

Copyright © 1993, by the author(s).
All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

**A PARALLEL ARCHITECTURE FOR
HIGH-DATA-RATE DIGITAL RECEIVERS
IN SCALED CMOS TECHNOLOGY**

by

Timothy Hak-Ting Hu

Memorandum No. UCB/ERL M93/62

26 July 1993

**A PARALLEL ARCHITECTURE FOR
HIGH-DATA-RATE DIGITAL RECEIVERS
IN SCALED CMOS TECHNOLOGY**

Copyright © 1993

by

Timothy Hak-Ting Hu

Memorandum No. UCB/ERL M93/62

26 July 1993

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Abstract

A Parallel Architecture for High-Data-Rate Digital Receivers in Scaled CMOS Technology

by
Timothy Hak-Ting Hu

Doctor of Philosophy
in
Engineering-Electrical Engineering
and Computer Sciences

University of California at Berkeley

Professor Paul R. Gray, Chair

Reduction in cost of terminal electronics is essential in order for fiber optic technology to penetrate LAN and telephony subscriber loop applications extensively. Currently, high-data-rate fiber transceiver electronics are implemented principally with the multichip approach, with bipolar and gallium arsenide technology.

The objective of this research is to develop a new parallel architecture to use the lower cost and higher integration of scaled CMOS technology to address the cost problem in fiber transceivers. The problem is first attacked by integrating the parallel-to-serial conversion, automatic gain control (AGC), decision, and clock recovery functions in CMOS at rates about 500Mb/s. The inherently slower speed of CMOS compared to bipolar and GaAs is compensated through the use of a high degree of parallel signal processing in the signal path.

To verify the main idea about the new parallel architecture proposed in this thesis, an experimental prototype with 8 parallel channels was designed and fab-

ricated in a $1.2\mu\text{m}$ CMOS technology. A bit rate of 480Mb/s is achieved with a minimum peak-to-peak input voltage of $18\text{mV}_{\text{p-p}}$. The area of the chip is 160 mil X 160 mil (4mm X 4mm) and it consumes 900mW of power.

This thesis arrives at three main conclusions. First a minimum device f_T to data rate ratio of 4:1 can be achieved with the parallel architecture enabling CMOS technology to be used for Gb/s digital optical fiber receivers. Second, clock recovery can be done, through a decision directed scheme with two times oversampling, by inserting one or more timing channels in between the parallel data channels. Finally, the area and power consumption of the parallel architecture are comparable to implementations in bipolar and GaAs.



Chairman of Committee

Table of Contents

Chapter 1 Introduction	1
1.1 Background and Motivation	1
1.2 Thesis Organization	2
Chapter 2 Optical Fiber Communication System	3
2.1 Introduction	3
2.2 Transmitter	4
2.2.1 Light Source	4
2.2.1.1 Light Emitting Diode (LED)	5
2.2.1.2 Semiconductor Junction-Diode Laser	6
2.2.1.3 Comparison of LED and Laser	8
2.2.2 Coder (MUX)	10
2.2.3 Driver	11
2.2.3.1 LED Driver	11
2.2.3.2 Semiconductor Laser Driver	13
2.3 Optical Fiber	14
2.3.1 Modes of Propagation	14
2.3.2 Attenuation and Dispersion	15
2.3.2.1 Attenuation.....	16
2.3.2.2 Dispersion.....	16
2.3.3 Single-Mode Fibers	16
2.3.3.1 Single-Mode vs Multimode	16
2.3.3.2 Dispersion in SM Fibers.....	17
2.3.3.3 Attenuation in SM Fibers	18
2.3.4 Erbium-Doped Fiber Amplifiers (EDFAs)	20
2.4 Receiver	21
2.5 Probability of Error and Quantum Limit	23
Chapter 3 Traditional Receiver Architecture	25

3.1 Introduction	25
3.2 Photo-Detectors (PDs)	26
3.2.1 Semiconductor Photodiodes	26
3.2.2 Responsivity and Quantum Efficiency	26
3.2.3 PIN Photodiode	28
3.2.4 Avalanche Photodiode (APD)	30
3.2.5 Photodiode Equivalent Circuit	32
3.2.6 PIN Photodiodes vs APDs	33
3.3 Low Noise Preamplifier	33
3.3.1 Low-impedance Preamplifier	34
3.3.2 High-impedance Preamplifier	35
3.3.3 Transimpedance Preamplifier	37
3.4 Integrated Optoreceiver	40
3.5 Main amplifier	42
3.5.1 Limiting Amplifiers	42
3.5.2 AGC Amplifiers	43
3.6 Clock Recovery	47
3.6.1 Spectral-Line Method	49
3.6.2 Phase-Locked Loop (PLL)	50
3.6.3 Charge-Pump Phase-Locked Loop (CPPLL)	52
3.6.4 Wide-Band Clock Recovery	53
3.6.5 Circuit Requirements	53
3.7 Decision Circuit	54
3.8 Demultiplexer (DEMUX)	55
3.9 Limitations of Traditional Architecture	56
Chapter 4 Parallel Receiver Architecture	58
4.1 Introduction	58
4.2 Analog DEMUX	59
4.3 Multi-Phase Clock	60
4.4 Parallel Channel	61
4.5 Clock Recovery	62
4.6 Improvement in Performance	63
Chapter 5 Parallel Receiver Implementation	65
5.1 Introduction	65
5.1.1 Input Demultiplexing	67

5.1.1.1 CMOS Sample-and-Hold	67
5.1.1.2 Bandwidth Related Error.....	67
5.1.1.3 Accuracy Related Error.....	69
5.1.1.4 Bandwidth vs Accuracy Trade-off	70
5.1.1.5 Circuit Implementation	71
5.1.2 Multiple Phase Clock Edges	72
5.1.3 Channel Amplifier	73
5.1.3.1 MOS Variable Gain Amplifier.....	73
5.1.3.2 DC and AC Analysis	74
5.1.3.3 Channel Amplifier with Input DC Offset Cancellation	77
5.1.3.4 Distortion Analysis	78
5.1.4 Clock Recovery	82
5.1.4.1 Designing a Charge-Pump Phase-Lock Loop	82
5.1.4.2 Inductive Clock Recovery	85
5.1.4.3 Decision Directed Phase Detection	87
5.1.4.4 Phase Detector Implementation	88
5.1.4.5 Ring Oscillator Implementation.....	89
5.1.4.6 Charge-Pump PLL Loop Dynamics	91
5.1.4.7 Initial Acquisition.....	94
5.1.4.8 Effect of DC offset and Pulse Distortion	95
Chapter 6 Experimental Results	97
6.1 Experimental prototype	97
6.2 Voltage Control Oscillator	97
6.3 AGC Biasing	100
6.4 Bit Error Rate Measurements	102
6.4.1 Minimum Input Voltage vs Bit Rate	105
6.4.2 BER vs Input Voltage	106
6.4.3 Waveform Dependence of BER	107
6.4.4 Channel Dependence of BER	109
6.4.5 Supply Dependence of BER	109
6.5 Jitter Performance	111
6.6 PLL Performance	115
6.7 Performance Summary	115
Chapter 7 Conclusion	118
7.1 Summary of Research Results	118

7.1.1 Traditional Architecture vs Parallel Architecture	118
7.1.2 Special Issues in Designing a Parallel Receiver	121
7.2 Projected performance in Scaled Technologies	121
7.3 Future Work	122
References	124

Acknowledgments

I wish to express my deepest gratitude and appreciation to Professor Paul R. Gray, my research advisor, for his continuous support and encouragement throughout the course of my Ph.D. study. His vast knowledge and insight in the field of analog integrated circuit design have never ceased to amaze and enlighten me. It has been my most rewarding experience working under him.

I would also like to thank Professor Robert G. Meyer, and Professor Donald O. Pederson for the helpful discussions and suggestions during these years. The help of Professor Joseph M. Kahn in giving suggestions and loaning the equipments for testing the prototype is also gratefully acknowledged.

The graduate students in the analog integrated circuit group made these years of research a lot more enjoyable and memorable, especially the "inner cubicle gang" - Ken Nishimura, Gregory Uehara, and Weijie Yun. Ken provided numerous technical and CAD supports. Greg gave so much moral support and spent a lot of his time coaching me for my talks. Weijie helped broadening my knowledge in integrated circuit processing and improving my tennis. Discussions with other colleagues - Gani Jusuf, Robert Neff, Cormac Conroy, and David Cline, was always fruitful and enlightening.

Last but not least, I thank my parents, Annie and Hung-Nyie Hu, for their support and patience. Their strong belief in me was an important driving force that

carried me through the tough times.

**The research was supported by the National Science Foundation grant
MIP-9101525, MICRO, Texas Instruments, and Level One Communications.**

Chapter 1 Introduction

1.1 Background and Motivation

The first optical communication system was probably the "photophone", a pattern awarded to Alexander Graham Bell in 1880 [1]. It was a system for transmitting voice over a few hundred meters by modulating reflected sunlight with a vibrating mirror and the receiver was simply a photocell. It was never a commercial success. The breakthrough didn't come until the invention of laser in 1960. With continuous research and development of lasers, photodetectors and materials for optical transmission, the bandwidth for communication is gradually increasing and the attenuation over distance is decreasing. Optical fiber communication is already a major technical and commercial success.

The field of lightwave datacommunication has experienced explosive growth in the past ten years, both in terms of its commercial importance and in terms of the research effort devoted to it in academic and industrial laboratories. High speed integrated circuits (IC) have played a key role in this phenomenal growth because they are essential to interface with the high speed light sources and photodetectors at both ends of a fiber span [2].

A major problem limiting the deployment of optical fiber communication systems in the range of 500Mb/s and above is the cost of the optical components and the terminal electronics attributed to the high cost technology, package and assembly required. Electronic interfaces in general involve complex, low noise linear and nonlinear elements for amplification, threshold detection, and timing extraction. High speed electronic circuits have been among the bottlenecks in the realization of practical high-data-rate systems.

The application of VLSI technology to communication systems has played an important role in reducing the cost of many types of communication systems in use today and will play an increasingly important role in the future. The objective of this research is to develop a new architecture to relax the speed requirement imposed on the technology for implementation, so that the lower cost and higher integration of scaled CMOS technology can be used to address the cost problem in the implementation of high-data-rate terminal electronics.

1.2 Thesis Organization

An examination of the use of demultiplexed sampling and parallel processing in analog signal data path for a CMOS high-data-rate receiver is presented in this thesis. The three basic building blocks of optical fiber communication systems are the transmitter, the fiber, and the receiver. Chapter 2 reviews the transmitter and fiber in detail, and give a brief description of the receiver. The building blocks in the traditional architecture of a fiber optic receiver is described in detail in chapter 3, together with their circuit implementations. The disadvantages and limitations of traditional receiver architecture are pointed out. In Chapter 4, a new parallel receiver architecture is proposed to relax the speed requirement imposed on the technology for implementation. The key building blocks involved in this new parallel receiver architecture is also discussed in the same chapter. In Chapter 5, the design of a prototype implementing the parallel architecture is described in detail with the circuit requirements and circuit solutions. Experimental results from a prototype implemented in a 1.2- μm CMOS technology are presented in chapter 6. Chapter 7 summarizes the research results, and conclusion is drawn with a discussion of further improvement and future work.

Chapter 2 Optical Fiber Communication System

2.1 Introduction

The development of the laser and the optical fiber has brought about a revolution in communication system design. In the 1960s, the laser evolved from a laboratory curiosity to become a versatile and widely applied family of devices and systems. In the 1970s, the optical fiber was developed from an idea with some promise to be a proven communication channel capable of carrying high data rates with low attenuation over distances far exceeding those used in coaxial and microwave systems. In the 1980s, optical fiber communication systems technologies are achieving a broad and still expanding range of successful commercial applications and are expanding the horizons of optical fiber systems capabilities.

An optical fiber communication system consists of three principle parts: the transmitter, the fiber, and the receiver, as shown in Figure 2-1. When binary information is transmitted, one format may be converted to another; e.g. Non-Return-to-Zero (NRZ) data may be converted to Return-to-Zero (RZ) or M-tary patterns, or the line rate may be raised by transmitting extra bits for framing and coding. These modification facilitate demultiplexing, error detection, and clock recovery. For transmitting analog information, the input signal may be predistorted to compensate for non-linearities in the light source, or digitally encoded to subcarrier frequency modulated to circumvent non-linearities. A good summary of coding and analog modulation techniques can be found in [3] [4]. This chapter will concentrate on direct intensity modulation of the light source for transmission, as it is employed in most practical high-data-rate systems.

The following sections will give a brief description for each of the three main parts, and

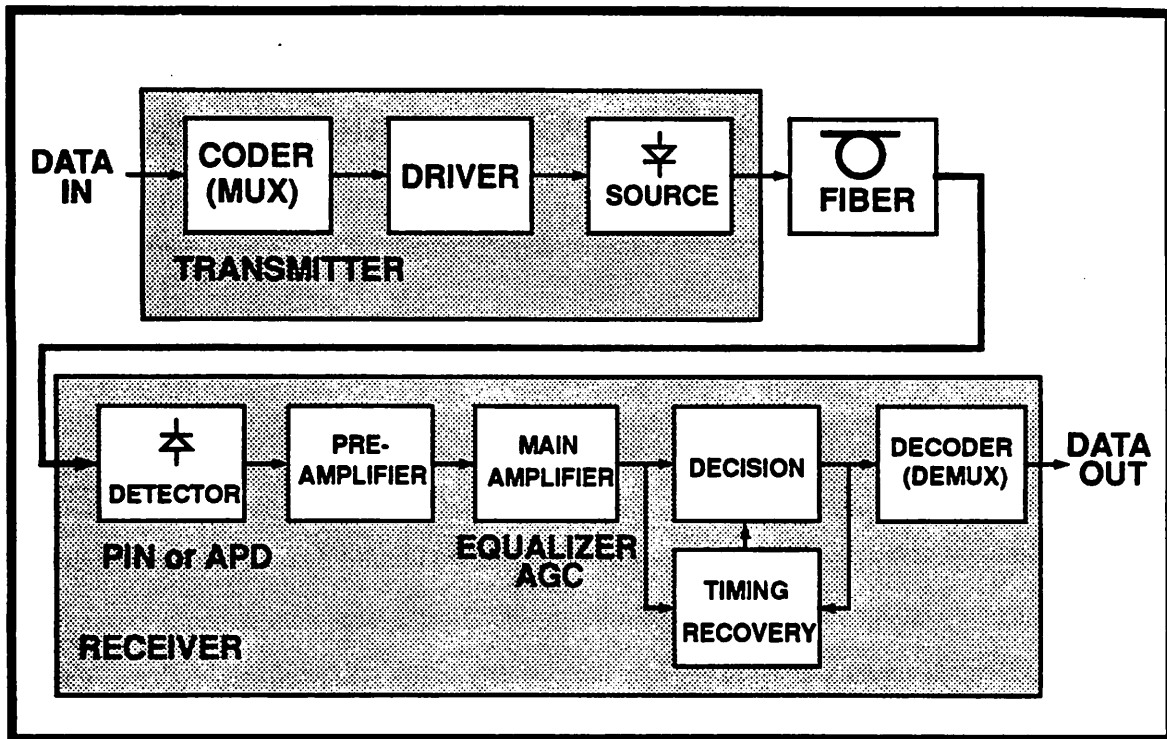


Figure 2-1 Block Diagram of a Basic Optical Fiber Communication System

the limitations that each part imposes on the total system performance. Receiver Implementation is the main focus in this thesis and will be described in detail in Chapter 3.

2.2 Transmitter

A digital transmitter converts the electrical signals into corresponding light-intensity envelopes. Such direct modulation is intended to affect only the average optical power. Any phase or frequency information incidently imparted to the optical carrier itself is not used at the receiver. This information is however used in another class of systems which employ coherent transmission [5].

The main components in a transmitter are a coder (with multiplexing function), a device driver, and a light source.

2.2.1 Light Source

While optical fiber transmission uses light energy to carry the information, at the present

state of the art, the signals are generated and manipulated electrically. This implies an electrical--to-optical conversion at the input to the fiber medium is needed. Two semiconductor devices, light-emitting diodes (LEDs) and semiconductor lasers, are suitable for use in terms of device dimensions, speed, efficiency, electrical characteristics and reliability.

2.2.1.1 Light Emitting Diode (LED)

Optical power of a LED is produced in a forward-biased p-n junction diode by the radiative recombination of holes and electrons. The power is proportional to I , the input current, and hf , the energy of the photon, where h is the Planck constant and f the frequency of light emitted. The constant of proportionality is the quantum efficiency η .

$$P = \eta \frac{I}{q} hf \quad (\text{EQ 2-1})$$

Because the relationship between P and I is linear, the LED can be intensity modulated by modulating the input current. Optical frequency and phase modulation are not feasible when the light source is a LED because the LED output is incoherent; i.e. its spectrum is spread over a range of wavelengths which is large with respect to the bit rate for transmission. The spectrum can be narrowed by using a double-heterostructure instead of a single p-n junction. The double-heterostructure (DH) LED as shown in Figure 2-2, consists of a narrow-bandgap material, constituting the active region, sandwiched between two wide-bandgap materials. The wavelength of the radiation is determined by the narrow-bandgap energy. The DH LED has several advantages over the single p-n junction diode. The recombination region, from which light is emitted, is well defined by the structure of the device. The internal quantum efficiency can be higher than that of the homojunction diode. Because the wide-bandgap material has lower index of refraction, the double heterostructure constitutes a dielectric waveguide that confines the lightwave and results in higher external quantum efficiency. The wide-bandgap material can be made transparent at the wavelength of the lightwave, thus reducing attenuation and further increasing the internal quantum efficiency.

LED devices are available in two basic types, edge-emitting and surface-emitting. The

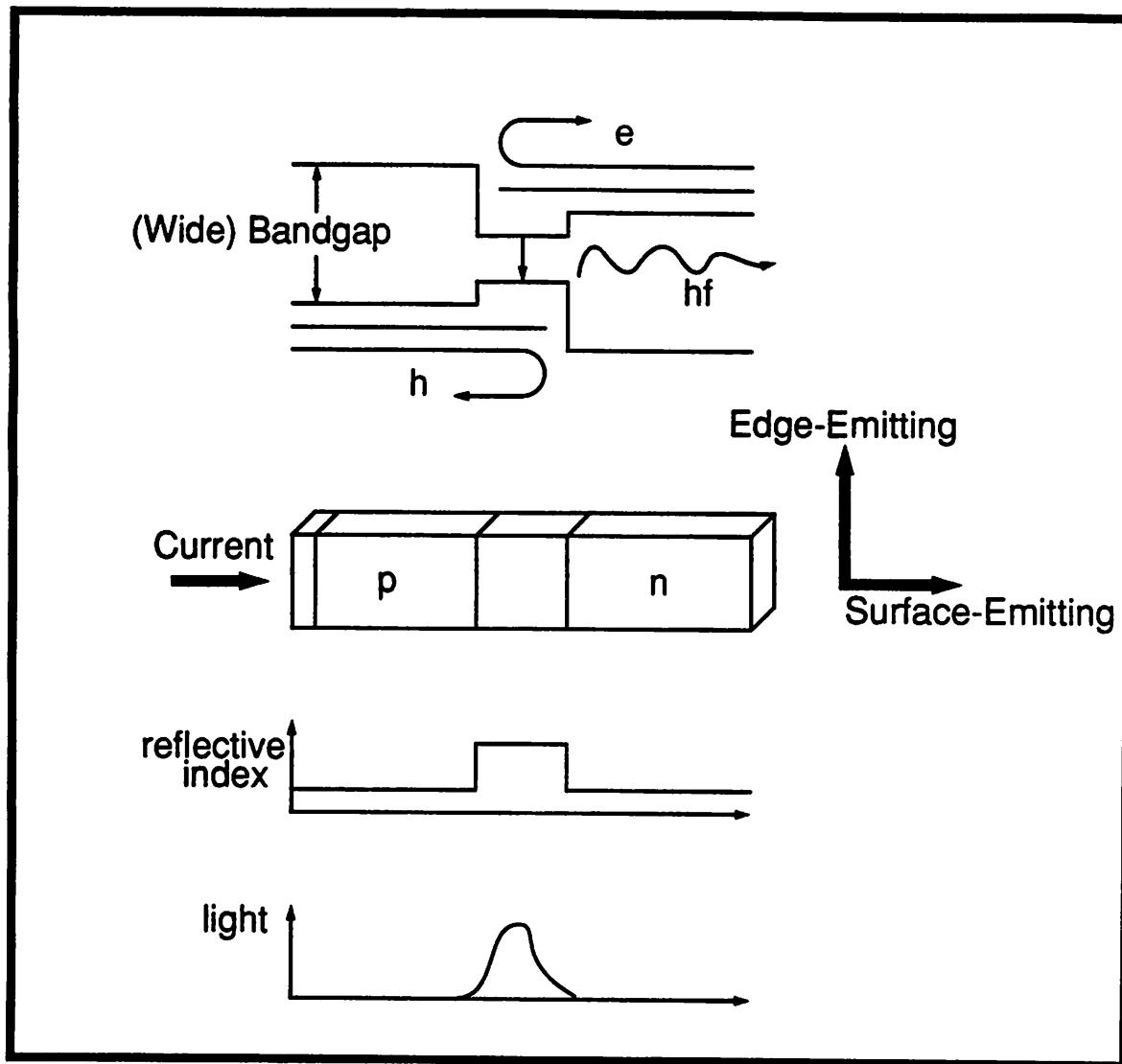


Figure 2-2 Double Heterostructure LED

edge-emitting LED is usually a stripe-contact device. Light generated in the active region propagates parallel to the stripe and is emitted from one end of the active region. Light propagating in other directions is lost. The surface-emitting LED is usually designed to couple light directly into an optical fiber. The surface through which light is emitted is circular, with a diameter similar to that of the fiber. Because the output is taken from a side of larger area, the external quantum efficiency of the surface-emitter can be larger than that of the edge-emitter.

2.2.1.2 Semiconductor Junction-Diode Laser

The semiconductor junction-diode laser is the dominant laser in optical fiber communica-

tion systems. The laser consists of an optical cavity resonator, and an amplifying mechanism compensating for the losses of the passive resonator. In the semiconductor laser, the optical resonator is normally a Fabry-Perot interferometer with two plane mirrors. Amplification in the laser is provided by the stimulated emission of radiation. Electrons at excited energy levels can be stimulated to decay to a lower-energy state by an incident lightwave, emitting a photon in the process. The energy of the photon and the frequency of the lightwave are related by

$$E = hf = \frac{hc}{\lambda} \quad (\text{EQ 2-2})$$

where c is the speed of light.

One of the requirement for lasing is to have the gain, as lightwave propagates in an amplifying medium, larger than the losses in the passive resonator which includes absorption, scattering, and transmission through the mirrors. This requires population inversion of the minority carrier concentration in the medium. A population inversion requires some kind of non-equilibrium condition to create and sustain it. Means for sustaining a population inversion are called pumping. With an adequate pump rate, the population inversion will produce enough gain to cause lasing to begin, and the laser field will increase in strength. Saturation ultimately causes the gain to decrease until the gain and total loss are equal; this is the steady-state condition.

The frequency of oscillation is determined by the resonant frequencies for which the gain is greater than the total losses. There can be several modes of oscillation occurred at the same time. However, single mode operation can be achieved by the integration of wavelength selectivity directly into the semiconductor laser structure. As an example, a distributed-feedback (DFB) laser integrated a grating region in the pumped part of the gain region for wavelength selectivity.

In the semiconductor laser, pumping is done by providing a direct current in the forward direction in the junction diode. The steady state photon density can be expressed as [6]

$$\phi_s = \left(\frac{\tau_{ph}}{qd} \right) (J - J_{th}) \quad (\text{EQ 2-3})$$

where τ_{ph} is the characteristic lifetime of photon, d is the length of the resonator and J_{th} is

the threshold current density for lasing. The power density can then be calculated from the photon density by recognizing that each photon carries energy hf , and are moving with velocity c/n where c is the speed of light and n is the refractive index of the medium.

Modulation of the laser is similar to modulation of the LED. The linear P versus I characteristic provides linear intensity modulation of the laser output powered by modulating the input current. However, even for pulse signals, the biasing current is often set at a value near the threshold current because of the time required for the pulse to reach its peak value can be made shorter if the current and the population of electrons in the conduction band do not have to build up from zero.

2.2.1.3 Comparison of LED and Laser

The LED output is incoherent, therefore only intensity modulation is possible. The wide spectral width of the LED also sets a limit on the bit-rate distance product through material dispersion. The laser output is much more coherent, meaning that it has a single frequency with linewidth small compared with the signal bandwidth, making it the only choice if coherent modulation and demodulation is needed. Shown in Figure 2-3 are some typical output spectra of different kinds of LEDs and semiconductor lasers.

The LED emits light spread over a much larger solid angle than a laser. The poor directivity of the LED implies that a smaller proportion of the power is coupled into the fiber, particularly for a single mode fiber which accepts light only at smaller incident angles. Thus, the laser is necessary for single mode fiber except for short distance where the attenuation is low.

A device driver is needed to control the current flowing through the light source in order to modulate the optical power emitted. LED is characterized by a single pole frequency response and can be modulated at frequencies up to 100's MHz. A semiconductor laser source is much more complicated. The maximum modulation frequency depends on the frequency response and turn-on dynamics of the laser. A semiconductor laser can be modulated up to 10 GHz depending on the structure and bias point.

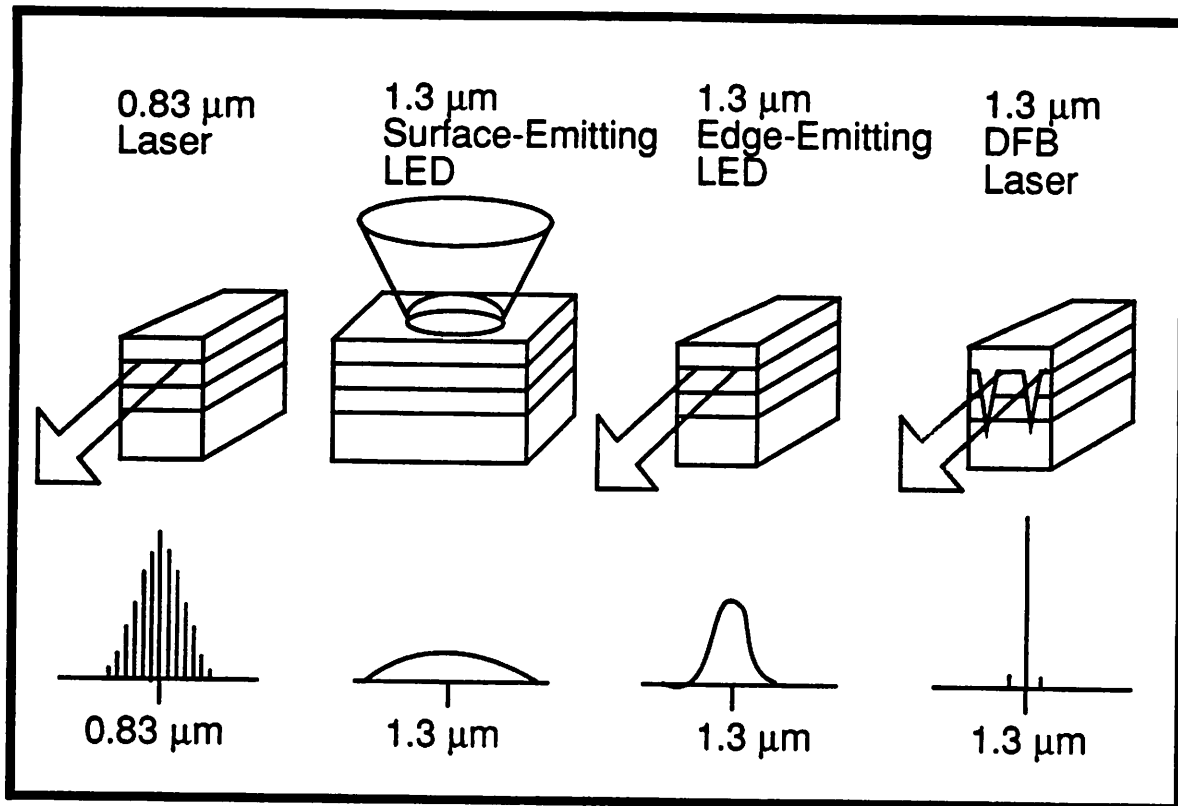


Figure 2-3 Light Sources and Spectra

The maximum power emitted by a LED is about $100\mu\text{W}$ (-10dBm) while the laser can go up to 10mW (10 dBm) with 1mW (0dBm) being typical. However, the light output of the laser is very temperature dependent as shown in Figure 2-4, and hence it is generally necessary to monitor the light output and control the driving current using a feedback circuit.

On the other hand, LED is much cheaper and reliable than semiconductor laser and the driver is much simpler with lower biasing currents.

With modest launched power and speed, and broad spectral width, LED is used for short-distance, medium-bit-rate (50 to 500Mb/s) applications. With high reliability, LED is also used in high-temperature or uncontrolled environments such as for data links within high-speed equipments, in local area networks, or in the telephone subscriber loops.

Laser provides a higher-quality lightwave source than does the light-emitting diode. For many optical communication system applications, the advantages that the laser can offer are valuable. Given the $10\text{-}30\text{dB}$ additional launched power levels compared with LED, higher speeds and

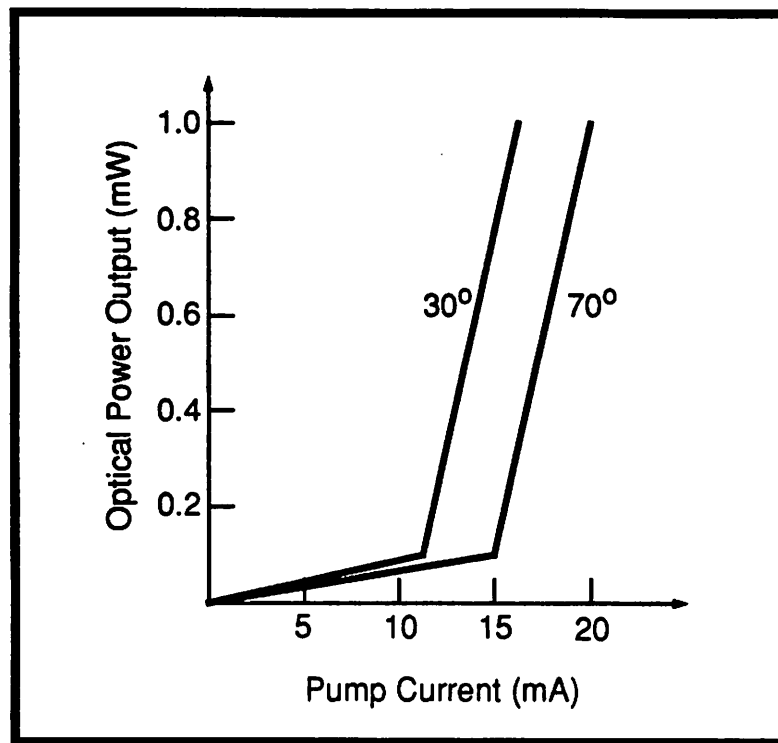


Figure 2-4 Effect of Temperature on Laser Threshold Current

narrower spectral widths, lasers are useful over a significantly wider bit-rate · length region. For the long term, the semiconductor laser is the source of choice, especially when single mode fiber is dominating these days. However the laser is considerably more expensive than the LED and also has poorer reliability.

2.2.2 Coder (MUX)

In digital data communications, the optical fiber has bandwidth much greater than most applications required. Most of the time the input to the driver is a time-division-multiplexed (TDM) version of many input bit streams. e.g. Synchronous Optical NETWORK (SONET) [7]. SONET has basic transmitting rate of 51.84Mb/s. The communication lines are called OC-n where n is an integer and the lines is transmitting at a rate of n times the basic rate. e.g. OC-9 is a line transmitting at 466.56Mb/s which may be a TDM version of three OC-3 signals, each has a bit rate of 155.52Mb/s. High speed multiplexer, for time division multiplexing, with data rates in the range of 30Gb/s has been realized [8] with a record set at 34Gb/s [9].

For some application, scrambling may be required. There are many reasons for scram-

bling the signal. It can be due to security reasons which one does not want to have unauthorized access to the transmitting data. Another reason may be system requirements. Some transmitting and receiving systems require the transmitted signal to be randomized or having 50% duty cycles (no DC content). Redundant bit(s) may be incooperated into the transmitted bit stream for error detection, or control framing for demultiplexing, or control the line spectrum for clock recovery. As a result, sometimes a coder is required at the transmitter. Coding does not require high speed circuitry because it is done in low data rates before being multiplexed up. However because of the added redundancy, it lowered the effective bit rate of the transmitted data.

2.2.3 Driver

The basic function of the driver for digital transmission is to provide a high current pulse to turn the light source on and off. Since there are two types of light source and they have different characteristics, two types of drivers are needed.

2.2.3.1 LED Driver

LEDs are frequently operated with a small forward bias in the "off" state to overcome the turn-on delay associated with the space-charge capacitance of the junction. To this bias, the transmitter adds a high-speed drive current of 25 to 200mA to reach the "on" state. With series resistance of a few ohms (series resistance r_d of a forward biased p-n junction), LED drive current can be provided by simply using a line-driver IC from logic family appropriate for the speed.

A generic block diagram is shown in Figure 2-5. An important characteristic of LEDs is that their optical outputs decrease with temperature exponentially. Therefore, to stabilize the output power, a temperature compensated control is needed to increase the driving current with rising temperature [10]. An alternative way is to stabilize the temperature of the LED is using a Peltier active-cooling device. Such a thermoelectric cooler (TEC) is usually integral to the LED package.

For high-data-rate operation with bit rate above 100Mb/s, special compensation or speed up networks are required. The problem is that a high current pulse must be generated which drives a series connection of a low-ohmic load and a relatively large parasitic inductance mainly caused

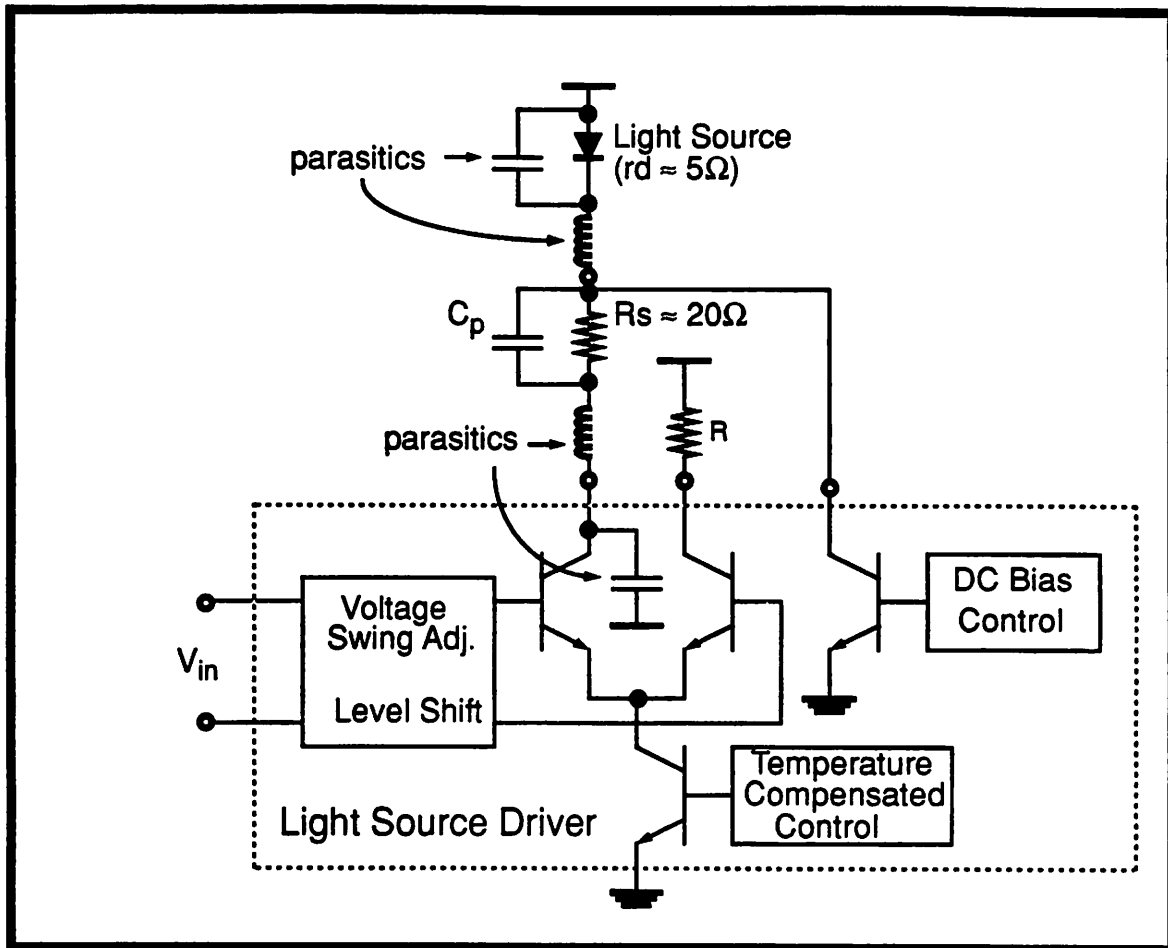


Figure 2-5 Generic Block Diagram of a Light Source Driver

by the bond-wires, which is paralleled by the output capacitance of the laser driver caused by the junction capacitance of the large output transistor. As a consequence of these facts, severe ringing may occur, which also causes large time jitter. Therefore, ringing must be reduced by appropriate circuit and transistor design and by damping the parasitic resonance circuit at the output by using additional resistors. However, these techniques degrade pulse steepness and thus reduce maximum bit rate. Therefore, as an important precondition for achieving high-data-rate operation, the parasitic inductance should be minimized. This can be overcome if the LED can be integrated with the driver on a single chip. High DC bias levels or circuits that actively remove stored charges are also used [11] to speed up the charging process. By using a very-high-speed LED and using GaAsFET as driver, operation up to Gb/s has been reported [12].

2.2.3.2 Semiconductor Laser Driver

The laser driver has the same problems of a LED driver. On top of everything mentioned in Section 2.2.3.1, there is an additional problem of the lasing threshold as explained in Section 2.2.1.2.

The threshold current I_{th} is the forward injection current at which optical gain in the laser cavity exceeds losses. Additional injected current is converted efficiently to light through the processes of stimulated emission. Once lasing occurs, a lower driving current is needed for the same optical output as for the LEDs.

The turn-on time for the laser is increased dramatically when the DC biasing current used in the off-state is close to this threshold current. A bias at or above threshold also reduces overshoot and ringing associated with laser turn-on. As the DC biasing current is increased beyond the threshold, laser frequency response rises. However, this advantage is offset by reduced extinction ratio, which results in degraded sensitivity at the receiver. Therefore, the design goal is to provide bias near threshold that optimizes the turn-on characteristics for the specific application.

The major design problem is that the threshold current varies with temperature as shown in Figure 2-4. Therefore to stabilize the laser operating point with regard to both temperature and time, feedback regulation of the DC biasing current is needed. Stabilization is accomplished by monitoring the light output of the laser with a photodiode. Bias adjustment using low-frequency circuitry stabilizes the photocurrent of the monitor, hence the laser operating point [13].

As in the LED driver case, for high-data-rate operation, interactions among all circuit elements, interconnections, and packaging have to be considered. Operation up to 3Gb/s with a current swing of 30mA has been reported [14]. In the future, monolithic integration of the laser with the driver should result in highest speed performance as interconnection lead inductance severely limits laser performance.

2.3 Optical Fiber

Optical fibers have become the preferred transmission medium for datacommunication because they are capable of transmitting light for long distances with high bandwidth and low attenuation. Not only this but they offer freedom from external interference, immunity from interception by external means, and inexpensive with abundant raw materials.

2.3.1 Modes of Propagation

An optical fiber can be considered as a waveguide based on total internal reflection for light propagating inside the fiber. If we assume for the moment that the conditions for total internal

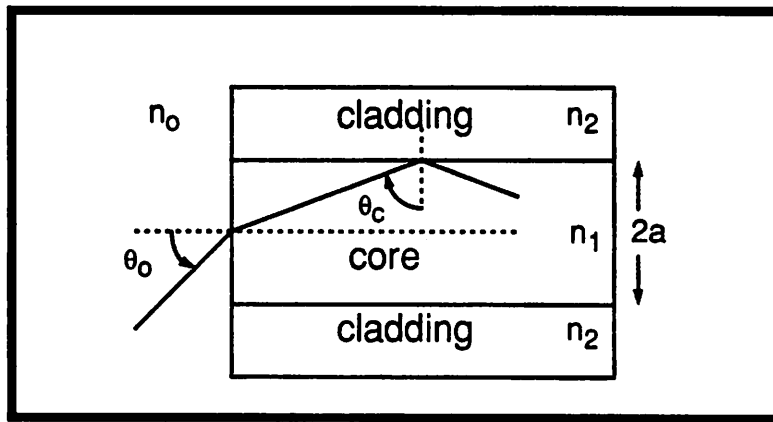


Figure 2-6 Numerical Aperture of a dielectric waveguide

reflection at the boundaries of the waveguide are requirements for propagation, we can consider the manner in which a lightwave must enter the guide in order to satisfy these conditions. The geometric relationships are illustrated in Figure 2-6. By applying Snell's law to the wave incident on the end of waveguide, it can be shown that total internal reflection can take place inside the guide for

$$n_0 \sin \theta_0 \leq n_1 \cos \theta_c \quad (\text{EQ 2-4})$$

where θ_c is the critical angle for total internal reflection and the maximum value of $\sin \theta_0$ that can satisfy this equation is called the numerical aperture, NA, of the waveguide. Specifying

the NA is the standard way of specifying the range of angles of incidence over which the waveguide will accept input wave. If the outside medium is air then $n_o = 1$ and

$$NA = (n_1^2 - n_2^2)^{\frac{1}{2}} \quad (\text{EQ 2-5})$$

While the ray model gives some insight into the behavior of light in an optical fiber waveguide, this model is inadequate to give an accurate description since in practice, the radial dimension “a” of the fiber is on the order of wavelength of the light. The ray model predicts that there is a continuum of angles for which the light will bounce back and forth between core-cladding boundaries indefinitely. A more refined model using Maxwell’s equations to predict the behavior of light in the waveguide, finds that in fact there are only a discrete and finite number of angles at which light zigzags indefinitely.

A set of eigenvalues can be found to satisfy the Maxwell’s equations. Each eigenvalue defines a “mode” that will, if excited, be sustained in the waveguide. Also, there is a cut-off frequency at which there will be no solution. This cut-off frequency can be used to decide how many modes can propagate in the fiber. The normalized cutoff frequency is

$$V = 2\pi \frac{a}{\lambda} (n_1^2 - n_2^2)^{\frac{1}{2}} = 2\pi \frac{a}{\lambda} (NA) \quad (\text{EQ 2-6})$$

where V is the normalized cutoff frequency, a is the diameter for the core and λ is the wavelength of the propagating light. As the radius of the core is reduced, V is reduced and fewer and fewer modes are accommodated. It can be shown that for $V < 2.405$, there can only be one single mode for propagation. Based on this, optical fibers can generally be classified as single-mode fiber where only one mode is allowed to propagate, and multi-mode fibers where more than one mode is allowed to propagate.

2.3.2 Attenuation and Dispersion

The bandwidth or bit rate which can be transmitted through an optical fiber for a given length is basically limited by two major factors, the attenuation of the signal inside the fiber and the dispersion of the signal. With high-data-rate in the range of Gb/s, the maximum distance without

repeater is limited by dispersion and for low data rate below Gb/s, the maximum distance is basically limited by attenuation [15].

2.3.2.1 Attenuation

Attenuation in the optical fiber is the loss in signal power that inevitably results as light travels down an optical waveguide. There are four major sources of attenuation - scattering of light by inherent inhomogeneities in the molecular structure of the glass crystal, absorption of the light by impurities in the crystal, losses in connectors, and losses introduced by bending of fiber. Generally these losses are affected by the wavelength of the light, which affects the distribution of power between core and cladding as well as scattering and absorption mechanisms. The effect of these attenuation mechanism is that the signal power loss in dB is proportional to the length of the fiber.

2.3.2.2 Dispersion

Dispersion is the difference in time of arrival of the signal resulting in a form of distortion. It limits the rate at which data can be transmitted through the medium. Dispersion in optical fiber can be related to the frequency dependence of the index of refraction. This implies the velocity of propagation is a function of frequency. Dispersion can arise from three major sources, material dispersion, waveguide dispersion and modal dispersion. Material dispersion results when the dielectric constant, and therefore the index of reflection is a function of frequency. Waveguide dispersion results when the propagation constants for the waveguide are functions of frequency. For multimode fibers, modal dispersion results because each mode will have a characteristic group velocity and corresponding propagation delay. The modal dispersion is the difference between the longest and shortest propagation delay.

2.3.3 Single-Mode Fibers

2.3.3.1 Single-Mode vs Multimode

The first low-loss fibers fabricated by Corning Glass Works in 1970 were single-mode (SM) fibers, since the transmission characteristics of such fibers were expected to be best suited

for datacommunication. However, to get a single mode fiber, the normalized cutoff frequency V must be smaller than 2.405 which implies a small fiber radius a and small difference in the reflective index $\Delta n = n_1 - n_2$. This makes the fiber more difficult to manufacture and it also make the numerical aperture NA small and coupling lightwaves into the fiber more difficult. As a result, in the early stage, development efforts were first concentrated on large-core multimode (MM) fibers.

With the development of semiconductor light sources and the growing demand of longer distance, wider bandwidth transmission, SM fibers was reintroduced and have become the most widely used optical transmission medium in datacommunication. their major advantage over MM fibers is the absence of modal dispersion and modal noise, making it the lowest cost, largest bandwidth transmission medium available. MM fibers are only used in short distance applications these days.

Important design considerations for SM fibers are low transmission loss, suitable dispersion characteristics, low splice loss, and low bending loss.

2.3.3.2 Dispersion in SM Fibers

In SM fibers, modal dispersion is eliminated. The two other types of dispersion remains are the material dispersion and waveguide dispersion. The total dispersion can be approximated by [16]

$$D_T = \frac{\Delta\tau}{L \times \Delta\lambda} = D_M + D_W \quad (\text{EQ 2-7})$$

$$D_M = \frac{1}{c} \cdot \frac{dN_2}{d\lambda}$$

$$D_W = -\frac{N_1 - N_2}{\lambda c} \left[V \cdot \frac{d^2}{dV^2}(Vb) \right] \equiv -\frac{1.984N_2}{(2\pi a)^2 2cn_2^2} \lambda$$

Where D_M is the material dispersion and D_W is the waveguide dispersion. n_2 is the reflective index, and N_2 is the group reflective index of the cladding material defined by c/V_g ; c is the speed of light and V_g is the group velocity of the light pulse; λ is the wavelength used, V is the normalized cutoff velocity, a is the radius of the core and b is the normalized propagation constant.

Multiplying D_T by the distance travelled and the width of the light pulse, the spread in time can be calculated.

The important result from (EQ 2-7) is that the waveguide dispersion is of opposite sign of the material dispersion. As a result, by adjusting the fiber core radius a , the material dispersion can be controlled to give a null at a desired wavelength. A further refinement in the design of SM fibers is the use of more complex index profiles. One possible design is the quadruple-clad step index profile. By using such profiles, it is possible to achieve relatively low total dispersion over a wide range of wavelengths and there can be two wavelengths of zero total dispersion. The profiles and dispersion characteristics of these SM fibers are shown in Figure 2-7.

2.3.3.3 Attenuation in SM Fibers

Attenuation in state-of-the-art SM fibers is primarily caused by two fundamental physical phenomena that establish a lower limit to the fiber attenuation. In the short wavelength end of the range, the limit is due to the intrinsic Rayleigh scattering of the doped fused silica, which decreases with the inverse of 4th power of wavelength. The lower limit on attenuation at the longer wavelength is determined by atomic absorption. For wavelength longer than $1.6\mu\text{m}$, the absorption is due to the intrinsic infrared tail of the Si-O and/or Ge-O vibrations.

Excess loss caused by waveguide imperfections and metallic impurities are negligible even in today's mass-produced fibers made by any of the commonly-used preform fabrication techniques and minimum losses can be obtained at a wavelength of $1.55\mu\text{m}$. An additional peak of loss at $1.38\mu\text{m}$ is due to the Hydroxyl ion (OH^-) absorption and can be practically minimized through suitable dehydration of the preform and limited to below 2dB/Km. The range of spectral losses is shown in Figure 2-7.

In addition to the fundamental limits on scattering and absorption losses, there are other practical factors that can introduce additional losses such as micro-bending losses, splicing losses, and connecting losses. However these can be kept small as evident from Figure 2-7.

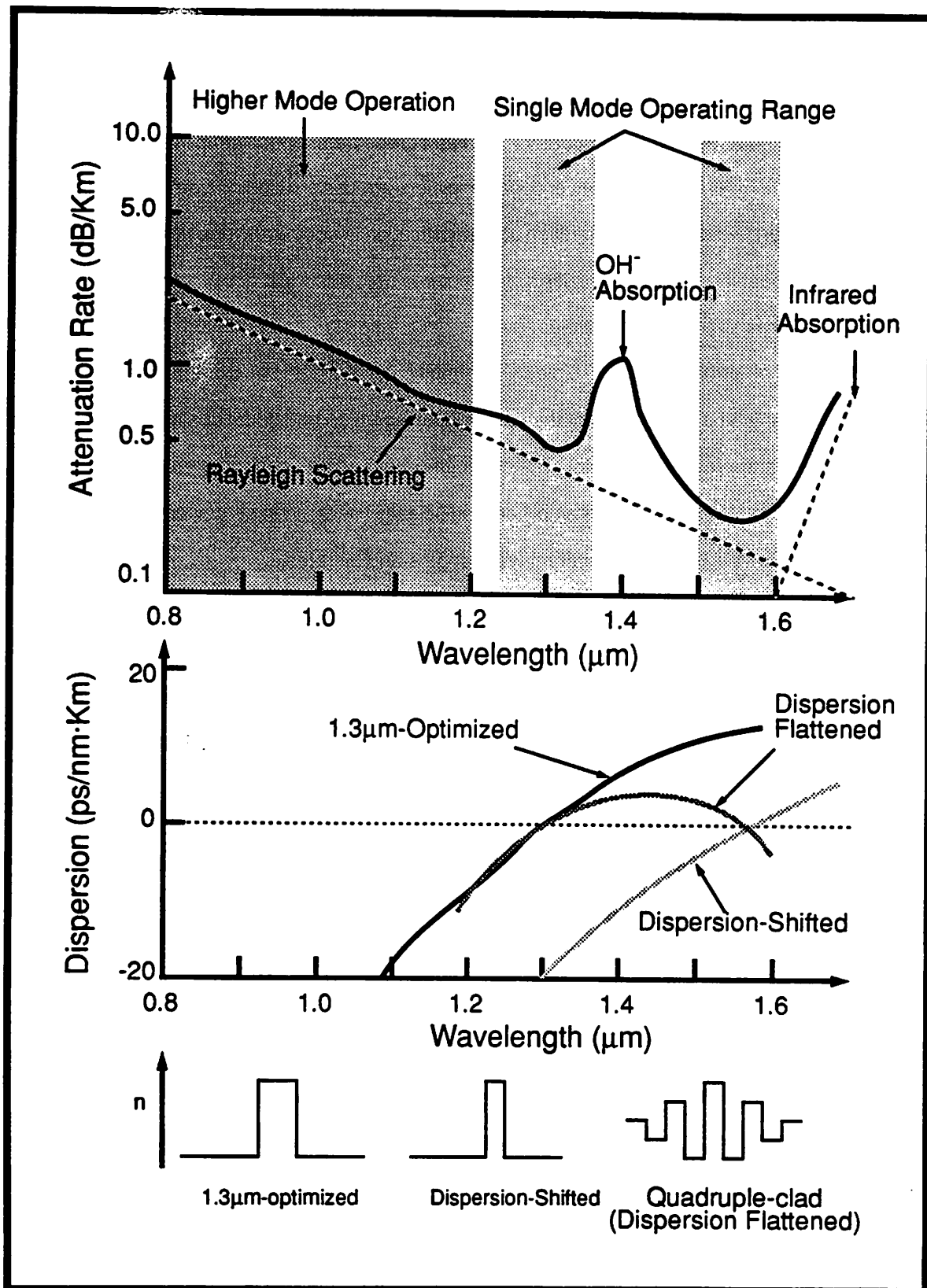


Figure 2-7 Attenuation, Dispersion, and Reflective Index Profiles of SM Fibers

2.3.4 Erbium-Doped Fiber Amplifiers (EDFAs)

Coherent lightwave techniques can be used to construct long-distance digital transmission systems without regenerative repeaters. Instead of repeaters, optical amplifiers are placed at intervals along the fiber, much as conventional amplifiers are used in analog coaxial-cable systems. The difference between regeneration and amplification is not merely one of nomenclature. A regenerative repeater as currently implemented requires optoelectronic devices for source and detector, as well as substantial circuitry for pulse slicing, retiming and reshaping. The optical amplifier, on the other hand, is in principle much simpler; it is a single component that delivers at its output a linearly amplified replica of the optical input signal. In addition to simplicity, the advantage to this approach is flexibility. The same amplifier can be used for any modulation scheme at any bit rate. Indeed, if the amplifier is sufficiently linear, a single device can simultaneously amplify several signals at different wavelengths and bit rates.

There are two promising approaches to optical amplification: semiconductor amplifiers, which utilize stimulated emission from injected carriers [17] [18], and fiber amplifiers, in which gain is provided by stimulated Raman or Brillouin scattering of fiber dopants [19] [20]. The semiconductor optical amplifiers have the advantages of smaller size and lower power consumption due to direct injection pumping. However, they are sensitive to polarization and have a large connection loss when connected to transmission fibers. As a result, they are mostly suitable for use when combined with optical integrated circuits and optoelectronic integrated circuits. On the other hand, fiber amplifiers are directly connected to the transmission fiber and connection loss is small. Although there are many attractions to using Raman gain mechanism, high pump power is required and the pump and signal polarization have to be aligned for higher efficiency. The Erbium-doped fiber amplifiers (EDFAs), on the other hand, have the extra advantage of polarization insensitive and low power level of the pump power. As a result, EDFAs will be the choice as optical amplifiers for future all-optical networks.

An EDFA mainly consists of an Erbium-doped fiber (EDF), an optical coupler, and a pump light source. There are three basic configurations of EDFA, classified mainly by their pump light propagation direction. Forward pumping gives the best noise performance and is shown in

Figure 2-8. The type of EDFA used depends on the applications.

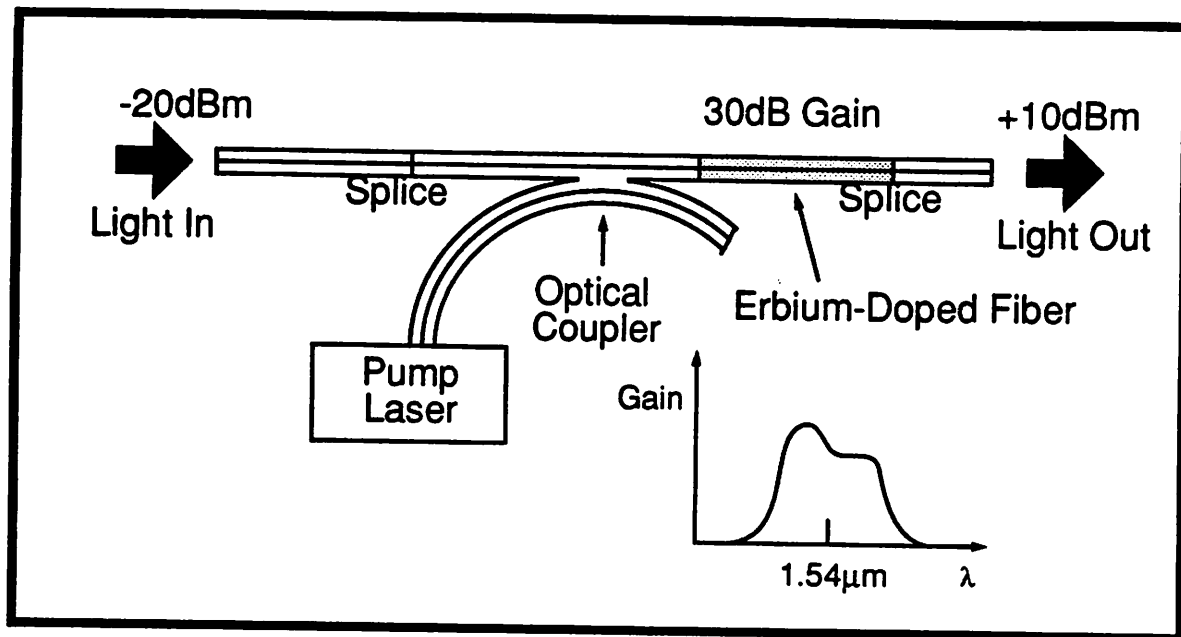


Figure 2-8 Forward Pumped EDFA Configuration

2.4 Receiver

The purpose of an optical receiver is to convert a modulated optical signal to an electrical signal and to recover from electrical signal whatever information had been impressed on the optical carrier. This information may be digital or analog, and the optical carrier may be modulated in a variety of ways, including amplitude modulation (AM), frequency modulation (FM), and phase modulation (PM). However, because a simple optical detector implemented by either PIN diode or avalanche photodiode (APD) is an ideal AM envelop detector, but is insensitive to phase or to small changes in wavelength, it is amplitude modulation that is used extensively in present optical communication systems. Frequency and phase modulation are being investigated in many laboratories for use in coherent optical transmission [21] but in this thesis, we will concern with the conventional approach of amplitude-modulated optical signals and the design of appropriate direct-detection receivers.

Analog transmission is not commonly used in optical systems because of the non-linear characteristics of optical sources (laser and LEDs). Digital modulation, also known as on-off key-

ing, is easily accomplished with these sources by modulation of bias current as explained in Section 2.2.1. The simplicity of modulation plus the extremely wide bandwidth of optical fibers have made binary digital transmission the method of choice for the majority of optical communications systems. A direct detection digital transmission consists of a stream of light pulses, where the presence of a pulse corresponds to the transmission of a binary "1", and the absence of light corresponds to the transmission of a binary "0".

Shown in Figure 2-1 is a block diagram of a basic datacommunication receiver. First, there is a photodetector and a low-noise preamplifier that converts the input photon into photo-current, and subsequently into a low level voltage. These two blocks form the front-end for the receiver and is sometimes called collectively as a photoreceiver. It is then followed by a main amplifier that serves several purposes - equalization of roll-off in the front end (not always necessary), low pass filtering to limit the noise bandwidth to the minimum required, and high gain amplification with limiting amplifier or automatic-gain-control (AGC) amplifier. For optimal performance, it is necessary to extract timing information from the received signal in order to synchronize the process of making decision on the noisy main amplifier output. For this purpose, the output of the main amplifier is also fed to a clock recovery circuit that generates a clock signal at the baud rate synchronized with transitions in the received data. The decision circuit then use this timing information to compare the input with a fixed threshold level set to the center of the received eye pattern to give an equal probability of error for decisions on both 1's and 0's. Once the digital data is regenerated from the analog waveform, the high-data-rate bit stream is then deserialized back to its parallel form by a demultiplexer, and passed to a digital data link control for high level manipulation of data received.

For most lightwave communication systems, the receiving function limits the maximum data rates for transmission. The optical fiber itself has a wide bandwidth, larger than 40GHz, larger than any electronic components can handle. The transmitters using high speed technology and semiconductor lasers can operate above 10 Gb/s. The digital blocks of the receiver, such as the demultiplexers and decision circuits, can also operate above 10Gb/s, but the analog blocks such as the low noise preamplifier, main amplifier, and clock recovery circuit, todate, still cannot break

through the 10Gb/s barrier even with the fastest technology available.

High-data-rate receiver design is the main topic in this thesis and will be discussed with more detail in Chapter 3.

2.5 Probability of Error and Quantum Limit

If one consider an ideal photodetector followed by a noiseless electronic amplifier, it is possible to derive the minimum received optical power required for acceptable bit-error-rate (BER) performance in a digital transmission system. This minimum received power level, known as the quantum limit, is a result of the statistically random nature of light absorption in a material medium. The probability of receiving exactly n photons during a pulse interval T , when the average number of photons received during this interval is N , is given by a Poisson distribution $P[n]$

$$p[n] = \frac{N^n e^{-N}}{n!} \quad (\text{EQ 2-8})$$

The probability of making an error can be expressed as

$$p[\text{error}] = p[\text{error}|1]p[1] + p[\text{error}|0]p[0] \quad (\text{EQ 2-9})$$

When "0" is transmitted, no photon will be received and the decision will be made with probability of error equals to zero. When "1" is transmitted, even though the average number of photon transmitted is N , there is a finite probability that none will be received, causing an error. If we assume there is an equal probability of transmitting a "1" and a "0", then the probability of error is $p[\text{error}] = p[n=0] \cdot 1/2 + 0 \cdot 1/2 = 1/2e^{-N}$. For a BER of 10^{-9} , $N = 20$. This implies an average of 10 photons per bit is the quantum limit provided everythings are perfect.

Although the quantum limit is of interest as a basis for evaluating the performance of specific systems, it is not a realistic measure of the sensitivity of such systems. Two assumptions are unrealistic; the first one is that when no signal is transmitted, the output of the detector will be zero, and the second assumption that when a signal is transmitted, the only output is due to signal photons from the transmitter. The major source of violation comes from the fact that there is noise both

with the photodetector and the amplifier following it.

Noise is characterized with a Gaussian probability density function

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \quad (\text{EQ 2-10})$$

where μ is the mean and σ the rms deviation from the mean value.

For pulse code modulation, a binary "1" would be represented by a signal voltage of mean value V_o and a binary "0" by zero volt. In each case, noise will cause a rms deviation from the signal voltage of σ volts. If the threshold is placed at $V_o/2$, and let

$$Q(x) = P[X > x] = \int_x^{\infty} p(x) dx \quad (\text{EQ 2-11})$$

where X follows a standard normal distribution with $\mu = 0$ and $\sigma = 1$. Assuming the probability of receiving a "1" is the same as receiving a "0", and the noise mechanism is the same in both cases, the probability of error can be calculated from (EQ 2-9) to be

$$Q\left(\frac{V_o}{2\sigma}\right) = Q\left(\frac{\sqrt{\text{SNR}}}{2}\right) \quad (\text{EQ 2-12})$$

For a BER of 10^{-9} , $V_o = 12\sigma$, or a signal-to-noise (SNR) ratio of 21.6dB; for a BER of 10^{-11} , a SNR of 22.5dB is needed. For high BER or high SNR, $Q(x)$ can be approximated by

$$Q(x) \approx \frac{1}{\sqrt{2\pi}x} e^{-\frac{x^2}{2}} \quad (\text{EQ 2-13})$$

On a log plot of the BER vs the input Voltage, the input referred SNR can be calculated from the slope of the curve. If two such curves are measured with different $(x-\mu)$'s, then the magnitude of the rms deviation, or the rms noise, can also be calculated.

Chapter 3 Traditional Receiver Architecture

3.1 Introduction

The traditional architecture of an optical fiber receiver is shown in Figure 3-1. It can be

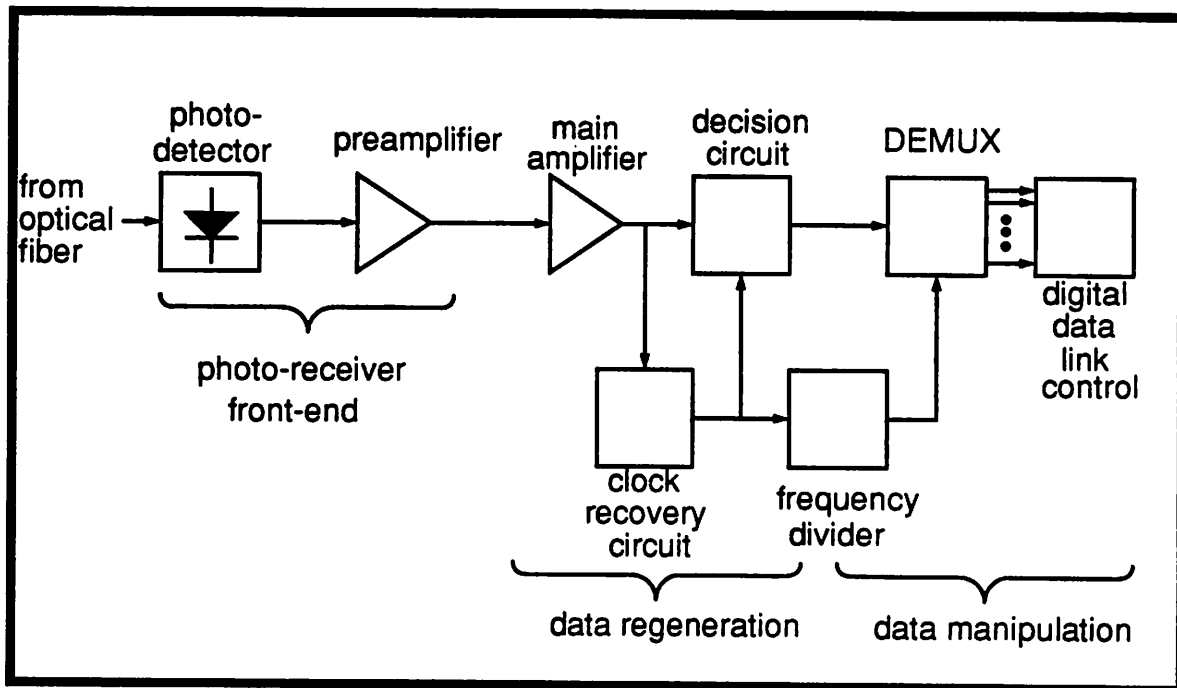


Figure 3-1 Traditional Architecture for Digital Fiber Optic Receivers

basically divided into a photoreceiver front-end and a data regeneration and manipulation back-end. This chapter will look into each functional block in more detail and identify the performance limits.

3.2 Photo-Detectors (PDs)

The function of the photodetector is to convert the normally weak optical signal into a corresponding weak electrical signal. Subsequent stages in the receiver provide amplification and signal processing.

There are quite a variety of photodetectors such as PIN photodiodes, avalanche photodiodes (APDs), Metal-Semiconductor-Metal (MSM) photodiodes, phototransistors, photoconductors. However only PINs and APDs will be described in detail herein, as they have demonstrated the best characteristics over the photoresistors and photoconductors for datacommunication applications, and MSM detectors are only used in integrated optoreceivers.

3.2.1 Semiconductor Photodiodes

Both PINs and APDs are semiconductor photodiodes. As an electromagnetic wave propagate through a semiconductor, the wave may deliver enough energy to a valence electron to free it from the covalent bond that holds it in its place in the crystal structure. When this occurs, an electron-hole pair is created and a photon is removed from the wave. The electron that has been excited into the conduction band and the hole in the valence band are now free to move under the influence of an electric field. If an electric field is present in the region where electron hole pairs have been created, the electron and the hole are accelerated in opposite directions and are swept out of the region. This motion of charges in the semiconductor induces a current that can be detected in the external circuit. The charges released by the incident photons are called photocarriers. The external current due to the motion of these charges is the photocurrent.

3.2.2 Responsivity and Quantum Efficiency

The responsivity of a photodiode R is defined as the output photocurrent produced per unit of incident optical power. An incident optical power of P with frequency f is equivalent to P/hf photons per second, where h is the Planck's constant. Let η be the ratio of the average number of electrons excited into the conduction band to the number of incident photons. The average number of electrons per second, or the photocurrent will then be

$$I_p = q\eta \frac{P}{hf} \quad (\text{EQ 3-1})$$

and the responsivity is

$$R = \frac{I_p}{P} = \frac{q\eta}{hf} = \frac{\eta\lambda}{1.24} \quad (\lambda \text{ in } \mu\text{m}) \quad (\text{EQ 3-2})$$

The term η is the quantum efficiency of the photodiode. Its value is less than unity and is determined by both the properties of the semiconductor material(s) and the physical structure of the device. It can be expressed as

$$\eta = (1 - r) (e^{-\alpha d}) (1 - e^{-\alpha_p W}) \quad (\text{EQ 3-3})$$

The first term is due to the reflection loss from the front surface of the photodiode, the optical power that passes through the front surface and into the photodetector is then $(1-r)P$. The second term is due to the absorption loss of the input optical power when it gets through the region of length d where no electrical field exists. Charges produced will recombine back without any significant effect on the photocurrent. The wave then enters the high field depletion region of width W where the usual absorption takes place. The attenuation in this region is represented by the third term. The first two terms are due to the physical structure of the device and can be design to make r small for low reflection loss and d small for low absorption loss. The third term is both structure dependent and material dependent. W can be made large but then it takes more time for the carriers to be swept through this region, so there will be a speed vs gain trade-off as usual. α and α_p are essentially the same, it is determined by the bandgap energy of the material(s) used.

Some typical photodiode spectral response is shown in Figure 3-2. Silicon is the material most commonly used in photodiodes for wavelength shorter than $1\mu\text{m}$. It is not used for $1.3\mu\text{m}$ and $1.55\mu\text{m}$ wavelengths because the bandgap energy of Silicon is too large for these long wavelength spectral regions. From Figure 3-2, it appears that Germanium and compound semiconductors such as InGaAs should be useful in longer wavelengths.

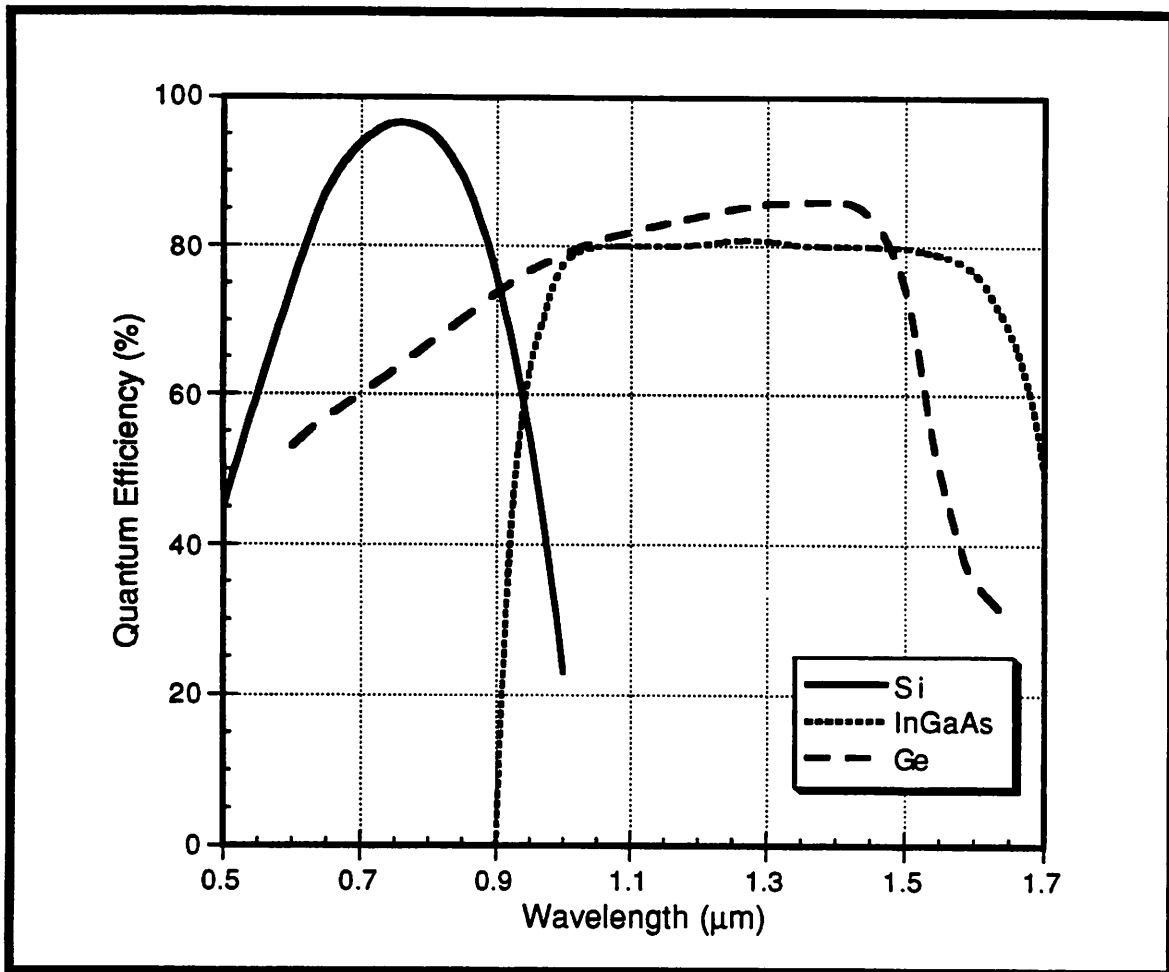


Figure 3-2 Typical Photodiode Spectral Response

3.2.3 PIN Photodiode

The most commonly used semiconductor photodetector in long wavelength optical communication systems is the PIN photodiode. The P-I-N photodiode is a p-n junction structure with a very lightly doped "intrinsic" region between the normal p and n type regions. The normal mode of operation of the PIN photodiode is by applying a reverse biased voltage, the intrinsic region is completely depleted, thus establishing a high electric field region with no free carrier. This minimizes the dark current flow in the absence of light. The dark current in a PIN photodiode is typically in the range of 10nA. With state-of-the-art design, the lowest dark current reported is below 0.1nA [22].

To maximize the quantum efficiency and hence responsivity, photon absorption in the

depletion region should be maximized, and minimized elsewhere. From (EQ 3-3), this can be done by increasing the width W of the depleted intrinsic region, and by using a material with large absorption coefficient α_p . However, if α is approximate the same as α_p , loss will increase before the optical wave reaches the intrinsic region. α can be made smaller than α_p by applying a hetero-structure similar to those used in the light sources. If the material through which the optical wave enters the device can be selected to have a bandgap larger than hf , it will not absorb energy from the wave as it passes through this section of the photodiode. Absorption is thus confined to the region where it is effective in producing photocarriers. With such design, quantum efficiency as high as 0.95 can be achieved with InGaAs PIN photodiodes, making it the most frequently used device for long wavelength communication systems.

The rate at which the photodiode can respond to changes in the intensity of the optical input signal is a measure of the maximum information rate or bandwidth of the photodetector. The ultimate bandwidth of the device is limited by the carrier transit time, the carrier diffusion time, and the hole trapping at heterojunction interfaces; with the carrier transit time being the dominating factor. The transit time of the PIN photodiode is limited by the high-field, saturated carrier velocity of the slowest carrier. The bandwidth of efficient InGaAs PIN diodes has been found to be greater than 20GHz [22] with a quantum efficiency of 0.8. If the intrinsic region's width W is allowed to be narrow, frequency responds as high as 67GHz has been reported [23]; although at these width, a significant loss in quantum efficiency is expected.

In practise, because of the wide bandwidth of the PIN photodiode, the detector's bandwidth is limited by the extrinsic RC time constant, where R is the effective resistance of the external bias circuit, and C is the photodiode capacitance. The photodiode capacitance depends on the area of the photodiode, typically in the range of 0.5pF. When the area of the photodiode is small, typically below 40 μ m in junction diameter, the bandwidth is limited by the transit time, and for larger diameters, it is limited by the extrinsic RC delay. There are several techniques to reduce the parasitics at the interface of the photodetector and the amplifier. One approach is to use microwave design to compensate for the parasitics [24], bandwidth of 16GHz is achievable. Another approach to integrated design is to fabricate the device elements at the closest distance possible

so that the introduction of any parasitic reactances is minimized. Monolithic integration of optoelectronic integrated circuits (OEICs) is an attractive prospect, and recent advances have shown encouraging receiver sensitivity improvement. It will be discussed later in Section 3.4. While OEICs will become increasingly important, the fabrication technology of OEICs has not grown quick enough to bear many practical applications, primarily due to the necessity of integrating highly different devices on one chip. For near future applications of integrated receivers, a more practical approach can be found in flip-chip integration, in which two existing elements are joined directly by bonding. A bandwidth of 21GHz is achievable using this technique [22].

The generation of photocurrent is a sequence of discrete events that includes the creation of electron hole pairs and the motion of these charges under the influence of local electric field. Each electron hole pair will result in a pulse of current. The total current is the sum of many pulses. The total current is not a smooth continuous flow but has variation about an average value. This variation is the "shot" noise of the photodiode. The mean-squared value of the shot noise associated with the photocurrent is

$$\overline{i_n^2} = 2qIB \quad (\text{EQ 3-4})$$

where I is the total photocurrent and B is the equivalent noise bandwidth.

3.2.4 Avalanche Photodiode (APD)

The photocurrents produced in the photodiode of optical communication receivers are usually very small. Amplification of these weak electrical signals is necessary before useful signal levels are established. One approach to amplify the signal is the use of an avalanche phenomenon to provide current amplification within the photodiode. The device that provides this current amplification is the avalanche photodiode (APD).

APDs are similar to PIN photodiodes in that they are operated under reverse bias, and therefore in the absence of large background dark currents. Unlike PIN photodiodes, APDs are operated at sufficient high reverse voltage such that photocurrent gain due to impact ionization of carriers with the lattice atoms occurs. Detail studies of the avalanche mechanism can be found

in [25]. The important result is that a current gain factor M , which is a function of the applied reversed voltage, results from the avalanche mechanism, thus, improving the responsivity of an APD by M times over the PIN photodetectors. Most long-distance high-data-rate lightwave transmission systems operating near $1.3\mu\text{m}$ or $1.55\mu\text{m}$ have utilized InP/InGaAsP/InGaAs APDs in the front-ends of the receiver because of the excess gain factor M .

The bandwidth of APDs is limited by the carrier transit time, hole trapping at the hetero-junction interface, and extrinsic RC time constant just as in the PIN photodiodes. As a result, for low gain setting, the bandwidth is constant just as a PIN photodiode. At high gain setting, the avalanche build-up time dominates. The bandwidth decreases proportionately with increasing gain M , resulting in a constant gain-bandwidth (GBW) product. A GBW product as high as 70GHz has been reported in [26] though the constant bandwidth at low gain setting is only 8GHz due to a higher carrier transit time because more layers are required than for the PIN photodiodes.

The avalanche current-multiplication process is a random process in that M is only the average, but not absolutely fixed multiplication factor. The number of secondary electrons and holes that results from any individual injected electron or hole may differ from this average value M . The multiplication process has its own fluctuations that are superimposed on any fluctuations inherent in the primary photocurrent. The excess noise is usually represented by an excess noise factor F which in itself a function of M . The mean-squared total noise current is therefore

$$\overline{i_n^2} = 2qI_{ph}FM^2B \quad (\text{EQ 3-5})$$

where I_{ph} is the primary photocurrent without multiplying by M . It was shown in [27] that the excess noise factor can be expressed by

$$F = kM + (1 - k) \left(2 - \frac{1}{M}\right) \quad (\text{EQ 3-6})$$

where k is the carrier ionization ratio. Note that F increases with increasing k as the randomness increases with higher ionization ratio. As expected, when $k=0$ and $M=1$, $F=1$ and the equation degenerates to be the same as that of the PIN photodiode.

3.2.5 Photodiode Equivalent Circuit

Figure 3-3 shows an AC equivalent circuit for both the PIN photodiodes and APDs. The

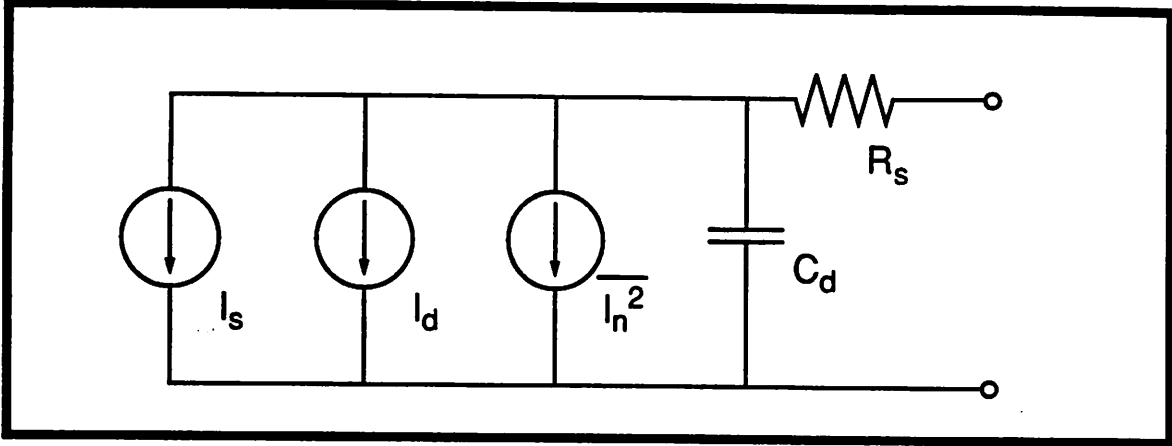


Figure 3-3 AC Equivalent Circuit of Photodiode

current source I_s represents photocurrent resulting from the detection of an optical signal. I_d is the background dark current that flows with no input signals. I_n^2 is the shot noise current of the photodiode.

$$I_d = I_{du} + M I_{dm} \quad (\text{EQ 3-7})$$

$$I_n^2 = 2q [I_{du} + M^2 F(M) I_{dm}] B \quad (\text{EQ 3-8})$$

The back ground dark noise has two components, one is the unmultiplied dark current I_{du} and the other one is the multiplied dark current I_{dm} multiplied by the avalanche gain M . For a PIN diode, M is simply one. The shot noise term also has two components, one is due to the unmultiplied dark current and the other is due to the avalanche gain M . Because of the randomness of the multiplication process for APDs, there's an extra excess noise factor F which is a function of M that has to be multiplied with M^2 to get the variance in the multiplication gain factor M . For a PIN diode, $M = 1$ and $F(M) = 1$ indicating that as expected, there is no excess noise. C_d is the parasitic non-linear junction capacitance of the p-n junction and R_s is the series resistance which is usually very small (5-10 Ω) and can be neglected in most cases. C_d depends on the size of the photodiode and

is typically 0.5pF.

3.2.6 PIN Photodiodes vs APDs

A comparison of PIN photodiodes against APDs is listed in Table 3-1. While PIN diodes

Photo-diodes	Advantages	Disadvantages
PIN	Simple to Use Low Cost High Speed Reliable	Low Receiver Sensitivity Low Receiver Dynamic Range
APD	Higher Sensitivity (5 to 15 dB better than PIN) Wider Dynamic Range through Gain Variation	Gain Characteristic is Temperature Sensitive Requires High Voltage Power Supply Higher Cost Dark Current and Excess Noise of Long Wavelength APDs are still High, Limiting Their Usefulness at Low Bit Rates.

Table 3-1 Comparison of PIN Photodiodes Vs APDs

have all the advantages of a low power and low cost solution for a high speed photodetector, the only draw back is that it has a limited dynamic range and most of all, low sensitivity. However, with the advances made in both fiber and semiconductor optical amplifiers as mentioned in Section 2.3.4, there is a trend of replacing APDs with a combination of optical amplifiers and PIN photodiodes because of the lack of excess noise. This approach is even more attractive fueled by the integration of optical detector and amplifiers such as those in OEICs, as it is much more complicated to integrate an APD as compared to integrating a PIN photodiode. As a result, the future solution will be an optical amplifier, such as Erbium-doped fiber, followed by an OEIC with a PIN photodiode.

3.3 Low Noise Preamplifier

The preamplifier acts as an interface between the photodetector and conventional electronics. Once the signal is converted from the optical domain into the electrical domain, in the form of a low level photocurrent, a low noise preamplifier is needed to bring the signal and noise ampli-

tudes to a level where the noise produced in subsequent stages has a negligible effect on the overall signal-to-noise ratio. As a result, the preamplifier and photodetector have to be considered together to evaluate the performance and this combination is known as an optical receiver. The primary performance parameters that determines the usefulness of an optical receiver is the sensitivity, bandwidth, dynamic range, and the interface to the remaining electronics.

The preamplifiers can be classified into three broad categories, based on the equivalent resistance seen by the photodetector. They are the low-impedance, high impedance, and transimpedance preamplifier.

3.3.1 Low-impedance Preamplifier

Low-impedance preamplifiers typically consists of a photodiode operating into a low-impedance amplifier, often through a length of coaxial cable or other transmission line. A terminating resistor R_L equal to the transmission line impedance is generally included to suppress standing waves for uniform frequency response. A typical low-impedance preamplifier is shown in Figure 3-4.

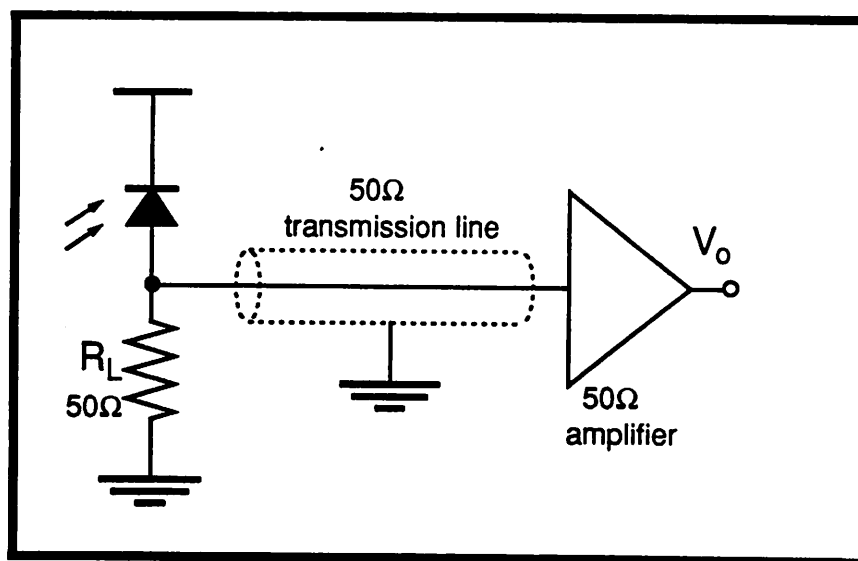


Figure 3-4 Low-impedance Preamplifier

The first advantage of the low-impedance preamplifier is that it is simple, commercially available 50Ω RF amplifier can be used. The second advantage is that it has a wide bandwidth as

the RC time constant is extremely low due to a low load resistance. It also has a wide dynamic range as the signal to the preamplifier input is a a very low level. However, the disadvantage is also due to the low load resistance. The noise of the low-impedance amplifier is dominating by the thermal noise of the low resistance

$$\overline{i_n^2} = \frac{4kTFB}{R_L} \quad (\text{EQ 3-9})$$

where F is the noise figure of the amplifier and B is the equivalent bandwidth. The sensitivity is degraded heavily because of the low load resistance. As a result, it is only useful in application such as instrumentation where sensitivity is not a major concern.

3.3.2 High-impedance Preamplifier

Because of current source nature of photodiodes, it is possible to increase signal voltage for a given photocurrent by operating the photodiode into an amplifier with higher input resistance. The signal voltage is therefore increased prior to the addition of noise from the following preamplifier, resulting in improved sensitivity. In a high-impedance preamplifier, as shown in Figure 3-5, the load resistance R_L is large ($\sim 5K\Omega$) and a high-impedance field-effect transistor (FET) amplifier is employed. The total input capacitance C_T is the sum of FET input capacitance, detector capacitance, and stray capacitance. As a result of large input RC time constant, the preamplifier bandwidth is small and acts like an integrator for high frequency input. Such integrating front-ends require equalization to compensate for their lack of bandwidth. Equalization is commonly provided by a simple RC circuit that also shown in Figure 3-5. The transfer function of the equalizer has a zero at f_1 and a pole at f_2 . The value of R and C are chosen such that the zero at f_1 cancels the pole of the high-impedance preamplifier, and the overall receiver bandwidth is extended to f_2 .

The biggest advantage of the high-impedance preamplifier is that it has the highest sensitivity as it has the best power transfer efficiency from the photodiode. Neglecting noise from the preamplifier and shot noise of the photodiode, the SNR can be expressed as

$$\text{SNR} = \frac{(I_s R_L)^2}{\frac{4kTB}{R_L} \times R_L^2} = \frac{I_s^2 R_L}{4kTB} \quad (\text{EQ 3-10})$$

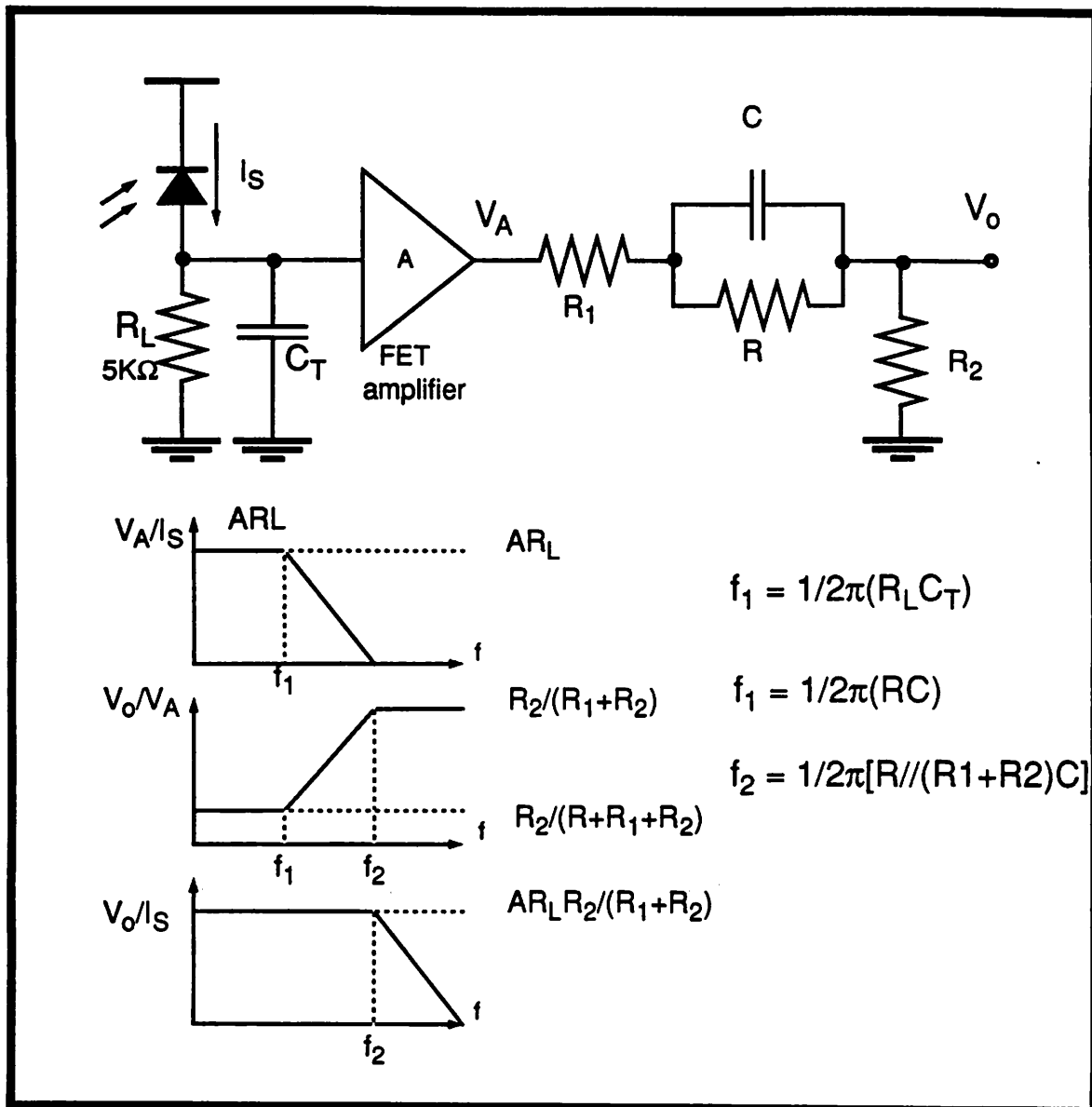


Figure 3-5 High-impedance preamplifier and Equalizer

The SNR increases with increasing R_L . It is ultimately limited by the input impedance of the preamplifier.

The disadvantages of the high-impedance amplifier are the need of an equalizer that increases the design complexity as it is hard to match the zero of the equalizer to the pole of the front-end; also, because of the front-end enhances the amplitude of low frequency signals at the preamplifier input, early saturation of the preamplifier may occur. As a result, the high-impedance preamplifier only has a limited dynamic range.

For practical implementations, the amplifier used does not have infinite frequency response and contributes significant amount of noise that has to be taken into consideration. Wide-band amplifiers are generally based on a cascaded amplifier configuration using mismatched interstage resistive load coupling [28]. The interstage load resistances are shunted by the parasitic capacitances of the FETs, giving rise to additional RC integrations which restricts the frequency response of the receiver. In the conventional design, receiver bandwidth can only be extended at the expense of receiver gain and signal-to-noise performance. To overcome this problem, inter-stage microwave matching networks and series inductive peaking between the photodiode and preamplifier are used [29] [24]. Applying the above technique, a 16GHz optical receiver, operating at 11Gb/s, with -19.8dBm sensitivity results. It was fabricated as a hybrid integrated circuit, with a high speed InGaAs p-i-n photodiode coupled to a three stage GaAs HEMT preamplifier.

3.3.3 Transimpedance Preamplifier

A configuration that provides improved dynamic range is the transimpedance front end as shown in Figure 3-6. The load resistor is connected as a feedback resistor around an inverting

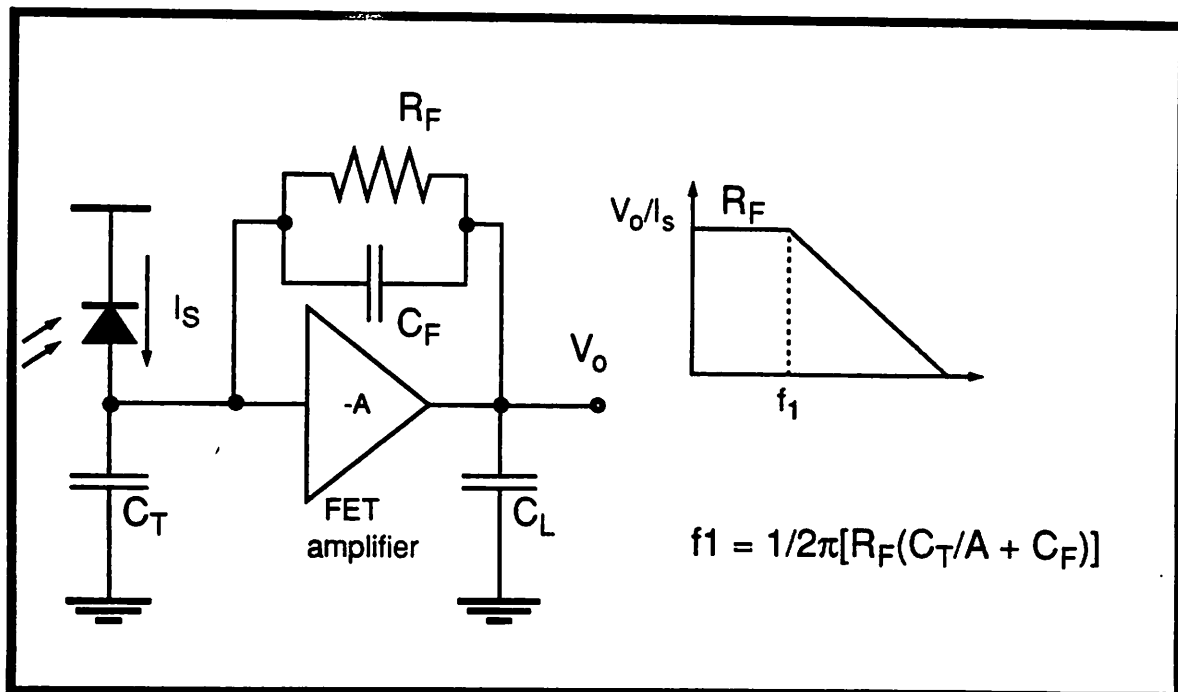


Figure 3-6 Transimpedance Preamplifier

amplifier with gain A. The bandwidth of the preamplifier is increased roughly by the factor A compared with a high-impedance preamplifier with the same load resistance and input capacitance. As a result, there is no need for equalization in general. The dynamic range is improved because of the flat response with no low frequency signal enhancement.

The noise of the preamplifier is a combination of the thermal noise of the feedback resistance and thermal noise of the input transistor. For high transimpedance gain and low input referred current noise, a high value of R_f is desired just as in the high-impedance preamplifier case. However, the bandwidth is inversely proportional to R_f . As a result, there is a trade-off between the sensitivity and bandwidth of the preamplifier.

Another major consideration in the design of transimpedance preamplifier is stability because the amplifier used does not have infinite bandwidth. To understand this problem, one implementation of the transimpedance preamplifier is shown in Figure 3-7. For small C_F , it can be shown that the transfer function is a two pole system with complex poles where

$$|s| = \frac{1}{\sqrt{\frac{C_T C_L R_F}{g_m}}} \quad (\text{EQ 3-11})$$

$$\sigma = -\frac{C_F}{2(C_L + C_F) \frac{C_T}{g_m}} \quad (\text{EQ 3-12})$$

Note that $|s|$ is independent of C_F and σ is a function of C_F , therefore a proper value of C_F should be used to ensure enough phase margin.

The opened-loop gain A of the amplifier and the transimpedance gain R_f is set by the corner frequency since C_T , the total input capacitance, is usually fixed. For a fixed bandwidth, in order to have a high transimpedance gain, the opened-loop gain A must also be high. A single gain stage as in Figure 3-7 can only provide a gain value of less than 20 [30]. A simple solution for higher opened-loop gain is to cascade more gain stages to get the high gain required [31]. However, each gain stage contributes additional phase shift, making the overall loop more unstable. A lower value of R_f has to be used to lower the feedback ratio for loop stability, resulting in lost of sensitivity.

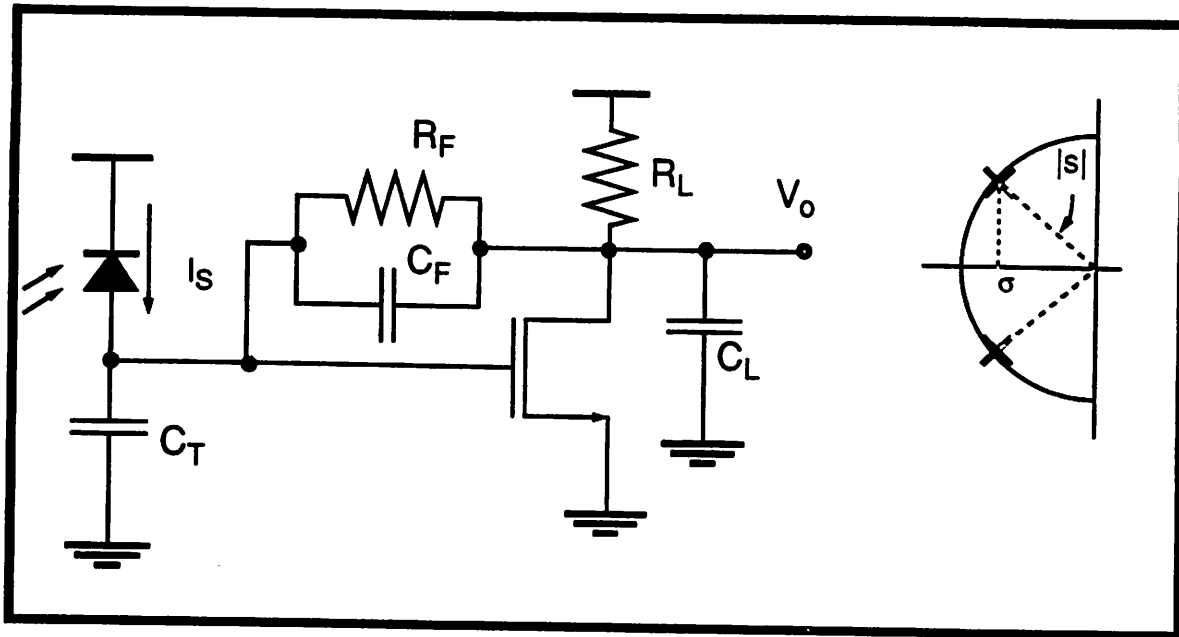


Figure 3-7 Single-stage Transimpedance Preamplifier

Other two stage feedback transimpedance preamplifiers have been introduced to circumvent the stability problem so that the highest value of R_f can be used. Shown in Figure 3-8 are

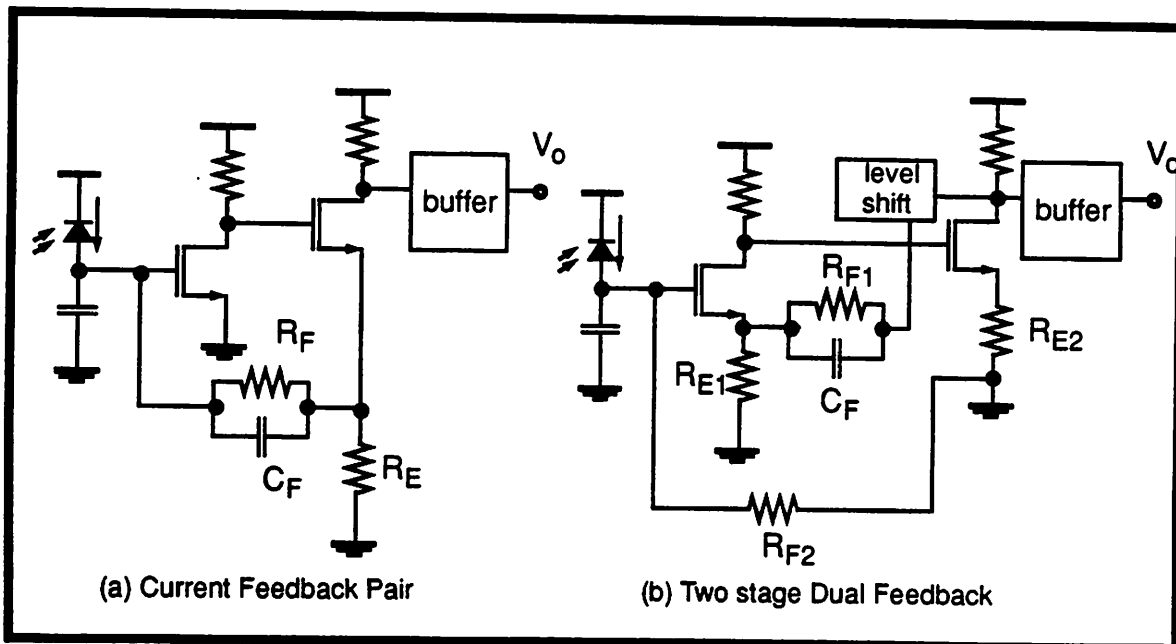


Figure 3-8 Simplified Schematics of Two-stage Transimpedance Preamplifiers

some examples of two-stage transimpedance preamplifier. In Figure 3-8 (a), the preamplifier is based on the current feedback pair design [32]. A feedback zero provided by C_F is used to com-

pensate for stability. In Figure 3-8 (b), the preamplifier is based on the dual current and voltage feedback loops introduced by [33]. The feedback zero C_F in the voltage feedback loop is used to reduce the high frequency peaking and extend the 3dB bandwidth of the preamplifier [34].

It is hard to compare the transimpedance preamplifier of different designs as it depends on the technology and one can trade sensitivity for bandwidth using different value of R_f . Results from different technologies are listed in Table 3-2. In cases where device f_T is not available, the minimum channel length of the device is listed.

BW (GHz)	TZ Gain(dB Ω)	Technology	Device f_T (GHz)	Reference
13	47	AlGaAs/InGaAs/GaAs	51	Suzuki [32]
11.2	53.4	Si BJT	40	Suzaki [34]
2.3	67	GaAs	25	Scheinberg [35]
0.92	22	NMOS	(0.85 μ)	Abidi [31]
0.26	69	BiCMOS	(0.45 μ)	Lim [30]

Table 3-2 Performance of Transimpedance Preamplifier of Different Technologies

The advantages of transimpedance preamplifier are the wider dynamic range, better sensitivity than the low impedance preamplifier, and no need for an equalizer. The disadvantages are lower sensitivity than high-impedance preamplifier, and stability problem due to the negative feedback prohibiting the use of a maximum feedback resistance. However, because of its flexibility, it is the most popular type of preamplifier unless extremely high sensitivity is required.

3.4 Integrated Optoreceiver

So far, the optoreceivers discussed have been a hybrid of a photodetector and a preamplifier. Optoelectronic integrated circuits (OEICs), circuits that monolithically integrate optical and electronic components on a single semiconductor chip, offer significant advantages over hybrid circuits in compactness, reliability, possible performance improvements resulting from reduced parasitics, and potentially significant reductions in cost, particularly in the case of arrays. However, despite these many potential advantages, todate OEICs have not outperformed hybrid circuits per-

forming similar functions. This is generally recognized to be due to the difficult materials and materials fabrication challenges presented by integration of very different devices on a single chip.

The photodetectors used in an integrated optoreceiver are the PIN photodiodes as used in the hybrid approach. APDs are too complex to be integrated into OEICs. However, a new type of photodetectors, the metal-semiconductor-metal (MSM) photodetectors have been used widely in integrated optoreceivers because of their ease of integration with IC process, low capacitance, and high speed [36]. The major disadvantage of MSM detectors is the low responsivity compared to PIN photodiodes due to the MSM detector's interdigitated finger structure that blocks part of the incident light.

The MSM detector works just like a PIN photodetector. It consists of an interdigitated pattern of metal fingers deposited on a semiconductor substrate. During normal operation, a potential is applied between alternate fingers in this structure creating an electric field which serves to sweep photogenerated electron hole pairs to the positive and negative electrodes, respectively as shown in Figure 3-9. As a result of the finger structure, the responsivity is approximately lower than that of a PIN photodetector by a factor of $s/(w+s)$.

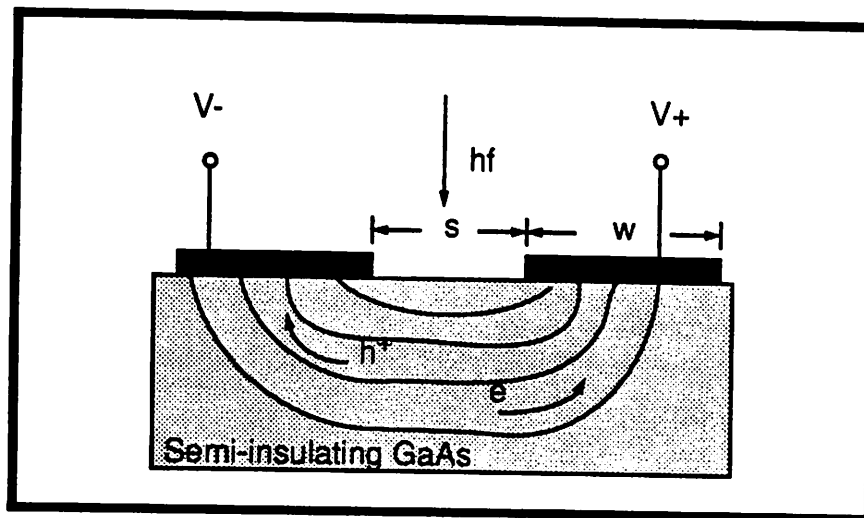


Figure 3-9 MSM Detector

GaAs and InP are the two mostly used materials in OEICs. The high electron drift velocity and high electron mobility of these systems, as well as the ability to form hetero-junctions exhibit-

ing high quantum efficiency, combine to make these materials highly desirable for high-speed electronic and optoelectronic devices. A collection of some of the recent published results are collected in Table 3-3.

BW (GHz)	Sensitivity (dBm)	Bit Rate (Gb/s)	Wave-length (μm)	Material	PD	Pre-amp	Yr.	Reference
8.2	N/A	10	0.84	AlGaAs/GaAs HEMT	MSM	TZ	91	V. Hrum, et al. [37]
6	-21.2	8	1.3	InP HEMT	PIN	TZ	92	H. Yano, et al. [38]
1	-30.4	1.2	1.3	InP HEMT	PIN	HZ	90	H. Yano, et al. [39]
0.5	-26.1	1	1.5	InGaAs/InP HBT	PIN	TZ	91	S. Chandrasekhar, et al. [40]
0.4	-34.7	0.622	1.3	InP/InGaAs JFET	PIN	TZ	91	N. Uchida, et al. [41]
3	-49	0.2	1.5	InGaAs/InP HBT	PIN	TZ	91	S. Chandrasekhar, et al. [42]

Table 3-3 Performance of Integrated Optoreceivers

3.5 Main amplifier

Another basic component of optical fiber receivers is the main amplifier which has to amplify the output voltage of the low-noise preamplifier to a value required by the succeeding decision and clock-recovery circuits. Apart from high operating speed and high gain, an important demand on this circuit is to keep the output voltage swing constant despite great changes in the input voltage swing, i.e. a wide dynamic range. This demand can be met either by automatic gain control (AGC) or limiting amplifiers.

3.5.1 Limiting Amplifiers

The advantages of limiting amplifiers are the higher operating speed, lower power consumption, lower supply voltage, easier design, smaller chip area and fewer external components. However, the disadvantages of the limiting amplifier is that the output no longer has a linear relationship with the input, resulting in higher jitter for clock recovery and hard to do input offset cancellation because of the non-linearity.

In order to achieve the high gain required at high data rate, cascade of intermediate gain stages is usually used. One approach is the used of transadmittance (TAS) and transimpedance (TIS) stages as proposed by [43] and shown in Figure 3-10. A limiting amplifier using a cascade of

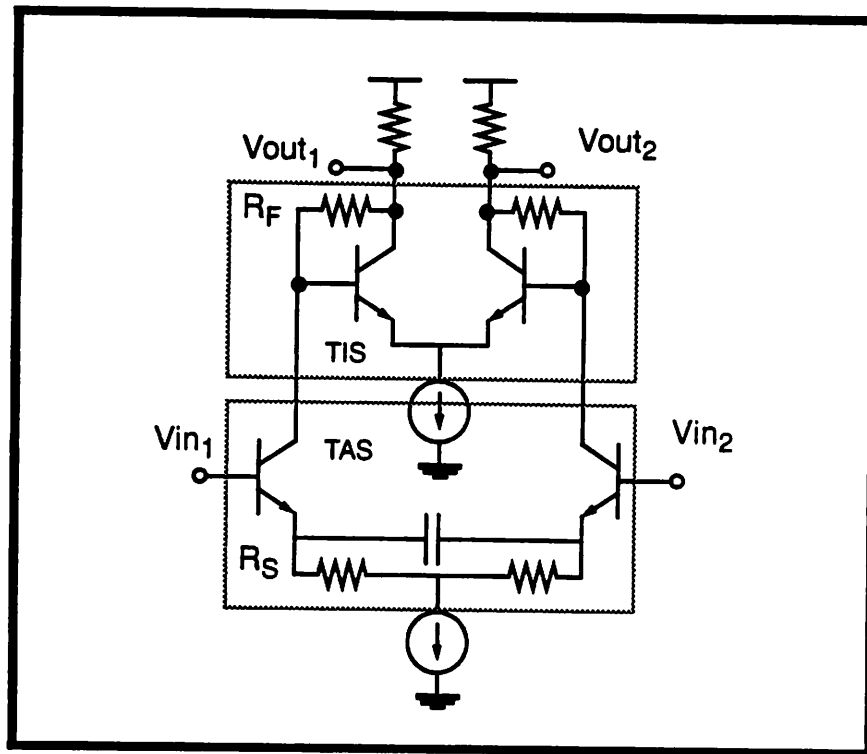


Figure 3-10 Basic Cell for Wide-band Amplifiers

the dual feedback gain stage discussed was reported by [44]. It uses a cascade of input follower with 3 amplifier cells similar to the one in Figure 3-10 (omitting R_S), and followed by an output buffer stage. With an 8.2GHz bipolar technology, the limiting amplifier can operate at about 4Gb/s with maximum voltage gain of 54dB, 52dB dynamic range (1-400mV_{p-p} input), and a fixed constant output of 400mV_{p-p}.

3.5.2 AGC Amplifiers

For high performance optical fiber datacommunication systems, linearity has to be maintain throughout the linear channel. It is required for low jitter clock recovery, coherent optical reception, and other techniques requiring further processing of the received analog signal. In order to maintain a constant output voltage, wide dynamic range AGC amplifiers are needed.

The major disadvantage of AGC amplifiers are more complex circuit, and the need of extra circuitry forming a feedback loop to control the amount of gain needed, to maintain a constant output voltage.

Most wide-band variable-gain amplifiers can be derived from four high-performance amplifiers which all consists of a quadruple of transistors driven by an input pair. They are the AGC amplifier [45], the multiplier [46], the variable series feedback amplifier [47], and the Gilbert variable-gain quad [48] as shown in Figure 3-11. All of them are operating in the principle of switching the loading seen by a differential pair through the switching of the current path of the loading devices.

In field-effect-transistor (FET) technology, direct application of the above principles results in extremely poor distortion performance because of the square law nature of a FET [49]. Most FET variable gain amplifiers utilize the fact that by operating the FET in the triode region, a voltage-controlled variable resistor can be realized. Several realizations of grounded and floating resistors can be found in [50]. Variable gain amplifiers can be realized with the voltage-controlled resistors as loading devices. For an example, the small signal resistance of a MOSFET biased in the triode region can be expressed as

$$\frac{1}{r_{ds}} = \frac{\partial I_D}{\partial V_{DS}} = \mu C_{ox} \frac{W}{L} (V_{GS} - V_t - V_{DS}) \quad (\text{EQ 3-13})$$

When V_{DS} is small, the resistance is inversely proportional to $V_{GS} - V_t$. The range of resistance that can be realized is thus limited by the practical range of $V_{GS} - V_t$ that can be applied. For $V_{dd} = 5V$, $V_t = 1V$, max. $(V_{GS} - V_t) \approx 3V$, min. $(V_{GS} - V_t) \approx 0.5V$, there can only be a 6:1 variation in resistance. Therefore the drawback of this scheme is a low variable gain range. For a larger gain range, cascade of several variable gain stages is needed [51]. Nevertheless, the resistance is highly linear and if driven symmetrically with differential input, it can be shown that the resistance is linear.

Two commonly used configurations are shown in Figure 3-12 (e) and (f).

For example, Figure 3-12 (e) is used in [51]. The gain is $gm_1 r_{ds}$ where r_{ds} is the variable

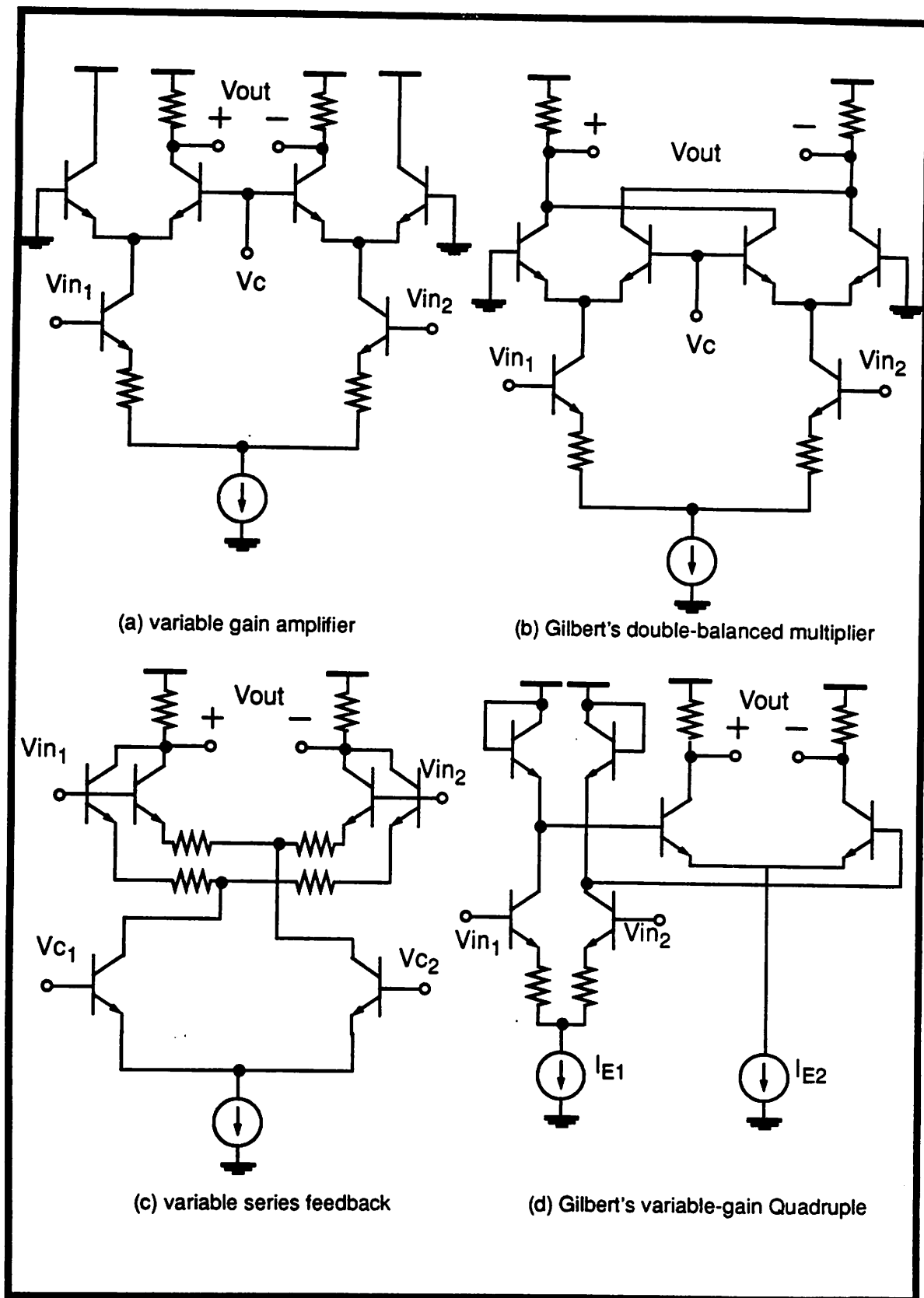


Figure 3-11 Principles for Gain Control in Wide-band Amplifiers

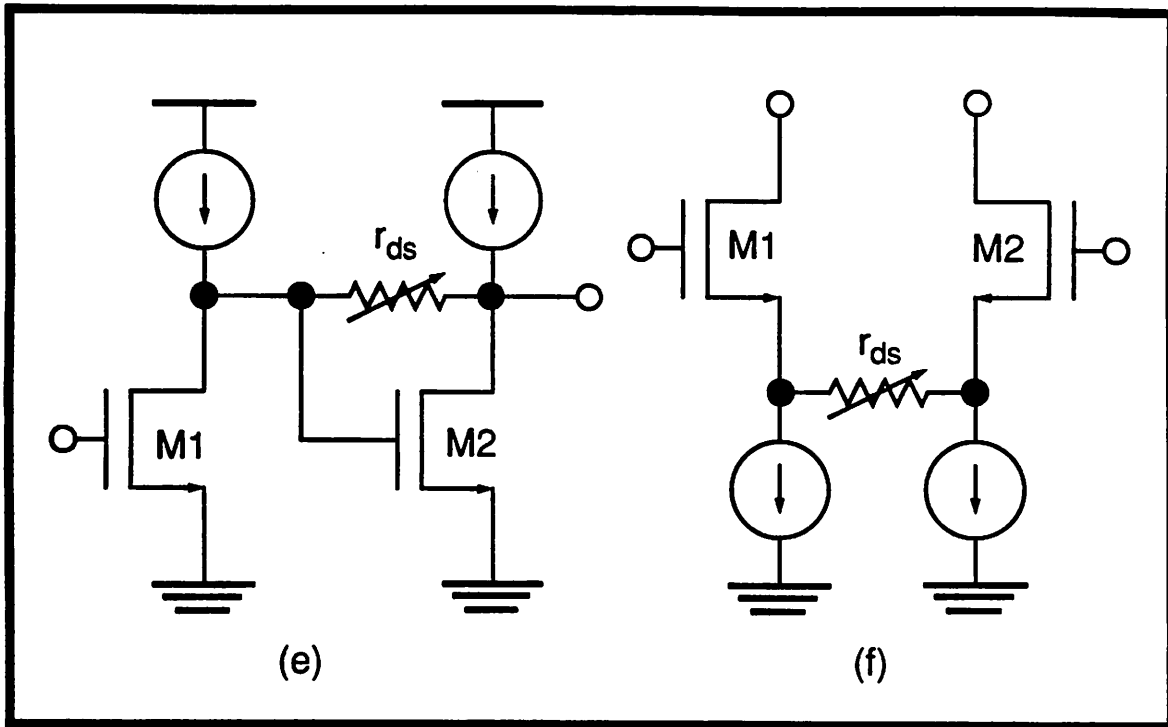


Figure 3-12 Commonly used FET variable gain amplifier building blocks

resistor implemented by a MOSFET biased in the triode region. The drawback of this circuit is that it is single-ended and a DC feedback must be used to bias up the input MOSFET M1. Therefore it can only be used by AC coupling the input. A differential version can be realized but the circuit is much more complicated and is not an area efficient implementation. The circuit in Figure 3-12 (f) operates differently. It is modulating the source degeneration to change the G_m of the differential pair. The G_m can be expressed as

$$G_m = \frac{g_{m1}}{1 + 2g_{m1}r_{ds}} \cong \frac{1}{2r_{ds}} \quad (\text{EQ 3-14})$$

This can be used to convert an unbalanced inputs into differential signals (as input for the differential version of Figure 3-12 (e)), and the input can be DC coupled in. The drawback is that the distortion is higher and additional loading devices must be used to convert the differential currents back to voltages.

Examples of AGC amplifiers using different technologies and different principles are listed in Table 3-4.

BW (GHz)	Dynamic Range (dB)	Output (mV _{p-p})	Technology	Device f_T (GHz)	Bit Rate (Gb/s)	Method	Reference
2.5	40	400	Si BJT	8	3	(b)	R. Reimann [52]
N/A	40	N/A	CMOS	2.5	0.05	(a)+(b)	D. M. Pietruszynski [53]
11.4	20	800	Si BJT	40	10	(a)	T. Suzaki [34]
1.3	26	N/A	Si BJT	25	1.6	(c)	Y. Akazawa [54]
1	70	250	NMOS	9	0.88	(e)	R. P. Jindal [51]
1.6	13	N/A	GaAs	25	N/A	(e)	T. Imai [55]

Table 3-4 Performance of AGC Amplifiers in Different Technologies

A separate feedback loop, as shown in Figure 3-13, is needed to generate the control voltage for the variable gain amplifier to keep the output voltage constant. This is usually done by a peak detector, generating a signal corresponding to the peak-to-peak voltage difference of the output. It is then compared with a desired reference signal. The difference between the two is passed through a low-time-constant integrator to generate the control voltage.

3.6 Clock Recovery

The purpose of clock recovery in the optical receiver is to recover a clock at the input bit rate or a multiple of the input bit rate for use in the synchronization of the decision circuit to regenerate the digital data from the analog waveform received. The strength of the timing information in a signal depends on the statistics of the signal, the line code, and the pulse shape. Unfortunately, no practical clock recovery circuits can duplicate the clock used at the transmitter perfectly. The best one can do is to have the average frequency of the recovered clock equal to the average frequency of the transmitted signal. In a real system, pulses arrive at times differ from the integer multiple of the bit rate T . These unwanted pulse position modulation of the pulse stream is called jitter [56]. The recovered clock has an accurate average frequency, but in itself, also has instantaneous phase jitter which can be reduced to any desired level with a cost. The difference between

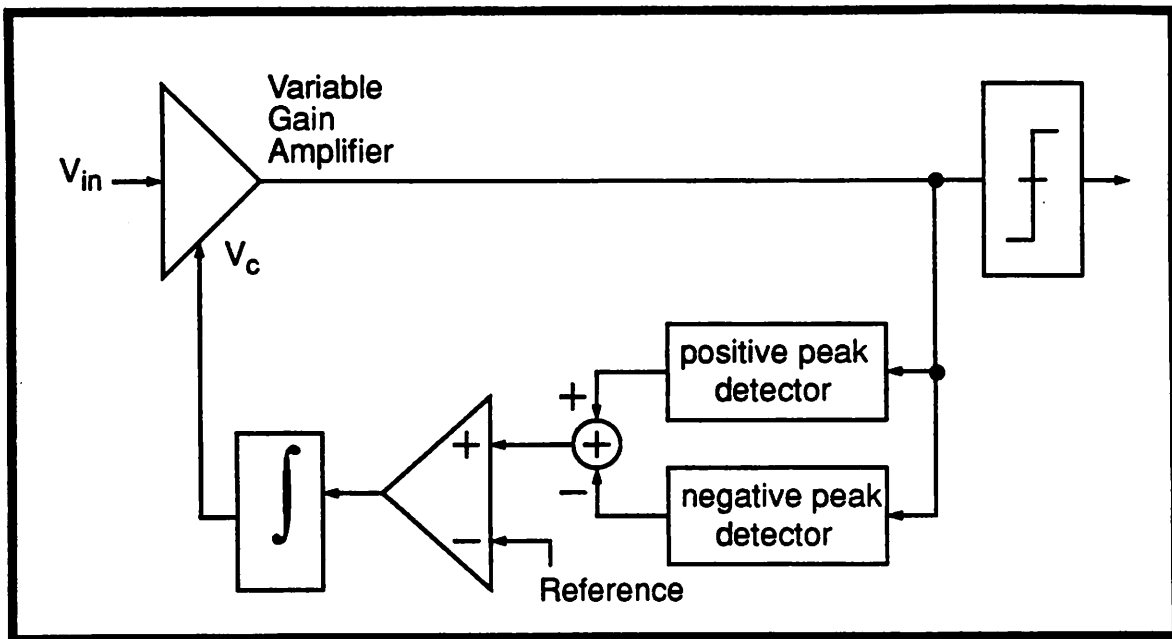


Figure 3-13 Automatic Gain Control (AGC) Loop

these two jitters is called the alignment jitter. It is this alignment jitter that affects the decision making process, and if not controlled, can cause decision errors.

Making a correct bit decision is dependent on the received pulse shape, the decision threshold, the bit sequence before and after the present bit decision, the alignment jitter, the static phase offset, and of course, noise. The main degradation caused by alignment jitter is that the input analog signal is sampled at a non-optimal point in the eye, decreasing the magnitude of the sampled data, increasing the intersymbol interference (ISI) and thereby reducing noise immunity. For pulse code modulation (PCM) transmission, the timing jitter of the recovered clock causes non-uniform sampling results in distortion of the reconstructed signal. For long distance transmission, repeaters are used. Each repeater introduces its own timing jitter. The accumulation of timing jitter after a number of repeaters must be accounted for in the design of clock recovery circuits.

Clock recovery can be done basically in two ways - deductive and inductive [57] as shown in Figure 3-14. Deductive clock recovery directly extracts from the incoming signal a timing tone which has an average frequency exactly equal to the input bit rate. Inductive clock recovery does not directly process the received signal to get a timing tone, but rather uses a feedback loop. A current estimate of the timing tone is used to sample the analog signal input. Then the timing

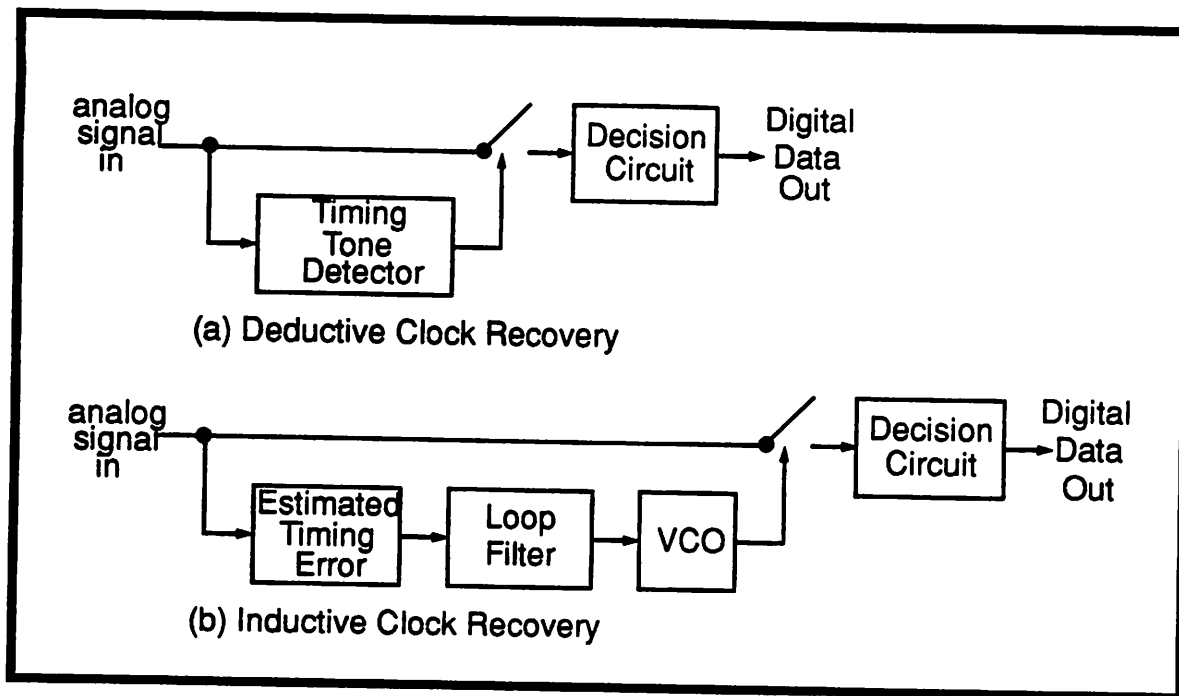


Figure 3-14 Generic Clock Recovery Methods

error is estimated and the timing tone estimated is updated. This is effectively a phase-locked loop [58] that will be discussed in detail in Section 3.6.2.

3.6.1 Spectral-Line Method

The commonly used deductive method is the spectral-line method as shown in Figure 3-15. Since most transmitted signal has zero DC value in order to reduce power penalty, the timing

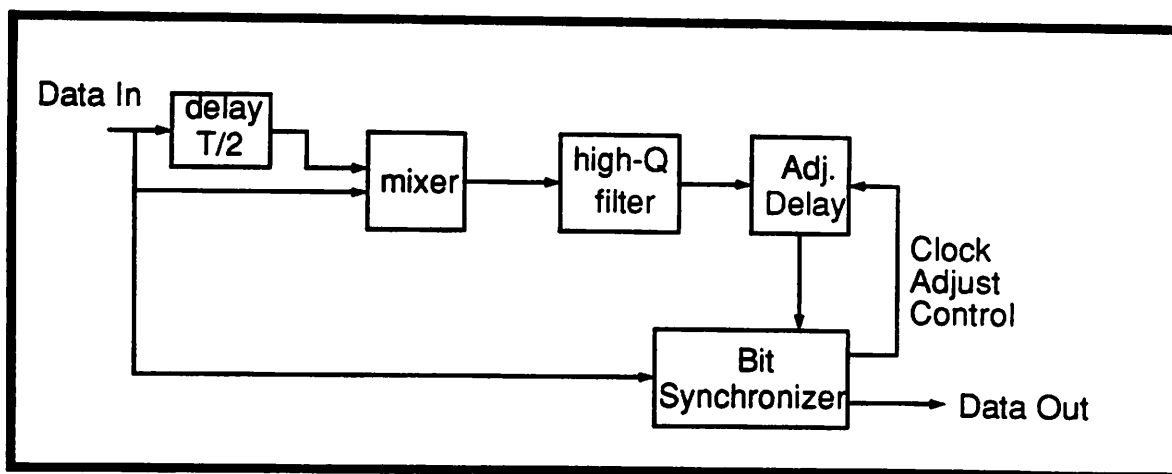


Figure 3-15 Spectral Line Method

signal cannot be deduced from the transmitted signal itself but from its higher moments. As a result, a non-linear circuit such as a mixer must be used. Timing tone can be obtained by letting the input signal mixing with itself and then passing the result through a high-Q bandpass filter [59]. There are quite a few disadvantages associated with this method. For high-data-rate receiver, a wideband mixer with bandwidth close to or exceeding the input data rate is required because of the excess bandwidth needed to get a stronger timing signal. The high-Q bandpass filter is usually implemented by a high-Q tank circuit [60] or surface acoustic wave (SAW) filter [61]. A high-Q filter is expensive, cannot be integrated with the receiver chip, requires large amount of power to drive, and can only be tuned to one center frequency. An adjustable delay is also needed to align the sampling instant because the output is not phase aligned with the input data [62]. The advantage of this approach is that it is opened-loop and therefore, simple and high speed. Also, because of the narrow bandwidth provided by the high-Q filter, it has a high jitter rejection.

3.6.2 Phase-Locked Loop (PLL)

The advantage of inductive method is that most of the timing recovery can be done in discrete-time. The disadvantage is that the sampling rate of the received signal might have to be higher than the input bit rate in order to estimate the timing error. Baud-rate technique is available but results in more complicated design.

The inductive methods are essentially phase-locked loops with different ways to implement the phase detector. A phase-locked loop contains three basic components, a phase detector (PD), a loop filter and a voltage-controlled oscillator (VCO) as shown in Figure 3-16.

The phase detector compares the phase of a periodic input signal against the phase of the VCO; output of the PD is a measure of the phase difference between its two inputs. The difference voltage is then filtered by the loop filter and applied to the VCO. Control voltage on the VCO changes the frequency in a direction that reduces the phase difference between the input signal and the local oscillator. When the loop is locked, the control voltage is such that the frequency of the VCO is exactly equal to the average frequency of the input signal.

There are two broad categories of phase detectors: analog and digital phase detectors.

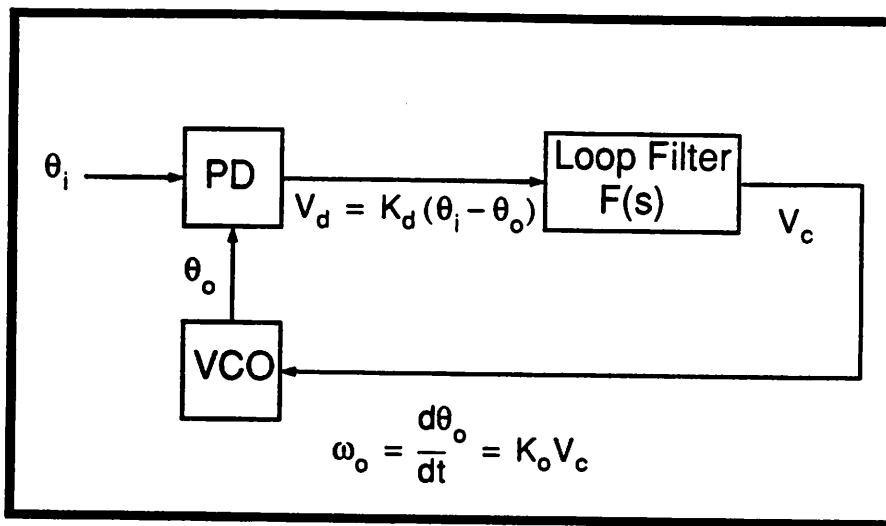


Figure 3-16 Phase-Locked Loop

An analog phase detector is usually a wide-band mixer. When the input signal is mixed with the local oscillator, a DC error term proportional to the phase error can be obtained. The simplest form of a digital phase detector is an XOR. When the signal is strong, the multiplier can be replaced by a digital XOR gate giving the same result. Sequential circuits can also be used as phase detectors. A sequential phase detector operates on the zero crossings of the signal and the local oscillator.

Using a PLL for clock synchronization has two major advantages. One is the fact that it can be implemented as a monolithic IC, and second, its dynamic properties can easily be adjusted by a proper choice of the loop filter. The major disadvantage of a PLL is that the pull-in range, which is the range of frequency in which the PLL can capture and lock on, is not much greater than the noise bandwidth. For high jitter rejection, as that obtained through the SAW filter approach, a narrow bandwidth has to be used. As a result, an acquisition aid is indispensable for clock recovery. This leads to a clock recovery approach based on a phase and frequency detector (PFD) as part of a phase- and frequency-locked loop (PFLL).

Again there are two ways to implement a PFD: analog and digital. Analog methods usually based on correlations of the inphase and quadrature component of the signal and the internal VCO [63]. Digital methods rely on memories of past crossing events to do frequency detection [64]. For most cases, the digital method is a lot simpler because digital circuits can be used. However, for high-data-rate frequency detection, the digital sequential circuits approach is

usually too slow and cannot be used.

3.6.3 Charge-Pump Phase-Locked Loop (CPPLL)

When the PLL is locked, the control voltage is such that the frequency of the VCO is exactly equal to the average frequency of the input signal. To maintain the control voltage needed for lock, it is generally necessary to have a nonzero output from the phase detector. Consequently, the loop operates with some static phase error present. One implementation of the PLL known as the Charge-Pump Phase-Locked Loop (CPPLL) can eliminate this static phase error. A Charge-Pump Phase-Locked Loop is shown in Figure 3-17.

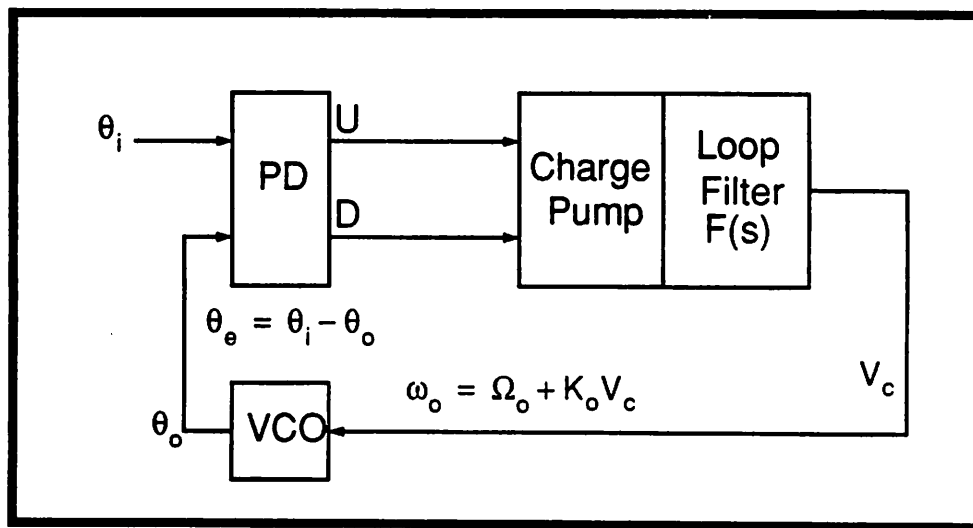


Figure 3-17 Charge-Pump Phase-Locked Loop

Just as a conventional PLL, a lot of the times, the PD used in the CPPLL is capable of detecting frequency error by using a phase/frequency detector (PFD) [65]. There are three states for the output of the PD; Up (U), Down (D) and neutral (N) when the output is neither U nor D. The output of the PD is then used to control a charge pump to convert the digital output into analog signals suitable for controlling the VCO. The advantage of the CPPLL is the frequency-aided acquisition, the extended tracking range, the ease of implementation, and no static phase error.

3.6.4 Wide-Band Clock Recovery

So far, the clock recovery methods mentioned can all be classified as narrow-band systems. A “narrow-band” system is refer to the narrow bandwidth of the PLL or the High-Q filter. The advantage of a narrow bandwidth system is the high jitter rejection. However, because of the narrow bandwidth, it requires a lot of transitions to lock. In applications where jitter rejection is not a major concern but the fast tracking is the desired goal (e.g. burst-mode channels like ATM networks), a wide-band system can be used. A Wide-band system can be a PLL with a much wider bandwidth but this approach is not suitable for high speed systems. For high data-rate, an opened-loop system like [66] is preferred.

3.6.5 Circuit Requirements

No matter what method is used, clock recovery is a very demanding function for a high-data-rate receiver. A wide-band mixer must be used for phase detection or generating correlations. A wide-band voltage-controlled oscillator (VCO) is needed to generate a clock at the frequency of the input bit rate. High speed digital circuits are needed if the phase detector (PD) or phase-frequency detector (PFD) is implemented digitally. For example, some clock recovery designs for bit synchronization reported are list in Table 3-5.

Bit Rate (Gb/s)	Technology	Device f_T (GHz)	Recovery Method	PD/PFD Method	Reference
10	AlGaAs/GaAs/AlGaAs FET	45	SL	digital	P. Wennekers [70]
8	Si BJT	12	PLL	analog	A. Pottbaker [69]
2.5	GaAs	28	PLL	analog	H. Ransijn [71]
2.3	Si BJT	30	PLL	analog	M. Soyuer [76]
1.5	GaAs	20	SL	none	P. Wallace [75]
1.1	NMOS	8	PLL	analog	S. K. Enam [72]
0.66	Si BJT	10	PLL	digital	B. Lai [67]
0.66	CMOS	3	opened-loop	digital	M. Banu [66]
0.3	Si BJT	2.5	SL	none	G. E. Andrews [73]

Table 3-5 Clock Recovery Circuits in Different Technologies

Bit Rate (Gb/s)	Technology	Device f_T (GHz)	Recovery Method	PD/PFD Method	Reference
0.266	CMOS	5	PLL	analog	D. L. Chen [77]
0.155	Si BJT	3.5	PLL	digital	L. DeVito [68]
0.12	Si BJT	5	PLL	digital	J. Tani [74]

Table 3-5 Clock Recovery Circuits in Different Technologies

3.7 Decision Circuit

The function of the decision circuit is to regenerate the digital bits from the amplified analog output of the main amplifier. This can be done by either a regenerative latch strobed by the recovered clock, or in a limiting amplifier whose output is sampled with the appropriately delayed recovered clock. The most common way is to use a Master-Slave-Flip-Flop (MSFF) with an optional preamplifier as shown in Figure 3-18. Improvement can be made by using other

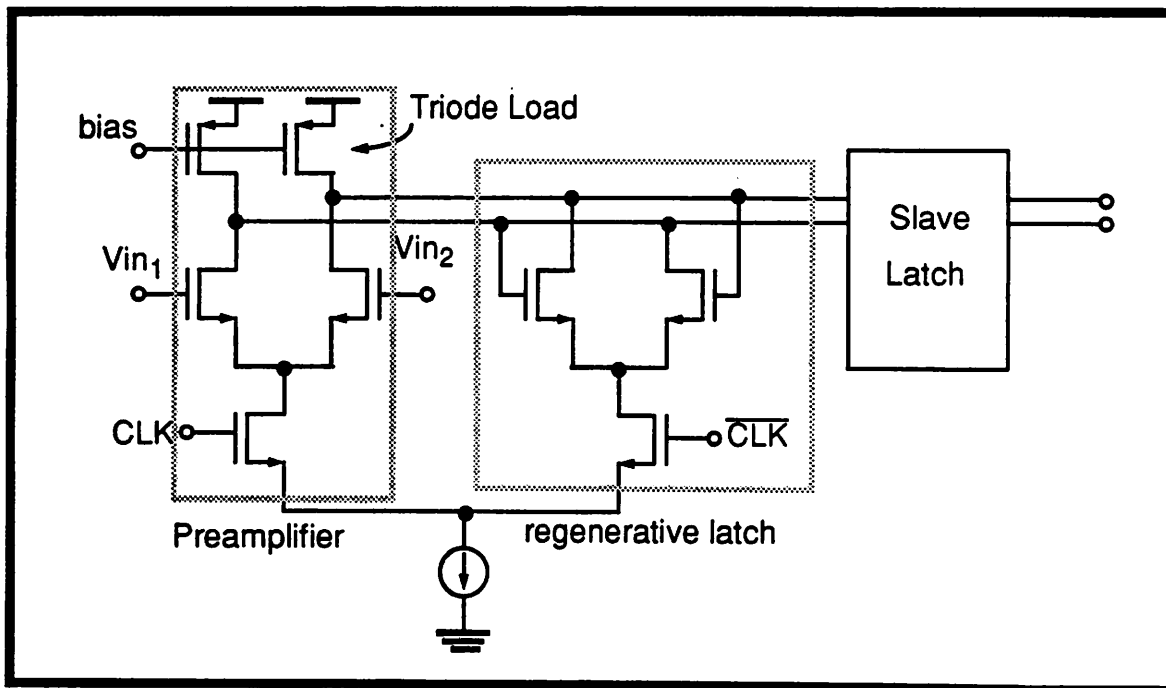


Figure 3-18 Master-Slave-Flip-Flop with Preamplifier

configurations [72] for implementing the MSFF, and interleaving the operation of two data regenerator [78]. This functional block can be implemented at very high speed and is usually not a

bottleneck for the overall data rate of the receiver. The highest bit rate reported so far is a 25Gb/s decision circuit implemented in a 45GHz Silicon Bipolar technology, using just a simple MSFF [9].

3.8 Demultiplexer (DEMUX)

After regeneration, the high-data-rate bit stream is deserialized into a parallel form by a demultiplexer. The demultiplexing can be done by a serial-input parallel-output shift register with the MSFF as basic cell. The speed of this configuration is limited by the fact that the clock frequency of the serial operating MSFFs must equal the input bit rate. Higher bit rate can be achieved by using a one-stage demultiplexing scheme as shown in Figure 3-19. Even higher data rate can

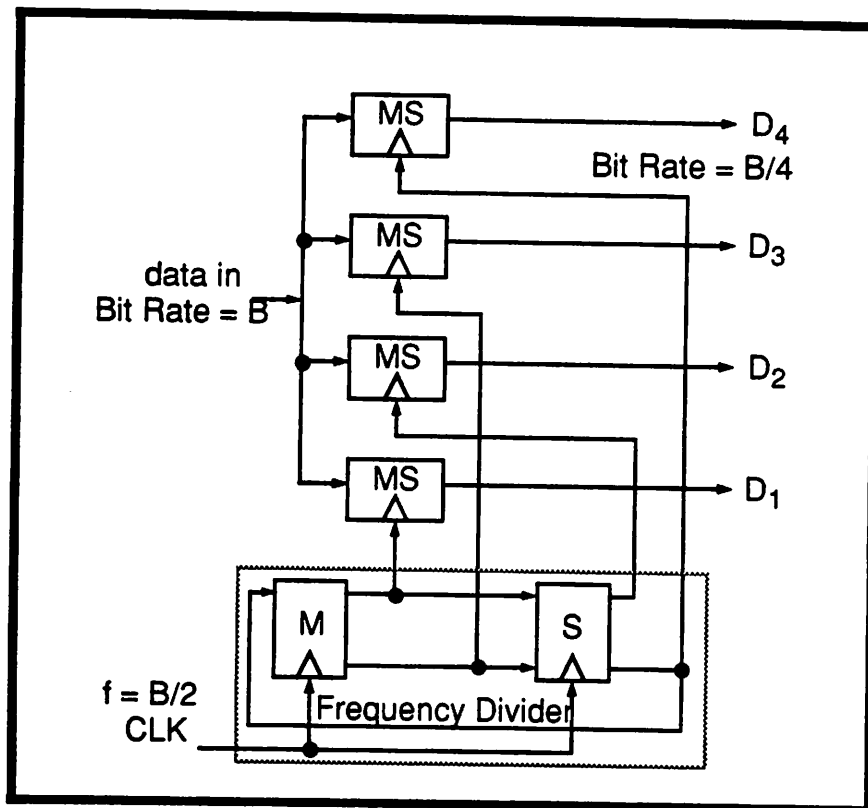


Figure 3-19 Block Diagram of a One-stage 1:4 Demultiplexer

be achieved by using a two-stage demultiplexing scheme, but two more MSFF and an adjustable delay line have to be used [79]. A frequency divider is needed to provide the right clocking for this demultiplexing scheme. A frequency divider is nothing but a cascade of DFFs to divide the input clock frequency. Therefore, the frequency divider's maximum speed is the same as the decision

circuit [9]. The maximum bit rate for a 1:4 demultiplexer is 40Gb/s with the same 45GHz Silicon Bipolar technology [9].

3.9 Limitations of Traditional Architecture

The digital components of traditional receivers (demultiplexers, decision circuits, and frequency dividers), are easily realizable even at very high data rate with a moderate speed technology. However, the analog components (photoreceiver, main amplifier, and clock recovery circuits), impose a design challenge. The key problem is that the bandwidths of every analog blocks have to be about the same as that of the input bit rate.

The operating bandwidth of the analog blocks in a receiver depends on the speed of the technology used, and the circuit technique used. One way to make relative comparison of different circuit implementations is to normalize the data rate achieved with the device f_T of the technology. A plot of the “normalized” bit rate against the bit rate of different implementations of the clock-recovery function listed in Table 3-5 is used as an example, and is shown in Figure 3-20 . Care has to be taken while reading this plot. Only clock recovery circuits with on-chip VCO, using PLLs are included but they may still have different system requirements. Also, the device f_T of a FET technology depends on biasing conditions. However, the plot is still useful for comparing different circuit techniques independent of the technologies used. The higher a data point is on the plot, the better the technology is utilized.

For most implementations (all except [72]), a device f_T to data rate ratio of at least 12:1 is needed. This implies high speed technologies such as III-V semiconductor technologies (GaAs), high f_T Si bipolar technologies, and other special high speed technologies (fine-line-NMOS) are needed to implement a high-data-rate receiver. The draw back of these technologies are high cost, and low level of integration. As a result, present solutions to a complete receiver function are either a multi-chip approach or a hybrid approach with a high manufacturing cost due to the cost of the high speed technology, package, and assembly required.

One solution for lowering the cost of a receiver is to utilize the lower-cost and higher-integration of CMOS technology. This is a preferred solution for low-data-rate receivers [77]. However,

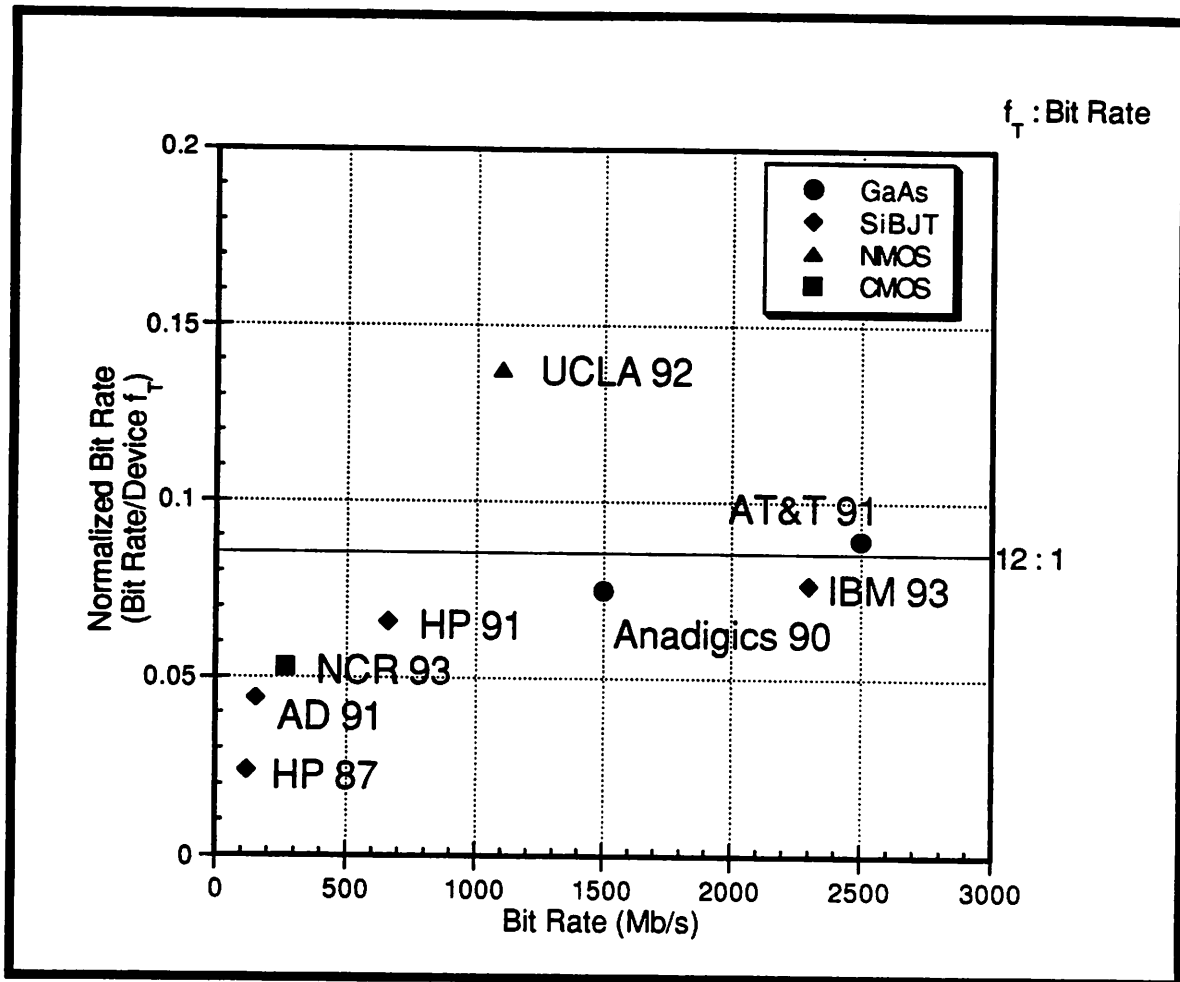


Figure 3-20 Plot of Normalized Bit Rate vs Bit Rate for Different Technologies

for high-data-rate receivers, the inherent speed of a production CMOS technology is usually 2 to 3 times slower than that of the bipolar counter part. None of the existing clock recovery circuits utilizing PLLs, implemented in CMOS technology, can break the 500Mb/s barrier (even with existing sub-micron technologies). This limits the application of CMOS technology in the implementation of Gb/s receivers. A new architecture must be derived to relax the speed requirement imposed on the technology for implementation.

Chapter 4 Parallel Receiver Architecture

4.1 Introduction

One way to improve throughput of a low speed technology is the use of parallelism. The bottleneck in the speed performance of the traditional architecture is in the analog blocks, namely the photoreceiver, the main amplifier, and the clock recovery circuit. Demultiplexing in the digital domain can be performed at a much higher data rate. As a result, if the demultiplexing function can be moved as early as possible to the receiver front-end, the speed requirement on the following analog blocks is greatly relaxed.

A single chip solution for the whole receiver function with CMOS technology is impossible for long-wavelength optical fiber communication systems because of the low responsivity of Silicon in the 1.3- to 1.55- μm wavelength range. As a result, the photodetector cannot be implemented with Silicon efficiently. An OEIC implementing the photoreceiver may be the best solution as discussed in Section 3.4, and will be assumed in this work. Therefore, the demultiplexing function should be moved to right after the photoreceiver.

A complete solution may be an OEIC implementing the photoreceiver, followed by a CMOS mixed-signal VLSI chip implementing the rest of the receiving functions and the high level data link control as shown in Figure 4-1. The key idea is to move the demultiplexing function right to the front-end after the photoreceiver. Unlike in the traditional architecture, the demultiplexing is now done in the analog domain because the input is a low level voltage signal. The high-data-rate bit stream is demultiplexed right-a-way into several low-data-rate bit streams and then processed by different parallel channels. Each parallel channel has its own amplifier and decision circuit. The

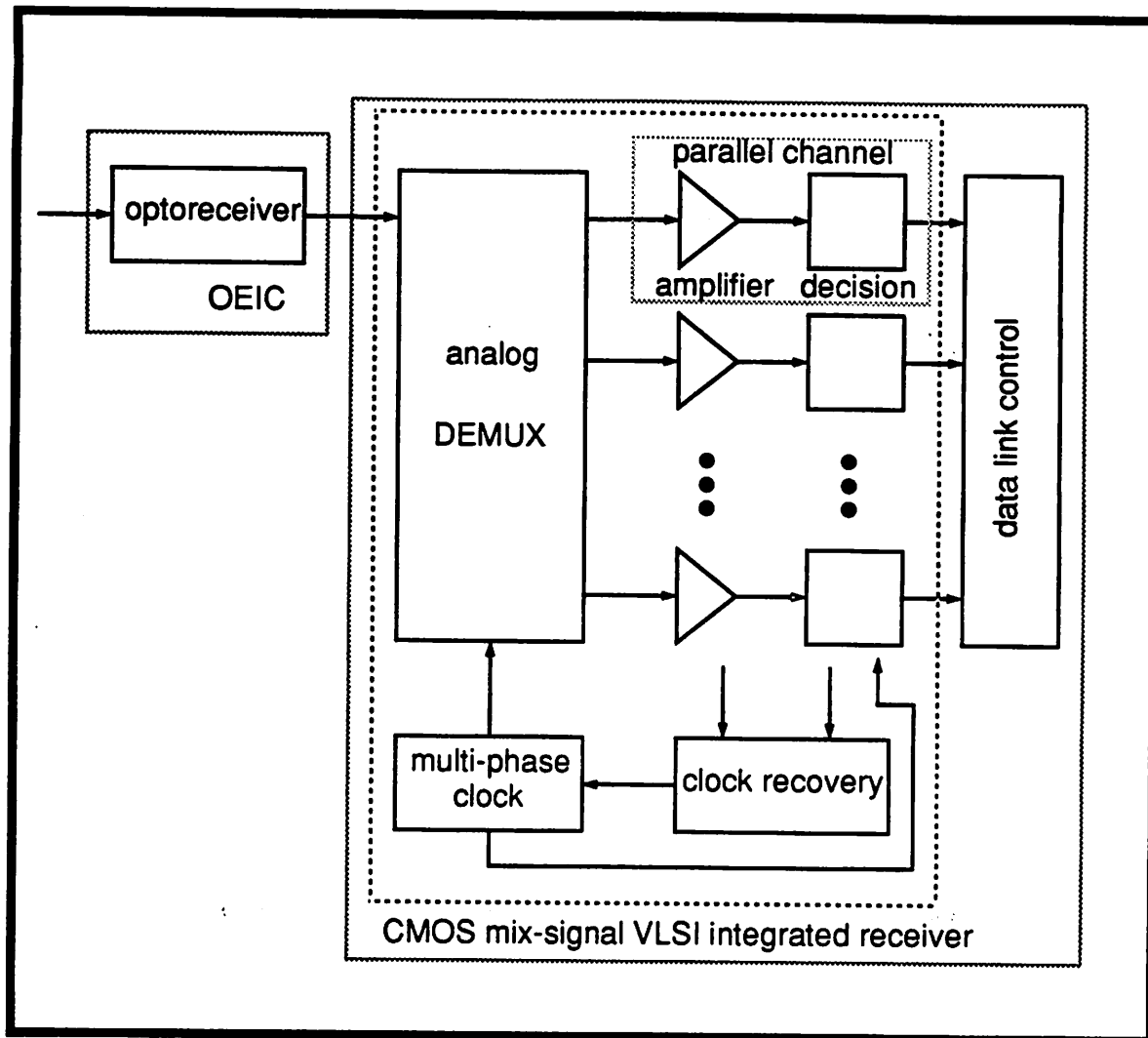


Figure 4-1 Block Diagram of a Parallel Receiver

output of each parallel channel is passed to the data link control for high level manipulation of the data received. Clock recovery is done based on information from the parallel channels. A multi-phase clock is used to strobe the input demultiplexer and the decision circuits in each parallel channels.

4.2 Analog DEMUX

For most direct detection digital datacommunication applications, a Non-Return to Zero (NRZ) code is usually used. The analog input demultiplexing of the output from the low noise preamplifier can be done by a series of sample-and-hold circuits, controlled by multi-phase clock edges, each separated by exactly one bit period of the input data apart. When the multi-phase

clock edges are aligned perfectly with the input data, the analog sample-and-hold front-end will sample each input data bit at the center of the bit cell.

The output voltage of the preamplifier is at a very low level. Special care has to be taken to demultiplex signals at this low level at high data rate. Also, the number of parallel stages used depends on the application. For example, if the target is for Synchronous Optical Network (SONET) application [7], the basic bit rate is 51.84Mb/s. For a SONET OC-12 application, the bit rate is then $12 \times 51.84\text{Mb/s} = 622.08\text{Mb/s}$. For such application, 12 parallel channels each running at a demultiplexed bit rate of 51.84Mb/s should be used. The bit rate that can be achieved with the parallel architecture is ultimately limited by the sample-and-hold front end and how closely spaced the clock edges can be generated.

The input sample-and-hold circuit is the only function in the whole architecture that is required to have a bandwidth exceeding the input data rate. Every other functions can be implemented with less bandwidth.

The input referred offset on the sampling node is mainly due to the sample-and-hold offset and the channel amplifier offset. The whole analog DEMUX and channel amplifier have to hold this total offset to below the minimum signal output by the preamplifier. The output from the preamplifier is usually in the range of 5mV to 15mV. As a result, the circuit has to hold the total input offset on the sample node to below 5mV for sensing the low level output from the preamplifier.

A Bit Error Rate (BER) of 10^{-9} to 10^{-11} is usually required in a lot of the system applications. As a result, a signal to noise ratio of 25dB or above is needed to achieve this BER. On top of that, it also has to handle the coupling noise due to the on-chip VCO, drivers, and digital circuits.

4.3 Multi-Phase Clock

A low-jitter, voltage-controlled, evenly-spaced, multi-phase clock has to be generated to strobe the analog demultiplexer. It can be done as in traditional architecture using a high frequency VCO with a frequency divider. However, this defeats the purpose of eliminating high frequency components in the parallel architecture. A perfect solution will be the use of a ring

oscillator. The multi-phase clock edges are taken from taps of the ring oscillator. The oscillation frequency will then only be $1/n$ of the input bit rate where n is the number of parallel stages used.

The ring oscillator must be able to provide even number of edges. This requirement alone dictates the use of differential ring oscillators. The clock edges must be evenly spaced for sampling at the center of the bit cell for every channel to reduce the effect of inter-symbol-interference (ISI). The jitter of the sampling clock can cause degradation in decision, the same way as in a traditional architecture. Assuming a band-limited channel so that the input data is sinusoidal, an alignment jitter of 40° causes a 0.5dB reduction in the magnitude of the data sampled. As a result, the jitter of the ring oscillator must be below 30° to reduce the contribution to the alignment jitter.

4.4 Parallel Channel

Once the high-data-rate bit stream is demultiplexed into low-data-rate bit streams, the parallel channel works the same way as in the traditional architecture, with the exception that now it is working at a much lower rate than the input bit rate. With the parallel architecture, the demultiplexed bit rate of each parallel channel is limited by the amplify-and-latch circuit in the data channel. A fast implementation with wide dynamic range is desired, just as in traditional architecture, to reduce the number of parallel stages required to achieve a certain input bit rate.

The parallel architecture relaxes the requirement in speed for the amplify-and-latch circuit. In the best case, the overall throughput should be increase by n times. However, due to practical implementation of the analog demultiplexing and clock recovery, the factor is usually reduced to $n/2$. The AGC amplifier should be able to handle a bit rate $2/n$ times the input bit rate.

Using the data from the previous chapter, the device f_T to data rate ratio of a demultiplexer and a decision circuit is about 2:1; the ratio for the AGC amplifier with reasonable gain and dynamic range is about 11:1. As a result, the best a parallel receiver can do is to use 12 parallel stages with an overall ratio of 2:1 as limited by the demultiplexer. However, this is not completely right as demultiplexing in digital domain is different from demultiplexing in analog domain where both speed and sensitivity have to be considered. A more realistic overall ratio will be 4:1, again as limited by the analog demultiplexer.

The received signal power of a fiber optic communication channel can vary as much as 40dB in electrical range. A dynamic range of 40dB or more is desired. In order to accommodate the low level output voltage signal of the preamplifier which can be as low as a few mVs, the maximum gain must be at least 100.

The input offset of the amplifier and the offset from the input sample-and-hold must also be handled by the channel amplifier. An offset cancellation scheme must be used to lower the effective total input offset when the input signal is small.

The variable gain amplifier must be simple and small in area as it is replicated many times in the parallel architecture. The implementation must be differential to minimize the common mode and supply noise and can be DC coupled.

4.5 Clock Recovery

The parallel architecture also relax the requirement of clock recovery by decoupling it from the data channel. Phase detection doesn't have to be done in one bit period as in the traditional architecture. However, the input waveform is demultiplexed and sampled. As a result, only samples of the input waveform is available for clock recovery. A new algorithm has to be derived to align the clock edges of the multi-phase clock to the center of the bit cells. The approach should be a closed-loop narrow bandwidth architecture for flexibility and high jitter rejection. Jitter tolerance of 0.15UI is expected at jitter frequency as high as the input bit rate.

Phase detection using a narrow bandwidth closed-loop approach results in a narrow capture range. Self-acquisition of frequency tends to be slow and often unreliable. Means must be provided for initial acquisition. This can be accomplished by many means such as frequency sweeping, frequency discriminators, and bandwidth widening.

Initial acquisition by frequency sweeping is done by sweeping the frequency of the VCO until the VCO frequency gets close to the desired frequency (signal frequency divided by the number of parallel channels for the parallel receiver) and the loop locks up. This can be done by generating a sweep current into the loop filter to generate a ramp output voltage to the VCO. Slew shut

off is necessary once lock is acquired. This method is slow and susceptible to false lock for wide variation of frequency and is therefore not recommended.

Bandwidth widening is done simply by using a much larger bandwidth during initial acquisition to widen the capture range and switch back to a narrow bandwidth once lock is acquired. This can be achieved by changing either the loop gain or the loop filter. However, the lock range is still limited.

The most commonly used method is frequency discrimination. It is done by introducing another loop that can detect the frequency. Phase locking occurs when the frequency error is brought within the lock limit. This method requires a separate frequency detector [64]. This can also be done by locking the VCO to an external known reference which has the average frequency of the input data divided by the number of channels and transfer the control back to data upon data reception. With the extra frequency detector, the lock range is then only limited by the range of the VCO.

4.6 Improvement in Performance

By identifying the speed difference in demultiplexing and analog processing implemented in the same technology, the parallel architecture removes the bottleneck of the traditional architecture by reducing the bandwidth requirement of the main AGC amplifier; eliminating the need of a high frequency VCO and frequency divider; decoupling the clock recovery from the high-data-rate bit stream so that it doesn't have to be done in one bit period. This architecture can greatly reduce the device f_T to data rate ratio required for the implementation of high-data-rate receivers, thus allowing Gb/s receivers implemented in CMOS technology. The ultimate device f_T to data rate ratio that can be achieved with reasonable sensitivity is about 4:1. With a given technology, the speed performance should be comparable to an opened-loop SAW filter implementation, while at the same time has all the advantages of a closed-loop system.

The disadvantage of this architecture is that more hardware is needed results in larger chip size and higher power dissipation. Nevertheless, because of the use of highly integrated CMOS technology with most of the circuits operating at a much lower rate, the chip area and

power consumption is comparable with implementations in other high speed technologies.

In order to prove the concept of the parallel architecture, a prototype implementing the functions in the dotted box in Figure 4-1 was design and fabricated. A detail description of the implementation issues and experimental results will be discussed in the following chapters.

Chapter 5 Parallel Receiver Implementation

5.1 Introduction

This chapter describes in detail the implementation of a 480Mb/s serial-to-parallel-conversion/AGC-amplifier/decision/clock-recovery prototype in a 1.2- μ m CMOS technology. A simplified, single-ended equivalent block diagram of the prototype implemented with 8 parallel channels is shown in Figure 5-1. The whole system may consist of a combination of an Opto-Electronic Integrated Circuit (OEIC), implementing the photodetector and the low noise preamplifier. It is then followed by a CMOS VLSI chip that implements the rest of the functions all on one chip.

For this experiment, the prototype implements the serial-to-parallel conversion, the variable gain amplification, decision, and clock recovery functions all on one chip. It uses a technology that is fully compatible with digital VLSI technology so that a back-end data link controller can easily be integrated with the parallel receiver to form a low-cost, high-integration solution to the implementation of terminal receivers.

The output waveform from the low noise preamplifier is also shown in Figure 5-1. The analog input demultiplexing of the output from the low noise preamplifier is done by a series of CMOS sample-and-hold circuits with switches controlled by multiphase clock edges, each separated by exactly one bit period of the input data apart. For example, when ϕ_1 goes low, the input bit at that instance is sampled and held for further amplification, sliced and latched. With a 50% duty cycle clock, instead of having to finish in one bit period, now it has 4 bit periods till ϕ_1 goes back to high again. On the surface, it seems that the throughput can be further improved using a clock with lower than 50% duty cycle. A maximum of 7 bit periods for 8 parallel channels should be available

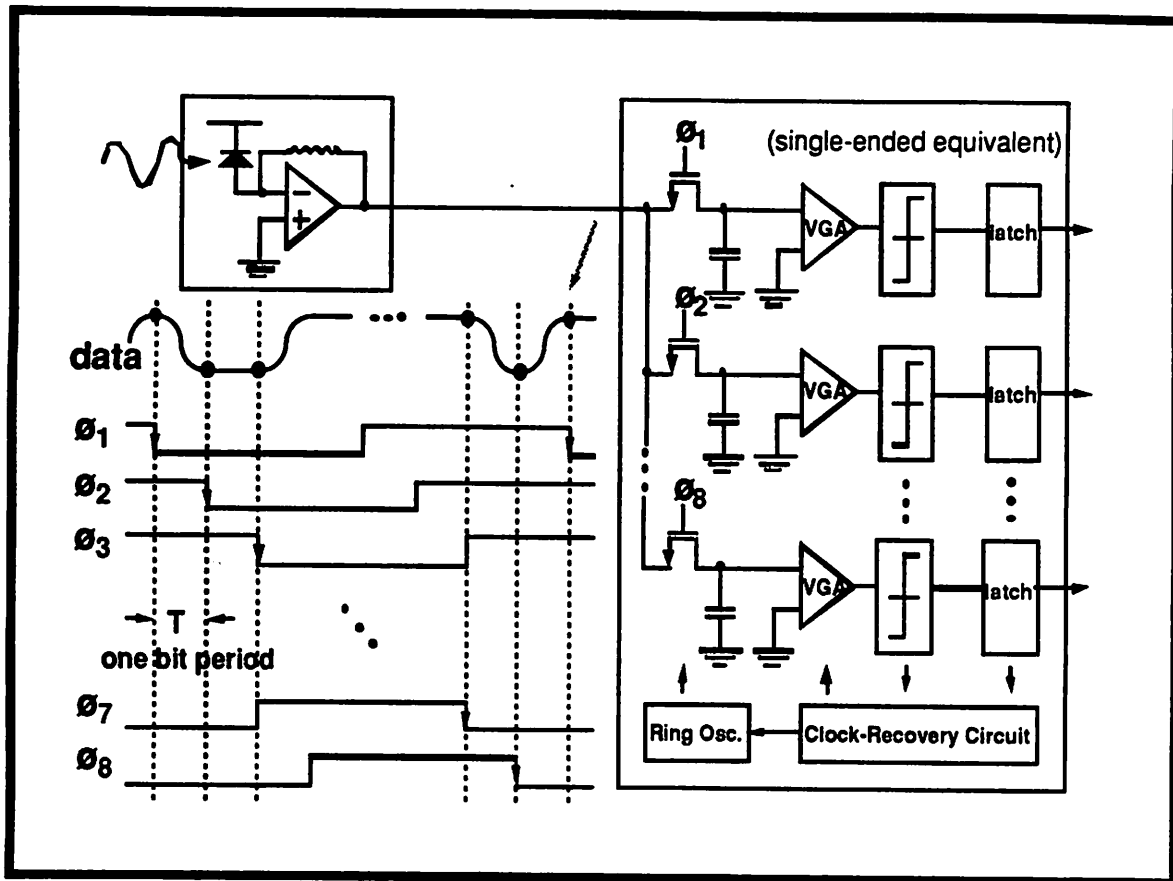


Figure 5-1 Parallel Receiver Implementation

using a 12.5% duty cycle clock instead of just 4 bit periods using a 50% duty cycle clock. However, amplification of the sampled data from one or more channels must be completed and the corresponding decisions must be valid at the same time for clock recovery. This together with other house keeping processes limit the maximum throughput to just about 4 to 1 and therefore a 50% duty cycle clock is used.

Note that the clock frequency of the clock that generates the multiple phase clock edges is only 1/8 of that of the input data rate and not the same or higher than the input data rate as in the traditional architecture. This eliminates the need for a high frequency VCO. Also, more time is available for clock recovery. This eliminates the need for a high frequency mixer.

This parallel approach while potentially allowing higher data rate in a given technology, it also introduces special design issues. The following sections will describe these special design issues one by one.

5.1.1 Input Demultiplexing

The input analog demultiplexing function is implemented by a series of sample-and-hold circuits. It is the only function in the whole architecture that is required to have a bandwidth of twice the input data rate (will be shown in Section 5.1.1.2).

5.1.1.1 CMOS Sample-and-Hold

The key factor enabling parallel receiver implementation to improve throughput is the fact that in MOS technology, the sample-and-hold function can be realized with higher bandwidth than the amplification function. This is true because MOS switches can utilize higher $V_{GS}-V_t$ and do not suffer from velocity saturation because of low V_{DS} .

With MOS technology, zero-offset switches and virtually zero gate current eliminate sample-mode offset and hold mode droop that appears in other technologies. However there are other major sources of error. A simple CMOS sample-and-hold circuit is shown in Figure 5-2 together with its key error sources [80]. There can be error due to the non-zero acquisition time for the sampling mode, the amplitude error because of the finite sample mode bandwidth, and there is a sample to hold mode transition error because of clock feed-through and channel charge injection of the switch. On top of these, there is also aperture delay which is the time difference between the sample to hold mode transition and the real sampling instant. Aperture delay is not a problem by itself but the variation of it because of timing jitter is. The effect of jitter will be discussed in Section 5.1.4. In the following subsections, error sources limiting the speed and accuracy of the sample-and-hold function will be examined.

5.1.1.2 Bandwidth Related Error

If the input is not bandlimited, then the input can be closely approximated by square pulses with finite rise and fall time. If we assume the input is a step function, then only the finite acquisition time is a concern. The sampled voltage is then

$$V_s = A \left(1 - e^{-\frac{t}{\tau}} \right) \quad (\text{EQ 5-1})$$

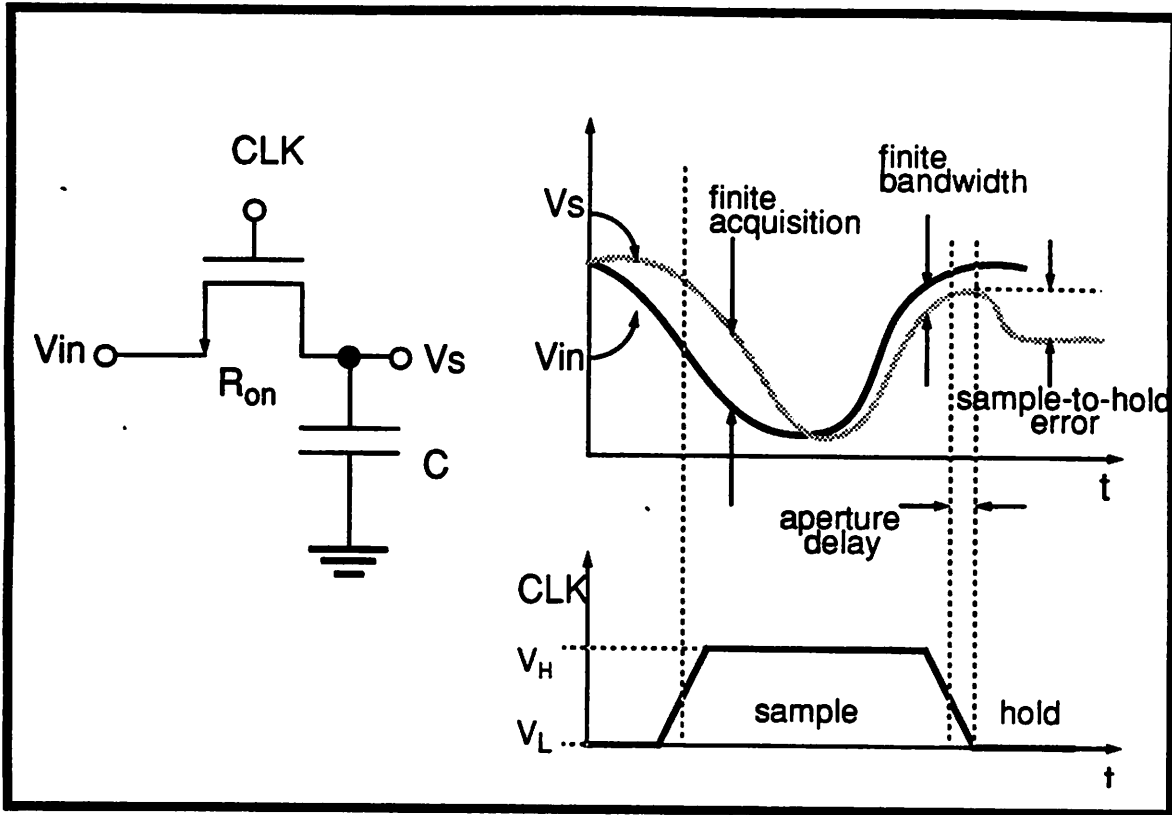


Figure 5-2 MOS Simple Sample-and-Hold Circuits and Error Sources

where A is the peak-to-peak amplitude of the square pulse and $\tau = R_{on}C$ where R_{on} is the on-resistance of the MOS switch and C is the sampling capacitor. For $t_a = 4\tau$, the sample voltage is about 98% of the final value, results in a 0.16dB lost in amplitude which is acceptable. Therefore, with t_r as the rise time of the square pulse, and T the bit period, τ should be chosen such that $t_r + t_a \leq T$. For example, with a 500Mb/s input bit rate, $T = 2\text{ns}$; assuming a rise time of 0.4ns, then τ should be less than 0.4ns. This corresponds to a minimum bandwidth of about 400MHz, which is 0.8 times the input bit rate. However, the output of the preamplifier is usually bandlimited as it is difficult to design a wideband low-noise preamplifier and it is also desirable to limit the noise bandwidth before sampling. As a result the output of the preamplifier resemble a sinusoidal waveform more than square pulses.

If we assume the input to be $A\cos(\omega t)$ for $t \geq 0$, then it can be shown that

$$V_s(t) = -A \left[\frac{\omega_\tau^2}{\omega_\tau^2 + \omega^2} \right] e^{-\omega_\tau t} + A \left[\frac{\omega_\tau}{\sqrt{\omega_\tau^2 + \omega^2}} \right] \cos \left[\omega t - \text{atan} \left(\frac{\omega}{\omega_\tau} \right) \right] \quad (\text{EQ 5-2})$$

where $\omega_\tau = 1/\tau$. The first term is the acquisition time and the second term is the steady state respond. Since the sample mode has a finite bandwidth, there will be an error both in the magnitude and phase of the sample waveform, resulting in a reduction in the sampled amplitude and in turn lowering the signal to noise ratio. From (EQ 5-2), it is obvious that lowering the phase shift can lower the amplitude at the same time. For $\omega_\tau = 4\omega$, the reduction in the magnitude of the data sampled is 0.52dB which is acceptable. For 500Mb/s input data rate, $\omega = (2\pi)250\text{MHz}$, therefore, $\omega_\tau = (2\pi)1\text{GHz}$. A sample mode bandwidth twice the input bit rate is needed. The acquisition time error is negligible in this case.

5.1.1.3 Accuracy Related Error

There are two types of accuracy error in the MOS sample-and-hold function. The first one is a random noise component due to the thermal noise in the channel of the MOS switch. It can be shown that the rms noise voltage sampled on the capacitor is $\sqrt{\frac{kT}{C}}$ where k is the Boltzmann's constant and T is the absolute temperature [81]. For example, if $C = 1\text{pF}$ and $T = 300\text{K}$, then V_{rms} is $64\mu\text{V}_{\text{rms}}$.

The second type of accuracy error is deterministic. It is due to the channel charge injection of the sampling switch and the clock feed-through by the parasitic capacitors. It can be shown that for fast clock transition, the error voltage V_d is approximated by

$$V_d \equiv \frac{C_{ol} + \frac{C_{gate}}{2}}{C} (V_H - V_{in} - V_t) + \frac{C_{ol}}{C_s} (V_{in} + V_t - V_L) \quad (\text{EQ 5-3})$$

where the gate capacitance $C_{gate} = WLC_{ox}$, C_s is the sampling capacitor, C_{ol} is the overlap capacitance, $C = C_{gate} + C_{ol} + C_s$ [82]. It can be written as

$$V_d = a + bV_{in} \quad (\text{EQ 5-4})$$

$$\text{where } a = \frac{C_{ol} + \frac{C_{gate}}{2}}{C} (V_H - V_t) + \frac{C_{ol}}{C_s} (V_t - V_L) \quad (\text{EQ 5-5})$$

$$b = -\frac{C_{ol} + \frac{C_{gate}}{2}}{C} + \frac{C_{ol}}{C_s}$$

where the first term is a DC offset term independent of the input and the second term is dependent on the input voltage which can be viewed as a gain error term.

5.1.1.4 Bandwidth vs Accuracy Trade-off

The on resistance of a MOS switch with small V_{DS} can be written as

$$\frac{1}{R_{on}} \equiv \mu C_{ox} \frac{W}{L} (V_H - V_{in} - V_t) \quad (\text{EQ 5-6})$$

If we assume $C_s \gg C_{gate} \gg C_{ol}$, then (EQ 5-3) can be approximated by

$$V_d \equiv \frac{WLC_{ox}}{2C} (V_H - V_{in} - V_t) \quad (\text{EQ 5-7})$$

Some observation can be made from (EQ 5-6) and (EQ 5-7). For high data-rate operation, the sample mode bandwidth must be high. For high sample mode bandwidth, a low R_{on} and low C must be used. However, the minimum C that can be used is limited by the kT/C noise of the sampling switch. Therefore, the only way to get higher bandwidth is through lower R_{on} . However, if we use a higher W or V_H , the error voltage V_d due to charge injection will be increased. Therefore there is a trade-off between sample mode bandwidth and accuracy. In theory, one can use as wide a device as possible for the switch, until limited by parasitics, to obtain the desired sample mode bandwidth as long as there is a way to reduce the error due to charge injection offset and clock feed-through.

Combining (EQ 5-6) and (EQ 5-7), we can find that

$$\tau \times V_d = \frac{L^2}{2\mu} \quad (\text{EQ 5-8})$$

Both the sample mode bandwidth and the offset error can be improved with a short channel length and high mobility technology. For a $1.2\mu\text{m}$ CMOS technology implementing a 500Mb/s receiver, assuming the sample mode bandwidth is $2 \times (2\pi)500\text{MHz}$ and a mobility μ of $500 \text{ cm}^2/\text{Vs}$, and a effective channel length of $1\mu\text{m}$, the error voltage calculated from (EQ 5-8) is as high as 60mV! Obviously something has to be done to reduce the charge injection offset.

The results from the error analysis can be summarized as a bandwidth against accuracy trade-off. The sample-mode bandwidth must be about twice of that of the input bit rate because of the bandwidth related error. The accuracy related error is also directly proportional to the sample-mode bandwidth. As a result, for a fixed technology, once the amount of error V_d that the system can tolerate is fixed, the maximum bit rate is also fixed. One can trade the sensitivity for higher bit rate or vice versa. For a reasonable sensitivity, a device f_T to data rate ratio of about 4:1 is needed for the implementation of the analog DEMUX. This is the fundamental limit on the overall throughput of the parallel receiver.

5.1.1.5 Circuit Implementation

The offset is mainly due to the input offset of the channel amplifier and the channel charge injection and clock feed-through of the switches. As calculated in Section 5.1.1.4, just the channel charge injection offset alone is in the range of 60mV. Means have to be derived to lower this to below 5mV. Also, care has to be taken to pick the size of the sampling capacitor to reduce the kT/C noise contribution. The sample mode bandwidth required as calculated in Section 5.1.1.4 is about twice the input bit rate. Minimum device width that can achieve the required bandwidth should be used to reduce the clock feed-through and channel charge injection.

From (EQ 5-5), the charge injection offset and clock feed-through can be decomposed into a DC offset part and a signal dependent gain error part. A fully differential sample-and-hold stage is used to minimize error due to the channel charge injection and clock feed-through. Let the differential input be V_{in1} and V_{in2} and use the notation that $x = (x_1 + x_2)/2$, $\Delta x = x_1 - x_2$, the differential offset error can be written as

$$\Delta V_d = (\Delta a + \Delta b V_{in}) + b \Delta V_{in} \quad (\text{EQ 5-9})$$

The first part is an offset due to matching error which is small for proper layout and the second part is the same signal dependent gain error part as before. Bottom-plate sampling can be used to minimize the gain error part [80], but a more complicated clocking scheme must be used. However, this is first not feasible and secondly not needed. It is not feasible because the multiple

phase clock edges are operating on gate delays, it is hard to generate the extra clock edges for bottom-plate sampling without increasing the minimum spacing between clock edges, resulting in a slower sampling rate. Secondly, it is not needed because the channel amplifier following it employs gain control. Therefore the signal dependent gain error part is never an issue and a simple differential sample-and-hold pair is used.

The remaining offset due to matching error and the input offset of the channel amplifier can still be a major problem if not handled properly. These offsets are handled by the channel amplifier with a DC input offset cancellation scheme that will be discussed later in Section 5.1.3.3

The fully differential sample-and-hold is followed by a fully differential datapath to minimize the coupling noise from the on-chip VCO, drivers, and digital circuits. It also helps to minimize any common mode noise and supply noise. Input followers are used in each parallel channel to isolate the sampling node from the noise of the outside world and also serves to lower the input capacitance.

The inputs to the chip contain high frequency components and must be terminated to reduce the reflections seen as a result of transmission line effects. This can be done off-chip with two 50Ω resistors but a better way to do it is to terminate the inputs on-chip. By terminating the input transmission lines on-chip, it can also account for the extra phase shift due to the bonding pads and bonding wires. On-chip termination is done differentially by connecting both inputs with a PMOS transistor biased in the triode region to give a series resistance of about 100Ω .

5.1.2 Multiple Phase Clock Edges

A fully differential ring oscillator is used to generate the multiphase clock edges required for the input sample-and-hold. The jitter of the ring oscillator must be low to reduce the contribution to the total jitter error. Layout has to be done in such a way that each tap of the ring oscillator sees the same loading, at least in the first order, to generate the evenly spaced clock edges. Detail design of this will be described in Section 5.1.4.5.

5.1.3 Channel Amplifier

A dynamic range of 40dB with a maximum gain of 100 is desired. As a result, the target gain range is from 1 to 100. The parallel architecture relaxes the requirement in speed for the amplify-and-latch circuit. For a prototype implemented with 8 parallel channels and a 50% duty cycle clock, 4 bit periods are available instead of just one in the traditional architecture. For a 500Mb/s receiver, the amplify-and-latch function has to be completed in 8ns at the maximum gain of 100, instead of just 2ns with traditional architecture.

5.1.3.1 MOS Variable Gain Amplifier

High speed variable gain amplifier design has primarily used high speed technologies such as silicon bipolar and GaAs. The design activity in this area has primarily used bipolar technology with limited activity in GaAs. High gain can be realized but the dynamic range is usually limited [55] [54]. There are also limited activity in design using fine-line NMOS [51] but extremely little in CMOS because of the performance achieved is limited by the speed of the technology.

Several variable gain principles for MOSFET circuits have been introduced in Section 3.5.2. However they cannot fulfil the speed and area requirements for the parallel receiver architecture. A faster and simpler configuration is used in this research. The circuit is shown in Figure 5-3.

The variable gain resistors (M5-M6) are used as the loading devices of a differential pair (M1-M2). The result is a fast and simple variable gain unit that can have unbalanced inputs DC coupled in. The variable resistance is controlled by a replicated biasing circuit (M7-M9) that sets up the correct gate voltage to bias up the output common mode voltage of the differential pair. The line L_{bias} can be shared with all variable gain units. The result is a fast and simple variable gain unit with minimum area for implementation. The draw back of this approach is that the DC operating point of the V_{DS} of the variable resistor is not 0. For maximum gain, a larger resistance is required with a lower V_{GS} and at the same time, V_{DS} is larger because of the larger IR drop across the load. Care has to be taken to ensure that the loading device is still in the triode region. This fur-

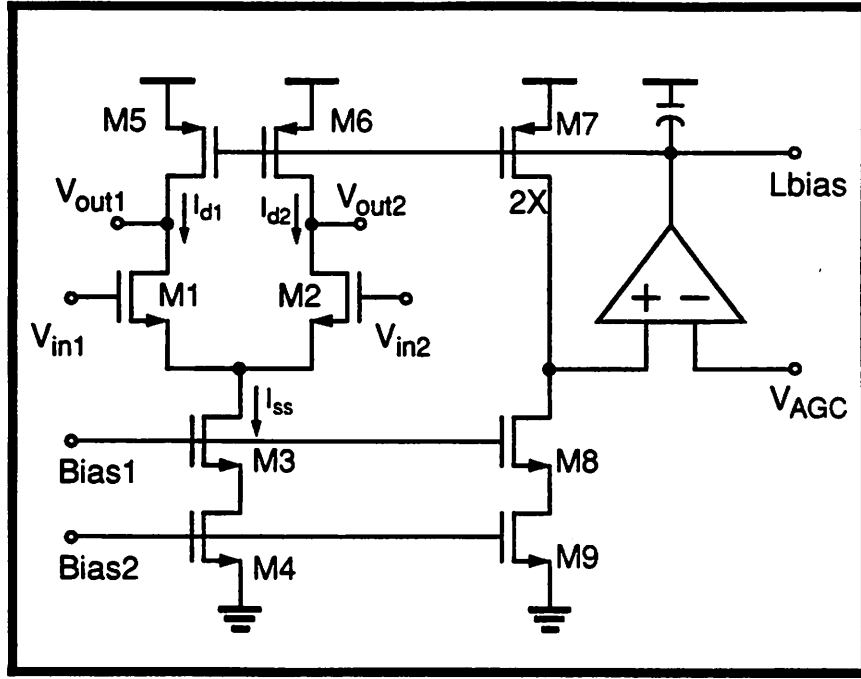


Figure 5-3 MOS basic variable gain unit

ther reduce the range of variation of the resistance and will be shown in Section 5.1.3.2.

5.1.3.2 DC and AC Analysis

First, the large signal conductance and small signal conductance in the linear region are defined as

$$g_s \equiv \frac{1}{r_{ds}} = \frac{\partial I_D}{\partial V_{DS}} = \mu C_{ox} \frac{W}{L} (V_{GS} - V_t - V_{DS}) \quad (\text{EQ 5-10})$$

$$g_L \equiv \frac{I_D}{V_{DS}} = \frac{\mu C_{ox} W}{2} \frac{W}{L} (2(V_{GS} - V_t) - V_{DS}) \quad (\text{EQ 5-11})$$

Define another variable $0 \leq r \leq 1$ to be

$$V_{DS} = r(V_{GS} - V_t) \quad (\text{EQ 5-12})$$

Therefore, when r is close to 0, the transistor is deep in the linear region and when r is close to 1, the transistor is closer to the saturation region. Put this into (EQ 5-10) and (EQ 5-11), we have

$$g_s = \frac{2(1-r)}{(2-r)} g_L \quad (\text{EQ 5-13})$$

The small signal gain of the differential pair is

$$A_v = g_{m_N} r_{ds} \quad (\text{EQ 5-14})$$

Putting (EQ 5-13) into (EQ 5-14), it can be shown that

$$A_v = f(r) \cdot \frac{V_{DS_P}}{V_{dsat_N}} \quad f(r) = \frac{2-r}{1-r} \quad (\text{EQ 5-15})$$

Note that V_{DS_P} is set by the replica biasing circuit. Therefore, the gain for each variable gain unit is proportional (almost linearly if r is small) to V_{AGC} , the gain control voltage.

Next, the maximum amount of gain range that can be achieved as constrained by the DC biasing condition is examined. On a crude approximation, treating the MOS variable resistor whose resistance varied linearly with the DC biasing V_{DS_P} (good approximation if r is small). Define V_s to be the desired output swing plus some margin (e.g. for a swing of $\pm 0.3V$, V_s is set to be $0.5V$). For low gain setting, V_{DS_P} is low and with a constant current biasing, V_{GSP} is high, so that r is small. V_{DS_P} is set to be V_s so that there is no clipping. If the lowest gain is set to unity, from (EQ 5-15), $f(r) \approx 2$, so $V_{dsat_N} \approx 1V$. For high gain setting V_{DS_P} is higher and with a constant current biasing, V_{GSP} is lower. There are two limits on how high V_{DS_P} can be. First, for the loading devices to remain in the linear region, V_{DS_P} plus V_s cannot be higher than $V_{GSP} - V_t$. The second constrain is that for the input differential pair and the current source to remind in the saturation region, the input common mode voltage as set by V_{DS_P} minus V_s cannot be less than $V_{dsat_N} + V_{tN}$ (with body effect) + V_{min} (current source) = $1 + 1.5 + 0.5 = 3V$. This implies the maximum V_{DS_P} that can be used is $1.5V$ for $V_{DD} = 5V$. As compared with the low gain setting with a V_{DS_P} of $0.5V$, the maximum gain range possible is only 1:3.

For the prototype, a design target of gain range of 1:2.5 is set.

At low gain setting, let V_L be the $V_{GS} - V_t$ of the load device in the lowest setting and V_{DS_P} is set to be V_s . The maximum V_L possible is $V_{DD} - V_t = 4V$ and a value of $3.5V$ is used, setting r to

be $0.5/3.5 = 1/7$. The lowest gain is set to unity, From (EQ 5-15), V_{dsatN} is found to be 1.085V. The biasing condition can be written as

$$I_d = \frac{\beta_P}{2} (2V_L V_s - V_s^2) \quad \beta = \mu C_{ox} \frac{W}{L} \quad (\text{EQ 5-16})$$

$$I_d = \frac{\beta_P}{2} (3.25) \quad (\text{EQ 5-17})$$

At high gain setting, let V_H be $V_{GSP} - V_t$ at the high setting, we have the following constraint;

$$V_{DSP} + V_s = xV_H \quad 0 < x < 1 \quad (\text{EQ 5-18})$$

Also, in this setting we have

$$I_d = \frac{\beta_P}{2} [2V_H V_{DSP} - V_{DSP}^2] \quad (\text{EQ 5-19})$$

Combine (EQ 5-19) with (EQ 5-17) and setting the gain A_v to be 2.5 in (EQ 5-15), we can solve for $V_{DSP} = 0.966V$, $V_H = 2.165V$ and $r = 0.446$. Therefore for the high gain setting, a V_{DSP} of 1V is used. Further more, $V_{DSP} + V_s = 1.5V < 2.165V = V_H$ and (EQ 5-18) is satisfied with $x = 0.7$ and the load device is still well in the triode region.

With the DC biasing problem solved, the known values are

$$V_{dsatN} = 1.085V \quad (\text{EQ 5-20})$$

$$\frac{I_d}{\beta_P} = 1.625 \quad (\text{EQ 5-21})$$

From these, we have $\beta_N/\beta_P = 2.76$. Once the biasing current is determined, the devices' sizes are known. Basically, this is a trade-off of speed and power consumption. With a cascade of variable gain units to form a variable gain amplifier and handle a demultiplexed bit rate of $2 \times 60\text{Mb/s} = 120\text{Mb/s}$ (because of 50% duty cycle clock), each variable gain unit must have a bandwidth of 300MHz at a maximum gain of 2.5. Each variable gain unit is basically a single pole ampli-

fier with the corner frequency determined by $r_{ds}C$ where $C = C_{gsN} + (1 + 2.5)C_{gdN} + C_{dsN} + C_{dsP}$ assuming driving an equivalent unit. From iterating the above requirement with (EQ 5-20) and (EQ 5-21), the minimum current required is 300uA and 400uA is used in case of process variations. Computer simulations are used to fine tune the W/L ratio for the differential pair and loading devices. The values used in the final design is summarized in the table below;

parameter	value
$(W/L)_N$	28/1.2
$(W/L)_P$	30/2.0
gm_N	1/1K
$r_{ds} (@ V_L = 3.5V)$	1K
$r_{ds} (@ V_H = 2.2V)$	2.5K
I_d	400uA

Table 5-1 Biasing conditions for variable gain unit

5.1.3.3 Channel Amplifier with Input DC Offset Cancellation

The variable gain amplifier consists of a cascade of the variable gain units described in Section 5.1.3.1. Cascading low gain amplifiers had been studied extensively and analysis of optimal number of stages used for a target overall gain was done [51] [84]. For a gain of 100, the optimal number of stages is 9 and a gain per stage of 1.8. Nevertheless, the normalized gain bandwidth product exhibits a fairly broad minimum for $5 \leq n \leq 13$. Computer simulations were used to decide the minimum number of stages used without significant degradation of performance to save power and area. Six stages are used in the prototype with a maximum gain per stage of 2.15 (6.7dB).

The main design objective for the channel amplifier is high speed. Most input offset cancellation schemes add parasitics to the signal datapath and slow down the amplifier. For this prototype, input offset cancellation is done by a continuous feedback path from the outputs back to the inputs through an auxiliary stage. The auxiliary stage adds very little parasitic to the first input stage and maintain the maximum bandwidth possible for the technology. However, this impose a requirement on the input data to be a zero DC random bit stream. The Gm ratio between the input

stage and the auxiliary stage has to be chosen carefully. This is because a constant voltage has to be applied to the auxiliary stage to cancel the input offset. The auxiliary stage has to have a much lower G_m than the input stage so that the constant input voltage is larger as compared to the input offset of the auxiliary stage itself. However, this voltage is superimposed on the output as an output offset and may cause error in latching the digital data. As a result, a G_m ratio of 1 to 5 is used. The corner frequency of the RC feedback is chosen to be extremely low, around 10 kHz, so that it won't cause any stability problem. Each channel has its own feedback, and the capacitor cannot be too large because of area occupied. Since the actual value of the RC product is not important, the resistors used in the feedback is implemented by a distributed long channel MOSFET to get the high resistance required. The channel amplifier is shown in Figure 5-4

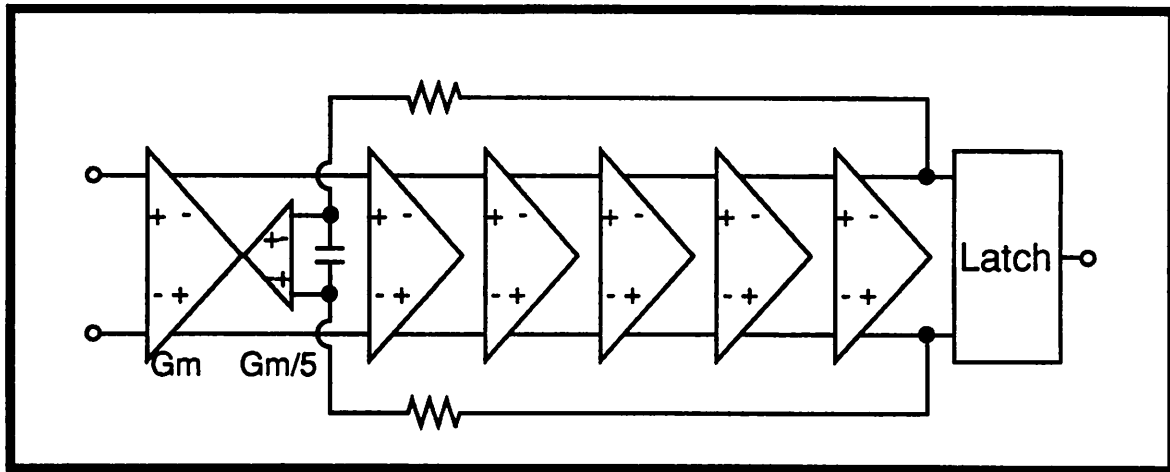


Figure 5-4 Channel amplifier with input offset cancellation

5.1.3.4 Distortion Analysis

Distortion is not an important factor in this application because the input is a digital waveform. The major factor is how fast the input can be amplified to a value that the latch circuit can utilize and regenerate the digital data without error. However, it is interesting to take a look at the distortion performance when used as a linear amplifier for other applications.

For a MOSFET differential pair, output differential current can be expressed as [83]

$$i_o = \frac{\beta}{2} v_i \sqrt{\frac{4I_{ss}}{\beta} - v_i^2} \quad (\text{EQ 5-22})$$

fier with the corner frequency determined by $r_{ds}C$ where $C = C_{gsN} + (1 + 2.5)C_{gdN} + C_{dsN} + C_{dsP}$ assuming driving an equivalent unit. From iterating the above requirement with (EQ 5-20) and (EQ 5-21), the minimum current required is 300uA and 400uA is used in case of process variations. Computer simulations are used to fine tune the W/L ratio for the differential pair and loading devices. The values used in the final design is summarized in the table below;

parameter	value
$(W/L)_N$	28/1.2
$(W/L)_P$	30/2.0
gm_N	1/1K
r_{ds} (@ $V_L = 3.5V$)	1K
r_{ds} (@ $V_H = 2.2V$)	2.5K
I_d	400uA

Table 5-1 Biasing conditions for variable gain unit

5.1.3.3 Channel Amplifier with Input DC Offset Cancellation

The variable gain amplifier consists of a cascade of the variable gain units described in Section 5.1.3.1. Cascading low gain amplifiers had been studied extensively and analysis of optimal number of stages used for a target overall gain was done [51] [84]. For a gain of 100, the optimal number of stages is 9 and a gain per stage of 1.8. Nevertheless, the normalized gain bandwidth product exhibits a fairly broad minimum for $5 \leq n \leq 13$. Computer simulations were used to decide the minimum number of stages used without significant degradation of performance to save power and area. Six stages are used in the prototype with a maximum gain per stage of 2.15 (6.7dB).

The main design objective for the channel amplifier is high speed. Most input offset cancellation schemes add parasitics to the signal datapath and slow down the amplifier. For this prototype, input offset cancellation is done by a continuous feedback path from the outputs back to the inputs through an auxiliary stage. The auxiliary stage adds very little parasitic to the first input stage and maintain the maximum bandwidth possible for the technology. However, this impose a requirement on the input data to be a zero DC random bit stream. The Gm ratio between the input

stage and the auxiliary stage has to be chosen carefully. This is because a constant voltage has to be applied to the auxiliary stage to cancel the input offset. The auxiliary stage has to have a much lower G_m than the input stage so that the constant input voltage is larger as compared to the input offset of the auxiliary stage itself. However, this voltage is superimposed on the output as an output offset and may cause error in latching the digital data. As a result, a G_m ratio of 1 to 5 is used. The corner frequency of the RC feedback is chosen to be extremely low, around 10 kHz, so that it won't cause any stability problem. Each channel has its own feedback, and the capacitor cannot be too large because of area occupied. Since the actual value of the RC product is not important, the resistors used in the feedback is implemented by a distributed long channel MOSFET to get the high resistance required. The channel amplifier is shown in Figure 5-4

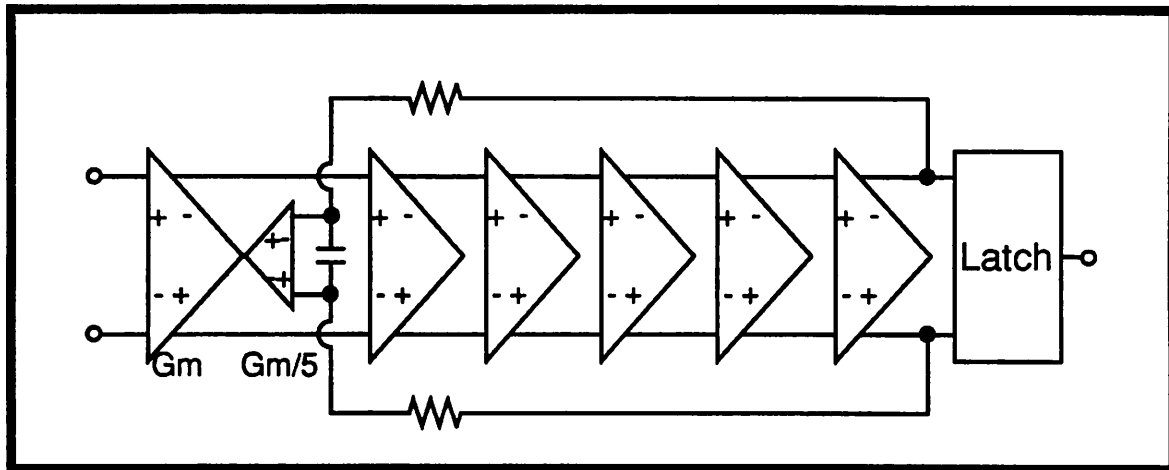


Figure 5-4 Channel amplifier with input offset cancellation

5.1.3.4 Distortion Analysis

Distortion is not an important factor in this application because the input is a digital waveform. The major factor is how fast the input can be amplified to a value that the latch circuit can utilize and regenerate the digital data without error. However, it is interesting to take a look at the distortion performance when used as a linear amplifier for other applications.

For a MOSFET differential pair, output differential current can be expressed as [83]

$$i_o = \frac{\beta}{2} v_i \sqrt{\frac{4I_{ss}}{\beta} - v_i^2} \quad (\text{EQ 5-22})$$

where $i_o = I_{d1} - I_{d2}$ and $v_i = V_{gs1} - V_{gs2}$. Let

$$K_1^2 = \frac{4I_{ss}}{\beta} \Rightarrow K_1 = 2(V_{GS} - V_t) \quad (\text{EQ 5-23})$$

and expand the square root with the Taylor series and dropping high order terms, it can be shown that

$$i_o = a_1 v_i + a_2 v_i^2 + a_3 v_i^3 \quad (\text{EQ 5-24})$$

$$a_1 = \frac{\beta}{2} K_1 \quad (\text{EQ 5-25})$$

$$a_2 = 0 \quad (\text{EQ 5-26})$$

$$a_3 = \frac{\beta}{4K_1} \quad (\text{EQ 5-27})$$

Similar analysis can be done for the load devices in the triode region, it can be shown that

$$i_i = -\frac{\beta}{2} v_o \sqrt{4V_B^2 - \frac{4I_{ss}}{\beta} - v_o^2} \quad V_B = V_{GS} - V_t \quad (\text{EQ 5-28})$$

where $i_i = I_{d1} - I_{d2}$ and $v_o = v_{out1} - v_{out2}$. Let

$$K_2^2 = 4V_B^2 - \frac{4I_{ss}}{\beta} \Rightarrow K_2 = 2(V_{GS} - V_t - V_{DS}) \quad (\text{EQ 5-29})$$

Expanding (EQ 5-28) with a Taylor series and dropping higher order terms,

$$i_i = a_1 v_o + a_2 v_o^2 + a_3 v_o^3 \quad (\text{EQ 5-30})$$

$$a_1 = -\frac{\beta K_2}{2} \quad (\text{EQ 5-31})$$

$$a_2 = 0 \quad (\text{EQ 5-32})$$

$$a_3 = \frac{\beta}{4K_2} \quad (\text{EQ 5-33})$$

But what is useful is

$$v_o = b_1 i_i + b_2 i_i^2 + b_3 i_i^3 \quad (\text{EQ 5-34})$$

A series reversal can be used to get

$$b_1 = \frac{1}{a_1} = -\frac{2}{\beta K_2} \quad (\text{EQ 5-35})$$

$$b_2 = -\frac{a_2}{a_1^3} = 0 \quad (\text{EQ 5-36})$$

$$b_3 = \frac{2a_2^2}{a_1^5} - \frac{a_3}{a_1^4} = -\frac{4}{\beta^3 K_2^5} \quad (\text{EQ 5-37})$$

Cascading the two series, we can find the series representation of v_o in terms of v_i .

$$v_o = c_1 v_i + c_2 v_i^2 + c_3 v_i^3 \quad (\text{EQ 5-38})$$

$$c_1 = a_1 b_1 = -g_m r_{ds} \quad (\text{EQ 5-39})$$

$$c_2 = b_1 a_2 + b_2 a_1^2 = 0 \quad (\text{EQ 5-40})$$

$$c_3 = b_1 a_3 + 2b_2 a_1 a_2 + b_3 a_1^3 = \frac{1}{2} \frac{\beta_N}{\beta_P} \left[\frac{1}{K_1 K_2} + \frac{\beta_N K_1^3}{\beta_P K_2^5} \right] \quad (\text{EQ 5-41})$$

Since β_N/β_P is fixed by the biasing condition, the distortion depends on K_1 and K_2 . K_1 is fixed by the V_{dsatN} from biasing, so distortion can be reduced by increasing K_2 using the maximum gate bias and minimum drain-to-source voltage on the load. This agrees with the intuition that under such condition, the device in the triode region is much linear.

In our prototype, $\beta_N/\beta_P = 2.76$ and $K_1 = 2V$, K_2 depends on the gain setting. With the high gain setting, $A_v = 2.5$, $K_2 = 2.4V$ and with the low gain setting, $A_v = 1.0$, $K_2 = 3V$. Putting this into the distortion equations and fixed the output to be $\pm 500mV$, we can find that for the high gain setting,

$$HD_3 = \frac{1}{4} \frac{c_3}{c_1^3} S_o^2 = \frac{1}{4} \frac{(0.67)}{(2.5)^3} (0.5)^2 = 0.00268 \quad (\text{EQ 5-42})$$

and for the low gain setting

$$HD_3 = \frac{1}{4} \frac{c_3}{c_1^3} S_o^2 = \frac{1}{4} \frac{(0.3554)}{(1.0)^3} (0.5)^2 = 0.0222 \quad (\text{EQ 5-43})$$

Distortion is worse for the low gain stage because the non-linearity in the differential pair dominates when the input is large.

Cascading N of these variable gain units together, it can be shown that with the special case that there is no second harmonic component,

$$d_1 = c_1^N \quad (\text{EQ 5-44})$$

$$d_2 = 0 \quad (\text{EQ 5-45})$$

$$d_3 = c_3 c_1^{N-1} \sum_{i=0}^{N-1} c_1^{2i} \equiv \frac{c_3 c_1^{3N-1}}{c_1^2 - 1} \quad \text{for } c_1 > 1 \text{ and } N \gg 1 \quad (\text{EQ 5-46})$$

For the low gain setting, $c_1 = 1$, we have

$$HD_3 = \frac{1}{4} \frac{d_3}{d_1^3} S_o^2 = \frac{1}{4} N c_3 S_o^2 = N (HD_3)_{\text{unit}} = 0.1332 \quad (\text{EQ 5-47})$$

The result is very intuitive as the input differential pair dominates at low gain setting and unity gain is used, the HD_3 is therefore N times worse than the unit case.

For high gain setting, $c_1 = 2.15$ to get a total gain of 100 for 6 stages.

$$HD_3 = \frac{1}{4} \frac{d_3}{d_1^3} S_o^2 = \frac{1}{4} \frac{c_3}{c_1 (c_1^2 - 1)} S_o^2 = \frac{c_1^2}{(c_1^2 - 1)} (HD_3)_{\text{unit}} = 0.00342 \quad (\text{EQ 5-48})$$

Again as expected, when the gain is high, for the same output amplitude, distortion in the input differential pair is not as bad as in high gain setting. The triode load devices are very linear and the overall distortion of cascading the variable gain units is only slightly worse than the unit case.

5.1.4 Clock Recovery

Clock recovery is done by using a decision-directed two-times oversampling phase detection scheme and a CPPLL. The design equations used in a conventional CPPLL is reviewed first. Then an algorithm of phase detection is introduced together with the system implementation. It is followed by the circuit implementation of the ring oscillator. Based on the implementation of the phase detector, the CPPLL loop dynamic is calculated by adopting the conventional design equations to this new phase detector. Finally, the initial acquisition issue is addressed.

5.1.4.1 Designing a Charge-Pump Phase-Lock Loop

This is a summary of results presented in [85]. Interested reader should refer to the original paper. A continuous-time approach is not valid for the analysis of a CPPLL if the loop bandwidth approaches the input frequency. However, as in many applications, the state of the CPPLL changes by only a very small amount on each cycle of the input signal. i.e. the loop bandwidth is small compared to the signal frequency. In these cases we may apply a time average analysis where the behavior over many cycles is considered. Linear analysis can then be used.

The charge pump can be implemented by a voltage charge pump or a current pump. The loop filter attached to it can either be passive or active. The order of the loop depends on the loop filter used. The most commonly used current charge pump and passive loop filters are shown in Figure 5-5

The on time of the charge pump t_p is proportional to the phase error.

$$t_p = \frac{|\theta_e|}{2\pi} \times T \quad (\text{EQ 5-49})$$

where θ_e is the phase error and T is the input bit period. With d as the probability of update, the average error current over many cycles is then

$$i_d = I_p \times \frac{\theta_e}{2\pi} \times d \quad (\text{EQ 5-50})$$

The control voltage for the VCO, V_c is then

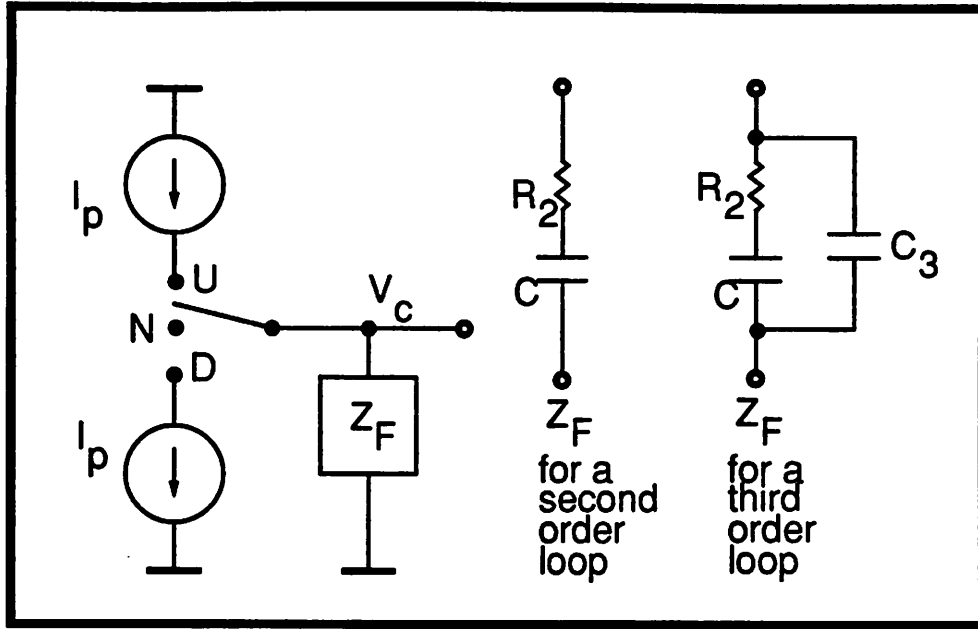


Figure 5-5 Charge Pump and Loop Filters

$$V_c = I_d Z_F = I_p d Z_F \frac{\theta_e}{2\pi} \quad (\text{EQ 5-51})$$

$$\theta_o(s) = K_o(s) \frac{V_c(s)}{s} \quad (\text{EQ 5-52})$$

where K_o is the VCO gain.

Therefore

$$\frac{\theta_i(s)}{\theta_o(s)} = H(s) = \frac{K_o I_p d Z_F(s)}{2\pi s + K_o I_p d Z_F(s)} \quad (\text{EQ 5-53})$$

For a second order loop as shown in Figure 5-5, It can be shown that [85]

$$\frac{\theta_i(s)}{\theta_o(s)} = H(s) = \frac{2\xi\omega_n s + \omega_n^2}{s^2 + 2\xi\omega_n s + \omega_n^2} \quad (\text{EQ 5-54})$$

$$\frac{\theta_e(s)}{\theta_o(s)} = 1 - H(s) = \frac{s^2}{s^2 + 2\xi\omega_n s + \omega_n^2} \quad (\text{EQ 5-55})$$

where

$$\tau_2 = R_2 C \quad (\text{EQ 5-56})$$

$$I_{pd} = I_p \cdot d \quad (\text{EQ 5-57})$$

$$\omega_n = \sqrt{\frac{K_o I_{pd}}{2\pi C}} = \sqrt{\frac{K}{\tau_2}} \quad (\text{EQ 5-58})$$

$$\xi = \frac{\tau_2}{2} \sqrt{\frac{K_o I_{pd}}{2\pi C}} = \frac{1}{2} \sqrt{K \tau_2} \quad (\text{EQ 5-59})$$

$$K = \frac{K_o I_{pd} R_2}{2\pi} = 2\xi \omega_n \quad (\text{EQ 5-60})$$

Where K is the loop gain, ω_n is the natural frequency and ξ is the damping factor. Note that any two parameters out of K , ξ , ω_n define the CPPLL. Also the static phase error for a step frequency input can be found by applying the final-value theorem to be

$$\theta_i(s) = \frac{\Delta \omega}{s^2} \quad (\text{EQ 5-61})$$

$$\theta_v = \lim_{t \rightarrow \infty} \theta_e(t) = \lim_{s \rightarrow 0} s \theta_e(s) = \frac{2\pi \Delta \omega}{K_o I_{pd} Z_F(0)} \quad (\text{EQ 5-62})$$

As long as there is no DC path through the loop filter, there will be no static phase error. In actual implementation, the DC input impedance of the VCO acts as a shunt impedance in parallel to the loop filter results in a finite Z_F . However, this can be designed to be extremely large and if the input of the VCO is a FET, it can approach the theoretical limit of infinity. Also we can see that the tracking range (lock range) is approximately the DC gain of the loop and it is infinite in this case, resulting in an infinite lock range in theory. Of course this is not true in practical implementation. The variation of oscillation frequency of the VCO is defined by a variation of the control voltage V_c , which has a limited range.

Due to the discrete nature of the actual implementation, caution has to be taken for two aspect - stability and ripple. Ripple comes from the charging current driven into the filter impedance Z_F , which responds with an instantaneous voltage jump of $\Delta V_c = I_p R_2$. At the end of the charging interval, the pump current switches off, and a voltage jump of equal magnitude occurs in the opposite direction. This causes a frequency shift of $|\Delta \omega_o|_2 = \frac{2\pi K}{d}$ and a corresponding phase jitter that is proportional to the phase error, which is undesirable. An easy solution to this is

to add a smoothing capacitor in parallel to the loop filter as in a third order loop. For C_3 that is small compared with C , the low frequency properties are identical to what was shown before and C_3 only has high frequency effects. With the addition of C_3 ,

$$H(s) = \frac{K \left(\frac{b-1}{b} \right) \left(s + \frac{1}{\tau_2} \right)}{s^3 \frac{\tau_2}{b} + s^2 + K \left(\frac{b-1}{b} \right) s + \frac{k(b-1)}{b\tau_2}} \quad (\text{EQ 5-63})$$

$$|\Delta\omega_o| = 2\pi K \left(\frac{b-1}{b} \right) \left[\frac{b-1}{b} \left(1 - e^{-\frac{b|\theta_e|}{\omega_i\tau_2}} \right) + \frac{|\theta_e|}{\omega_i\tau_2} \right] \quad (\text{EQ 5-64})$$

$$\text{where } b = 1 + \frac{C}{C_3}$$

For small $|\theta_e|$, we can find that the extra suppression of the ripple is

$$\frac{|\Delta\omega_o|_3}{|\Delta\omega_o|_2} \equiv \frac{(b-1)|\theta_e|}{\omega_i\tau_2} \quad (\text{EQ 5-65})$$

For stability, z-domain analysis can be used and the following is the limiting value for the normalized look gain $K' = K\tau_2$

$$K' < \frac{4(1+a)}{\frac{2\pi(b-1)}{b\omega_i\tau_2} \left[\frac{2\pi(1+a)}{\omega_i\tau_2} + \frac{2(1-a)(b-1)}{b} \right]} \quad (\text{EQ 5-66})$$

$$\text{where } a = e^{-\frac{2\pi b}{\omega_i\tau_2}}$$

As a rule of thumb, one may want the design to have $b > 10$ and $\frac{\omega_i}{K} > 15$ to 20.

5.1.4.2 Inductive Clock Recovery

Phase detection is the same process as estimating timing error in the inductive method for clock recovery. Most existing inductive clock recovery schemes use a variation of the maximum likelihood (ML) method which is approximating a minimum mean square error (MMSE) method. [86]. The method requires two times oversampling and complex processing, which is rather unattractive. Another interesting method is the early-late gate method [87] which includes two extra sampling instances, one prior to the sampling instance by $\Delta/2$ and once after the sampling instance by the same amount. The sampling instance is adjusted until the two extra samples

are equal. The method is simple and require very little processing. However, the parallel architecture is already working on gate delays of the technology to generate the multiphase clock edges. It is hard to generate the $\Delta/2$ delay accurately. A concept of clock recovery based on minimum likelihood is proposed in [88]. The minimum likelihood implies "least likely" for synchronization or an "orthogonal" timing condition which simply means that the locally generated clock is synchronized correctly, but with a delay of a half bit period! This can be combined with the early-late gate method and set Δ as the input bit period T . The whole clock recovery loop is shown in Figure 5-6

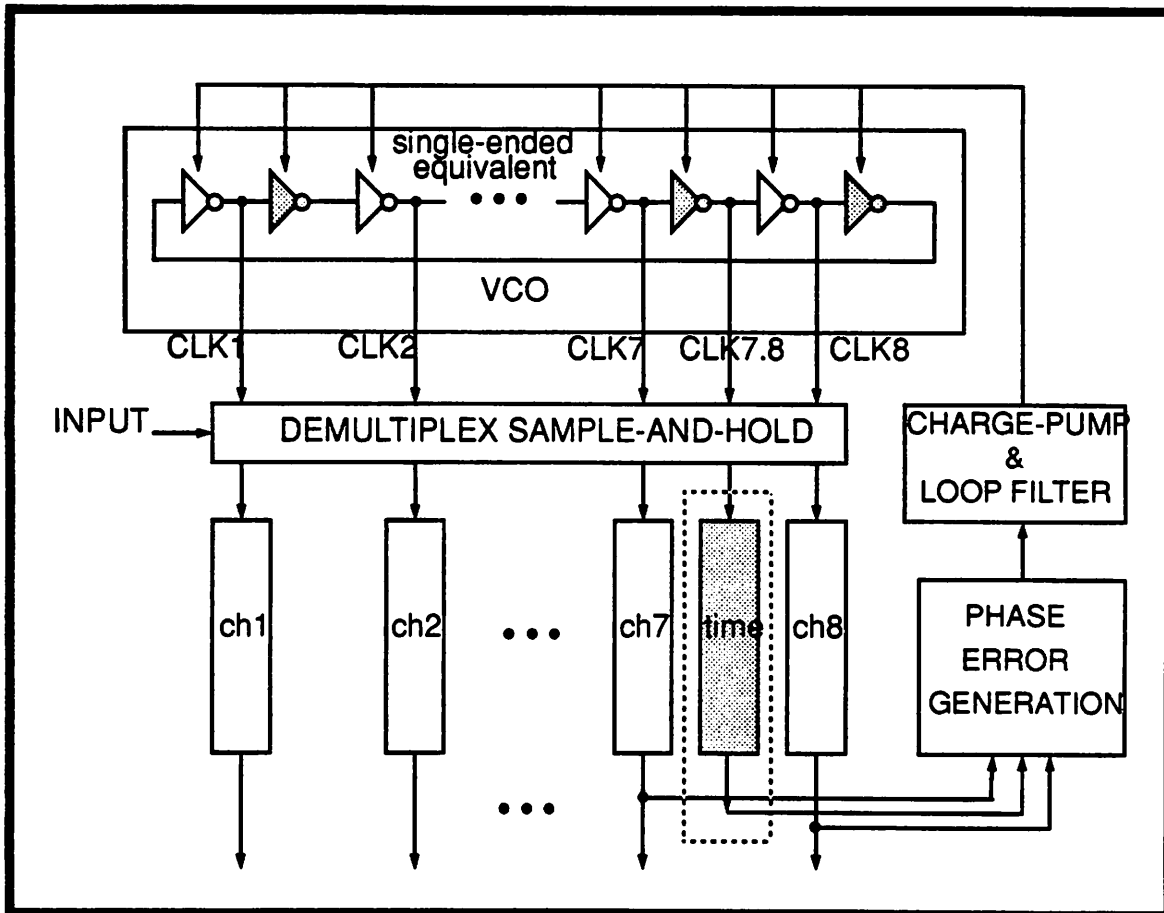


Figure 5-6 Minimum Likelihood Clock Recovery

Estimated timing error can be obtained by inserting an extra timing channel in between two adjacent data channels, effectively doing a two times oversampling. In the prototype, a timing channel is inserted between channel 7 and channel 8. An extra inverting stage was added to the ring oscillator to generate a sampling clock edge exactly in between the sampling clocks for the

two adjacent channels. In order for the ring oscillator to see the same amount of impedance at each tap, extra stages were added in between each pair of adjacent data channels. The phase error information is then passed to a charge-pump and then a loop filter to control the ring oscillator and complete the CPPLL. The sampling instances as determined by the multiphase clock edges from the ring oscillator is adjusted to force the output of timing channel to be zero.

5.1.4.3 Decision Directed Phase Detection

The phase error is generated as shown in Figure 5-7 By defining the error voltage as the

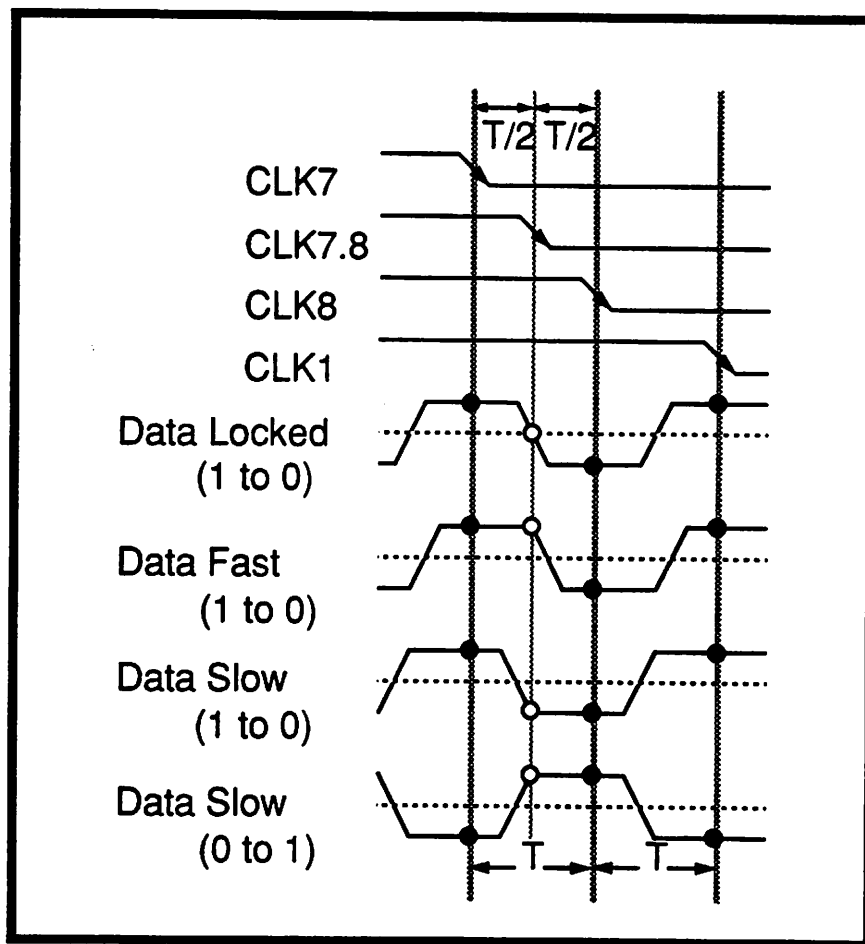


Figure 5-7 Examples of Decision Directed Phase Error Detection

output voltage of the timing channel, phase error can be generated. From the first waveform shown, the clock is locked onto the input data waveform. Channel 7 and Channel 8 is sampling in

the center of the bit cell and the error voltage is zero. In the following two waveforms, there are positive and negative phase errors, and the error voltages are positive and negative respectively. However, in the last waveform, the phase error is negative but the error voltage is positive because of a different transition. It can be seen that the magnitude of the error voltage will give the magnitude of the phase error but the polarity depends on the transition and must be decision directed. This is why information from adjacent data channels is required. This method is justified heuristically but there is no vigorous calculation to support its feasibility at this stage. It was proven by computer simulations in the design phase and empirically by evaluating the prototype.

5.1.4.4 Phase Detector Implementation

The essential part of the phase detector is repeated in more detail in Figure 5-8 Digital

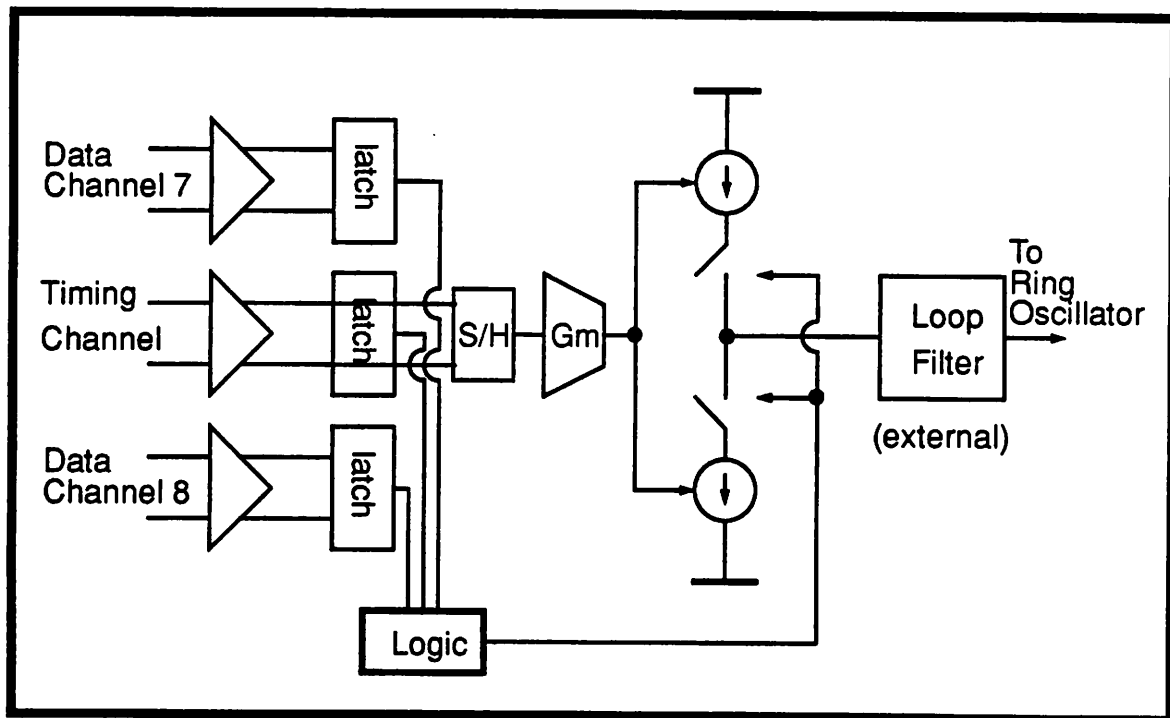


Figure 5-8 Pipeline Phase Detector with Charge-Pump and Loop Filter

information from channel 7 and channel 8 is latched and stored to determine if there is a transition and what kind of transition it is. The digital information of the timing channel is also used together with the transition information to control which charge pump to open. There is a sample-

and-hold stage to sample the magnitude of the final value of the timing channel (error voltage). This sampled voltage is converted to a current to control how much current there is to flow through the charge pump. Inserting the latches and the sample-and-hold stage pipelines the phase detection from the channel amplification to improve throughput. However, information from channel 7 and 8 together with the timing channel must be settled all at the same time for pipelining, thus preventing the input demultiplexing to use a clock with higher duty cycle. The loop filter of a CPPLL usually requires a large capacitance, and is hard to be integrated on the same chip. It is the only component that is required externally. However, the pin connected to the loop filter is directly used to control the VCO. In order to reduce the jitter of the VCO due to the noise coupling in from the I/O pad, a dummy set of loop filter is duplicated, and the signal is brought back on chip differentially to minimize the effect of pin-to-pin crosstalk and board/chip differential ground noise. The scheme is shown in Figure 5-9. The common mode voltage of OTA2 is set by $I_b R_b$. The free running frequency of the VCO is set by the voltage $I_o R_o$. The maximum deviation in frequency is set up by the maximum deviation in V_c , which in turn is set up by the maximum deviation in the current output of OTA2. This limits the lock range of the CPPLL.

The above setup minimizes the noise coupled into V_c because of the use of external loop filters and gives maximum flexibility for testing. C_3 is implemented on chip for a third order loop and filters out the ripple and high frequency noise coupled in from the I/O pad. As long as the ratio of C/C_3 is large, the loop will behave like a second order loop.

5.1.4.5 Ring Oscillator Implementation

The most important part of a ring oscillator is the unit variable delay element. Traditionally, variable delay elements in MOS technology is implemented by current starved inverters [89] as shown in Figure 5-10 (a). VCTRL modulates the on resistances of M1 and M2 in turn controls the current available for charging and discharging the load capacitance. Large value of VCTRL allows a large current to flow and a small delay. However, at small value of VCTRL (slightly larger than V_{th}), the current is extremely small, hence the term "current starved". The current mirror components M1 and M2 has high r_{ds} making the circuit nodes inside the delay element very susceptible to crosstalk and noise injection from the supply lines. One way to reduce the effect of this is to

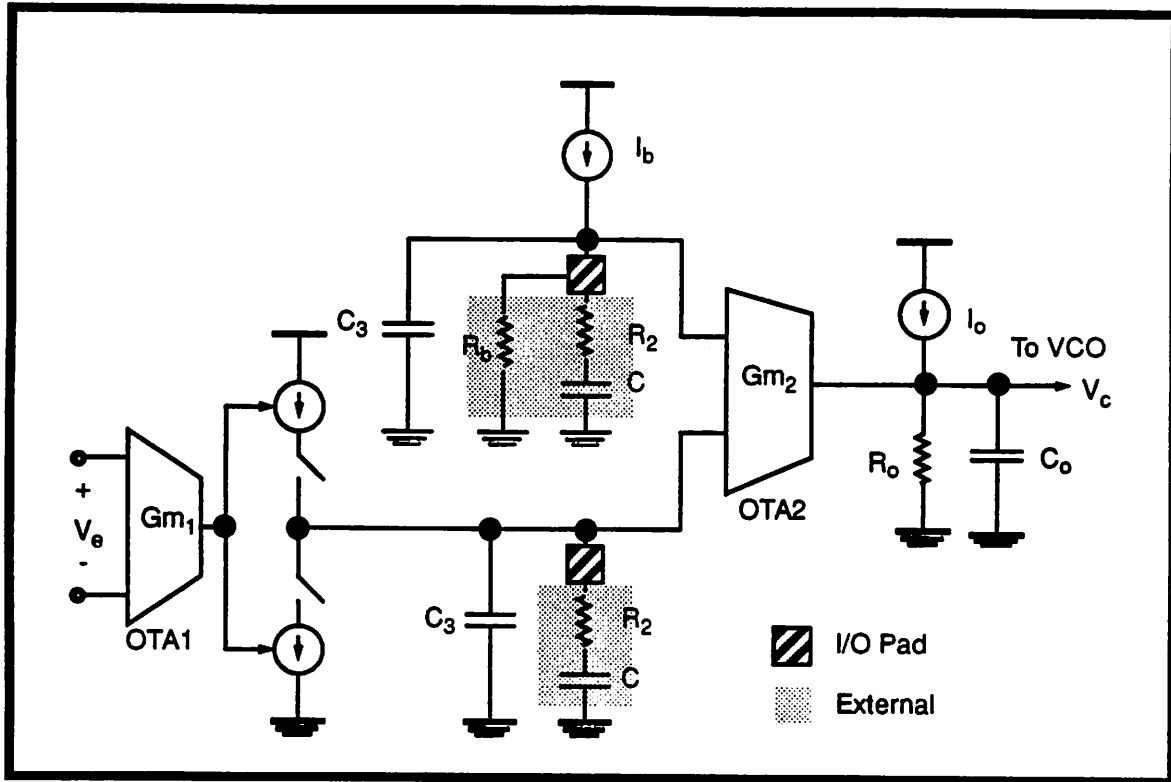


Figure 5-9 Charge Pump and Loop Filter Implementation

have an on-chip regulator to control the voltage across the delay element as in [90]. However, the delay varies sharply with the input voltage in this region. As a result of this, the VCO gain K_o varies tremendously and is undesirable for stable loop dynamics. Also, in this region of steep slope, any noise present on the control signal is also amplified, resulting in higher jitter.

Another type of delay element is the shunt capacitor delay stage [91]. This is shown in Figure 5-10 (b). VCTRL adjusts the resistance of a shunt transistor M1, which connects a large load capacitance to the output. The shunt resistor in essence control the amount of effective load capacitance seen by the driving gate. Large value of VCTRL decrease the resistance of the shunt resistor, so the effective capacitance at the output is increased, producing a larger delay. As long as VCTRL is larger than V_{th} , K_o is much more linear than the current-starved approach with no steep operating region. As a result, this configuration is used. Note that this configuration is highly susceptible to supply noise injection, as in the current-starved approach. If jitter due to supply noise injection is a concern, a fully differential approach as in [92] should be used to maximize supply rejection.

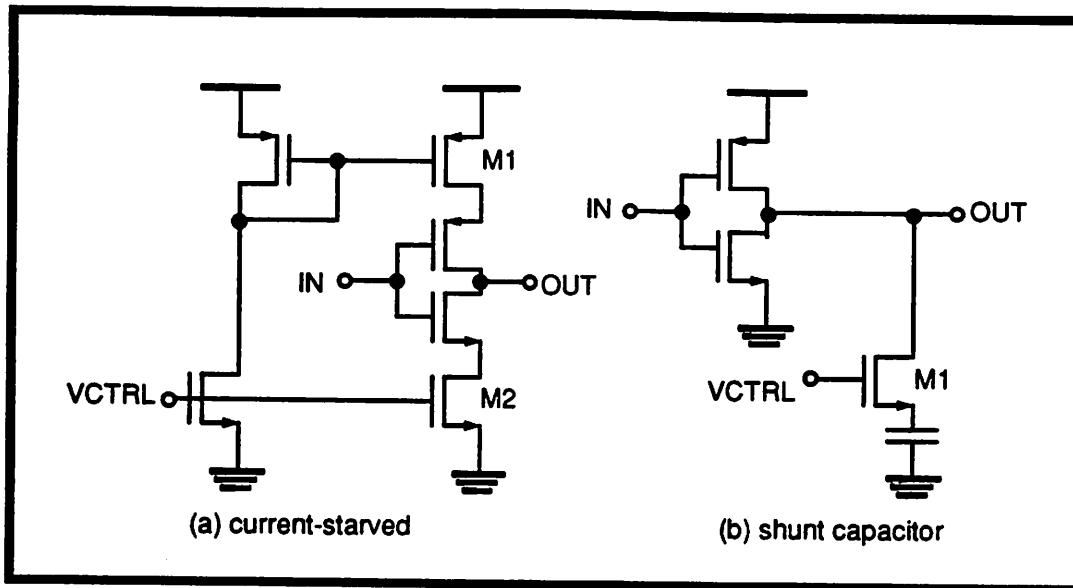


Figure 5-10 MOS Variable Delay Elements

The shunt capacitor variable delay element as shown in Figure 5-10 (b) cannot be used directly because it cannot generate an even number of multiphase clock edges. A differential input version can be derived easily from the single-ended configuration. As shown in Figure 5-11,

M1 and M2 is used to cross-couple two independent variable delay elements to form a single differential I/O delay element. They make sure that OUT1 and OUT2 are exactly 180° out of phase. The extra advantage of adding these two transistors is that they sharpen the fall edges of the outputs results in better resolution between adjacent clock edges. An extra pair of shunt capacitances are added and controlled by an external voltage EXT_VCTRL for the capability of testing in a slower mode.

5.1.4.6 Charge-Pump PLL Loop Dynamics

The loop dynamic of a CPPLL is completely characterized by the parameters K , ξ , and ω_n . Any two out of the three parameters defined the remaining one as shown in (EQ 5-58) to (EQ 5-60). For low jitter operation, a low bandwidth of ω_n is desired but that will lower the loop gain K results in a lower capture range. The capture range of the CPPLL is directly proportional to K with a multiplication constant depending on the input pulse shape, data statistics and the phase detection scheme.

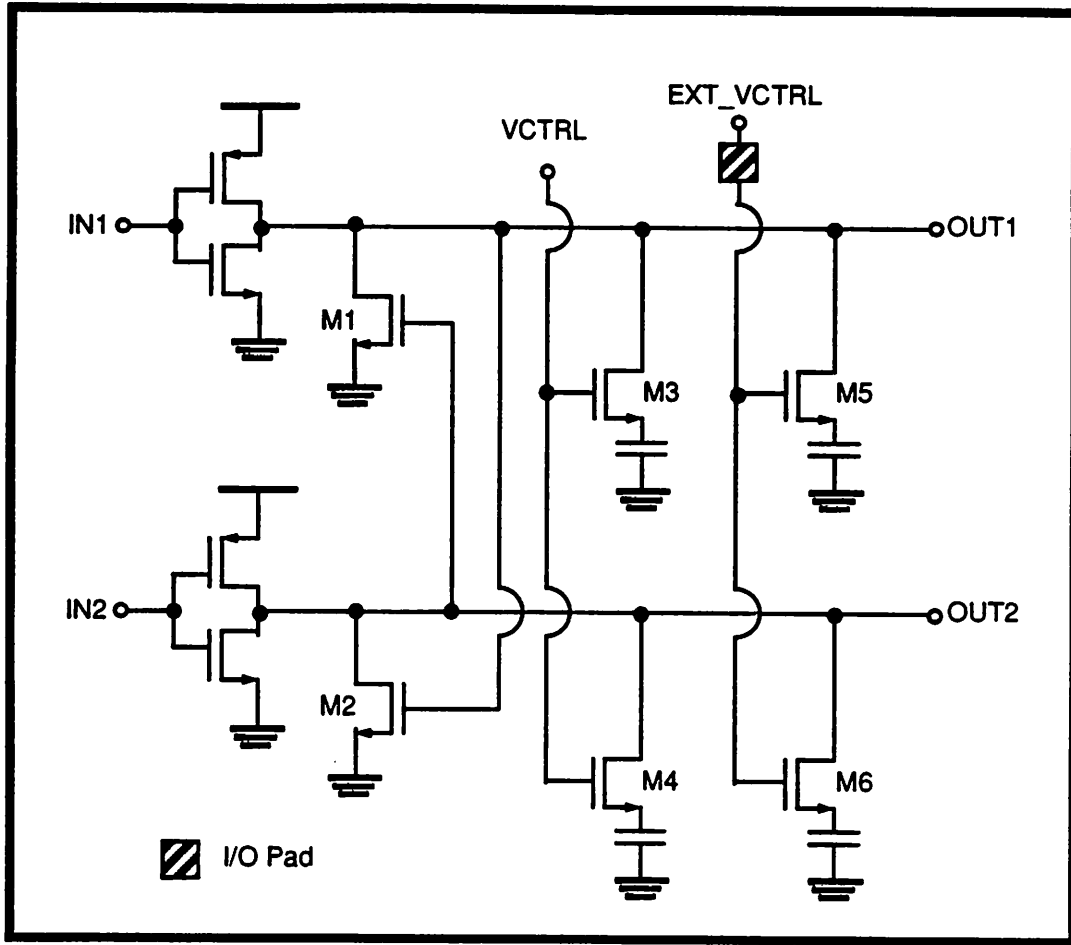


Figure 5-11 Differential Input-Output Shunt Capacitor Variable Delay Element

First, take a look at the phase detector. Traditionally, a CPPLL has a phase detector implemented with a wideband mixer (or an EXCLUSIVE-OR if implemented digitally) that generates the phase error in the time domain. The charge pump, with a fixed charge current, is then opened for a variable amount of time corresponding to the amount of phase error as reflected in (EQ 5-49) and (EQ 5-50). In the implementation of phase detector for the parallel receiver, the phase error information is generated by the error voltage in the voltage domain. Therefore, the charge pump is opened for a fixed period of time and charges with a current that is proportional to the phase error. An equivalent to the traditional CPPLL is derived so that the results can be applied directly.

Assuming the channel amplifier is linear, let V_e be the output of the timing channel, and $\frac{\partial}{\partial t} V_e(t) = m(t)$, then

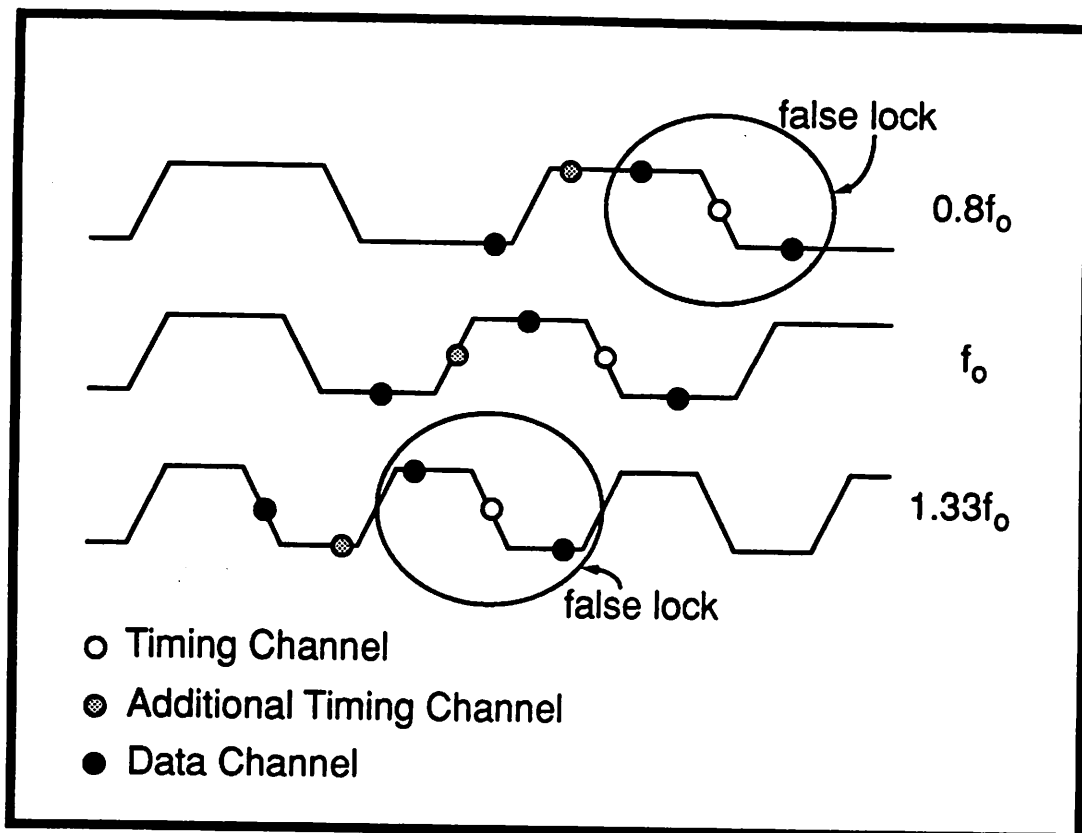


Figure 5-12 Frequency Detection with Additional Timing Channel

as in the prototype, false lock can happen to the two other examples with input frequency of $0.8f_0$ and $1.33f_0$ because the phase detector is looking at the one edge that produces zero error voltage and thinks that the loop is locked. An additional timing channel can provide extra information. In both cases where the frequency is not equal to f_0 , while the timing channel gives zero phase error, the additional timing channel gives a full error output. By considering the information from both timing channels and the adjacent data channels at the same time, the correct decision about the phase error can be made. This effectively increases the lock range of the CPPLL.

5.1.4.8 Effect of DC offset and Pulse Distortion

Input offset and pulse distortion can cause a degradation in jitter performance. Both of these effects effectively shifted the zero output point on the edges for timing recovery. As a result, when one kind of transition appears consecutively for time recovery, the clock edges are shifted resulting in a transient phase error because the CPPLL tries to move the edges of the clock to force a zero output from the timing channel. When the opposite kind of transition occurs, the phase

error generator will think that there is a huge phase error and try to correct it. This results in random wandering of the sampling edge for the timing channel resulting in a pattern dependent jitter. On the average, the frequency is still the same as the input frequency but that introduces an excessive amount of clock phase jitter even when the input has no jitter at all. The effect of the large DC output offset needed to cancel the huge input offset is graphically demonstrated by Figure 5-13

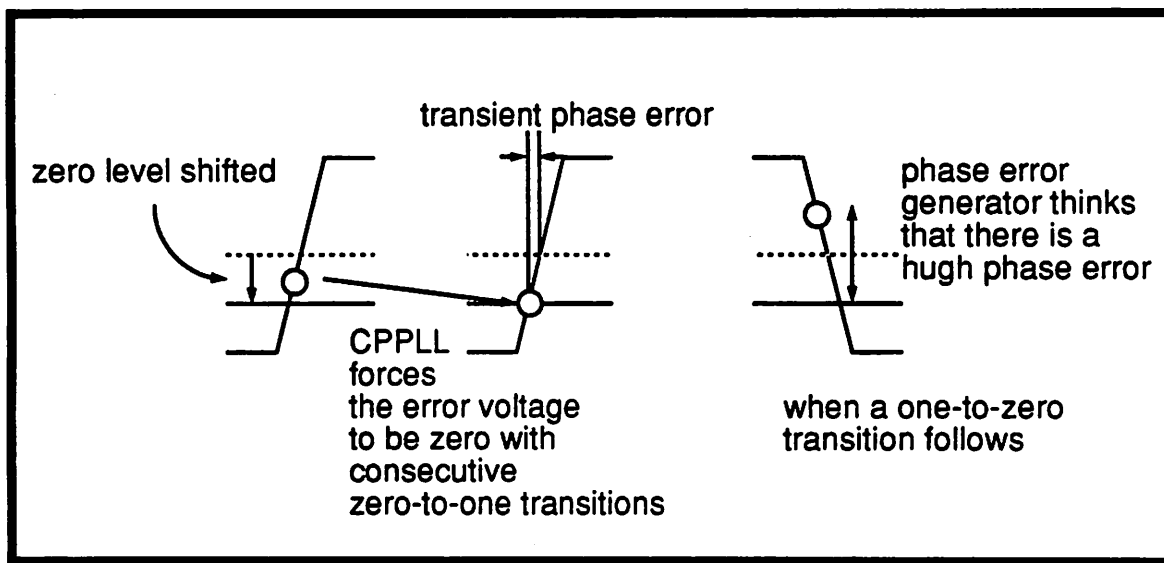


Figure 5-13 Excessive Phase Jitter due to Output Offset

Chapter 6 Experimental Results

6.1 Experimental prototype

The experimental prototype shown in Figure 6-1 is implemented in a double-poly, double-metal 1.2- μm n-well CMOS technology. Even though double poly is used, it is not necessary for the implementation, making it fully compatible with any digital single-poly, double-metal CMOS technology. The chip area is 4mm X 4mm with an active area of 3mm X 3mm. It consists of 10 parallel channels - 8 data channels, one timing channel and one extra dummy channel for testing. It is packaged in a 68-pin LCC package. The prototype operates with a single 5V supply. All testing of the prototype is done at room temperature.

6.2 Voltage Control Oscillator

The transfer characteristic of the period and oscillation frequency vs the control voltage VCTRL of the VCO are plotted in Figure 6-2. The measured results differ significantly from the simulation results. The measured results are consistently about 30% faster than simulation. This can be due to two reasons. The first one is that the extraction program over estimating the parasitic and line capacitances. This is so because "magic", the layout and extraction tool used, has a very poor model for metal line capacitance. For modern VLSI designs, the fringing capacitance is dominating for minimum width metal lines but not as significant for wider metal lines. In magic, the fringing capacitance can only be extracted as a constant multiple of the overlap. For worst case simulation, a factor corresponds to minimum line width is used and this grossly over-estimating the capacitance for wider metal lines. Secondly, the factor k' (μC_{ox}) is found to be about 20% larger in this run than the device we used for characterization and the simulation model provided by the fab-

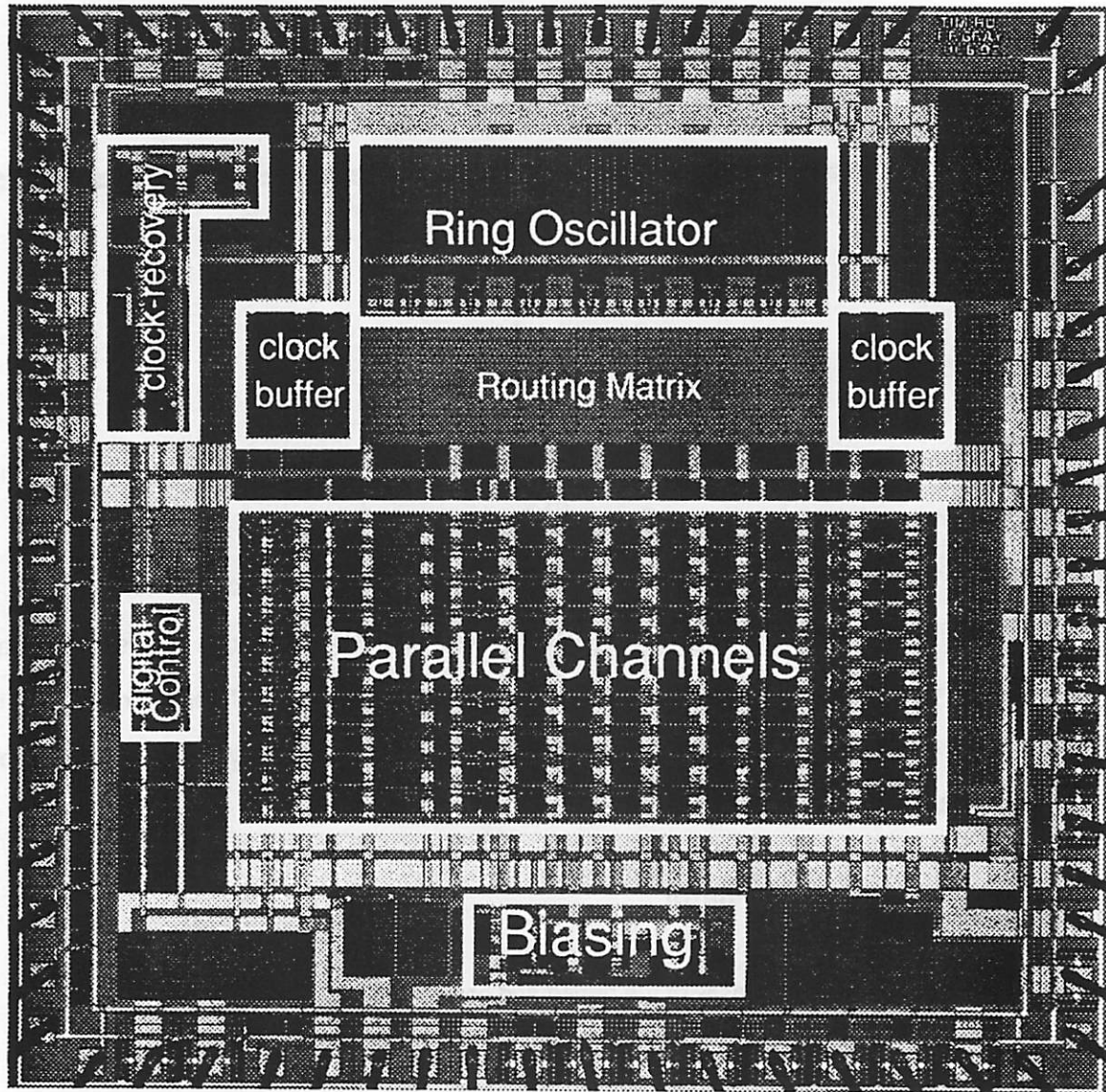


Figure 6-1 Die Photograph of Experimental Prototype

icator. The experimental result agrees well with simulation once these effects are corrected for further simulation.

The VCO gain K_o is also increase by approximately 30% of that of the simulation. The measured K_o is $(2\pi)12.3$ MHz/V when the oscillation frequency is 50MHz (corresponds to bit rate of 400Mb/s), and is $(2\pi)14.0$ MHz/V when the oscillation frequency is 60MHz (corresponds to bit rate of 480Mb/s).

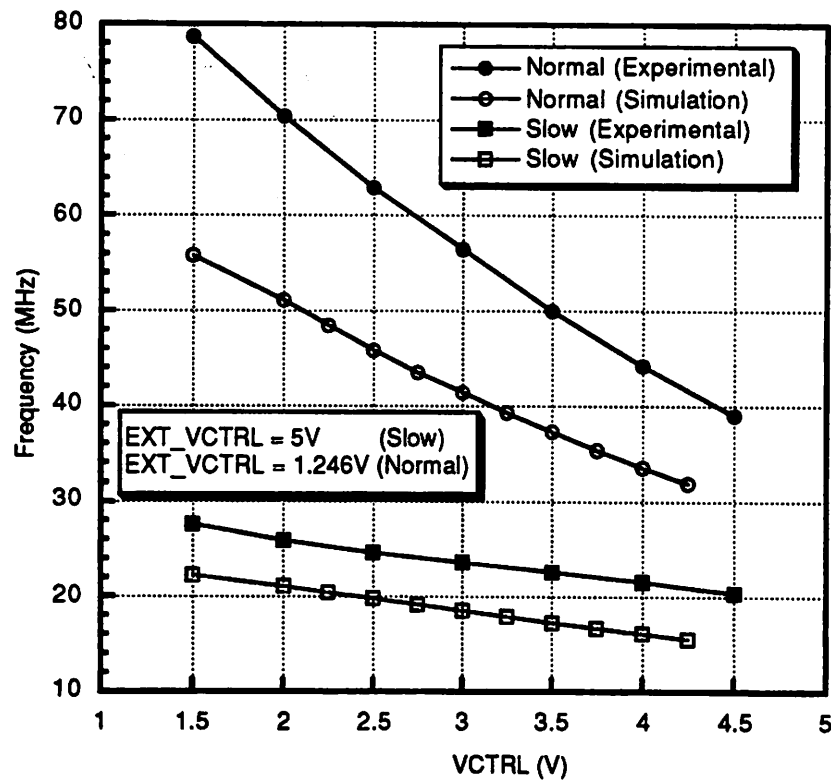
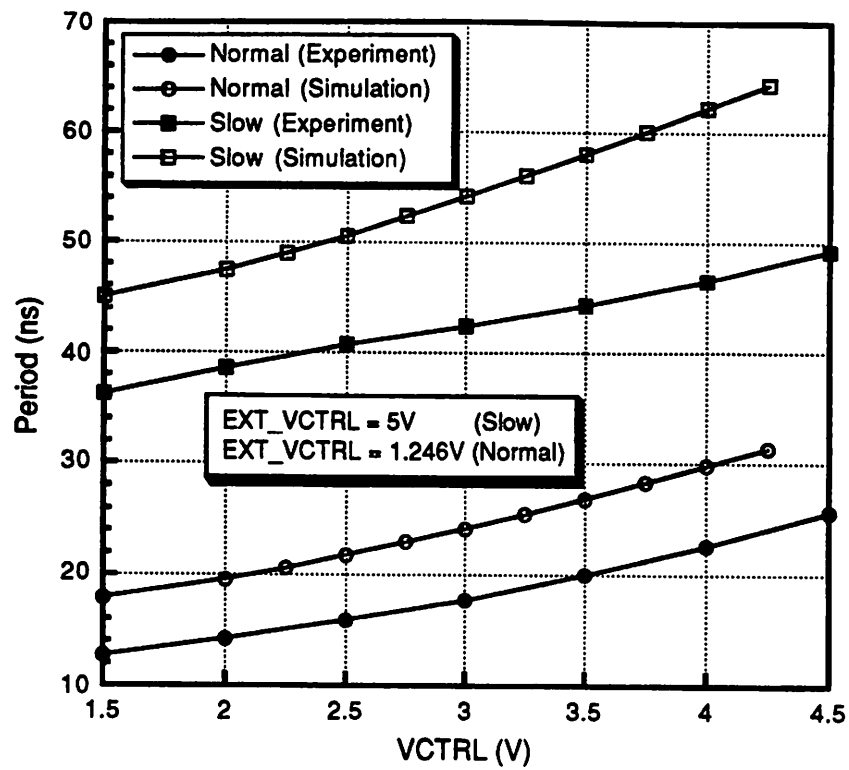


Figure 6-2 Period and Frequency vs VCTRL Plot

6.3 AGC Biasing

The variable gain unit was shown in Figure 5-3. The biasing voltage V_{AGC} is used to control the DC operating point of the two PMOS load devices who are operating in the triode region. The replica biasing circuit adjusts the corresponding gate voltage L_{BIAS} for the PMOS devices. Figure 6-3 Shows the plot of L_{BIAS} vs V_{AGC} for measurements from three different chips and the

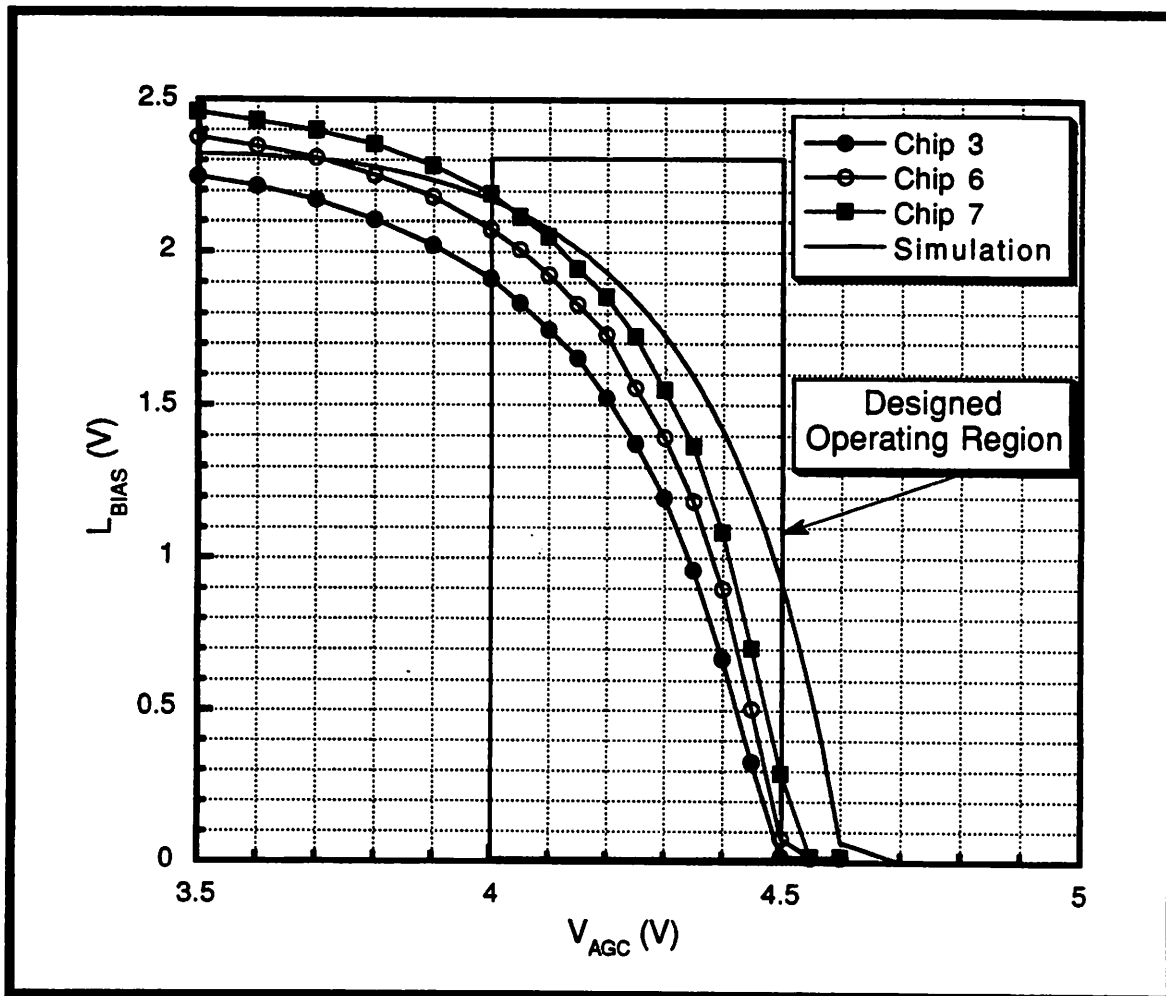


Figure 6-3 L_{BIAS} vs V_{AGC} Plot for Variable Gain Unit

simulation curve. It can be seen that the required L_{BIAS} varies significantly to get the desired DC operating point because of process variations. The replica biasing circuit gives a predictable control of the DC operating point and therefore, the gain of the variable gain units. The operating

region is designed to be for a V_{DS} of 0.5V to 1V for the PMOS load devices. This corresponds to a V_{AGC} of 4.5 to 4V. Note that for a high V_{AGC} , the PMOS devices are deep in the triode region and a high gate bias is required. The highest V_{GS} that can be applied is V_{DD} and force the L_{BIAS} to drop to GND. This imposes the limit on the lowest gain setting as described in Section 5.1.3.2. A good design is to keep this cutoff point as close to the desired operating point of the minimum gain setting as possible (a V_{DS} of 0.5V $\Rightarrow V_{AGC} = V_{DD} - 0.5 = 4.5V$) so that for the high gain setting, when the V_{DS} of the loading devices increases, there is still a high V_{GS} to force the load devices to operate in the triode region. This will achieve the maximum gain range possible for each variable gain units. From the plot shown in Figure 6-3, the design goal of maximum gain range is achieved.

The DC gain of the channel amplifier is measured with the dummy channel. The result for different V_{AGC} is shown in Figure 6-4. From the results obtained, a huge DC offset in the range of 40mV to 50mV appears at the input of the channel amplifiers alone. After further investigation, this DC offset occurs due to a layout error with no prior knowledge of the fabrication processes. During source and drain diffusion implant, the wafers were not spined resulted in a graded profile of source and drain diffusion, making the device asymmetrical. Since only mirror symmetry was used for layout with no step symmetry, the input differential pairs experience a huge DC offset. This is illustrated by Figure 6-5. The process variation gradient causes the pair with mirror symmetry to have different source and drain characteristics while the one with step symmetry has no such problem.

Even though DC offset cancellation is employed in the data channel, the design does not expect such a big offset and this degrades the performance of the chip in many aspects. The ratio of gm of the auxiliary feedback stage to the gm of the input stage is about 5:1. With a 40mV offset at the input, the output has to store an offset of 200mV to cancel this huge input offset. As a result, there is a referred input offset of 2mV for a gain of 100 and severely limit the sensitivity and BER performance of the parallel receiver. A second degradation of this huge DC offset is that it shifted the zero differential output voltage significantly and degrades the jitter performance. This will be discussed later in Section 6.5.

A measurement of gain vs V_{DS} of the loading devices is also made and shown in

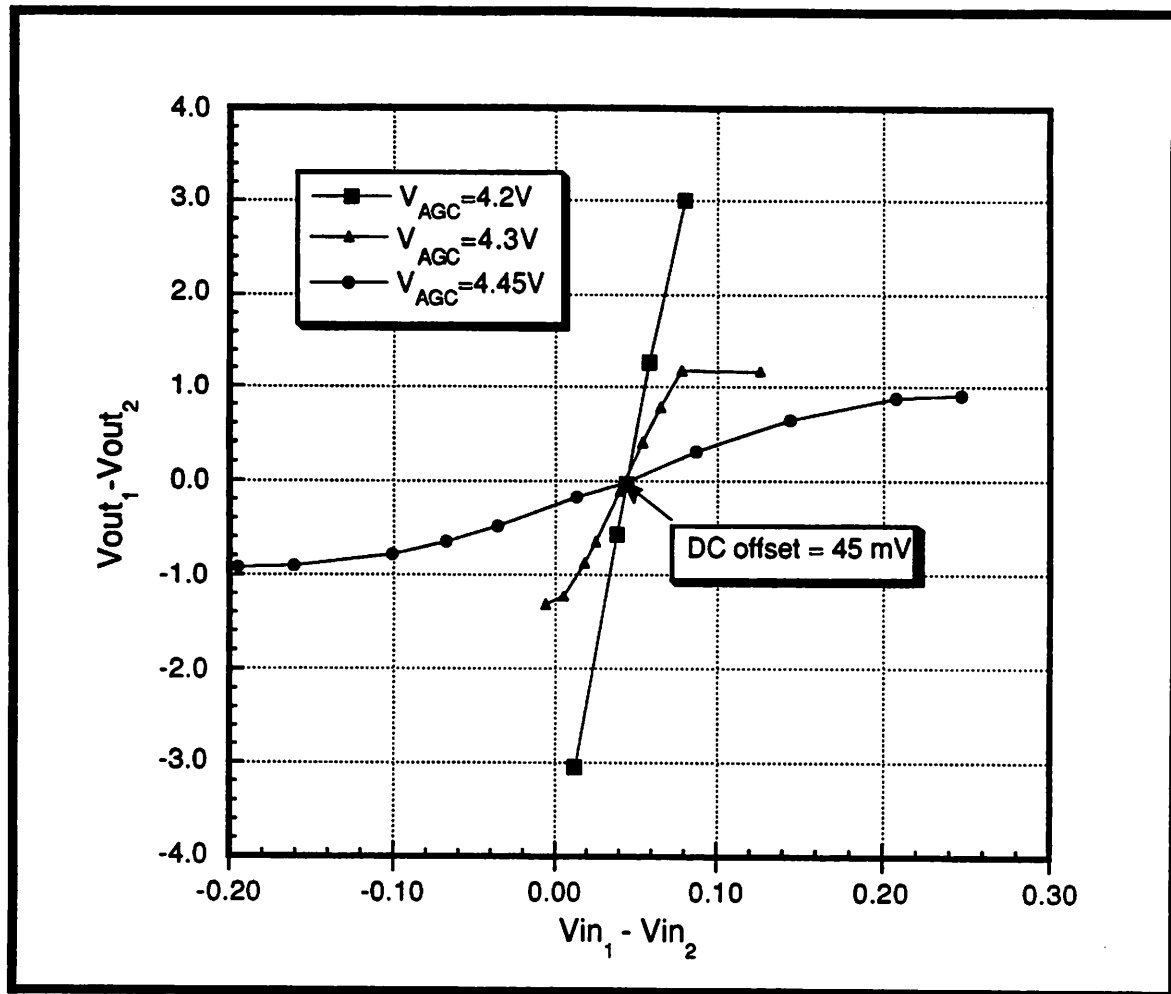


Figure 6-4 DC Gain of Channel Amplifier

Figure 6-6. Also shown on the same plot is the gain as predicted by (EQ 4-15). As k' of this run is higher than the value used for simulation and design by 20%, with the same biasing current, the $V_{dsatN} = V_{GS} - V_t$ of the input differential pair is decreased by about 10%, therefore a value of 0.88V is used for V_{dsatN} in (EQ 4-15). From the plot, it can be seen that the measured results agree closely with the theoretical calculation.

6.4 Bit Error Rate Measurements

From the viewpoint of a digital datacommunication system, the terminal receiver can be completely characterized by the input bit rate, the probability of making an error or the Bit Error Rate (BER), and the jitter performance. This section will present some of the measured performance of the BER under different experimental conditions. Section 6.5 will present some of the

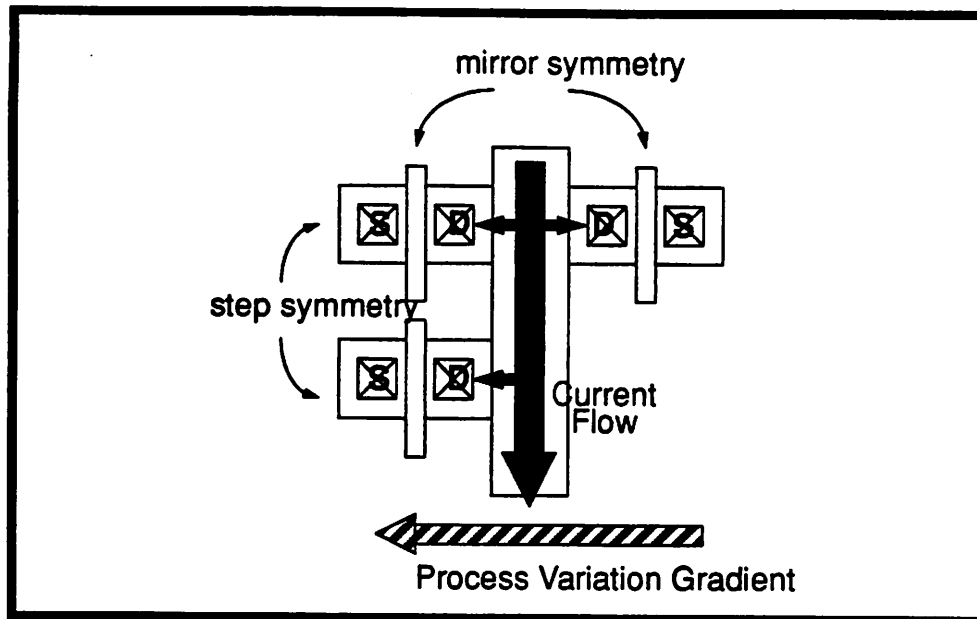


Figure 6-5 Source of High DC Offset

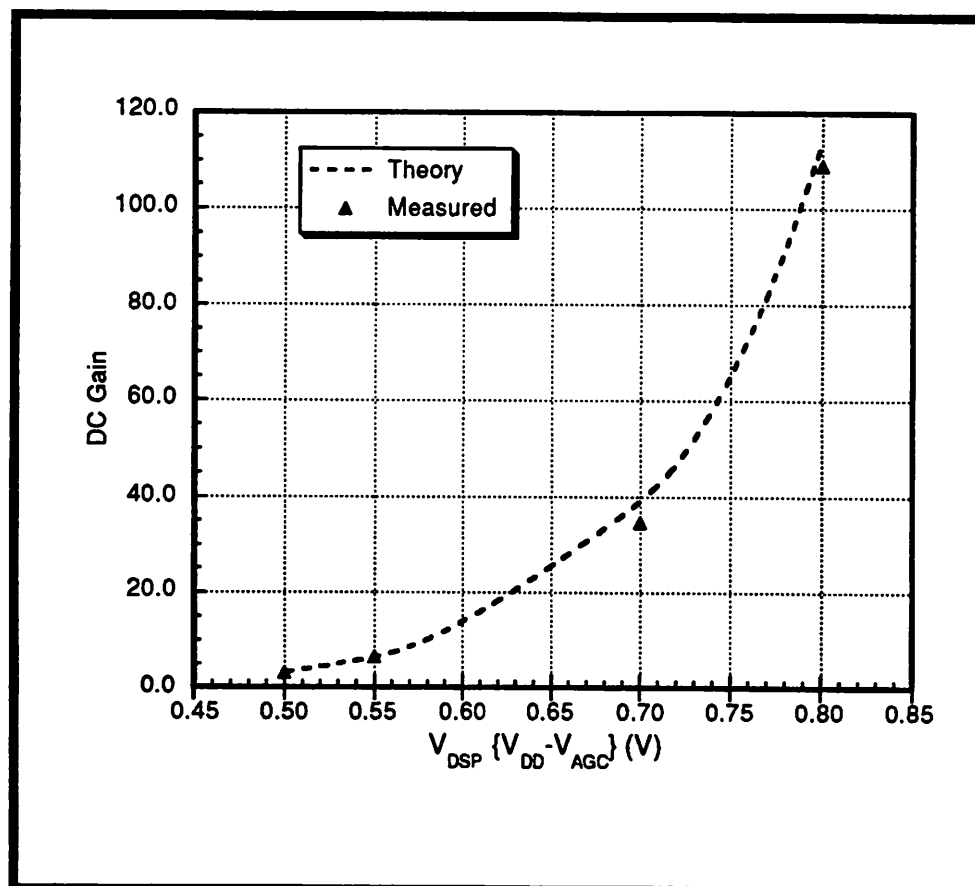


Figure 6-6 DC Gain vs V_{DSP} of Channel Amplifier

measured jitter performance.

The BER is measured using the HP71600 Series of Gb/s Testers. The input is a $2^{15}-1$ Pseudo Random Bit Sequence (PRBS) code which is generated as shown in Figure 6-7. The

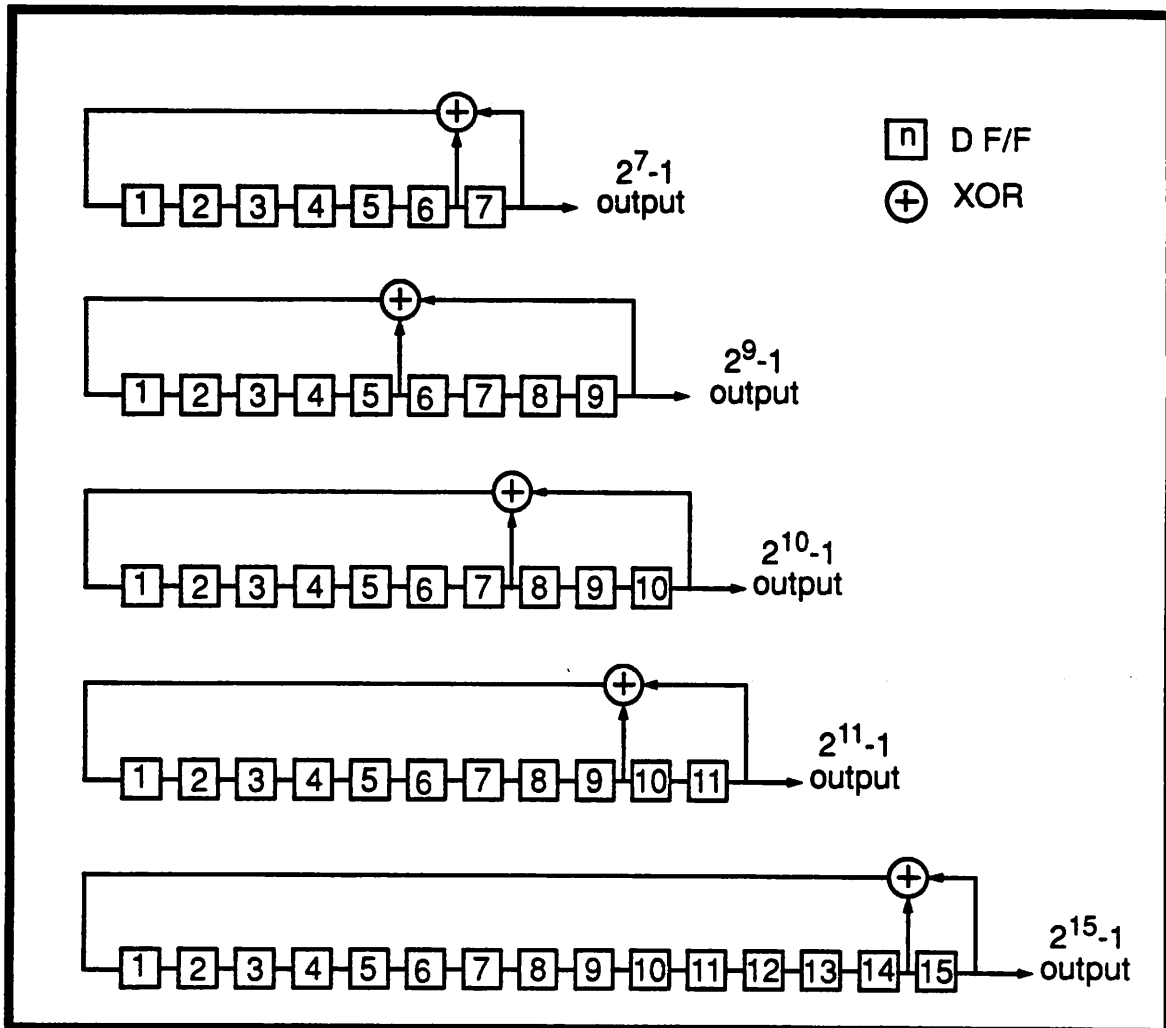


Figure 6-7 PRBS Generation Block Diagram

interesting point about the PRBS code is that if the input is demultiplexed into bit streams of 2^n , the demultiplexed bit stream will have the same pattern (but delayed in time) as the input bit stream itself! This allow measurements of BER of each individual output of the parallel receiver without having to multiplex the outputs back into one high-data-rate bit stream. In order to trick the BER tester to think that the data is transmitted at the demultiplexed rate, a clock that is $1/8$ of the input

data rate is required. Fortunately, this can be taken from the output of one tape of the ring oscillator. The setup for testing BER is shown in Figure 6-8.

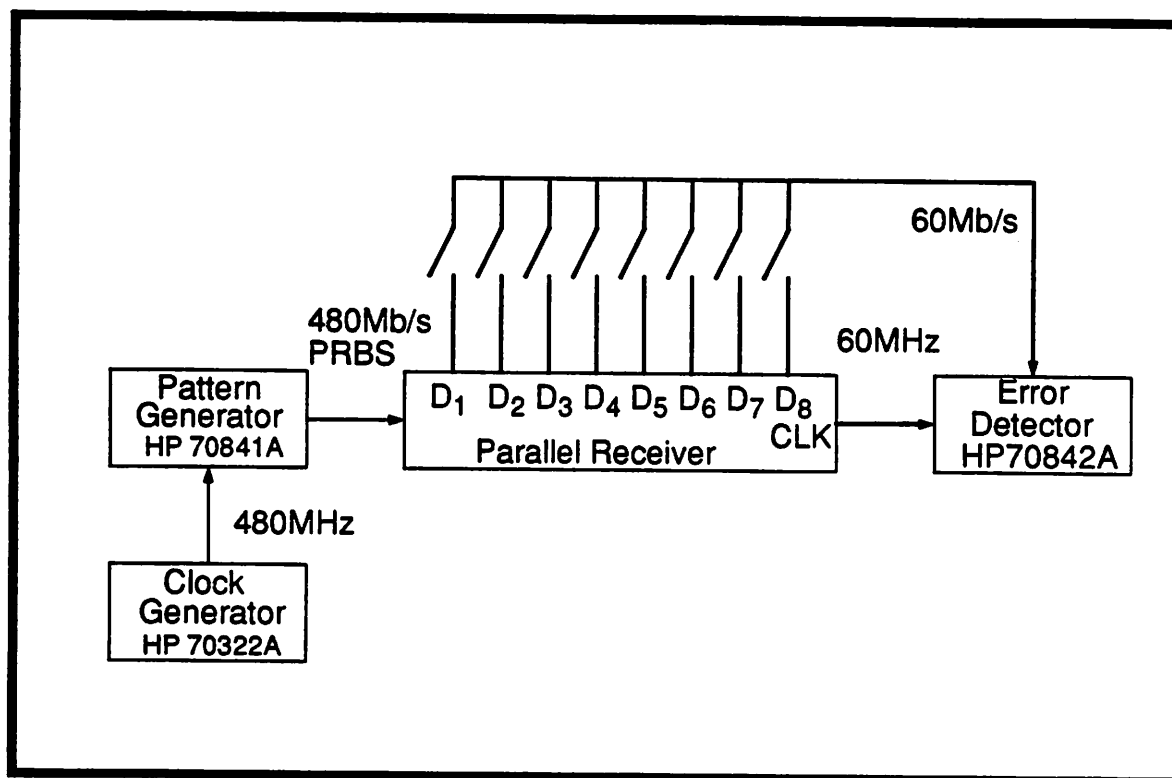


Figure 6-8 Setup for Bit Error Rate Experiment

6.4.1 Minimum Input Voltage vs Bit Rate

An important aspect of a terminal receiver is the minimum peak-to-peak input voltage required to achieve a fixed BER at different bit rates. A plot of this is shown in Figure 6-9. In order to achieve a BER of 10^{-11} , a minimum signal of about $6\text{mV}_{\text{p-p}}$ is required for input bit rate of 400Mb/s and below. It is basically limited by the noise floor which is assumed to be a white gaussian noise. That is why the minimum signal required is independent of the input bit rate. Operation above 400Mb/s requires a much larger input voltage because the receiver was originally designed to run at a maximum bit rate of 400Mb/s. In order to minimize the channel charge injection and clock feed-through, the minimum gate width that can achieve a sample mode bandwidth of 250MHz is used. The prototype receiver is able to achieve a higher bit rate than 400Mb/s because the k' is about 20% larger than expected. However for an input bit rate larger than 400Mb/s, the

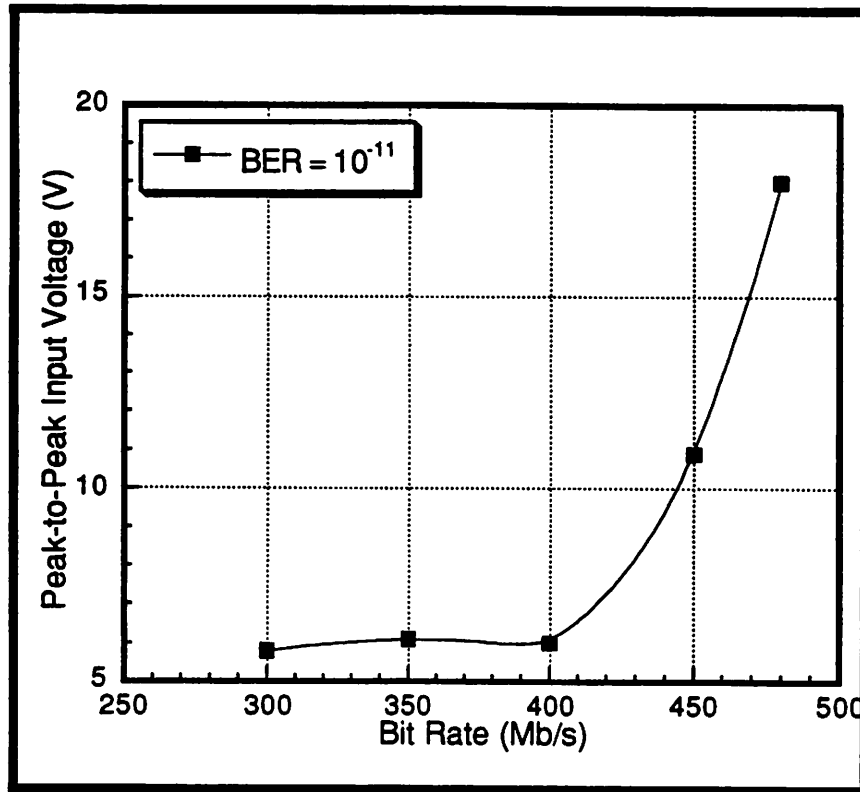


Figure 6-9 Minimum Input Voltage vs Input Bit Rate

performance starts to get limited by the sample mode bandwidth and the speed of the channel amplifier. Higher signal level is needed to overcome the lost due to limited sample mode bandwidth and also lower the gain required by the channel amplifier to speed up time required for amplification. The maximum bit rate that can be achieved is at 480Mb/s with an input voltage of 18mV_{p-p}.

6.4.2 BER vs Input Voltage

Further investigation can be made by examining the BER vs different input voltages at fixed input data rates. From Figure 6-10, it can be seen that the two set of data points can be fitted by exponential curves as predicted by using a simple gaussian noise assumption. The slope of the curve is related to the noise power. Both curves gives approximately the same slope indicating that the noise power is about the same for both cases. This shows that the BER is not limited by the noise floor alone as the BER is now input bit rate dependent. The only explanation to this experimental result is that the input signal is attenuated at high data rate. This can be due to the

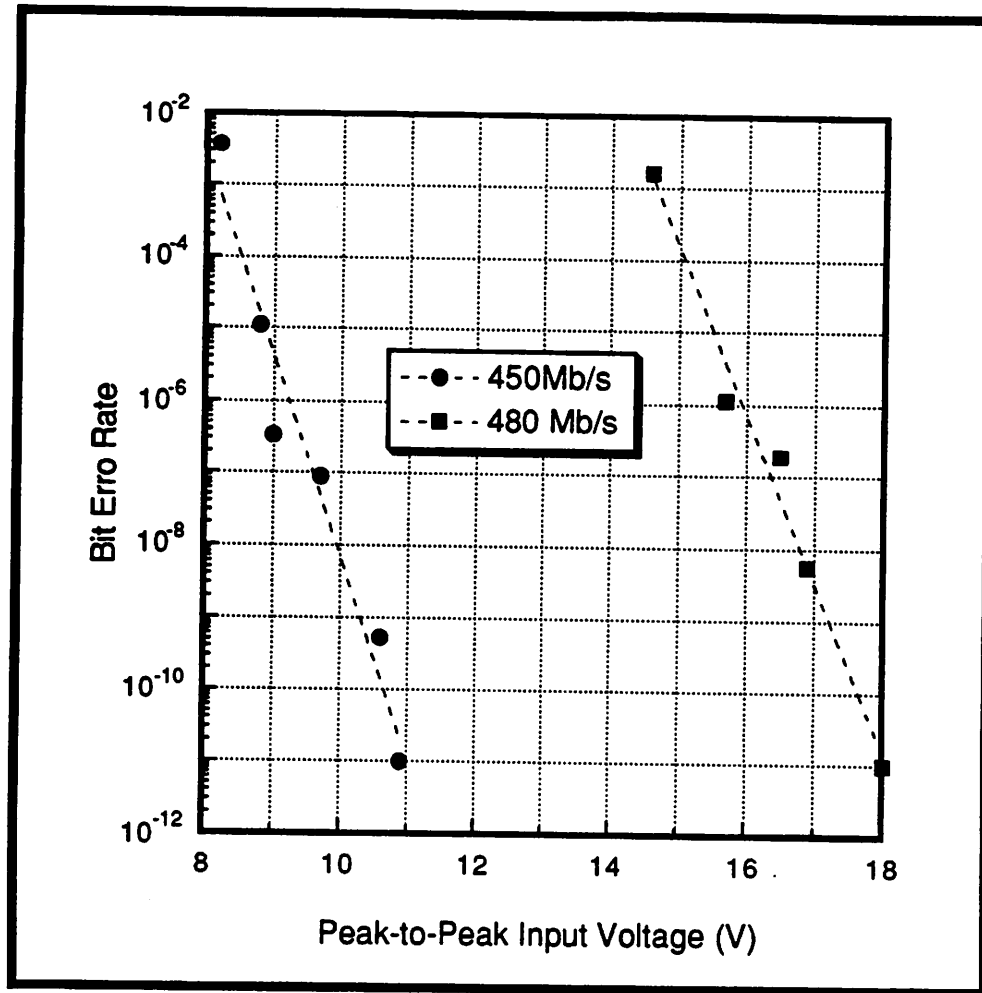


Figure 6-10 BER vs Input Voltage for 450Mb/s and 480Mb/s Input Data Rate limited sample mode bandwidth and the speed of the channel amplifier at high data rate.

An interesting side note is that the inferred root mean square voltage of the noise can be calculated from the slope of the curve. It is about $0.4\text{mV}_{\text{rms}}$ from both curves. In order to achieve a BER of 10^{-11} , the input must be at a level of 13.2 times or higher of that of the root mean square noise voltage. This means a voltage of 5.28mV or higher is required which is exactly what is observed for bit rate of 400Mb/s or lower.

6.4.3 Waveform Dependence of BER

For a digital datacommunication system, the transmitted signal through the data channel is still an analog waveform. If the data channel is not bandlimited, the transmitted signal approaches a square wave. The advantage of this is that the BER performance would not be sig-

nificantly affected by the clock recovery scheme. The sampling instance doesn't have to be at the center of the bit cell to sample the maximum input voltage. The disadvantage is that because the channel is not bandlimited, the noise power is a lot greater than a bandlimited channel.

An experiment is performed with an input bit rate of 480Mb/s, comparing the BER of a square wave input and the BER of a square wave input but coupled in though a 6th order low-pass Bessel filter with corner frequency at 270MHz. The result is shown in Figure 6-11. In high BER

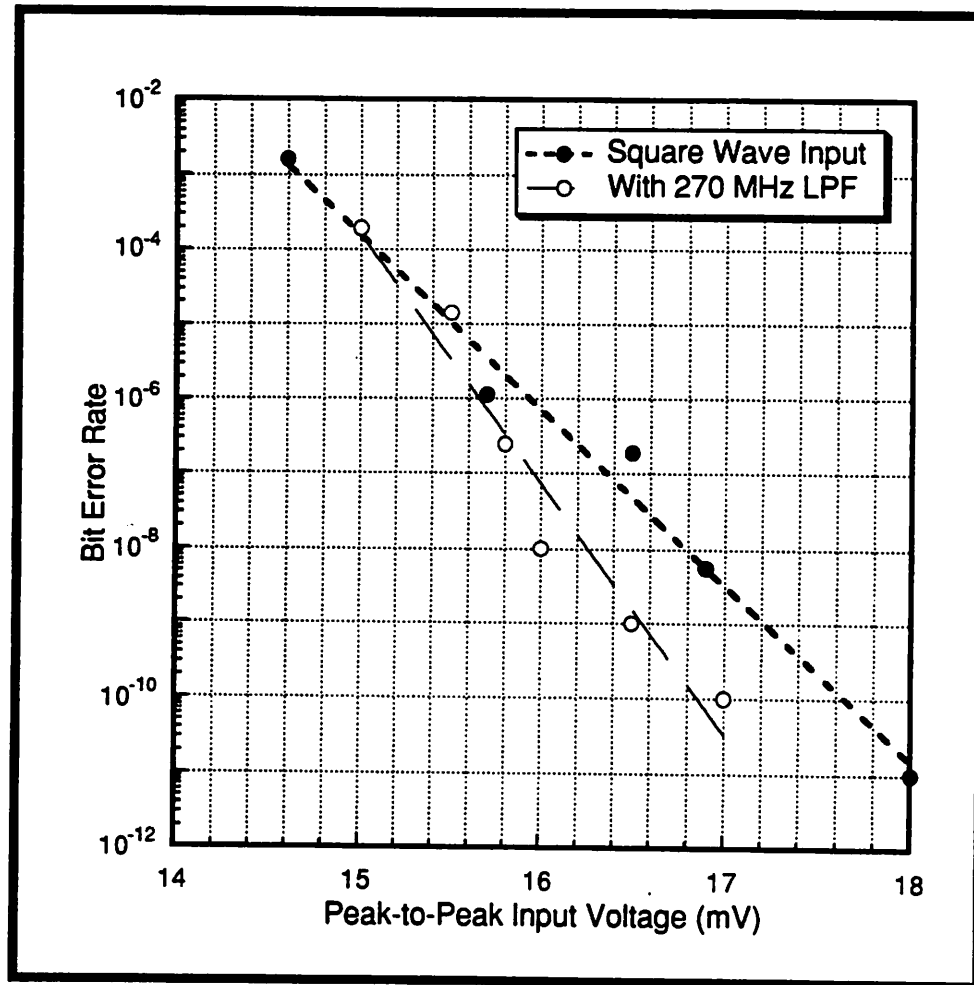


Figure 6-11 Waveform Dependence of BER at 480Mb/s Input Bit Rate

region, the performance of the bandlimited case is a lot better because clock recovery forces the sampling instances to be close to the center of the bit cell making the sampling input voltage for both cases to be the same. As reflected from the slope of the curve, the bandlimited case clearly

has a smaller root mean square noise voltage because the noise bandwidth is also limited by the low-pass filter. However at high BER, the situation reversed as the clock-recovery circuit starts to make wrong decisions because of high BER and the sampling instance is not at the optimal point for the bandlimited case, effectively lowering the input voltage sampled, results in a worse BER performance as compared to the case without the low-pass filter.

As a result of this, a low-pass filter should be added at the input of the receiver as an anti-aliasing filter to lower the noise bandwidth of the input signal for high BER operations.

6.4.4 Channel Dependence of BER

Even though channel matching is not an issue for the parallel receiver, it is still interesting to see the BER performance for each parallel data channel because the gain control loop updates the L_{bias} for all the parallel channels based on the output voltage of one single channel. A measurement of individual BER of each parallel channel will give an idea of how important matching between each channel is. Figure 6-12 shows the BER of each individual channel at an input rate of 480Mb/s. Data for two input levels are shown. There is no significant evidence to support the fact that any one channel is better or worse than any other. However, a mean square linear regression line shows that the mean value of the BER has an increasing trend with higher channel number. The parallel channels are numbered from bottom to top in the layout with channel 8 closest to the VCO and digital circuits. The slight increase of BER with higher channel number may be attributed to more coupling of the noise from the VCO and digital circuits through the substrate. Also, from this measurement, it can be seen that the channels are matched well with each other.

6.4.5 Supply Dependence of BER

Another important measurement is the Bit Error Rate performance of the parallel receiver with different supply voltage. It is shown in Figure 6-13.

With a higher supply voltage, the oscillation frequency of the ring oscillator increases with the same VCTRL as the load capacitance remains the same. The sample mode bandwidth increases as a higher turn-on voltage is available. The variable gain unit works the same way as

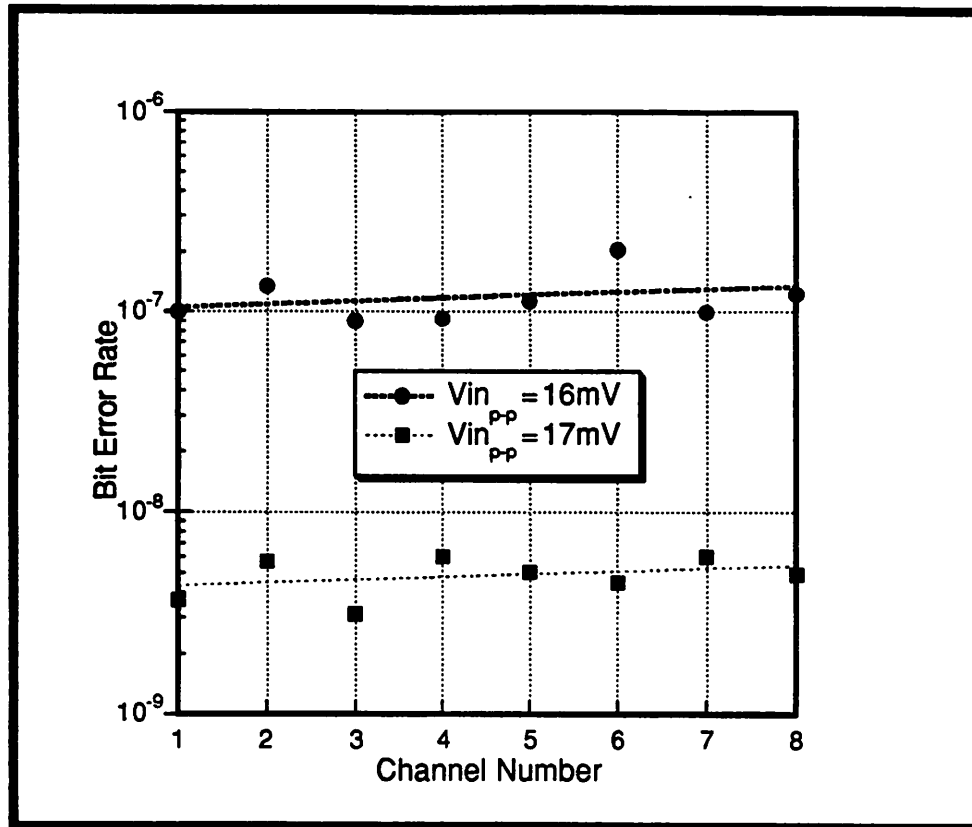


Figure 6-12 Bit Error Rate of Parallel Channels at 480Mb/s

before as every voltage is referenced to the supply voltage. A higher gate voltage can be applied to the load devices results in a higher gain range.

From the plot in Figure 6-13, the BER performance is only slightly better at a supply voltage of 5.25V. This shows that the limiting factor may not be the sample mode bandwidth but the speed of the channel amplifier. This is further supported by the fact that the slope of the curve is about the same as the 5V supply case indicating that the channel bandwidth is about the same.

For a lower supply voltage, we have the reverse effect of increasing the supply voltage, namely, lower VCRO frequency at the same VCTRL, lower sample mode bandwidth and a limited gain range.

The result of lowering the power supply is a significant decrease in BER performance. The limitation is dominated by the decreased sample mode bandwidth. From the slope of the curve, it can be seen that the root mean square noise voltage is actually smaller than the higher

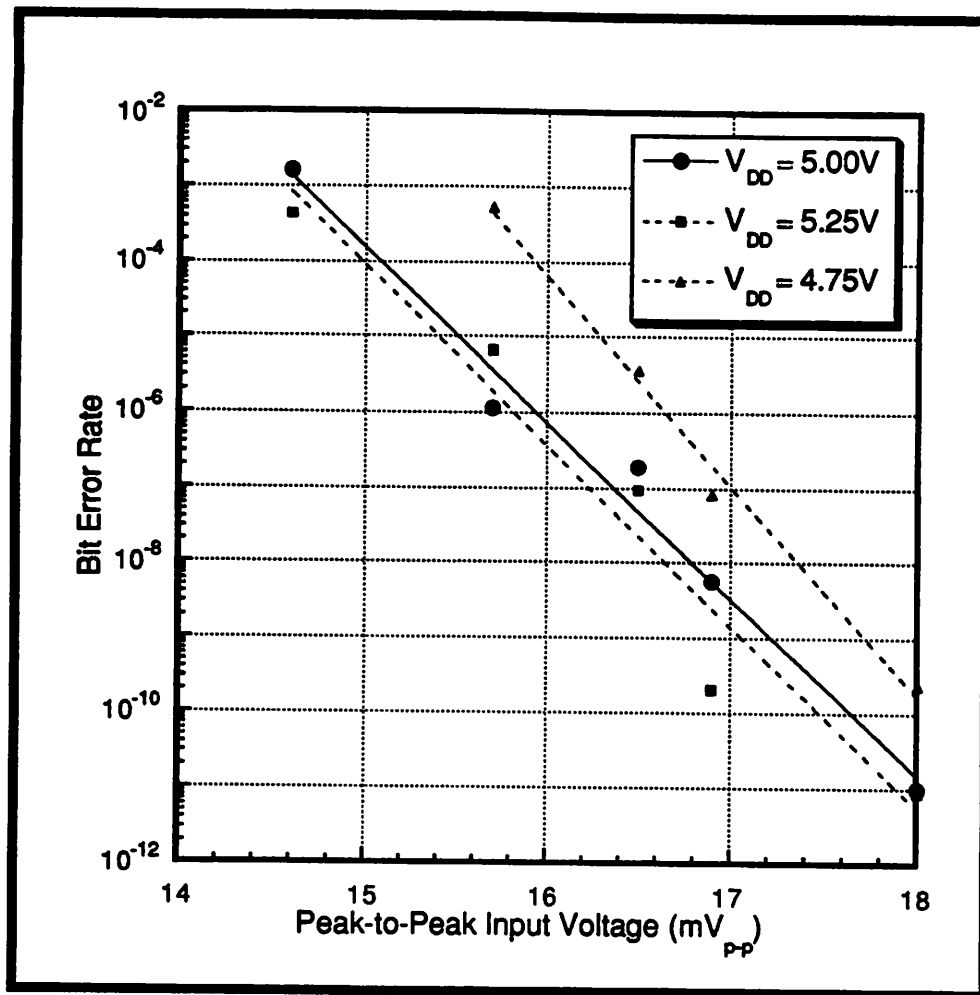


Figure 6-13 BER vs Supply Voltage at Input Bit Rate of 480Mb/s

supply cases, further supporting the fact that the channel bandwidth is lowered.

This plot gives a hint that for low power operation with low supply voltage, an internal DC-to-DC up-conversion is needed to provide a higher internal gate voltage to the input sampling switches. This can be done by charge pump techniques such as those in switching regulators and EPROM programming circuits.

6.5 Jitter Performance

For a digital datacommunication network, jitter is characterized by jitter transfer, jitter tolerance, and jitter generation. A linear, shift-invariant jitter model was shown to be experimentally valid for fiber optic regenerators. [94]. The jitter of a regenerator can be characterized by a jitter transfer function. The jitter transfer function is important for minimizing jitter generation and jitter

accumulation. A lack of proper equipments prevents the measurement of this jitter transfer function for the parallel receiver. However, for a terminal receiver where there is no further transmission of data received, a direct measurement of the jitter tolerance is more meaningful.

The jitter tolerance is an effective measurement of the parallel receiver's capability to tolerate incoming accumulated jitter. Excessive transmission penalty will not occur if the accumulated jitter does not exceed the parallel receiver's measured jitter tolerance template. The jitter tolerance is defined to be the magnitude of incoming jitter that results in a specific transmission penalty.

It can be shown that sinusoidal jitter has a distribution that causes a much larger penalty than jitter with a truncated Gaussian distribution [56]. Therefore, sinusoidal input jitter is used in the definition of jitter tolerance, since it can be considered as having a worst case distribution for most digital transmission systems. By selecting a BER penalty that is small and tolerable (e.g. 0.5dB), the peak-to-peak input jitter can be measured as a function of the jitter frequency which causes this tolerated BER penalty. This measurement will trace out a jitter tolerance template that, if not exceed, will assure that the receiver performs with less than the tolerated BER penalty.

The setup for the measurement of jitter tolerance is exactly the same as in Figure 6-8 except the fact that the output of the clock generator is FM modulated by another sine wave input with amplitude A and frequency f_{pm} . The FM modulated clock signal is therefore [95]

$$s(t) = A_c \cos \left[2\pi f_c t + \frac{k_f A}{f_{pm}} \sin(2\pi f_{pm} t) \right] \quad (\text{EQ 6-1})$$

where A_c is the clock amplitude, f_c is the clock frequency, K_f is the frequency modulation index. For most FM modulator, a peak frequency deviation $\Delta f = K_f A$ is usually used. Note that the phase of the clock is now modulated by a sinusoidal signal resulting in a controlled sinusoidal jitter with peak-to-peak input jitter $2K_f$ in degree where

$$2K_f = \frac{\Delta f}{f_{pm}} \times \frac{360^\circ}{2\pi} \quad (\text{EQ 6-2})$$

A low pass filter at 270MHz is added to the input to limit the bandwidth of the channel and round-off the input square wave for testing the degradation of performance due to alignment jitter.

With the initial jitter set to zero, the input voltage is adjusted to give a BER of 10^{-9} . Next the amplitude is increased by 0.5dB so that the BER is smaller than 10^{-9} . The jitter tolerance is measured by increasing the peak frequency deviation Δf until the BER gets back to 10^{-9} and the $2K_i$ is calculated from Δf and f_{pm} . The procedure is repeated at various f_{pm} . The plot obtained is shown in Figure 6-14.

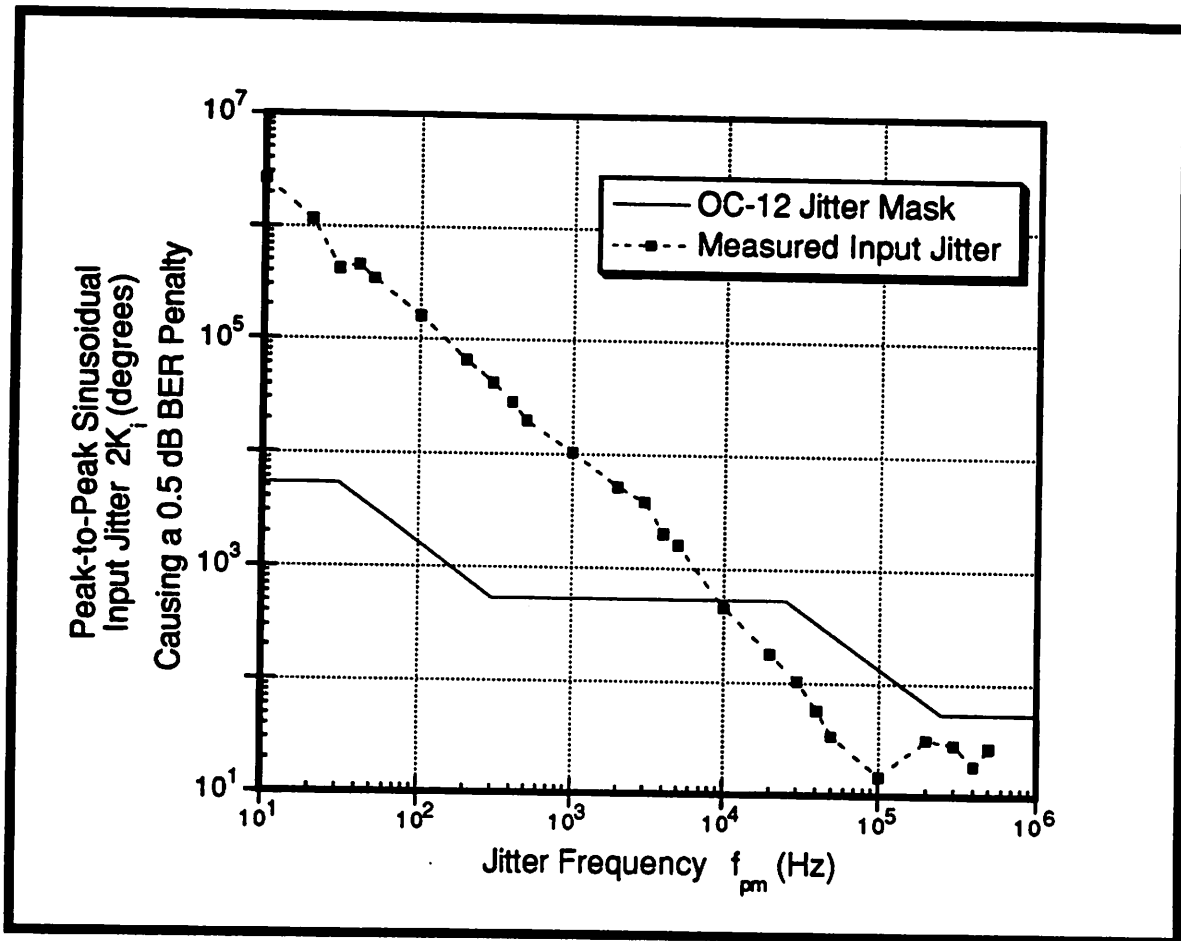


Figure 6-14 Peak-to-Peak Sinusoidal Input Jitter Tolerance at 480Mb/s

For most applications, a jitter mask will be set to ensure the performance of the whole system. An example which is also shown in Figure 6-14, is taken from the specification for the Synchronous Optical NETwork (SONET), Category II, layer OC-12, operating at 622.08Mb/s [96]. The jitter measured must be above this mask to meet the specification. The parallel receiver failed to meet the requirement at jitter frequencies higher than 10KHz.

The poor jitter performance is mainly due to the output offset stored to cancel the hugh

input offset. It is possible to estimate this extra jitter by making some simple assumptions. First, the maximum voltage deviation on VCTRL can be approximated by

$$\Delta V = \frac{I_{\max}}{C_3} \Delta t_c \times G_{m2} R_o \quad (\text{EQ 6-3})$$

where I_{\max} is the maximum charging current from the charge pump, C_3 is the capacitance of the external loop filter. Δt_c is the time the charge pump is on. With $I_{\max} = 400\mu\text{s}$, $\Delta t_c = 2T = 4\text{ns}$, $C_3 = 1.2\text{nF}$, $G_{m3}R_o = 0.6$; therefore $\Delta V = 0.8\text{mV}$. With the oscillator gain $K_o = 13\text{MHz/V}$, the Δf is therefore 20.8KHz, that corresponds to a Δt of 2.9ps. Normally, this would be the jitter for the clock with no huge DC output offset. Now with the offset, the PRBS = $2^{15}-1$ code can have 15 consecutive 0 to 1 transitions and 1 to 0 transitions for the timing channel. That means the sampling edge can be moved 15 times up and 15 times down the average point. Assuming the sampling edge has an equal chance of staying in any position allow by the shift, the rms jitter is therefore

$$\text{jitter}_{\text{rms}} = \sqrt{\frac{(30\Delta t)^2}{12}} = 25\text{ps} \quad (\text{EQ 6-4})$$

So assuming everything else perfect, the clock will have a 25ps rms jitter and a peak-to-peak jitter of 89ps just because of the stored output offset. The real degradation of course depends on the statistics of the input data bit stream and can be much worse than the approximation.

A jitter tolerance measurement at high input voltage to reduce the effect of offset is impossible to make because of the extremely low BER. However, a look at the output clock jitter can give a clue to what is going on.

The PC board required to test the prototype is too big making probing of the ring oscillator impossible. Two taps of the ring oscillator are available off-chip through digital output pad driver which is extremely noisy by themselves, so the root mean square jitter measured must be taken with a grain of salt. However, it is still useful to compare the jitter with different levels of input. A processing scope is used to measure the root mean square jitter at 480Mb/s. With an 18mVp-p input, the measured rms jitter of the clock is about 78ps!! With an 100mV input, the measured rms jitter dropped to 24ps!!! This gives a strong evidence that the excessive amount of jitter is due to

the high offset of the input differential pairs of the channel amplifiers.

6.6 PLL Performance

Initial acquisition is done by sweeping the current source I_o in Figure 5-9 to sweep the free running frequency of the ring oscillator to close to the demultiplexed bit rate. The CPPLL is then locked to the input PRBS bit stream. The loop bandwidth is close to 200KHz as evidence from the corner frequency of the jitter tolerance plot in Figure 6-14. The capture range of the parallel receiver is about $\pm 500\text{KHz}$ while the lock range is about $\pm 1\text{MHz}$ for an input bit rate of 480Mb/s. False lock is detected when a PRBS = 2^7-1 is used at 480Mb/s. The loop locks at 60.472MHz which is exactly 128/127 of 60MHz. The PRBS = 2^7-1 code has a run length of 127 bit which shows that this is a pattern dependent false lock. No such case is observed with a PRBS = $2^{15}-1$ code.

The capture range and the lock range is increased tremendously when a training sequence of 1111111100000000 is transmitted before the random data. In such case, each channel receives consecutive 0 and 1 and there is a 0 to 1 and 1 to 0 transition each time for clock recovery. The capture range increases to $\pm 5\text{MHz}$ because of this frequency assisted pull-in. The lock range increased to practically the range of the ring oscillator. This increase in capture and lock range is due to the fact that only one edge is available for clock recovery and it occurs at every 8 bits making the timing information clear as a whistle to the phase detector. The added advantage of sending a training sequence is the fact that word framing can be done for free because the end of word frame is lying exactly in the timing channel.

6.7 Performance Summary

The performance of the experimental prototype is summarized in Table 6-1.

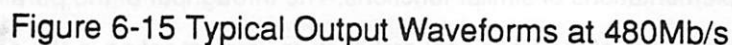
$T = 25^{\circ}$, $V_{DD} = 5V$, PRBS = $2^{15}-1$	
Maximum Input Bit Rate	480Mb/s
Minimum Input Amplitude for BER = 10^{-11} @ 400Mb/s @ 480Mb/s	18mVp-p 6mVp-p
Output Data Amplitude	5Vp-p CMOS level
Power Dissipation	900mW
AGC Gain Range	0dB to 40dB
Chip Size	4mm X 4mm
Active Area	3mm X 3mm
Technology	1.2 μ m CMOS
Package	68-pin LCC

Table 6-1 Achieved Performance Summary

The total power dissipation of the chip is 900mW. 100mW is dissipated on the pad I/O drivers to carry the digital output from the eight parallel channels and two taps of the ring oscillator off-chip. 300mW is dissipated on the ring oscillator. The high power dissipation in the ring oscillator is because of the number of inverting stages needed to generate the required sampling edges and the extra inverting stages needed to generate the sampling edge for the timing channel. The device sizes of the unit delay element is larger than minimum because it has to drive a routing matrix to equalize the loading of each tap in order to get the evenly spaced sampling edges. Each parallel channel is consuming about 45mW. For ten parallel channels, the power consumption is about 450mW. The remaining 150mW is dissipated on the clock recovery, digital and biasing circuits. If integrated with a VLSI digital data link control, the pad I/O drivers can be eliminated and the dummy channel can also be eliminated, resulting in a reduction of 150mW in power consumption.

The prototype chip is packaged in a 68-pin LCC chip carrier. Even though 68 pins are used, most of them are power supply pins and testing features.

An oscilloscope picture of some typical output waveforms are shown in Figure 6-15. The bottom waveform is the reference clock at 480MHz used by the pattern generator to transmit the



117

Chapter 7 Conclusion

7.1 Summary of Research Results

- It has been demonstrated that the parallel architecture can be used for the implementation of high-data-rate, digital direct detection, optical fiber receivers with a relatively low speed technology. Operation at 480Mb/s is demonstrated with a 1.2 μ m commercially available CMOS technology [97].
- With non-optimal design, a device f_T to data rate ratio of about 5 to 1 can still be achieved as shown in Figure 7-1. This is the fastest result reported so far as compared with other CMOS implementations of similar functions. The throughput of the parallel architecture is at least 2 times that of the traditional architecture, implementing in the same technology.
- Clock recovery can be done by inserting an extra parallel channel for phase detection. More parallel channels can be added for frequency detection to improve the performance of the clock-recovery circuit.

The research results are elaborated in the following two subsections. The first part compares the parallel architecture with traditional architecture, and the second part outlines the special issues involved in the design of parallel receivers.

7.1.1 Traditional Architecture vs Parallel Architecture

The main disadvantage of the traditional architecture is that all circuitry has to have bandwidth close to the input data rate. A wide-band AGC amplifier is needed to amplify the low-level signal output from the preamplifier for further processing. A high frequency mixer is needed for clock-recovery. A high frequency, voltage-controlled oscillator, oscillating at the input bit rate is

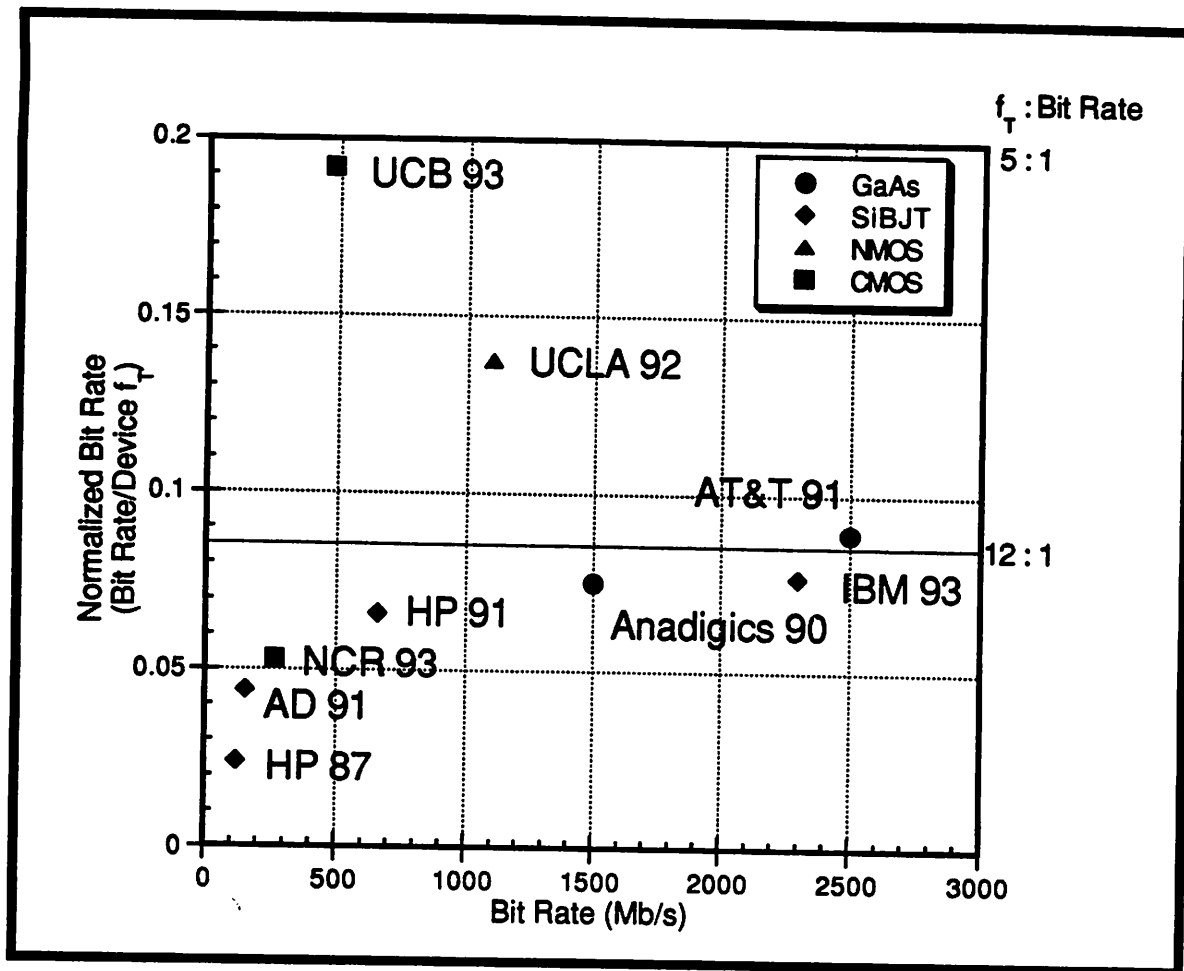


Figure 7-1 Comparison of Parallel Receiver with Other Architectures

needed to generate the clock edges for the decision circuit. The implementation of these three major high frequency parts requires a bandwidth at least half of that of the input data-rate. This imposes a limit on how high a data-rate can be achieved with a given technology with the traditional architecture.

The speed of receivers depends on the circuit techniques used and the speed of the technologies for implementation. By normalizing the bit-rate achieved with the device f_T of the technology, comparison between different circuit implementations can be made as in Figure 7-1. For most of the implementation of the traditional architecture, a device f_T to data-rate ratio of 12:1 is required. As a result, for Gb/s operations, a technology with device f_T of 12GHz or above is needed.

All high-data-rate receivers nowadays are implemented in high speed technologies such

as GaAs or high speed Si BJT because of the speed requirement. The drawbacks of these technologies are high cost, and low level of integration. All current solutions are either a multi-chip approach or a hybrid approach because of the low level of integration, and they are all expensive because of the cost of the technology, package, and assembly required.

The high cost of terminal receivers severely limits the deployment of high-data-rate optical fiber communication systems.

The advantage of the traditional architecture is that when high jitter rejection is not needed, an open-loop approach or a wide-band approach relaxes the speed requirement imposed on the clock recovery circuit allowing a much higher data-rate.

The parallel architecture, on the other hand, identifies the bottleneck and shift it to the high-data-rate demultiplexer which can be achieved even with a relatively slower technology. By demultiplexing the high-data-rate bit stream right-a-way relaxes the requirement imposed on the main amplifier and clock recovery. Also the use of taps from a ring oscillator oscillating at a much lower frequency, achieves the same as a combination of a high frequency VCO and a high speed frequency divider. This eliminates the need for the high frequency VCO. As a result, besides the input analog demultiplexer, there is no other high frequency component in the whole parallel architecture.

Lowering the speed requirement allows the use of the low-cost, high-integration scaled CMOS technology. The power consumption and chip area required to achieve a certain input bit-rate is comparable to implementations of traditional architecture by other high speed technologies because of the lack of high frequency components in the parallel architecture and the high level of integration of CMOS technology.

A complete solution can be as simple as an OEIC implementing the photoreceiver and then followed by a CMOS VLSI mixed-signal IC implementing the rest of the functions. Such solution can lower the cost of present day high-data-rate receivers tremendously.

The disadvantage of the parallel architecture is that since the input signal is demulti-

plexed and sampled, it is inherently a sampled data system, making the implementation of the open-loop approach impossible.

7.1.2 Special Issues in Designing a Parallel Receiver

The design of a parallel receiver is very flexible. The number of parallel stages used will depend on the system requirements. The ultimate bit rate that can be achieved depends on the speed of the CMOS technology used and the number of parallel stages used.

The ultimate device f_T to data rate ratio that can be achieved by the parallel architecture, based on circuit techniques used today, is about 4:1, as limited by the input analog demultiplexer. The maximum number of channels used, to take full advantage of the speed difference in the implementation of analog demultiplexing and amplification in the same technology is about 12. If the number exceeds 12, the performance will be limited by the speed of the input demultiplexer, resulting in no improvement in the total throughput.

The circuit design and implementation of the parallel receiver in this experiment is by no means optimal. Improvement can be made by several approaches. First is the use of a truly differential ring oscillator to minimize the jitter due to the supply noise coupling. Secondly, all the variable gain stages used in the AGC amplifier are of the same size and the same gain. Faster response can be achieved by sizing the cascading stages.

Based on the above assumptions, and the experimental results of the prototype, a 622Mb/s SONet receiver can be implemented with 12 parallel channels with the same 1.2 μ m CMOS technology!

7.2 Projected performance in Scaled Technologies

The device f_T of a scaled CMOS technology increases with decreasing minimum channel length. For the analog demultiplexing, there is a constant trade-off of speed and sensitivity as shown in (EQ 5-8) and repeated in here.

$$\tau \times V_d = \frac{L^2}{2\mu} \quad (\text{EQ 7-1})$$

A closer look at the constant term on the right-hand-side of the equation in (EQ 7-1) shows that it is directly related to the $\tau_T = 1/\omega_T$ of the technology. Therefore, the increase in speed is directly proportional to device f_T with the same sensitivity requirement. If the bit rate is set to below the maximum achievable, then higher sensitivity should be obtained. The required device f_T to bit rate ratio remains the same at 4:1. The speed of the AGC amplifier scales the same way. As a result, the overall device f_T to data rate ratio of 4:1 and a maximum of 12 parallel channels remain the same for scale technologies.

With the above assumptions, Gb/s operation is possible with the parallel architecture implemented in 1.0- or 0.8- μm CMOS technologies.

7.3 Future Work

An immediate need for the parallel architecture is a more robust clock recovery scheme. The present clock recovery scheme based on just one extra timing channel may be good enough for doing experiment with the parallel architecture but it is suffering from small capture range and false lock problems. It also have problems with large input offset and pulse distortion. A more robust algorithm based on two or more timing channel is possible and should be looked into for a practical implementation of the parallel architecture.

For short wavelength (0.8- to 1.0- μm) optical communication channels, Silicon can be used as photodetectors. The parallel architecture should be modified to push the analog demultiplexing all the way to after the photodetector and moved the preamplifier into the parallel channels. However, this impose a new problem as the signal that is demultiplexed is no longer a voltage but the photocurrent. The design of current demultiplexer can be challenging.

First, non-overlapped clock must be used for demultiplexing which is not necessary for demultiplexing a voltage signal. This reduces the maximum clock rate that can be achieved as extra gatings are needed from the ring oscillator to generate the non-overlapped clocks, and there must be guard time between them to account for clock skew.

Yet, another more difficult problem is that the photodetectors all have parasitic capaci-

tances associated with them. Demultiplexing the photocurrent is in effect charge sharing between the sampling capacitor and the parasitic capacitor. The residual charges remains on the parasitic capacitor is coupled into the next parallel channel resulting in inter-symbol-interference (ISI). A simple solution is to increase the size of the sampling capacitor so that most of the charges will remain on the sampling capacitor and reduce the effect of ISI. However, this will reduce the sample-mode bandwidth and greatly lower the maximum speed achievable. A better method is to use an equalizer to reduce the effect of ISI. A simple scheme that was looked into and shown to be successful from simulations is the use of decision feedback equalization to match the sampling capacitance with the parasitic capacitance. If the sampling capacitance is the same as the parasitic capacitance, then half of the signal is always remained on the parasitic capacitor. The ISI can easily be cancelled out by subtracting from the present sample voltage half of the previous sampled voltage! The penalty of this method is a 3dB reduction in SNR. The sampling capacitance can be altered by a digitally weighted capacitance array as the typical value of the photodiode capacitance is always known. A practical implementation should be investigated once the parallel architecture is proved to be feasible.

References

- [1] T. Li, "Advances in optical fiber communications: An historical perspective," *IEEE Journal on Selected Areas in Communications*, vol. SAC-1, n. 3, pp. 356-372, Apr. 1983.
- [2] N. K. Cheung, "Lightwave system requirements for integrated circuits," *Technical Digest of IEEE Optical Fiber Communications Conference*, pp.66-68, Jan. 1990.
- [3] M. Schwartz, "Modulation techniques," *Information Transmission, Modulation and Noise*, McGraw-Hill, New York, pp.209-316, 1980.
- [4] J. Gower, *Optical Communication Systems*, Prentice-Hall, London, 1984.
- [5] T. Okoshi, "Recent Advances in coherent optical fiber communication systems," *IEEE Journal of Lightwave Technology*, Vol. 5, No.1, pp. 44-52, Jan. 1987.
- [6] K. Y. Lau, A. Yariv, "High-frequency current modulation of semi-conductor injection lasers," Chapter 2, *Semiconductors and Semimetals*, vol. 22, Part B, Academic Press, Orlando, 1985.
- [7] R. Ballart, Y. C. Ching, "SONET: Now It's the Standard Optical Network," *IEEE Communication Magazine*, pp.8-15, Mar. 1989.
- [8] H-M. Rein, et al., "30 Gbit/s multiplexer and demultiplexer ICs in silicon bipolar technology," *Electronics Letters*, vol. 28, pp. 97-98. 1992.
- [9] A. Felder, et al., "25 to 40Gb/s Si ICs in Selective Epitaxial Bipolar Technology," *ISSCC Digest of Technical Paper*, pp. 156-157, Feb 1993.
- [10] A. L. Fisher, N. Linde, "A 50Mb/s CMOS optical transmitter integrated circuit," *IEEE Journal of Solid-State Circuits*, Vol.21, No. 4, pp.901-908, June 1986.

- [11] G. K. Chang, P. W. Chumate, "Novel high-speed LED transmitter for single-mode fiber and wideband lop transmission systems," *Electronic Letters*, Vol. 23, pp.1338-1340.
- [12] A. Susuki, et al., "Gb/s modulation of heavily Zn-doped surface-emitting InGaAsP/InP DH LED, *Electronic Letters*, Vol. 20, pp.273-274, 1984.
- [13] R. G. Schwartz, B. A. Wooley, "Stabilized biasing of semiconductor lasers," *Bell System Technical Journal*, Vol. 62, pp.1923-1936, 1983.
- [14] H-M. Rein, "Multi-Gigabit-Per-Second Silicon Bipolar IC's for Future Optical-Fiber Transmission Systems," *IEEE Journal of Solid-State Circuits*, Vol. 23, No. 3, pp. 664-675, June 1988.
- [15] D. B. Keck, "Single-Mode Fibers Output Perform Multimode Cables," *IEEE Spectrum*, Vol.20 No. 3 pp.30, March 1983.
- [16] D. Gloge, "Weakly Guiding Fibers," *Applied Optics*, Vol. 10, no. 10, pp. 2252-2258, Oct. 1971.
- [17] J. C. Simon, "GaInAsP semiconductor laser amplifiers for single mode fiber communications," *IEEE Journal of Lightwave Technology*, Vol. 5, no. 9, pp. 1286-1295, Sep.,1987.
- [18] T. Saitoh, et al., "Recent progress in Semiconductor Laser Amplifiers," *IEEE Journal of Lightwave Technology*, Vol. 6, no. 11, pp. 1656-1664, Nov. 1988.
- [19] K. Nakagawa, et. al., "Trunk and Distribution Network Application of Erbium Doped Fiber Amplifier," *IEEE Journal of Lightwave Technology*, Vol. 9, no. 2, pp. 198-208, Feb., 1991.
- [20] B. J. Ainslie, "A Review of the Fabrication and Properties of Erbium-Doped Fibers for Optical Amplifier," *IEEE Journal of Lightwave Technology*, Vol. 9, no. 2, pp. 220-227, Feb., 1991.
- [21] J. Salz, "Modulation and detection for coherent lightwave communications," *IEEE Communication Magazine*, vol. 24, pp. 38-49, June 1985.
- [22] O. Wada, et al., "High-Performance, High Reliability InP/GaInAs p-i-n Photodiodes and Flip-Chip Integrated Receivers for Lightwave Communication," *IEEE Journal of Lightwave Technology*, Vol. 9, No. 9, pp.

1200-1207, Sep. 1991.

- [23] J. E. Bowers, C. A. Burrus, F. Mitschke, "Millimeter-Waveguide-mounted InGaAs photodetectors," *Electronic Letters*, Vol. 22, pp. 633-635, 1986.
- [24] J. L. Gimlett, "Ultrawide Bandwidth Optical Receivers," *IEEE Journal of Lightwave Technology*, Vol. 7, No. 10, pp.1432-1437, Oct. 1989.
- [25] G. E. Stillman, C. M. Wolf, "Avalanche photodiodes," Chapter 5 in *Semiconductors and Semimetals*, Vol. 12, edited by R. K. Willardson and A. C. Beer, Academic Press, New York, 1977.
- [26] J. C. Campbell, B. C. Johnson, G. J. Qua, W. T. Tsang, "Frequency Response of InP/InGaAsP/InGaAs Avalanche Photodiodes," *IEEE Journal of Lightwave Technology*, Vol. 7, No. 5, pp. 778-784, May 1989.
- [27] S D. Personick, *Optical Fiber Transmission Systems*, Plenum Press, new York, 1981.
- [28] B. L. Kasper, J. C. Campbell, "Multigigabit-per-second avalanche photodiode lightwave receivers," *IEEE Journal of Lightwave Technology*, Vol. 5, pp. 1351-1364, 1987.
- [29] J. L. Gimlett, "Low-noise 8-GHz p-i-n/FET optical receiver," *Electronic Letters*. Vol. 23, pp. 281-283, 1987.
- [30] P. J. Lim, et al., "A 3.3-V Monolithic Photodetector/CMOS preamplifier for 531 Mb/s Optical Data Link Applications," *ISSCC Digest of Technical Paper*, pp. 96-97, 1993.
- [31] A. A. Abidi, B. L. Kasper, R. A. Kushner, "Fine Line NMOS Transresistance Amplifier," *ISSCC Digest of Technical Paper*, pp. 76-77, Feb. 1984.
- [32] Y. Suzuki, et al., "Pseudomorphic 2DEG FET IC's for 10-Gb/s Optical Communication Systems with External Optical Modulation," *IEEE Journal of Solid-State Circuits*, Vol. 27, No. 10, pp. 1342-1346, Oct. 1992.
- [33] R. G. Meyer, R. A. Blauschild, "A 4-Terminal Wide-Band Monolithic Amplifier," *IEEE Journal of Solid-State Circuits*, Vol. 16, No. 6, pp. 634-639, Dec. 1981.
- [34] T. Suzuki, et al., "Si Bipolar Chip Set for 10-Gb/s Optical Receiver," *IEEE Journal of Solid-State Circuits*, Vol. 27, No. 12, pp. 1781-1786, Dec. 1992.
- [35] N. Scheinberg, R. J. Bayruns, T. M. Laverick, "Monolithic GaAs

- Transimpedance Amplifier for Fiber-Optic Receivers," *IEEE Journal of Solid-State Circuits*, Vol. 26, No. 12, pp. 1834-1839, Dec. 1991.
- [36] D. L. Rogers, "Integrated Optical Receivers using MSM Detectors," *IEEE Journal of Lightwave Technology*, Vol. 9 No. 12, pp. 1635-1638, Dec. 1991.
 - [37] V. Hurm, et al., "8.2GHz Bandwidth Monolithic Integrated Optoelectronic Receiver Using MSM Photodiode and 0.5 μ m Recessed-Gate AlGaAs/GaAs HEMTs," *Electronic Letters*, Vol. 27, No. 9 pp.734-735, Apr. 1991.
 - [38] H. Yano, et al., "An Ultra-High-Speed Optoelectronic Integrated Receiver for Fiber-Optic Communications," *IEEE Transaction on Electron Devices*, Vol. 39, No. 10, pp. 2254-2259, Oct. 1992.
 - [39] H. Yano, et al., "Low-Noise Current Optoelectronic Integrated Receiver with Internal Equalizer for Gigabit-per-Second Long-Wavelength Optical Communications," *IEEE Journal of Lightwave Technology*, Vol. 9, No. 9, pp. 1328-1333, Sep. 1990.
 - [40] S. Chandrasekhar, et al., "A Monolithic Long Wavelength Photoreceiver Using Heterojunction Bipolar Transistors," *IEEE Journal of Quantum Electronics*, Vol. 27, No. 3, pp. 773-777, Mar. 1991.
 - [41] N. Uchida, et al., "A 622 Mb/s High-Sensitivity Monolithic InGaAs-InP pin-FET Receiver OEIC Employing a Cascode Preamplifier," *IEEE Photonic Technology Letters*, Vol. 3, No. 6, pp. 540-542, June 1991.
 - [42] S. Chandrasekhar, et al., "Monolithic Balanced p-i-n/HBT Photoreceiver for Coherent Optical Heterodyne Communications," *IEEE Photonic Technology Letters*, Vol. 3, No. 6, pp. 537-539, June 1991.
 - [43] E. M. Cherry, D. E. Hooper, "The design of wide-band transistor feedback amplifiers," *Proc. Inst. Elec. Eng.*, Vol. 110, pp.375-389, Feb. 1963.
 - [44] R. Reimann, H. M. Rein, "Bipolar High-Gain Limiting Amplifier IC for Optical-Fiber Receivers Operating up to 4Gb/s", *IEEE Journal of Solid-State Circuits*, Vol. 22, No. 4, pp. 504-511, Aug. 1987.
 - [45] W. R. Davis, J. E. Solomon, "A high-performance monolithic IF amplifier incorporating electronic gain control," *IEEE Journal of Solid-State Circuits*, Vol. 3, pp. 408-416, Dec 1968.

- [46] B. Gilbert, "A precise four-quadrant multiplier with subnanosecond response," *IEEE Journal of Solid-State Circuits*, vol. SC-3, pp. 365-373, Dec. 1968.
- [47] M. Ohara, et al., "Bipolar monolithic amplifiers for a gigabit optical repeater," *IEEE Journal of Solid-State Circuits*, Vol. 19, pp. 491-497, 1984.
- [48] B. Gilbert, "A precise four-quadrant multiplier with subnano-second response," *IEEE Journal of Solid State Circuits*, pp. 353-365, Dec. 1968.
- [49] R. G. Meyer, *Class Notes for Advanced IC for Communications*, Fall 1989.
- [50] R. Gregorian, G. C. Temes, *Analog MOS Integrated Circuits for Signal Processing*, pp.391-399.
- [51] R. P. Jindal, "Gigahertz-Band High-Gain Low-Noise AGC Amplifier in Fine-Line NMOS," *IEEE Journal of Solid-State Circuit*, vol. SC-22, no. 4, pp.512-521, Aug. 1987.
- [52] R. Reimann, H. M. Rein, "A Single-Chip Bipolar AGC Amplifier with Large Dynamic Range for Optical-Fiber Receivers Operating up to 3Gb/s," *IEEE Journal of Solid-State Circuits*, Vol. 24, No. 6, pp.1744-1748, Dec. 1989.
- [53] 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, pp. 1426-1433, Dec. 1988.
- [54] Y. Akazawa, et al., "A design and packaging techniques for a high-gain gigahertz-band single-chip amplifier," *IEEE Journal of Solid-State Circuits*, vol. SC-21 no. 3, pp.417-423, June 1986.
- [55] Y. Imai, et al., "A high-gain GaAs amplifier with AGC function," *IEEE Electron Device Letters*, vol. EDL-5, no. 10, pp. 415-417, 1984.
- [56] P.R. Trischitta, E. L. Varma, *Jitter in Digital Transmission Systems*, Artech House, Inc., Norwood, MA, 1989.
- [57] E. A. Lee, D. G. Messerschmitt, *Digital Communication*, Kluwer Academic Publishers, Boston, 1988.
- [58] F. M. Gardner, *Phaselock Techniques, 2nd Edition*, John Wiley & Sons, New York, 1979.
- [59] W. R. Bennett, "Statistics of Regenerative Data Transmission," *Bell System Technical Journal* 37, pp. 1501-1542, Nov. 1958.

- [60] W. I. Way, et al., "High Speed circuit technology for multi-gigabit/sec optimal communications systems," *IEEE ICC Conf. Rec.*, pp. 313-318, June 1988.
- [61] B. Fleischmann, W. Ruile, and G. Riha, "Rayleigh-mode SAW filters on quartz for timing recovery at frequencies above 1GHz," *Ultrasonics Symp. Conf. Dig.*, pp. 163-167, July 1984.
- [62] M. Kawai, et al., "Smart Optical Receiver With Automatic Decision Threshold Setting and Retiming Phase Alignment," *IEEE Journal of Lightwave Technology*, Vol. 7, No. 11, pp. 1634-1640, Nov. 1989.
- [63] R. R. Cordell, J. B. Fomey, C. N. Dunn, and W. G. Garrett, "A 50 MHz Phase- and frequency-locked loop," *IEEE Journal of Solid-State Circuits*, vol. SC-14, pp. 1003-1009, Dec. 1979.
- [64] D. G. Messerschmitt, "Frequency Detectors for PLL Acquisition in Timing and Clock Recovery," *IEEE Trans. on Communications*, COM-27, vol. 9 pp.1288, Sep. 1979.
- [65] J. I. Brown, "A digital phase and frequency-sensitive detector," *Proc. IEEE*, Vol. 59, p. 717, Apr. 1971.
- [66] M Banu, and A. Dunlop, "A 660Mb/s CMOS Clock Recovery Circuit with Instantaneous Locking for NRZ Data and Burst-Mode Transmission," *ISSCC Digest of Technical Papers*, p.102-103; Feb. 1993.
- [67] B. Lai, and R. C. Walker, "A Monolithic Extraction Data Retiming Circuit", *ISSCC Digest of Technical Papers*, p.144-145; Feb. 1991.
- [68] L. DeVito, et al., "A 52 MHz and 155MHz Clock-Recovery PLL," *ISSCC Digest of Technical Papers*, p.142-143; Feb. 1991.
- [69] A. Pottbacker, U. Langmann, and H. U. Schreiber, "A Si Bipolar Phase and Frequency Detector IC for Clock Extraction up to 8Gb/s," *IEEE Journal of Solid-State Circuits*, Vol. 27, No. 12, pp. 1747-1751, Dec. 1992.
- [70] P. Wennekers, "10-Gb/s Bit-Synchronizer Circuit with Automatic Timing Alignment by Clock Phase Shifting Using Quantum-Well AlGaAs/GaAs/AlGaAs Technology," *IEEE Journal of Solid-State Circuits*, Vol. 27, No. 10, pp. 1347-1352, Oct. 1992.
- [71] H. Ransijn, and P. O'Connor, "A PLL-biased 2.5-Gb/s clock and data

- regenerator IC," *IEEE Journal of Solid-State Circuits*, Vol. 26, No. 10, pp. 1345-1353, Oct. 1991.
- [72] S. K. Enam, and A. A. Abidi, "NMOS IC's for Clock and Data Regeneration in Gigabit-per-Second Optical-Fiber Receivers," *IEEE Journal of Solid-State Circuits*, Vol. 27, No. 12, pp. 1763-1774, Dec. 1991.
 - [73] G. E. Andrews, D. C. Farley, S. H. Kravitz, A. W. Schelling, "A 300Mb/s Clock Recovery and Data Retiming System", *ISSCC Digest of Technical Papers*, p.188-189; Feb. 1987.
 - [74] J. Tani, D. Crandall, J. Corcoran, T. Hornak, "Parallel Interface ICs for 120Mb/s Fiber Optic Links", *ISSCC Digest of Technical Papers*, p.190-191; Feb. 1987.
 - [75] P. Wallace, R. Bayruns, J. Smith, T. Laverick, R. Shuster, "A GaAs 1.5Gb/s Clock Recovery and Data Retiming Circuit", *ISSCC Digest of Technical Papers*, p.192-193; Feb. 1990.
 - [76] M. Soyuer, and H. A. Ainspan, "A Monolithic 2.3Gb/s 100mW Clock and Data Recovery Circuit," *ISSCC Digest of Technical Papers*, p.158-159, Feb. 1993.
 - [77] D-L. Chen, and R. Waldron, "A Single-Chip 266Mb/s CMOS Transmitter/Receiver for Serial Data Communications," *ISSCC Digest of Technical Papers*, p.100-101, Feb. 1993.
 - [78] D. Clawin, and U. Langmann, "Multigigabit/Second Decision Circuit," *ISSCC Digest of Technical Papers*, p. 222-223, Feb. 1985.
 - [79] H-M. Rein, R. Reimann, "6 Gbit/s multiplexer and regenerating demultiplexer ICs for optical transmission systems based on a standard bipolar technology," *Electronics Letters*, Vol. 22, pp. 988-990, Sep. 1986.
 - [80] S. H. Lewis, *Video-Rate Analog-To-Digital Conversion Using Pipelined Architectures*, Ph.D. Thesis, University of California, Berkeley, Nov. 1987.
 - [81] K. C. Hsieh, *Noise Limitation in Switched-Capacitor Filters*, Ph.D. Thesis, University of California, Berkeley, May 1982.
 - [82] C. C. Shih, *Precision Analog to Digital and Digital to Analog Conversion Using Reference Recirculating Algorithm Architectures*, Ph.D. Thesis, University of California, Berkeley, Jul. 1985.

- [83] P. R. Gray, "Basic MOS Operational Amplifier Design - An Overview" in *Analog MOS integrated Circuits*, IEEE press, New York, pp.28-49, Mar. 1980.
- [84] D. Soo, "High-Frequency Voltage Amplification and Comparison in a One-Micron NMOS Technology," *PhD thesis*, University of California, Berkeley, 1985.
- [85] F. M. Gardner, "Charge-Pump Phase-Lock Loops," *IEEE Trans. on Communications*, Vol. CMO-28, No.11, p.1849-1858, Nov. 1980.
- [86] Kobayashi, H., "Simulation Adaptive Estimation and Decision Algorithm for Carrier Modulated Transmission Systems," *IEEE Trans. on Communications Technology*, COM-19, pp. 268-280, June 1971.
- [87] R. D. Gitlin, J. Salz, "Timing Recovery in PAM Systems," *Bell System Technical Journal*, 50(5), pp.1645, May and June 1971.
- [88] J.H. Chiu, L. S. Lee, "The Minimum Likelihood - A New Concept for Bit Synchronization," *IEEE Trans. on Communications*, Vol. COM-35, No.5, May 1987.
- [89] D. K. Jeong, G. Borriello, D. G. Hodges, R. katz, "Design of PLL-based clock generation circuits," *IEEE Journal of Solid-State Circuits*, vol. SC-22, no. 2, pp. 255-261, Apr. 1987.
- [90] K. M. Ware, H. S. Lee, C. G. Sodini "A 200-MHz CMOS Phase-Locked Loop with Dual Phase Detectors," *IEEE Journal of Solid-State Circuits*, vol. 24, no. 6, 1560-1568, Dec. 1989.
- [91] M. Bazes, "A novel precision MOS synchronous delay line," *IEEE Journal of Solid-State Circuits*, Vol. SC-20, no. 6, pp 1265-1271, Dec 1985.
- [92] B. Kim, D. N. Helman, P. R. Gray, "A 30-MHz Hybrid Analog/Digital Clock Recovery Circuit in 2- μ m CMOS," *IEEE Journal of Solid-State Circuits*, vol. 25, no. 6, pp. 1385-1394, Dec 1990.
- [93] R. R. Cordell, J. B. Forney, C. N. Dunn, and W. G. Garrett, "A 50 MHz Phase- and frequency-locked loop," *IEEE Journal of Solid-State Circuits*, vol. SC-14, pp. 1003-1009, Dec. 1979.
- [94] P. R. Trishitta, P. Sannuti, "The Validity of the Linear, Shift-Invariant Model of Jitter for a Fiber Optic Regenerator," *IEEE Trans. on Communication*,

- 1988.
- [95] S. Haykin, *Communication Systems*, 2nd Edition, John Wiley & Sons Inc., 1983.
 - [96] *Synchronous Optical Network (SONET) Transport Systems: Common Generic Criteria*, Bellcore Technical Reference FR-NWT-000440, Dec. 1991.
 - [97] T. H. Hu, and P. R. Gray, "A Monolithic 480Mb/s Parallel AGC/Decision/Clock-Recovery Circuit in 1.2 μ m CMOS Technology", *ISSCC Digest of Technical Papers*, pp.98-99, 1993.