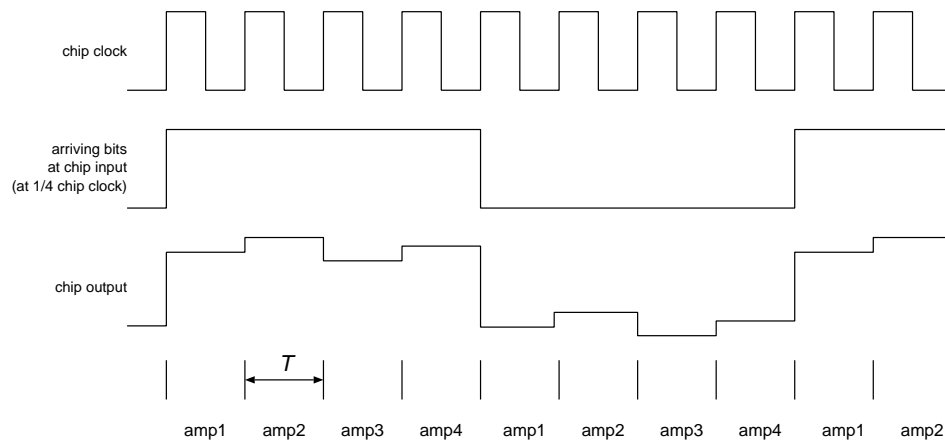


Figure 4.10. Method for clocking the integrate-and-dump receiver with four mismatched parallel amplifiers. By sending the bits at one fourth the chip clock, each of the four amplifiers will see the same bit once before the next bit transition occurs. Note that the chip input and output waveforms are lined up in the diagram for illustrative purposes. In reality, they will be delayed from each other by one clock cycle.

the four amplifiers in the parallel configuration.

By operating the chip in this fashion, we can generate an eye diagram as shown in Fig. 4.11, where the center portion of the “bit” contains a region with an open eye with just two logic levels. This portion of the eye opening will depend only on the thermal noise levels on the chip, without any ISI effects from the preceding bit. This waveform may then be connected directly to a bit error rate test for noise measurement.

Fig. 4.12 plots the bit error rate as a function of the received optical power. As more light is shined on the receiver, the error rate improves, as shown by the falling slope of the plot. The two sets of points in the plot show the performance of the integrate-and-dump receiver with and without the effect of ISI. This corresponds to measurement in the two regions of the eye of Fig. 4.11: “without ISI” corresponds to a decision time during the center bits of the waveform with only two logic levels, and “with ISI” corresponds to the outer bits of the waveform with the “split” logic levels. At the lowest light levels, the ISI degrades the performance by 1.1 dB at 10^{-9} BER. Table 4.2 lists the measured receiver sensitivity at three different bitrates without ISI.

The “with ISI” results are meant to represent the true performance of the test chip despite the mismatch between the four preamplifiers. Thus, neglecting mismatch, the best

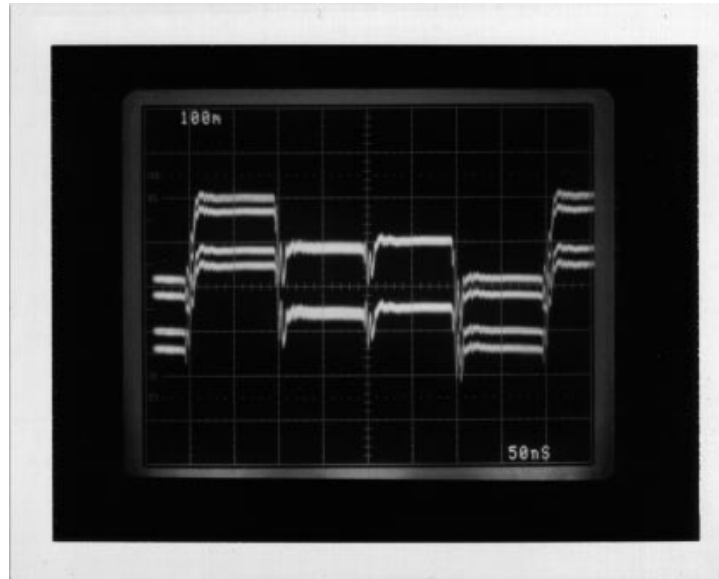


Figure 4.11. Eye pattern with the chip clock at 10 Mbps and the bit rate at 2.5 Mbps.

TABLE 4.2. Receiver sensitivity without ISI.

Bitrate (Mb/s)	Measured Sensitivity (dBm)	Predicted Sensitivity (dBm)
10	-49.4	-50.2
40	-42.0	-44.1
100	-35.0	-40.1

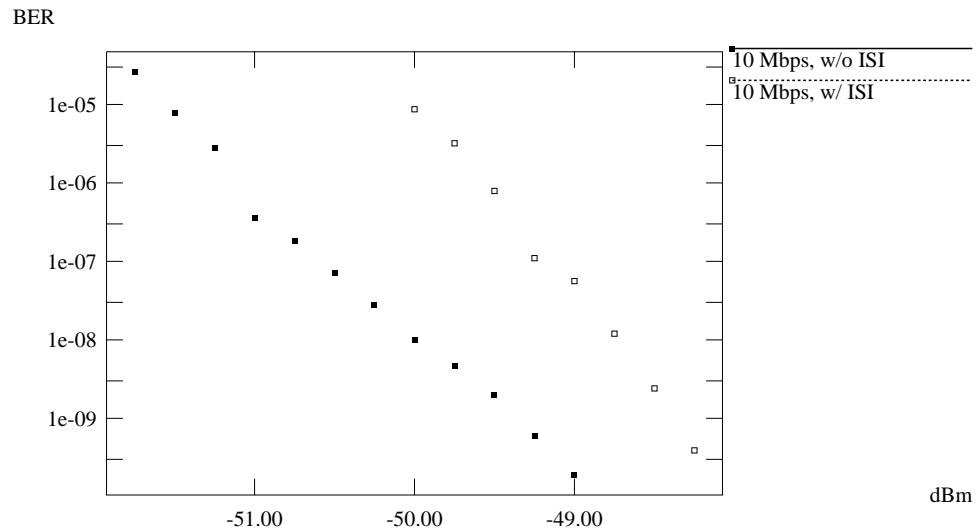


Figure 4.12. Bit error rate vs. received power at 10 Mbps.

case sensitivity of the test chip is -48.3 dBm. We can also compare the best-case performance (“without ISI”) with the prediction from Section 3.10.2. Because the prediction did not contain the ISI effect, this should be a valid comparison. This comparison is shown in Table 4.2 and plotted in Fig. 4.13.

It is apparent that at low bitrates (~ 10 Mb/s), the integrate-and-dump performance measures up well with the prediction. However, as the bitrate is increased, the sensitivity begins to deviate significantly from the prediction. The problem here is twofold: first, as previously shown in Fig. 4.8, at the higher bitrates, the transient portion of the bit transitions begin to dominate the total bit time and causes eye closure. In addition, the problem of power supply decoupling becomes more a problem due to the sharp transition times which are necessary on the reference clock input to the chip. These sharp transitions cause switching noise and ground bounce [48].

4.3.4 Dynamic Range

The dynamic range of the receiver is defined as the total range of optical power over which a minimum bit error rate may be achieved. Thus, it is the ratio between the *maximum* receivable power level and the *minimum* receivable power level for a given bit error

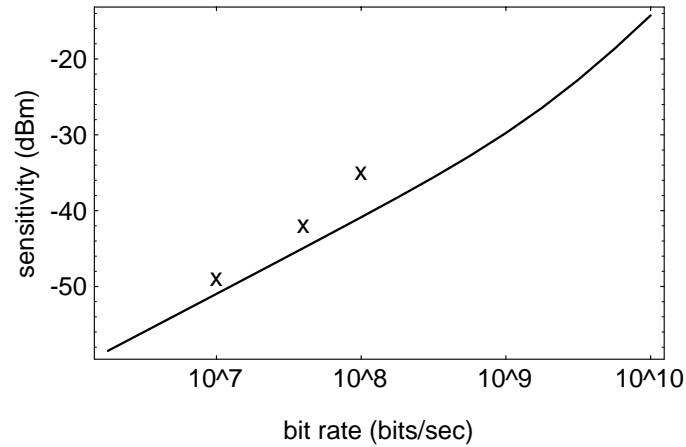


Figure 4.13. Theoretical (solid line) and actual performance (×-marks) of integrate-and-dump receiver chip. Measurements were taken at 10, 40, and 100 Mb/s.

rate. The dynamic range is difficult to predict by simulation because receivers often operate nonlinearly at high power levels and a given receiver may continue to operate adequately despite the nonlinearities (e.g. amplifiers saturating or transistors cutting off). Thus, dynamic range is usually measured in the lab.

The integrate-and-dump receiver operated satisfactorily at optical powers as high as -31 dBm for a BER of 10^{-9} . This corresponds to a dynamic range of $-31 - (-48.3) = 17.3$ dB. This number may be improved by lowering the gain by adding capacitance in feedback around the preamplifiers. As previously described, a 0.5 pF switchable feedback capacitor was included for the preamplifiers on the test chip specifically for this purpose. Adding this feedback capacitance allowed a maximum input optical power level of -23 dBm, thus giving allowing an overall dynamic range of $-23 - (-48.3) = 25.3$ dB.

4.4 IMPROVEMENT OF DESIGN

There are several areas in the prototype chip design in which a simple improvement in the circuit design could produce a drastic improvement in the circuit performance. First and foremost is the threshold mismatch between the four preamplifiers. This problem arises from the fact that the outputs of the four preamplifiers are combined together at the analog level and thus must share the same external decision circuit. This problem could be

remedied by using a separate decision circuit for each preamplifier (Fig. 4.14), thus causing the multiplexed outputs to be digital instead of analog. The four outputs may then be combined on the chip with a digital multiplexer. In practice, the multiplexer could be implemented with the same switching arrangement used for the analog case as in Fig. 3.10. Alternatively, the chip could have four demultiplexed outputs with separate output pins, thus performing a partial serial-to-digital conversion.

Another area for improvement of the design lies in the bandwidths of the preamplifiers. Because the switches and clocking limit the usable bitrate of the chip, the preamplifier bandwidths on the chip are unnecessarily high. Fig. 2.23 showed that the optimum ratio between the bandwidth and the bitrate for an integrate-and-dump preamplifier is $\gamma=0.9$. However, the preamplifiers on the test chip have a bandwidth of approximately 1.5 GHz. For a bitrate of 100 Mb/s, this corresponds to $\gamma=15$. Thus, the test chip is unoptimized with respect to bandwidth and could achieve significantly better performance by simply reducing the bandwidth. This is illustrated in Fig. 4.15, which shows the actual test chip performance, the predicted test chip performance, and the predicted performance with a preamplifier bandwidth of 90 MHz. The reduced bandwidth gives an improvement of about 5 dB in sensitivity at bitrates below 100 Mb/s.

A large problem with measuring the full performance of the test chip was due to noise coupling from the high-power output buffers to the sensitive preamplifiers. This effect was mainly due to coupling through the power supply lines. Methods to limit this noise included using a high-speed package and bonding decoupling capacitors directly in the well. While these measures caused visible improvement in the sensitivity performance (vs. the original 40-pin DIP package with external decoupling capacitors), they were still limited by the finite lead inductance of the bonding wires. A better chip layout could provide better performance by physically distancing the output drivers from the preamplifiers (to reduce substrate coupling), using on-chip decoupling capacitors, and by using separate power supply pads and wiring for the output buffers.

Another source of noise coupling was the reference clock input to the chip for the clock generator. This input was a single-ended rail-to-rail logic signal at one half the bitrate. A better format would have been a low-swing differential signal which would place balanced opposite loads on the supplies. The differential signal could then have been converted to standard CMOS logic levels on the chip.

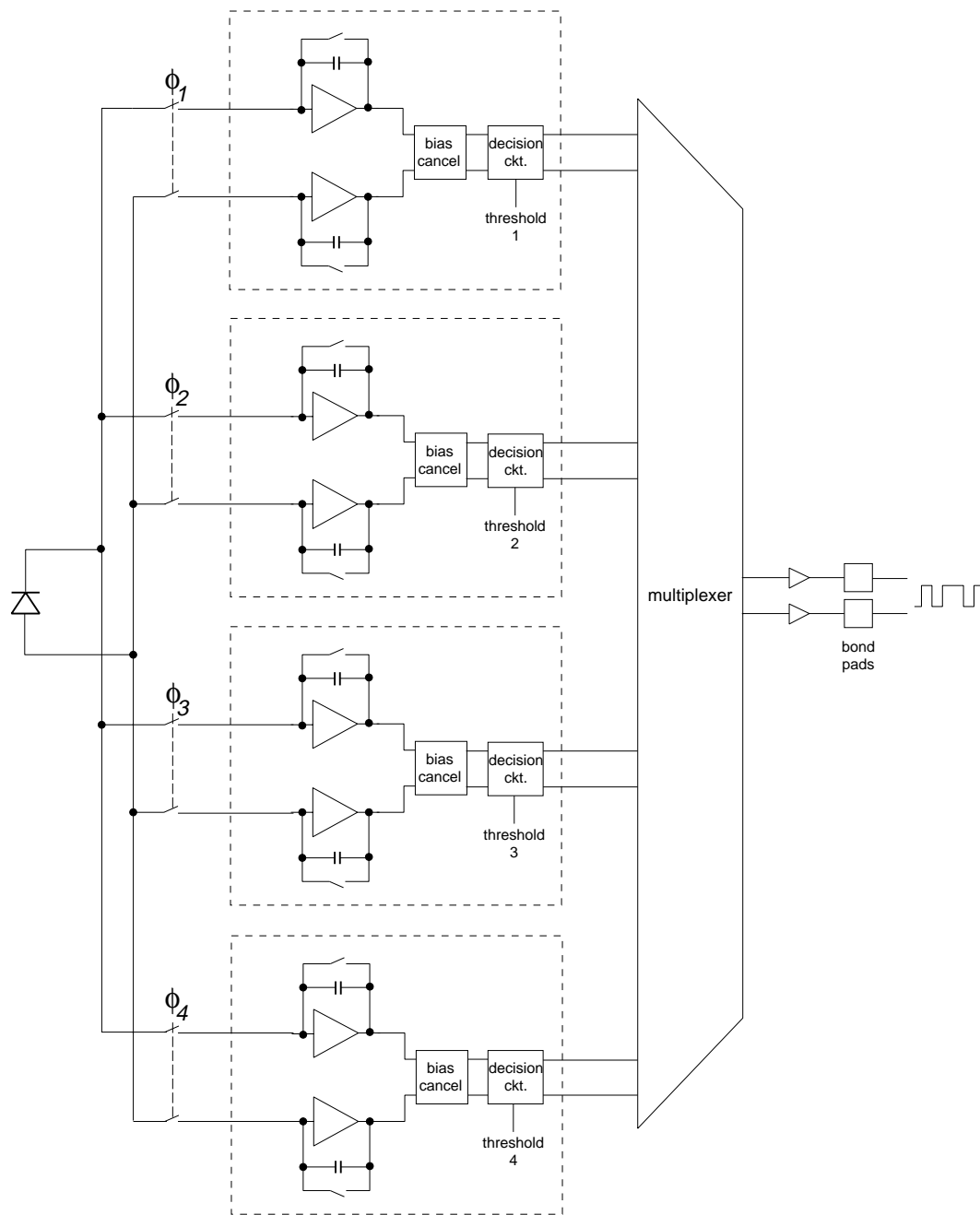


Figure 4.14. Proposed method for remedying the mismatch between the four parallel preamplifiers. Each preamplifier uses a dedicated decision circuit with its own threshold level. The digital outputs of the preamplifiers are then combined with a digital multiplexer.

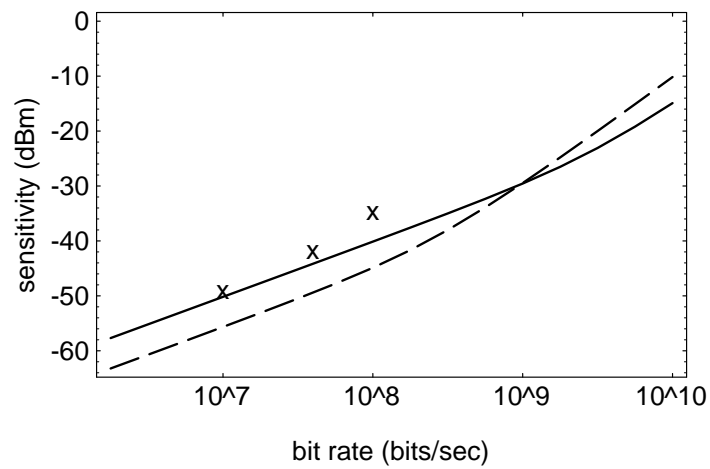


Figure 4.15. Actual test chip performance (\times -marks) and predicted test chip performance (solid line) of the integrate-and-dump receiver chip, and the predicted performance when the preamplifiers are bandlimited to 90 MHz (dashed line). The sensitivity is improved by approximately 5 dB at bitrates below 100 Mb/s.

CHAPTER 5

SUMMARY AND FUTURE WORK

5.1 THESIS SUMMARY

This thesis has examined the feasibility of building an integrate-and-dump optical receiver in CMOS technology.

Chapter 2 of this thesis examined the theoretical aspects of optical receivers and their internal components, including the photodetector, preamplifier, noise-shaping filter, and timing recovery circuit. The best-case signal-to-noise ratio was shown to depend on the optical bit waveform and the preamplifier noise power spectrum. An equation for the optimum noise filter, or *matched filter*, was derived.

Chapter 2 also examined the signal-to-noise performance of several other types of noise filters, including a simple second-order lowpass filter, a raised-cosine-output filter, and an integrate-and-dump filter. We also derived an expression for the receiver bit error rate as a function of the bitrate and signal-to-noise ratio for the four filtering topologies, assuming gaussian statistics. These were plotted against each other and also compared with the quantum limit for OOK optical receivers (Fig. 2.25).

The integrate-and-dump filter was shown to have better sensitivity (by several dB) than an ordinary second-order lowpass filter. The second-order filter is meant to represent the noise response of many existing optical receivers where the preamplifier frequency response is used as the noise filter. The implication here is that the sensitivity of a given optical preamplifier (transimpedance or high impedance design) could be improved by simply removing the load resistor, thus converting it to an integrator, with the added penalty of the extra circuitry required to reset the integrator after each bit. We also showed that the

optimum bandwidth of the amplifier in the integrate-and-dump configuration is 0.9 times the bitrate.

Chapter 3 of this thesis introduced an integrate-and-dump receiver circuit design which specifically takes advantage of the switches which are available in CMOS technology. The topology uses the concept of parallel analog signal processing in order to simplify the circuit design. By using four identical preamplifiers to simultaneously integrate, readout, bias cancel, and reset during a given bit, the requirements on each individual preamplifier may be relaxed. Thus, by switching the single photodetector among the four preamplifiers at the bitrate, each preamplifier operates on every fourth bit and has four bit periods to complete the necessary signal processing. A four phase nonoverlapping clocking sequence controls the switching among the preamplifiers.

Chapter 3 also introduced a method for biasing the photodetector by using a pair of preamplifiers at different dc bias levels. By utilizing the complementary devices available in CMOS technology, this idea was implemented by designing the two preamplifiers out of opposite types of transistors. A bias cancellation circuit was used to combine the outputs of the two preamplifiers into a single balanced signal.

A noise analysis of the parallel circuit design showed that the thermal noise of the preamplifier is doubled by the sampling operation of the bias cancellation circuit. In addition, the resistance of the switch at the input of the preamplifier adds thermal noise. We also showed that the correlated double-sampling operation of the bias cancellation circuit effectively removes the effect of MOSFET $1/f$ noise by placing a zero at the origin for the noise transfer function. A complete expression for the receiver sensitivity was derived in (3.33) which included the combined outputs of the dual preamplifiers, the fixed preamplifier bandwidth, the extinction ratio, the input switch resistance, and the doubling of the preamplifier thermal noise. The sensitivity was plotted in Fig. 3.22.

Chapter 4 described the implementation and performance of an integrate-and-dump test chip in $1.2\ \mu\text{m}$ CMOS. The chip consists of four parallel preamplifiers and bias cancellation circuits, a clock generator, and analog output buffers. For testing purposes, a PIN photodiode was mounted adjacent to the chip and connected with bond wires to the receiver input. An external reference clock was used to drive the four-phase nonoverlapping clock generator.

At a bitrate of 10 Mb/s, the receiver chip functioned well but was limited in performance by mismatch among the four parallel preamplifiers. This led to a variation in the eye

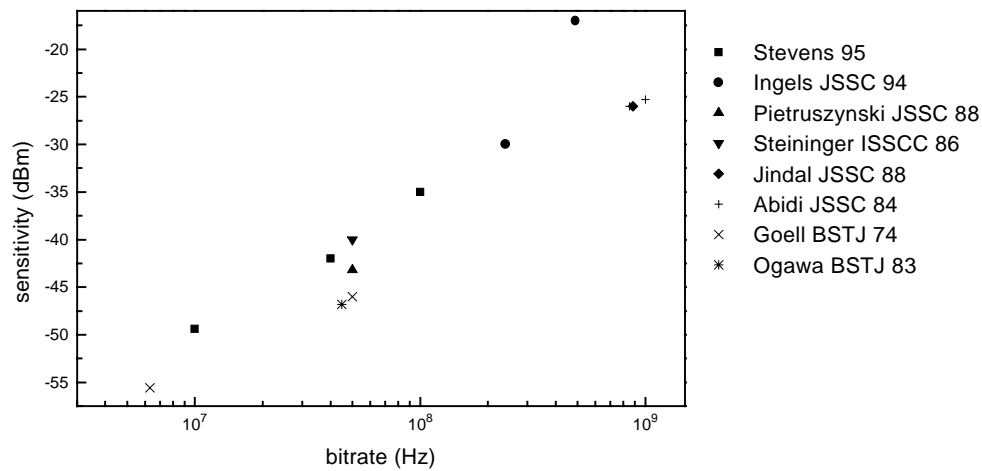


Figure 5.1. The results of the integrate-and-dump test chip compared with other PIN-MOSFET receivers from the literature.

threshold for every fourth bit in the output waveform. However, by reducing the input rate of bits to one fourth of the clocking rate, it was possible to measure the noise performance of the receiver despite the mismatch. At 10 Mb/s, the test chip sensitivity matched well with the theoretical prediction of -49.4 dBm. The receiver chip was also measured at 40 and 100 Mb/s, but had reduced sensitivity due to noise coupling in the power supply.

5.2 COMPARISON WITH PUBLISHED RESULTS

Fig. 5.1 compares the results of the integrate-and-dump test chip with other PIN-MOSFET receivers from the literature [10, 11, 12, 13, 33, 63, 64]. The integrate-and-dump performance is roughly in line with the other receivers. However, as previously stated, the sensitivity of the integrate-and-dump case would improve significantly by reducing the excess bandwidth of the preamplifiers.

5.3 AREAS FOR FURTHER STUDY

This thesis has demonstrated a practical circuit implementation of an integrate-and-dump optical receiver. However, many issues have been left unanswered, and many new questions have arisen. Some of these issues are:

Timing Recovery A single-chip solution is always desirable for any electronic system, and thus the integration of a timing recovery circuit directly onto the preamplifier chip would be a logical next step. Section 2.1.4 described a clock recovery circuit using an early-late gate synchronizer, and Fig. 2.13 showed a way to generate the early and late control signals by sampling the output of the preamplifier during the integration cycle. Other issues to be resolved would include the design of an appropriate VCO and loop filter.

Automatic Gain Control Fig. 3.12 showed a method for controlling the gain of the preamplifiers by adjusting the size of the feedback capacitor. However, some kind of automatic gain control circuit (AGC) would be needed to control the switches for these capacitors. The circuit would sense nonlinear operation in the preamplifier and increase the size of C_F until a linear operating point was found.

Preamplifier Biasing The dynamic biasing method for the preamplifiers described in Section 3.3 is simple and requires no additional circuitry. However, as seen in Fig. 4.6, the dynamic biasing causes a splitting of the logic levels in the eye diagram. This causes a reduction in sensitivity which could probably be recovered by using a dc bias for the preamplifiers.

BIBLIOGRAPHY

- [1] P.E. Green, *Fiber Optic Networks*, Englewood Cliffs, NJ: Prentice Hall, 1993.
- [2] D. Bayart, B. Clesca, L. Hamon, J.L. Beylat, "Experimental investigation of the gain flatness characteristics for 1.55 μm erbium-doped fluoride fiber amplifiers," *IEEE Photonics Technology Letters*, Vol. 6, No. 5, May 1994, pp. 613–615.
- [3] F. Tong, C.-S. Li, A.E. Stevens, Y.H. Kwark, "Characterization of a 16-channel optical-to-electronic GaAs receiver for packet-switched WDMA networks," *IEEE Photonics Technology Letters*, Vol. 6, No. 8, Aug. 1994, pp. 971–974.
- [4] J.B. Soole, et al., "Integrated grating demultiplexer and PIN array for high-density wavelength division multiplexed detection at 1.5 μm ," *Electronics Letters*, Vol. 29, No. 6, Mar. 18, 1993, pp. 558–560.
- [5] Thomas Schrans, private communication, 3/10/94.
- [6] V. Hurm, et al., "14 GHz bandwidth MSM photodiode AlGaAs/GaAs HEMT monolithic integrated optoelectronic receiver," *Electronics Letters*, Vol. 29, No. 1, Jan 7, 1993, pp. 9–10.
- [7] K.D. Pedrotti, et al., "Monolithic ultrahigh speed GaAs HBT optical integrated receivers, *13th Annual GaAs IC Symposium Technical Digest*, 1991, pp. 205–208.
- [8] A. Ketterson, et al., "A 10-Ghz bandwidth pseudomorphic GaAs/InGaAs/AlGaAs MODFET based OEIC receiver," *IEEE Transactions on Electron Devices*, Vol. 39, No. 11, Nov. 1992, pp. 2676–2677.
- [9] J.F. Ewen, A.X. Widmer, M. Soyuer, K.R. Wrenner, B. Parker, H.A. Ainspan, "Single-

- chip 1062Mbaud CMOS transceiver for serial data communication,” *IEEE International Solid-State Circuits Conference Digest of Technical Papers*, 1995, pp. 32–33.
- [10] M. Ingels, G. Van der Plas, J. Crols, M. Steyaert, “A CMOS 18 THzΩ 240 Mb/s transimpedance amplifier and 155 Mb/s LED-driver for low cost optical fiber links,” *IEEE Journal of Solid-State Circuits*, Vol. 29, No. 12, Dec. 1994, pp. 1552–1559.
- [11] D.M. Pietruszynski, J.M. Steininger, E.J. Swanson, “A 50-Mbit/s CMOS monolithic optical receiver,” *IEEE Journal of Solid-State Circuits*, Vol. 23, No. 6, Dec. 1988, pp. 1426–1433.
- [12] J.M. Steininger, E.J. Swanson, “A 50Mb/s CMOS optical data link receiver integrated circuit,” *IEEE International Solid-State Circuits Conference Digest of Technical Papers*, 1986, pp. 60–61.
- [13] R.P. Jindal, “Transimpedance preamplifier with 70-dB AGC range in fine-line NMOS,” *IEEE Journal of Solid-State Circuits*, Vol. 23, No. 3, June 1988, pp. 867–869.
- [14] D.L. Rogers, “Integrated optical receivers using MSM detectors,” *IEEE Journal of Lightwave Technology*, Vol. 9, No. 12, Dec. 1991, pp. 1635–38.
- [15] F. Tong, Y.H. Kwark, A.E. Stevens, “A four-channel monolithic optical/electronic selector for fast packet-switched WDMA networks,” *IEEE Photonics Technology Letters*, Vol. 6, No. 1, Jan. 1994, pp. 68–70.
- [16] R.G. Smith, S.D. Personick, “Receiver design for optical fiber communication systems,” in *Semiconductor Devices for Optical Communication*, 2nd ed., H. Kressel, ed., Berlin: Springer-Verlag, 1982, pp. 89–160.
- [17] V. Radeka, “Low-noise techniques in detectors,” *Annual Review of Nuclear and Particle Science*, Vol. 38, 1988, pp. 217–277.
- [18] F.M. Gardner, *Phaselock Techniques*, 2nd ed., New York: Wiley, 1979.
- [19] M.K. Simon, “Nonlinear analysis of an absolute value type of an early-late gate bit synchronizer,” *IEEE Transactions on Communication Technology*, Vol. COM-18, No. 5, Oct. 1970, pp. 589–596.

- [20] R. Gregorian, G.C. Temes, *Analog MOS Integrated Circuits for Signal Processing*, New York: Wiley, 1986.
- [21] B. Sklar, *Digital Communications*, Englewood Cliffs, NJ: Prentice Hall, 1988.
- [22] A.B. Carlson, *Communication Systems*, 3rd Ed., New York: McGraw-Hill, 1986.
- [23] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 2nd. ed., New York: McGraw-Hill, 1984.
- [24] S.D. Personick, "Receiver design for digital fiber optic communication systems, I," *Bell System Technical Journal*, Vol. 52, No. 6, July-Aug. 1973, pp. 843–874.
- [25] S.D. Personick, "Receiver design for digital fiber optic communication systems, II," *Bell System Technical Journal*, Vol. 52, No. 6, July-Aug. 1973, pp. 875–886.
- [26] B.L. Kasper, "Receiver design," in *Optical Fiber Telecommunications II*, S.E. Miller, I.P. Kaminow, eds., Boston: Academic Press, 1988, pp. 689–722.
- [27] G.F. Williams, "Lightwave Receivers," in *Topics in Lightwave Transmission Systems*, T. Li, ed., Boston: Academic Press, 1991, pp. 79–149.
- [28] T.V. Muoi, "Receiver design for high-speed optical-fiber systems," *IEEE Journal of Lightwave Technology*, Vol. LT-2, No. 3, June 1984, pp. 243–267.
- [29] E.A. Lee, D.G. Messerschmidt, *Digital Communication*, Boston: Kluwer Academic Publishers, 1988.
- [30] K. Ogata, *Modern Control Engineering*, Englewood Cliffs, NJ: Prentice-Hall, 1970.
- [31] *Foresight User's Manual*, Rev. 1.5, Orbit Semiconductor Inc., Sunnyvale, CA, July 1992.
- [32] R.P. Jindal, "Silicon MOS amplifier operation in the integrate and dump mode for gigahertz band lightwave communication systems," *IEEE Journal of Lightwave Technology*, Vol. 8, No. 7, July 1990, pp. 1023–1026.
- [33] A.A. Abidi, "Gigahertz transresistance amplifiers in fine line NMOS," *IEEE Journal of Solid-State Circuits*, Vol. SC-19, No. 6, Dec. 1984, pp. 986–994.

- [34] G.F. Williams, "Wide-dynamic-range fiber optic receivers," *IEEE International Solid-State Circuits Conference Digest of Technical Papers*, 1982, pp. 160–161.
- [35] E. Säckinger, W. Guggenbül, "A versatile building block: The CMOS differential difference amplifier," *IEEE Journal of Solid-State Circuits*, Vol. SC-22, No. 2, April 1987, pp. 287–294.
- [36] G.-K. Chang, T.P. Liu, J.L. Gimlett, H. Shirokman, M.Z. Iqbal, J.R. Hayes, K.C. Wang, "A direct-current coupled, all-differential optical receiver for high-bit-rate Sonet systems," *IEEE Photonics Technology Letters*, Vol. 4, No. 4, Apr. 1992, pp. 384–386.
- [37] S.J. Daubert, D. Vallancourt, Y.P. Tsvividis, "Current copier cells," *Electronics Letters*, Vol. 24, No. 25, Dec. 8, 1988, pp. 1560–1562.
- [38] R. Poujois, B. Baylac, D. Barbier, J.M. Ittel, "Low-level MOS transistor amplifier using storage techniques," *IEEE International Solid-State Circuits Conference Digest of Technical Papers*, 1973, pp. 152–153.
- [39] K.R. Stafford, R.A. Blanchard, and P.R. Gray, "A completely monolithic sample/hold amplifier using compatible bipolar and silicon-gate FET devices," *IEEE Journal of Solid-State Circuits*, Vol. SC-9, No. 6, Dec. 1974, pp. 381–387.
- [40] L.A. Glasser, D.W. Dobberpuhl, *The Design and Analysis of VLSI Circuits*, Reading, MA: Addison-Wesley, 1985.
- [41] N. Weste, K. Eshraghian, *Principles of CMOS VLSI Design*, Reading, MA: Addison-Wesley, 1985.
- [42] M.H. White, D.R. Lampe, F.C. Blaha, I.A. Mack, "Characterization of surface channel CCD image arrays at low light levels," *IEEE Journal of Solid-State Circuits*, Vol. SC-9, No. 1, Feb. 1974, pp. 1–13.
- [43] R.W. Broderon, S.P. Emmons, "Noise in buried channel charge-coupled devices," *IEEE Journal of Solid-State Circuits*, Vol. SC-11, No. 1, Feb. 1976, pp. 147–155.
- [44] W.B. Wilson, H.Z. Massoud, E.J. Swanson, R.T. George, Jr., R.B. Fair, "Measurement and modeling of charge feedthrough in n-channel MOS analog switches," *IEEE Journal of Solid-State Circuits*, Vol. SC-20, No. 6, Dec. 1985, pp. 1206–1213.

- [45] C. Tomovich, "MOSIS—A gateway to silicon," *IEEE Circuits and Devices Magazine*, Vol. 4, No. 2, Mar. 1988, pp. 22–23. Mosis may be reached by e-mail at mosis@mosis.edu.
- [46] R.N. Mayo, M.H. Arnold, W.S. Scott, D. Stark, G.T. Hamachi, "1990 DECWRL/Livermore Magic Release," *WRL Research Report 90/7*, Digital Equipment Corporation Western Research Laboratory, Sept. 1990. Magic is available by anonymous ftp from [gatekeeper.dec.com](ftp://gatekeeper.dec.com).
- [47] *HSPICE User's Manual*, Version H92, Campbell, CA: Meta-Software, 1992.
- [48] H.W. Ott, *Noise Reduction Techniques in Electronic Systems*, 2nd. ed., New York: Wiley, 1988.
- [49] W. M. Siebert, *Circuits, Signals, and Systems*, Cambridge, MA: MIT Press, 1986.
- [50] R.W. Lucky, J. Salz, E.J. Weldon, Jr., *Principles of Data Communication*, New York: McGraw-Hill, 1968.
- [51] W.B. Davenport, Jr., W.L. Root, *An Introduction to the Theory of Random Signals and Noise*, New York: IEEE Press, 1958.
- [52] V. Radeka, "Time-variant filters for high-rate pulse-amplitude spectrometry," in *Semiconductor Nuclear-Particle Detectors and Circuits*, W.L. Brown, *et al.*, eds., Washington, DC: National Academy of Sciences, 1969, pp. 553–569.
- [53] K. Kandiah, "Reduction of noise in nuclear-particle spectrometers by using switched time constants and gated integrators," in *Semiconductor Nuclear-Particle Detectors and Circuits*, W.L. Brown, *et al.*, eds., Washington, DC: National Academy of Sciences, 1969, pp. 545–552.
- [54] V. Radeka, "Optimum signal-processing for pulse-amplitude spectrometry in the presence of high-rate effects and noise," *IEEE Transactions on Nuclear Science*, Vol. NS-15, 1968, pp. 455–470.
- [55] F.S. Goulding, "Pulse-shaping in low-noise nuclear amplifiers: a physical approach to noise analysis," *Nuclear Instruments and Methods*, Vol. 100, 1972, pp. 493–504.

- [56] P.R. Gray, R.G. Meyer, *Analysis and Design of Analog Integrated Circuits*, 3rd ed., New York: Wiley, 1993.
- [57] S. Wolfram, *Mathematica: A System for Doing Mathematics by Computer*, 2nd ed., Reading, MA: Addison-Wesley, 1991.
- [58] R.G. Swartz and Y. Ota, "Electronics for high speed, burst mode optical communications," *International Journal of High Speed Electronics*, Vol. 1, Nos. 3 & 4, 1990, pp. 223–243.
- [59] Y. Ota, R.G. Swartz, V.D. Archer, III, "DC-1Gb/s burst-mode compatible receiver for optical bus applications," *IEEE Journal of Lightwave Technology*, Vol. 10, No. 2, Feb. 1992, pp. 244–249.
- [60] Y. Ota, R.G. Swartz, "Burst-mode compatible optical receiver with a large dynamic range," *IEEE Journal of Lightwave Technology*, Vol. 8, No. 12, Dec. 1990, pp. 1897–1903.
- [61] G.L. Turin, "An introduction to matched filters," *IRE Transactions on Information Theory*, June 1960, pp. 311–329.
- [62] B. Enning, E. Bødtker, G. Jacobsen, B. Jensen, "Design and test of novel integrate and dump filter (I&D) for optical Gbit/s system applications," *Electronics Letters*, Vol. 27, No. 24, Nov. 21, 1991, pp. 2286–2288.
- [63] J.E. Goell, "Input amplifiers for optical PCM receivers," *Bell System Technical Journal*, Vol. 53, No. 9, Nov. 1974, pp. 1771–1793.
- [64] K. Ogawa, B. Owen, H.J. Boll, "A long wavelength optical receiver using a short channel Si-MOSFET," *Bell System Technical Journal*, Vol. 62, No. 5, Part 1, pp. 1181–1188.

Appendix A

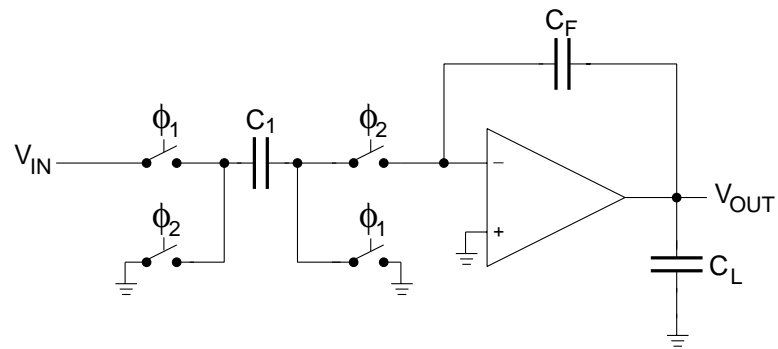
A HIGH-SLEW INTEGRATOR FOR SWITCHED-CAPACITOR CIRCUITS[†]

A.1 INTRODUCTION

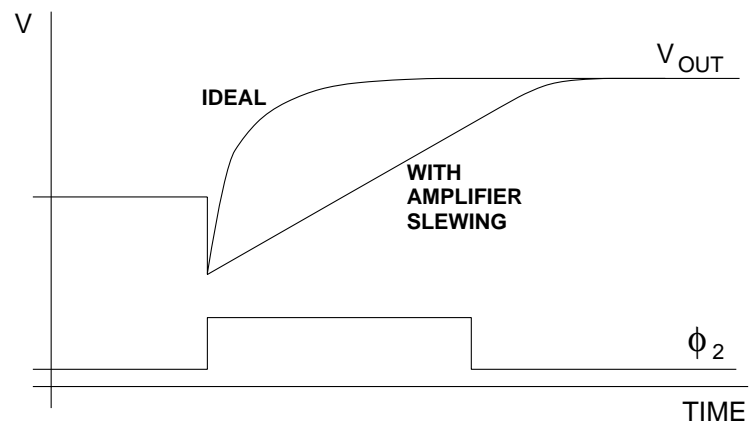
A common requirement in switched-capacitor circuits is a fast settling time on the amplifiers. In addition, the amplifier in such a circuit may need to drive a large capacitive load in order to reduce the effects of thermal (kT/C) noise. A direct result of these requirements is that a large output drive current is necessary on the amplifier in order to quickly change the voltage on the large capacitive loads and subsequently settle to the final value. Often, the amplifier slew rate will severely limit the settling time of the circuit. Fig. A.1a shows a typical non-inverting switched-capacitor integrator with non-overlapping clocks ϕ_1 and ϕ_2 . The worst-case capacitive load on the amplifier occurs during ϕ_2 and depends predominantly on the input capacitor C_1 , plus parasitics and loading effects of the next stage (collectively represented by C_L). The effect of the slewing is shown in Fig. A.1b, where the limited slew rate of the amplifier prevents the output voltage V_{OUT} from settling before the end of the ϕ_2 clock cycle.

Unfortunately, increasing the amplifier slew rate is often at odds with requirements for low power dissipation. The simplest method for increasing an amplifier slew rate is to increase the quiescent current in the input stage, thereby directly increasing the standby current in the entire amplifier. Other methods of enhancing the slew rate have involved

[†]A.E. Stevens, G.A. Miller, "A high-slew integrator for switched-capacitor circuits," *IEEE Journal of Solid-State Circuits*, Vol. 29, No. 9, Sept. 1994, pp. 1146–1149.



(a)



(b)

Figure A.1. (a) Switched-capacitor integrator. (b) Effect of slew rate on integrator transient response. If the slew rate is too low, the integrator will not settle before the end of the clock cycle.

dynamically boosting the input stage current [A.1,A.2] or using a class AB input stage [A.3], though both methods can cause voltage swing problems inside the amplifier due to the large peak currents. Another method boosts the output stage current by monitoring the amplifier differential input voltage [A.4], but is limited by the frequency response of the booster circuit.

In this paper, a new method is introduced which boosts the slew rate of a non-inverting switched-capacitor integrator. Since the integrator is an important building block in most switched-capacitor circuits, a complete high-slew integrator may be used to provide fast settling in lieu of using a high-slew amplifier. The new method boosts the current only at the output of the amplifier, and the extra current is only provided on an as-needed basis, thereby adding only a small amount to the total standby current of the circuit.

A.2 DESCRIPTION OF BOOSTED INTEGRATOR

A block diagram of the boosted integrator is shown in Fig. A.2a. This diagram shows a switched-capacitor integrator with an added external boost circuit which provides a dynamic current boost at the output of the amplifier. This is accomplished by measuring the input voltage and then injecting a proportionate amount of charge at the output of the integrator. The booster operates completely open loop with respect to the amplifier and reduces the amplifier to an error correction role. Thus, the requirements on the amplifier are relaxed because the boost stage does the bulk of the work. For example, if the booster circuit is 90% accurate, then the amplifier has to only settle the final 10% of the final value.

The booster circuit (Fig. A.2b) contains four parts: a sampling capacitor C_2 , a charge amplifier ($M1-M4$), a pair of current mirrors, and a shut-off switch. This sampling capacitor is ten times smaller than the sampling capacitor C_1 at the integrator amplifier input, which reduces the added input capacitance presented by the booster. The sampling capacitor is connected to the input devices $M1$ and $M3$ of the charge amplifier, whose input is at a virtual ground due to the biasing scheme (the sources of biasing transistors $M2$ and $M4$ are grounded). The current mirrors copy the charge amplifier output current to the booster output and also provide a gain of ten. The “DONE” switch is open during the quiescent state and prevents any undesirable loading effects on the integrator amplifier. The actual implementation of the “DONE” switch will be explained later.

A brief summary of the operation of the circuit is as follows: During the ϕ_1 clock phase,

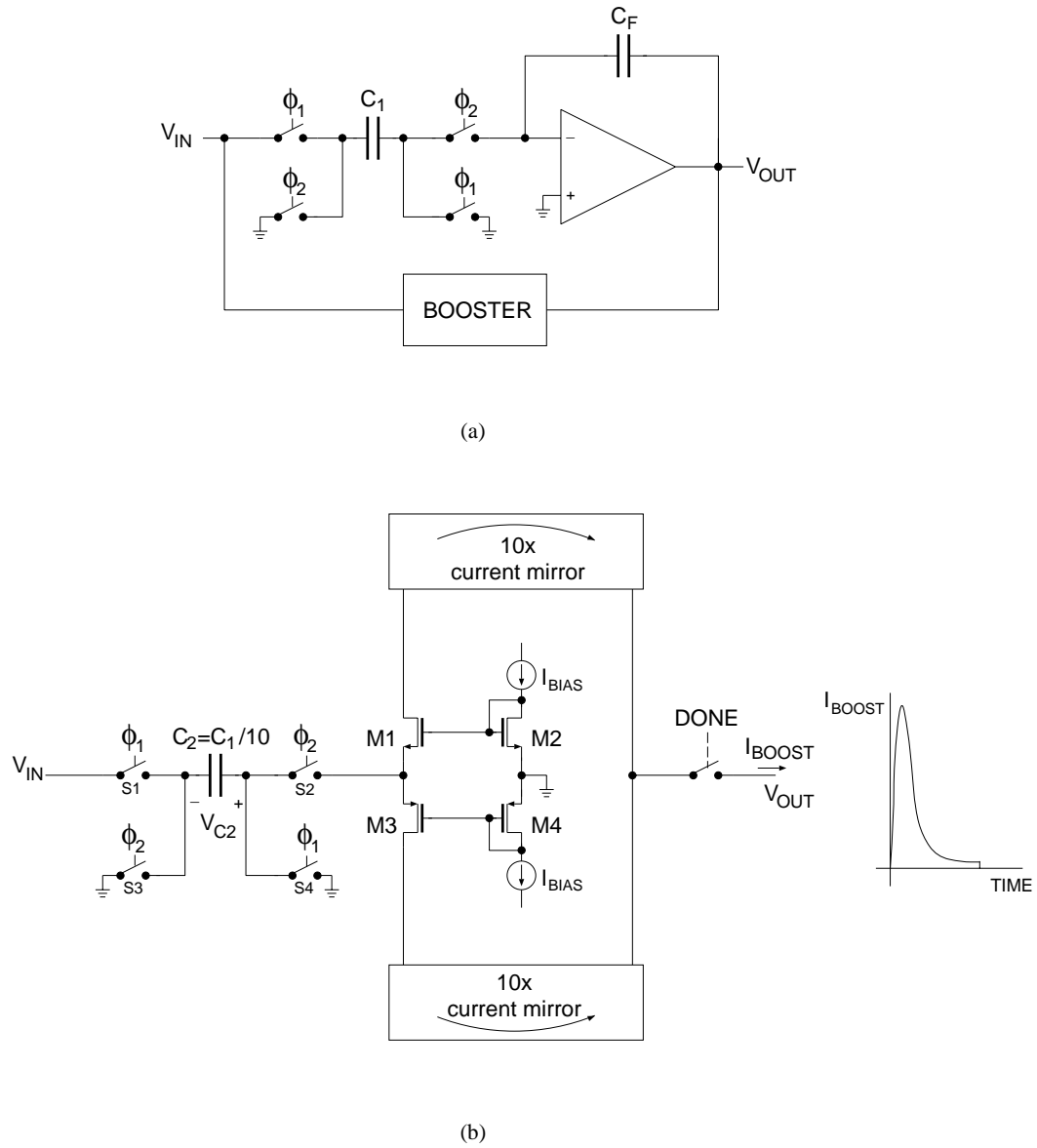


Figure A.2. (a) Slew booster is located external to integrator. (b) Schematic diagram of booster.

the input voltage V_{IN} is sampled onto capacitor C_2 . At the beginning of ϕ_2 , the “DONE” switch is closed. Assuming a positive input voltage, the voltage presented to the charge amplifier during ϕ_2 will be negative (because the input voltage is inverted by the switching of C_2). Thus, the source of $M1$ will be pulled down, $M3$ will turn off, and a large current will flow through $M1$. This current is mirrored and amplified by the top current mirror and then is dumped directly on the output of the integrator. This current will have some initial peak value, and will then continually decrease as C_2 is discharged and V_{GS1} drops. When the discharge of C_2 is complete, the circuit disconnects itself from the amplifier by opening the “DONE” switch. Note that depending on the polarity of V_{IN} , only half of the circuit (top or bottom) will operate during boosting.

A time-domain solution for the current through $M1$ may be found assuming a positive input voltage V_{IN} so that only the top half of the circuit operates (a corresponding solution for the current through $M3$ may be found for negative V_{IN}). Thus, a negative voltage is presented at the source of $M1$ during ϕ_2 via the negatively-charged C_2 . Since the source of $M1$ is fed directly to C_2 , the capacitor current may be equated to the drain current

$$I_{D1} = C_2 \frac{dV_{C2}}{dt} = \frac{\beta_1}{2} (V_{GS2} - V_{C2} - V_{Tn})^2 \quad (\text{A.1})$$

where I_{D1} is the drain current of $M1$, V_{C2} is the voltage across C_2 , $\beta_1 = (\mu C_{ox})(W/L)$ is the geometry factor of $M1$, V_{GS2} is the quiescent gate-source voltage of $M2$, and V_{Tn} is the threshold voltage of the n-channel device. Separating the variables and integrating yields

$$I_{D1}(t) = \frac{I_{D1,peak}}{\left(1 + \frac{g_{m1,peak}}{2C_2} t\right)^2} \quad (\text{A.2})$$

where $I_{D1,peak} = (\beta_1/2)(V_{GS2} + V_{IN} - V_{Tn})^2$ and $g_{m1,peak} = \beta_1(V_{GS2} + V_{IN} - V_{Tn})$ are the peak values for the drain current and the transconductance of $M1$ (at the beginning of the ϕ_2 clock phase). Equation 2 is an exact solution for the drain current of $M1$, neglecting drain-source resistances and assuming that only the top half of the total circuit is working. As C_2 reaches the very end of its discharge, the charge amplifier will return to its symmetrical quiescent state as $M3$ turns back on and the remaining charge on C_2 is bled off. The boost current I_{BOOST} will approximate (2), limited by the finite frequency response of the current mirror and amplified by the current mirror gain.

For design purposes, it is useful to approximate the transient performance of the booster circuit with a single RC time constant τ_{BOOST} determined by the resistance seen by C_2 .

By inspection, this resistance consists of the resistances of switches $S2$ and $S3$ (r_{s2} and r_{s3} , respectively), plus the resistance looking into the charge amplifier ($1/g_{m1}$ or $1/g_{m3}$, depending on which half of the circuit is working). Thus, a useful limit on the booster time constant τ_{BOOST} can be represented by

$$\tau_{BOOST} < C_2 (r_{s2} + r_{s3} + 1/g_{m, \text{worst case}}) \quad (\text{A.3})$$

where $1/g_{m, \text{worst case}}$ is the worst-case resistance looking into the input transistors $M1$ and $M3$ of the charge amplifier. The worst-case (largest) value for $1/g_{m1}$ will occur at minimum I_{D1} or I_{D3} , which, by inspection of (2), occurs at the end of the boost cycle (maximum t). Thus, $1/g_{m, \text{worst case}} = \max(1/g_{m1, \text{quiescent}}, 1/g_{m3, \text{quiescent}})$.

The control for the “DONE” switch is generated by sensing an “off” condition at either current mirror. Thus, if either current mirror is conducting zero current (indicating a half circuit shutdown during a boost), the DONE switch is closed and the boost current is conducted to the output of the integrator. When the boost is complete, both mirrors conduct a small standby current, and the DONE is switch opened, thereby disconnecting the booster from the integrator. The integrator amplifier then completes the charge transfer to high accuracy. In addition, since the booster is separated from the integrator by open switches, it does not affect the noise performance and stability of the complete integrator.

A.3 CIRCUIT DESIGN

A complete design for the booster circuit is shown in Fig. A.3. The charge amplifier is formed by $M1$ – $M4$. The two current mirrors are formed by $M5$ – $M10$ and $M11$ – $M16$. These cascode current mirrors contain a source follower between the input and output in order to extend the output swing [A.5]. The ratio between sampling capacitors C_1/C_2 is 8.5 (instead of ten) in order to ease the layout.

Transistors $M17$ – $M24$, switches $S5$ – $S8$, and the four logic gates form a shutoff circuit in order to perform the “DONE” function. Although conceptually modeled by a single switch in Fig. A.2b, this circuit actually works by operating four switches inside the current mirrors, thus avoiding any IR drop in the “DONE” switch. The quiescent power is also reduced because the output transistors $M9$ – $M10$ and $M15$ – $M16$ are off during the quiescent state. The circuit works as follows: both legs of the shutoff circuit (transistors $M17$ – $M20$ and $M21$ – $M24$) nominally carry a current of 20 μA . However, transistors $M19$ – $M20$ and

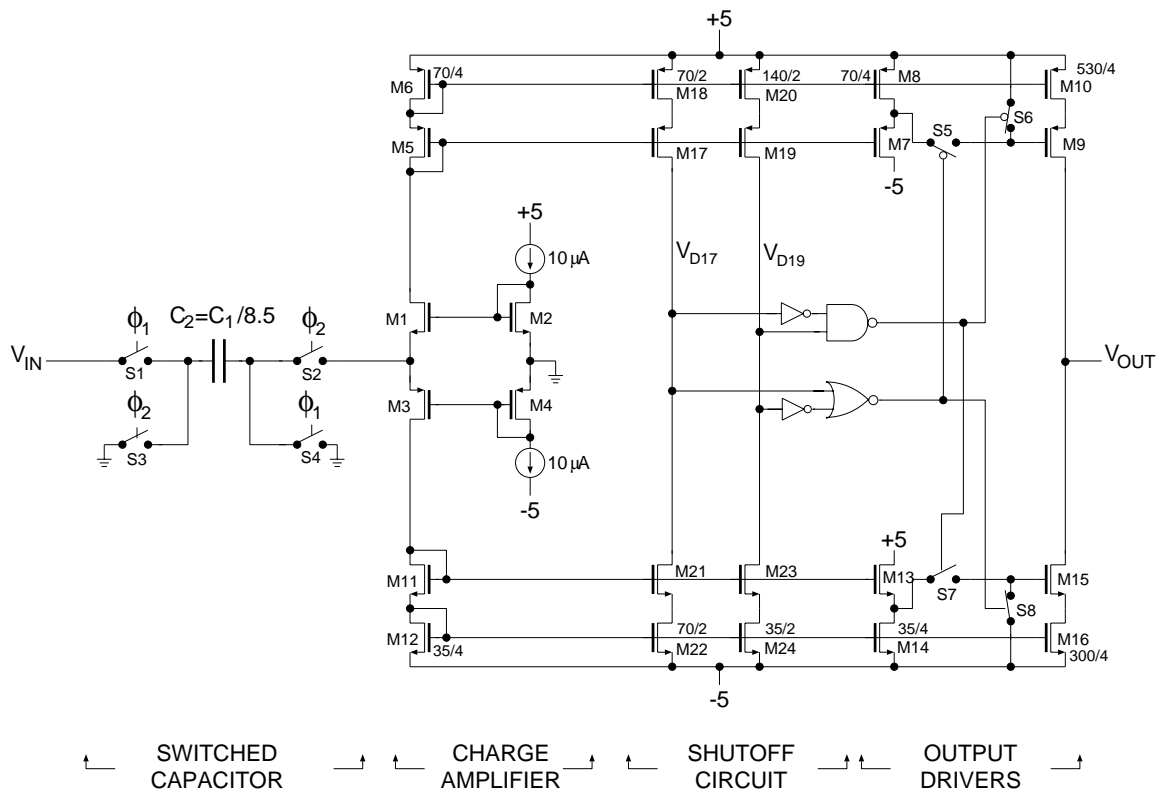


Figure A.3. Circuit diagram of booster circuit.

M_{21} – M_{22} are sized to carry twice as much current as their counterparts on the opposite rail. Thus, in the quiescent state, $V_{D17} = -5$ and $V_{D19} = +5$. The outputs of the logic gates then drive switches $S5$ – $S8$ in order to enable/disable the booster output transistors. In Fig. A.3, the switches are shown in their quiescent position, thereby turning off transistors $M9$ and $M15$ and effectively disconnecting the booster from the integrator. During booster operation, half of the circuit (top or bottom) will turn off, and the remaining $20\text{-}\mu\text{A}$ current source will charge its output node to its own rail. This change in the logic level (either V_{D17} or V_{D19} will change levels, depending on whether the top or bottom half shuts off) will subsequently close $S5$ and $S7$ and open $S6$ and $S8$, which will turn on mirror outputs $M9$ – $M10$ and $M15$ – $M16$ and thus connect the booster to the integrator. When boosting is complete, both halves of the circuit will be on, and the switches will return to their quiescent state.

The implementation of Fig. A.3 has several limitations due to the speed of the shutoff circuit and the finite bandwidth of the current mirrors. In particular, if the shutoff circuit is late turning on the output devices at the beginning of the clock cycle, a significant amount of charge may be lost at the output of the booster. The booster is particularly sensitive to this effect because the peak boost current occurs at the beginning of the cycle. Similarly, bandwidth limitations in the current mirror may cause the shutoff circuit to activate before the charge at the output is completely transferred. These effects were accounted for by adjusting the sizes of the output devices of the current mirrors in an attempt to optimize the overall performance. This was done at the simulation level. In addition, performance differences between the n-channel and p-channel transistors led to different gains in the two current mirrors (7.6X and 8.6X).

For small input voltages, the offset mismatch between the integrator amplifier and the booster may cause the booster to overcompensate and subsequently provide too much boost current. However, any added settling time due to overcompensation has no appreciable effect when compared to the longer settling time for larger input voltages (i.e., at input voltage magnitudes much larger than the offset voltage).

A.4 TEST RESULTS

Characteristics of the test chip are shown in Table A.1. The integrator amplifier is a single-stage folded-cascode design. For comparison purposes, the test chip contains an

TABLE A.1. CIRCUIT CHARACTERISTICS

Power supplies	± 5 V
Amplifier 3 dB bandwidth	3.6 MHz
Amplifier phase margin	72°
Amplifier slew rate	1.8 V/ μ s
Booster slew rate	14 V/ μ s
Value of C_1	85 pF
Value of C_F	155 pF
Value of C_2	10 pF
Die area	$2200 \mu\text{m} \times 2250 \mu\text{m}$
Technology	$2 \mu\text{m}$ CMOS double-poly

TABLE A.2. SUMMARY OF RESULTS

	unboosted	boosted
Settling time of integrator (1% accuracy, $V_{IN} = -3.5$ V)	2100 ns	600 ns
Settling time of integrator (1% accuracy, $V_{IN} = +3.5$ V)	2000 ns	750 ns
Die area of integrator	1.35 mm^2	1.65 mm^2
Static power of integrator	5.5 mW	7.5 mW

identical integrator without the boost circuit. As seen in Fig. A.4, the booster adds about 22% more area to the total integrator as compared to the unboosted integrator.

Fig. A.5 shows the performance of the unboosted and boosted integrators. In the unboosted case, amplifier slewing at 1.8 V/ μ s dominates the settling time. In the boosted case, the slew rate is increased to 14 V/ μ s during the boost (7.8X improvement), but the booster only provides about 70% of the total necessary charge. Thus, the amplifier must settle the final 30%, which includes a short period of slewing at 1.8 V/ μ s. We believe that this performance could be improved by a better design of the “DONE” shutoff circuit in the booster. Table A.2 shows the settling times to 1% accuracy.

By comparison, if the $200 \mu\text{A}$ idle current in the booster was instead applied directly to the amplifier, then the amplifier tail would be increased by roughly $100 \mu\text{A}$ (for a folded-cascode amplifier). This would only increase the amplifier slew rate by approximately 50% in an unboosted integrator configuration.

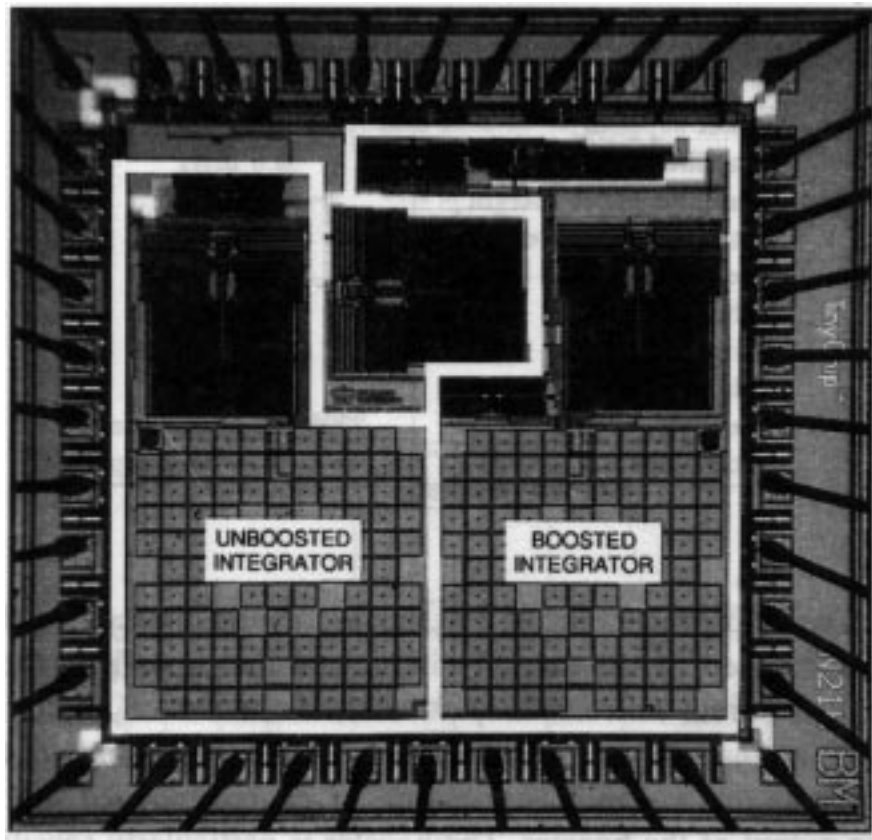


Figure A.4. Die photograph. The chip contains both boosted and unboosted integrators for comparison purposes.

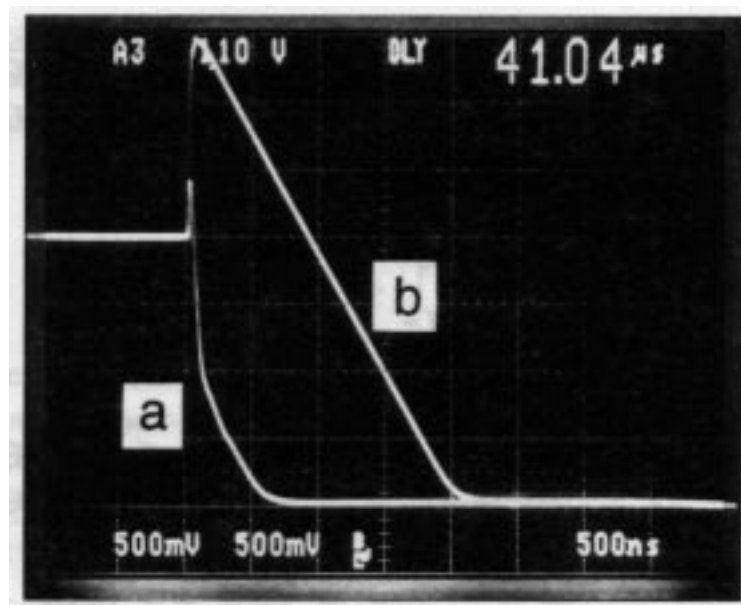


Figure A.5. Oscilloscope photograph showing the settling time of the boosted integrator (*trace a*). An unboosted integrator is shown for comparison (*trace b*). Horizontal scale: 500 ns/div. Vertical scale: 0.5 V/div.

A.5 CONCLUSION

A new method has been introduced for boosting the slew rate of a switched-capacitor integrator by boosting the integrator as a whole, thereby avoiding any redesign of the amplifier. In addition, since the booster circuit is disconnected from the integrator after boosting, there are no added noise or loading effects. The circuit is particularly useful in low-power and low-noise applications such as Σ - Δ modulators, data converters, and switched-capacitor filters.

A.6 REFERENCES

- [A.1] M.G. Degrauwe, J. Rijmenants, E.A. Vittoz, H.J. De Man, "Adaptive biasing CMOS amplifiers," *IEEE J. Solid-State Circuits*, Vol. SC-17, No. 3, pp. 522–528, June 1982.
- [A.2] J. Hosticka, "Dynamic CMOS amplifiers," *IEEE J. Solid-State Circuits*, Vol. SC-15, No. 5, pp. 887–894, Oct. 1980.
- [A.3] R. Castello, P.R. Gray, "A high-performance micropower switched-capacitor filter," *IEEE J. Solid-State Circuits*, Vol. SC-20, No. 6, pp. 1122–1132, Dec. 1985.
- [A.4] K. Nagaraj, "CMOS amplifiers incorporating a novel slew rate enhancement technique," *Proc. IEEE 1990 Custom Integrated Circuits Conference*, pp. 11.6.1–11.6.5, May 1990.
- [A.5] P.R. Gray, R.G. Meyer, *Analysis and Design of Analog Integrated Circuits*, 2nd ed., New York: Wiley, 1984.