2002

High performance CMOS amplifier and phase-locked loop design

Yonghui Tang
Iowa State University

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

High performance CMOS amplifier and phase-locked loop design

by

Yonghui Tang

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Computer Engineering

Program of Study Committee:
Randall L. Geiger, Major Professor
Robert J. Weber
Degang Chen
Chris Chong-Nuen Chu
Yuhong Yang

Iowa State University

Ames, Iowa

2002

UMI Number: 3073483

UMI Microform 3073483

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

Graduate College
Iowa State University

This is to certify that the doctoral dissertation of

Yonghui Tang

has met the dissertation requirements of Iowa State University

Signature was redacted for privacy.

Major Professor

Signature was redacted for privacy.

For the Major Program

To my new-born baby

To my beautiful wife

To my parents in China

TABLE OF CONTENTS

ACKNOWLEDGMENTS

The past five-year journey of my Ph.D. study has been the most impressive part of my life. I believe I would not be able to accomplish it without the help from a lot of people. First and foremost, I would like to express my greatest gratitude to my advisor, Professor Randall Geiger. I still remember vividly when I first read his email asking if I would like to study in analog and mixed signal VLSI area. I didn't even know what "VLSI" means at that time. During these five years, his broad knowledge and deep understanding in circuits and systems provided me invaluable guidance and insight. I would like to thank him for many inspiring conversations and advices on my research and teaching.

I enjoy the time I spend with the people in my group because of their invaluable assistance on all aspects. They always offer me help whenever I need it. My many thanks go to Mezyad Amourah, Huiting Chen, Saqib Malik, Kumar Parthasarathy, Mark Schlarmann, Kee-Chee Tiew, Yonghua Cong, Jie Yan, Mao-Feng Lan and Jing Ye. I especially would like to thank Jie, Huiting, Mark and Saqib for a lot of discussions and their generous help on my research projects.

I will never forget the time my colleagues and I travel together all over the US to attend various conferences. It was fun!

I would like to thank my committee members for their time. I would like to thank Professor Weber for his suggestion on designing the printed-circuit board.

I could not have come this far without the supports from my family. The love from them has always been the power to drive me forward. Especially I would like to express my appreciation to my wife, Mengting, for all the time we have been together, for her endless advices and helps, for sharing my frustration and success during these years.

ABSTRACT

Low voltage, high speed and high linearity are three different aspects of the analog circuit performance that designers are trying to achieve. In this dissertation, three design projects targeting these different performance optimizations are introduced.

The first work is a design of a low voltage operational amplifier. In this work, a threshold voltage tuning technique for low voltage CMOS analog circuit design is presented. A 750mV two-stage operational amplifier using this technique was designed in a standard 0.5μm 5V CMOS process with $Vtp \approx -0.9V$ and $Vtn \approx 0.8V$. The active area is 560μm ×760μm. It exhibits a 62dB DC gain and consumes 38μW of power. It works with supply voltages that range from 0.75V to 1V. Compared to its 5V counterpart consuming the same amount of current, it maintains nearly the same gain bandwidth product of 3.7MHz when driving 15pF load. This op amp is the FIRST strong inversion op amp that works at a supply voltage below the threshold voltage.

The second is a design of a high speed phase-locked loop for data recovery. A new non-sequential linear phase detector is introduced in this work. Most of the existing phase detectors for data recovery are based on state-machines. The performance of these structures deteriorates rapidly at higher frequencies because of the inadequate settling performance of the flip-flop used to form the state machine. The new phase detector has a speed advantage over the state-machine based designs because it is simple and easy to implement in CMOS technology. Using this phase detector, a PLL was designed in a 0.25μm CMOS process with an active area of 400μm ×290μm. Experimental results show it successfully locks to a 2.1Gbit/s pseudo-random data sequence at 2.3V. It is believed that the architecture is the fastest that has been introduced for data recovery applications.

The third work introduces the design of a highly-linear variable gain amplifier. It achieves high linearity with third harmonic distortion better than -60dB@Vopp=1V at 160MHz in a 0.25μm CMOS process. It has a precise gain step of 6.02dB that is controlled

digitally. The linearity performance is achieved with a linearized open loop amplifier configuration. Similar performance can only be achieved using feedback configuration before.

CHAPTER 1

INTRODUCTION

It was in the early 1980's that many experts predicted the demise of analog circuits because the emerging digital signal processing (DSP) algorithms were becoming more and more powerful. It was conjectured that all processing of the signal could be performed eventually more efficiently in digital domain. Yet the reality is, while much signal processing has indeed shifted to digital, our world is still an "analog" world and the demand for analog and mixed-signal circuits continues to grow. Most DSP relies heavily on interfaces to the analog world. The need for analog circuits in modern mixed-signal VLSI chips for multimedia, perception, control, instrumentation, medical electronics and telecommunications is very high. Analog and mixed-signal circuits are fundamentally necessary in many modern electronic systems.

For almost two decades, the dominant semiconductor technology has been shifted from bipolar to CMOS. This replacement happened first in digital market. Compared to bipolar or GaAs technology, static MOSFET logic dissipates power only when devices are in transition. It requires fewer devices to build comparable logic gates. Furthermore, the gate length can be shrunk much faster which results in higher speed, smaller die size and reduced power dissipation. The low fabrication cost and the possibility of integrating both analog and digital circuits on the same die make CMOS technology the technology of both choice and necessity for many applications. Nevertheless, bipolar and GaAs technology still find niche applications in high performance analog design because they have higher speed and lower noise than what can usually be achieved in CMOS technology.

All the technical contents in this dissertation is based on CMOS technology.

In modern CMOS analog design, engineers are facing all kinds of problems when they are designing high performance analog circuits. One of the most important problems is the power dissipation. With the shrinking channel length, more and more transistors are squeezed into a small die while the operating frequency is getting higher and higher. This trend results in a much higher power density on the chip which requires a way to cool it down. Low voltage design is a promising solution. Furthermore, the popularity of hand-held devices and mobile applications makes it even more attractive to develop low voltage circuits because they can work at a single battery cell and allow the hand-held device to operate much longer. Presented in Chapter 2 of this dissertation is a design of a 750mV operational amplifier (op amp). This op amp is the FIRST implementation of strong inversion op amp that works at a supply voltage below the threshold voltage in a standard CMOS process. It was implemented using a threshold voltage tuning scheme. This low voltage design technique can also be easily accommodated into the design of other low voltage analog circuits.

In the internet era, people always want to be connected in a higher bandwidth so they can communicate at higher speeds. The speed of the Ethernet rapidly evolved from 10Mb/s, 100Mb/s, 1Gb/s and now even to 10Gb/s. It is a never-ending challenge in integrated circuit design to continue pushing the speed/performance envelope. In any modern communication system, no matter whether it be wireline or wireless, the phase-locked loop (PLL) plays a vital role in determining the speed of the communications. In Chapter 3, a design of a high speed PLL for data recovery will be discussed. It employs a new non-sequential linear phase detector to achieve high speed operation. Compared to most of the existing full-rate phase detector structures, the new phase detector has a speed advantage because it is simple and easy to implement in CMOS technology.

Related to the design of the PLL, a short discussion on transient bit error rate (BER) analysis of data recovery systems using jitter models is given in Chapter 4. It co-relates the

acquisition behavior of the PLL to the BER of the recovered data which will be greatly helpful in system level design of the data recovery system.

Another problem most analog designers need to deal with frequently is the linearity performance of the circuits. Transistors are not perfect. Their input/output relationships are not linear. Short-channel effects in deep sub-micron process make the linearity performance even worse. There are certain applications in which the linearity of the amplifier is the key performance characteristic that determines the performance of the whole system. In Chapter 5, a high precision, highly-linear high speed variable gain amplifier (VGA) will be introduced. It has a precise gain step of 2 (6.02dB) that is controlled digitally. It has a third harmonic distortion better than -60dB@Vopp=1V for 160MHz inputs. The linearity performance was achieved using an open loop amplifier structure. Similar linearity performance has only been achieved previously by using feedback structures.

To have a better understanding of the linearity in open loop amplifiers and feedback amplifiers, an analysis of the effects of open loop nonlinearity on linearity of feedback amplifiers will be discussed in Chapter 6. The nonlinearity in feedback amplifiers is investigated quantitatively from several different aspects.

CHAPTER 2

A FULLY-INTEGRATED 750mV CMOS OPERATIONAL AMPLIFIER

2.1 Motivation

The design of low voltage low power CMOS analog circuits has become a subject of considerable interest in recent years. There are several driving forces to attract analog design engineers to persistently investigate the design of lower supply circuits.

The first reason is the constantly shrinking feature size of the modern CMOS processes. As the minimum channel length is approaching the 0.09μm level in the year of 2002, the thickness of the device gate oxide is becoming thinner. Since the gate oxide thickness is so small, the gate of the transistors can't withstand high voltages because of the high electric field strength in the gate oxide that is created by such voltage levels. In order to avoid gate breakdown and ensure device reliability, the power supply of the circuits has to be scaled down. With the shrinking channel length, the device threshold voltages are also decreasing. This threshold voltage change makes it possible to have lower supply voltages. Figure 2.1 shows an illustration of approximate relationships among minimum channel length, power supply voltages and the threshold voltages.

The need for low power supply voltage happened first in the digital design area. This is because digital circuits are much more compact and dense than most analog circuits. With more and more devices being integrated into a small die, power density has become a big problem and excessive power density will cause a part of the die to overheat.

Because there is almost no power dissipation through the digital circuits if there is no switching, most of the power consumed by digital circuits is dynamic power which is given by

$$P_d = C_{load} \cdot V_{DD}^{\ 2} \cdot f \qquad (2.1)$$

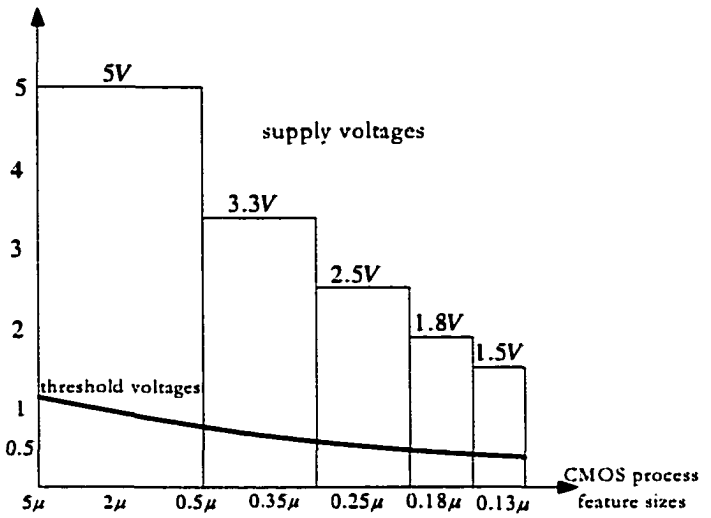


Figure 2.1 Migration of the CMOS process feature sizes, power supply voltages and threshold voltages

where C_{load} is loading capacitance, f is the operating frequency and V_{DD} is the supply voltage. We can see the one was to alleviate the power dissipation problem is to lower the supply voltage. If the load capacitance and speed remain constant, the total power consumption will be a quarter of what it was before if the supply voltage can be halved.

More recently, with the popularity of battery-powered devices for portable applications, low voltage design has become an even more attractive topic. A lot of people own cell phones, laptops and PDAs. One of the key performances to evaluate them is by battery life. Battery life strongly depends on the power dissipation of the chips. In a word, the longing for lower power consumption has always been the reason for low voltage design and this pursuit is going to continue for the foreseeable future.

Low voltage operation is always being paralleled with the scaling of threshold voltages. This trend makes it possible for digital circuits power supply to be decreased from 5V to about 1.5V nowadays. While it is relatively easy to accommodate the low supply voltage for digital circuits, these decreasing supply voltages often have a detrimental effect

on analog components in these systems. Moreover, the threshold voltage will not decrease significantly below what we already have now. Since many portable products operate from alkaline or rechargeable batteries, the operating supply voltage for these systems is migrating down to 0.9 V for a single battery cell. Therefore, circuits design techniques need to be improved in order to allow existing CMOS analog circuits to operate at a lower supply comparable to the threshold voltage while still maintaining key performance parameters at the levels achievable at higher supply voltages.

2.2 Low Voltage Design Techniques

The major issues in the design of low voltage analog circuits are:

1. The threshold voltage and saturation voltage (V_{dsat}) do not scale down linearly with power supply nor with smaller size technologies.
2. The designers can not use conventional cascode structures and other conventional design methodologies to maintain the performance for low voltage circuits.

As a fundamental building block in analog processing, the operational amplifier is a good test bed for developing low voltage design techniques. Quite some work has been done on CMOS low voltage analog design techniques. They can be categorized into three design strategies as discussed in the following three subsections.

2.2.1 Low Voltage Circuit Structures with Conventional Transistor Operation

The first strategy is to employ new circuit structures that use standard transistors to achieve low voltage operation without sacrificing much performance [2.2] [2.3] [2.4] [2.5] [2.6] [2.7] [2.8] [2.9]. Analog designers have invented a lot of methods to boost the performance of the circuits. Although these structures helped with supply migrates from 15V or higher down to the 1.5V range, most of them are proving not suitable for very low voltage design.

Examples of design approaches in this category include the use of rail-to-rail constant g_m complementary differential pair input stages [2.2] [2.3] [2.6] [2.9], dynamic biasing circuits [2.4], regulated-cascode transistors [2.2] [2.7] and low voltage transconductance stages [2.5] [2.8].

For op amps that will be used in the non-inverting configuration, a large input common mode voltage swing is required. Especially for a voltage follower which usually works as an output buffer, we need a rail-to-rail input common mode voltage range. For this reason, the rail-to-rail complementary differential pair input stage is quite popular in realizing low voltage op amps. Either P-input or N-input differential pairs are generally used as the input stage for op amps. Shown in Figure 2.2 is the typical input common mode voltage ranges for both the NMOS pair and the PMOS pair. For the NMOS input pair, the common mode input range is up to V_{dd}, but its lower end is limited by the V_{GS} of the input pair and the V_{dsat} of the current source. For the PMOS input pair, the common mode input range is down to $-V_{ss}$, but its higher end is limited by the V_{GS} of the input pair and the V_{dsat} of the current source. Neither of them has a rail-to-rail common mode input range. The standard approach for achieving rail-to-rail inputs is to connect the NMOS pair and PMOS pair in parallel so that it has rail-to-rail input common mode range shown in Figure 2.3. We can see the minimum supply voltage for this structure is

$$V_{sup} \geq 2V_{dsat} + V_{GS,M1/2} + V_{GS,M3/4} \tag{2.2}$$

i.e. $$V_{sup} \geq 4V_{dsat} + V_{tn} + V_{tp} \tag{2.3}$$

The required supply voltage for this structure is quite low. Almost all rail-to-rail input stages used in [2.2] [2.3] [2.6] [2.9] are similar to that of figure 2.3 except for some variations in performance enhancement methods to alleviate some limitations in this structure such as transconductance variations because of the overlapping of the common mode range for the NMOS pair and the PMOS pair. Similar methodologies can also be used to develop low voltage rail-to-rail output stages [2.3]. These will not be discussed here.

Another low voltage design technique is to use so called "self-cascode" or "regulated-cascode" structures. We all know cascoding in "normal" voltage analog circuits will usually enhance the performance by increasing the output impedance. But it is not useful in low voltage design because of its requirements for higher voltage headroom. To achieve a similar performance as the cascoding, the "self-cascode" scheme has been proposed for low voltage operation as shown in Figure 2.4.

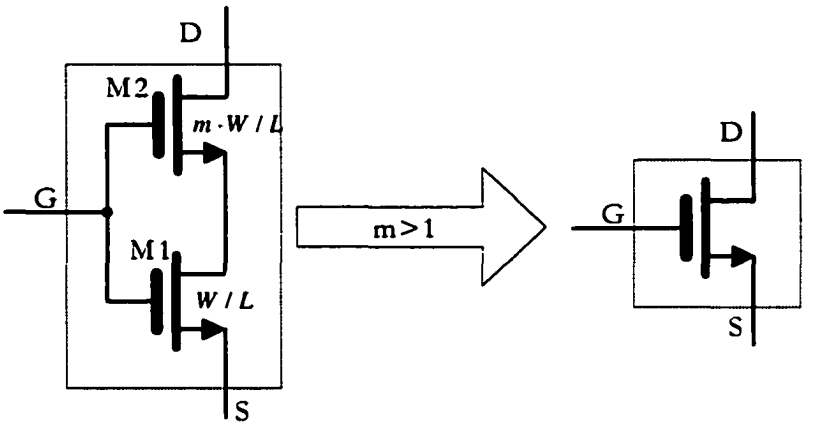


Figure 2.4 Self-cascode structure

This self-cascode structure consists of two NMOS transistors. It performs equivalently to a simple NMOS transistor with a much larger effective channel length [2.15] (thus higher output impedance). In practice, the optimal W/L ratio of M2 should be larger than that of M1, i.e. m>1. The lower transistor M1 is equivalent to a resistor, but this resistor is input dependent. The effective transconductance of the self-cascode transistor is approximately equal to the transconctance of M1 [2.15].

In the self-cascode structure, transistor M1 always operates in linear region while the top transistor operates in either saturation or the linear region. The voltage between the source and drain terminal of M1 is so small that there is no discernable V_{dsat} difference in both the self-cascode and simple transistors. Thus, the self-cascode structure can be used in low

voltage applications. Some other similar structures were also proposed to have an enhanced gain but not requiring additional voltage overhead.

While all the examples [2.2] [2.3] [2.4] [2.5] [2.6] [2.7] [2.8] [2.9] in this class were able to achieve comparable performance to traditional high voltage designs, they all failed to operate at a very low voltages. Usually, the minimum power supply required for this class of circuits is higher than $1.5V_t$. Table 2.1 summarizes the techniques and the supply voltages of the low voltage op amp designs. The lowest supply voltage they achieved was 1V supply in a CMOS process with threshold voltages of $V_{tn} = 0.6V$ $V_{tp} = -0.8V$.

2.2.2 Bulk-driven Transistors

The second strategy of low voltage design is to use bulk-driven transistors. This technique is suited for standard CMOS processes; nevertheless, only one kind of transistor can be used for bulk-driving in single-well processes, i.e. only P-channel devices can be bulk-driven in an N-well process.

The reason that bulk-driven transistor can be used for low voltage design is because the transistor exhibits some depletion mode characteristics when it is bulk-source driven, i.e. it conducts current at negative, zero or small forward bulk-source voltages.

Table 2.1 Low-voltage op amp designs using new circuits architectures

	Supply voltage	Process	Threshold voltage	Techniques
Coban [2.2]	2	2μm CMOS	0.9/-0.7	R-t-R input stage
Ferri [2.3]	1.3	0.7μm CMOS	0.7	R-t-R input/output stage
Giustolisi [2.4]	1.2	1.2μm CMOS	0.75	Dynamic biasing
Lee [2.5]	1	1.2μm CMOS	0.6/-0.8	Low voltage g_m stage
Lu [2.7]	1.3	0.8μm CMOS	0.72/-0.77	Regulated cascode
Palmisano [2.8]	1.5	1.2μm CMOS	0.8	Low voltage g_m stage
Hogervorst [2.9]	3	Custom CMOS	0.63/-0.77	R-t-R input stage

One major disadvantage of a bulk-driven MOSFET is that it has a substantially smaller g_m compared to a conventional gate-driven MOSFET [2.15]. For a conventional gate-driven MOSFET, the frequency response potential is described by its transitional frequency, f_T,

$$f_{T,gate-driven} = \frac{g_m}{2\pi C_{gs}} \tag{2.4}$$

For the bulk-driven MOSFET, f_T is given by

$$f_{T,bulk-driven} = \frac{g_{mb}}{2\pi(C_{bs}+C_{bsub})} = \frac{\eta g_m}{2\pi(C_{bs}+C_{bsub})} \tag{2.5}$$

where η is the ratio of g_{mb} to g_m and typically has a value in the range of 0.2 to 0.4.

For typical strong inversion MOSFET operation, the following approximation stands,

$$f_{T,bulk-driven} \approx \frac{\eta}{3.8} f_{T,gate-driven} \tag{2.6}$$

This will result in a lower gain bandwidth (GBW) and thus a more limited frequency response [2.15].

Examples of circuits included in this category include those of [2.10] [2.11] and [2.13]. The performance of these circuits is summarized in Table 2.2. Although the supply voltages are nominally lower than those of the circuits of Table 2.1, they can't work very close to a supply that is comparable to the threshold voltage. The best of them is that of [2.13]. It was able to work with a 0.9V supply with the help of both bulk-driven transistors and depletion-mode transistors which is not available in standard CMOS processes.

Table 2.2 Low-voltage op amp designs using Bulk-driven devices

	Supply voltage	Process	Threshold voltage	Note
Allen [2.10]	1	2μm CMOS	0.7~0.8	
Lasanen [2.11]	1	0.35μm CMOS	0.5/-0.65	
Stockstad [2.13]	0.9	Custom CMOS	N/A	Used depletion transistors

2.2.3 Process-dependent Transistors

A third strategy is to use special devices such as depletion-mode [2.13] and floating-gate transistors [2.12] [2.14]. Compared to normal single-gate transistor, the floating-gate transistor can be programmed to have a smaller effective threshold voltage which makes it suitable for low voltage operations. Depletion-mode transistor conducts current even at negative gate-source voltages.

The floating-gate device requires a critical very thin oxide to support the floating gate and this option is both costly and unavailable in most basic digital processes. The depletion-mode transistor is rarely available in standard CMOS processes. In [2.12], the authors presented a 1.2V op amp with a 0.85V threshold voltage process and floating-gate devices. With the help of both bulk-driven and depletion-mode transistors, the op amp given in [2.12] was able to work at 0.9V (the authors didn't mention the V_t value) with inferior performance compared to strong inversion op amps.

2.3 Threshold voltage tuning scheme

High threshold voltages are fundamentally limiting our ability to realize low voltage high performance analog and mixed-signal circuits. If the effective threshold voltage could be reduced using circuit design techniques, very low voltage analog circuits could be implemented. A threshold voltage tuning technique [2.16] [2.17] was introduced that allows strong inversion operation at supply voltages below the threshold voltage in any standard CMOS process. This technique is illustrated in Figure 2.5. Virtual transistors with lower effective threshold voltages are created by adding voltage sources in series with their gates. The effective threshold voltages for the virtual transistors are $V'_{tn} = V_{tn} - V_{dcn}$ for the NMOS devices and $V'_{tp} = V_{tp} + V_{dcp}$ for the PMOS devices. Both of them can be controlled by the bias voltage V_{dcn} and V_{dcp}. Assuming an ideal voltage source, the performance of the virtual transistor will be exactly the same as a normal transistor except it will have a lower effective

threshold voltage. Circuits built with virtual transistors can be designed to consume less power than those built with standard transistors because the supply voltage can be significantly reduced.

Figure 2.6 shows one method of implementation of the voltage sources V_{dcn} and V_{dcp} for both NMOS and PMOS. It employs a switch capacitor scheme. The idea is to keep a constant voltage across the capacitors. Due to leakage current, the capacitors need to be recharged periodically. In order to accomplish this, a bias voltage V_{dc} charges the capacitor C_1 when ϕ_2 is asserted. When ϕ_1 is asserted, C_1 is connected to the signal path and shares its charge with C_2. Because the current leakage is very small, the frequency of ϕ_1 and ϕ_2 can and should be very low in order to reduce the noise injected into the signal path during switching. To ensure correct operation, ϕ_1 and ϕ_2 must be non-overlapping. The use of C_2 is to provide a constantly-connected signal path.

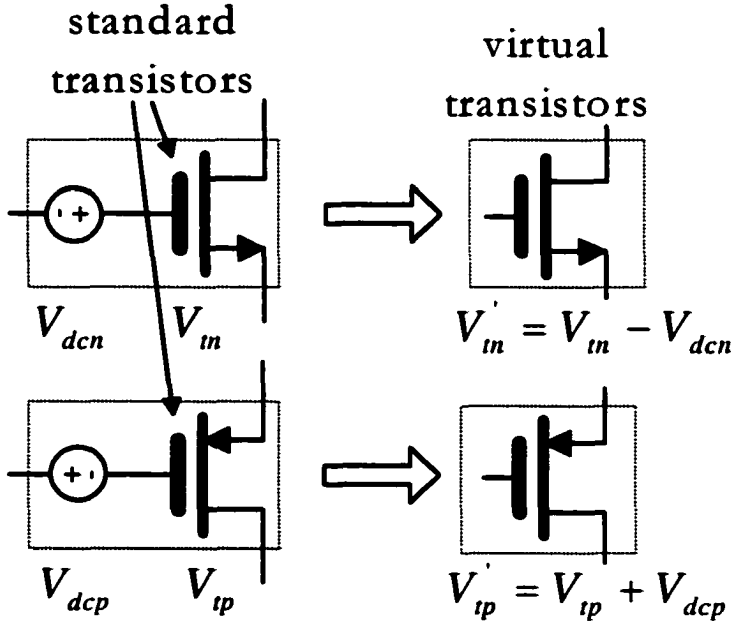


Figure 2.5 Threshold voltage tuning scheme

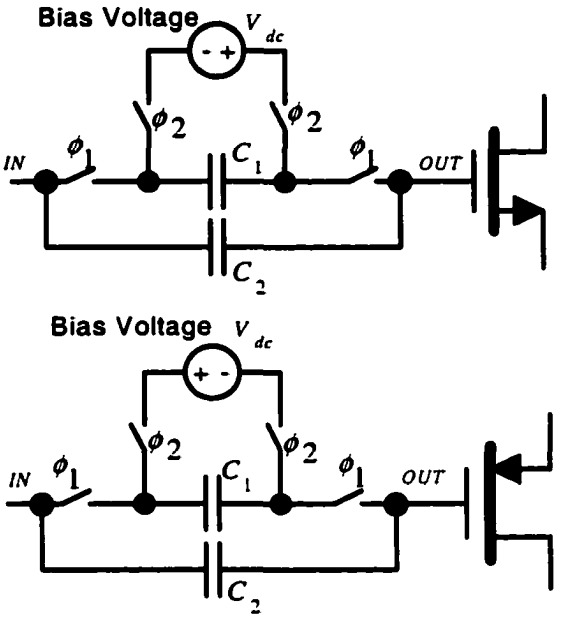


Figure 2.6 Threshold voltage tuning implementation

The required charging rate for the switched capacitors is determined by the leakage current. Leakage current is strongly process-dependent. For deep submicron process, the inherently larger leakage currents require a reduction in charging period.

2.4 Two-stage Op Amp

To demonstrate this technique, we designed a low voltage two-stage op amp. Its basic structure is shown in Figure 2.7. The first stage is a NMOS input differential pair with current mirror load. The second stage is a common-source amplifier. Miller capacitor C_C is used to compensate the op amp to ensure an acceptable phase margin. Resistor R_C is used to cancel the right-half-plane zero that is introduced by C_C.

Some well-known key performances of this op amp are given by

Slew Rate: $$SR = \frac{I_2}{C_C} \tag{2.7}$$

DC gain: $$Av = \frac{g_{m4,5}}{g_{ds4,5} + g_{ds6,7}} \cdot \frac{g_{m8}}{g_{ds8} + g_{ds3}} \quad (2.8)$$

Gain bandwidth product: $$GBW = \frac{g_{m4,5}}{C_C} \quad (2.9)$$

The low voltage op amp structure is based on this two stage op amp using the threshold voltage tuning scheme as shown in Figure 2.8. The transistors that have the same gate connection can share a single voltage source.

In order to compare the performance of the normal and the low voltage op amps, we designed two op amps with exactly the same structure, the same transistor sizing and the same quiescent current levels with the only distinction being a large supply voltage of 5V for one amplifier and a low supply voltage of 750mW for the other structure. They were fabricated on the same die as well. The transistor sizes for both op amps are given in Table 2.3. Our goal was to maintain the same GBW for the low voltage op amp as for the 5V op amp with the same structure and same current levels.

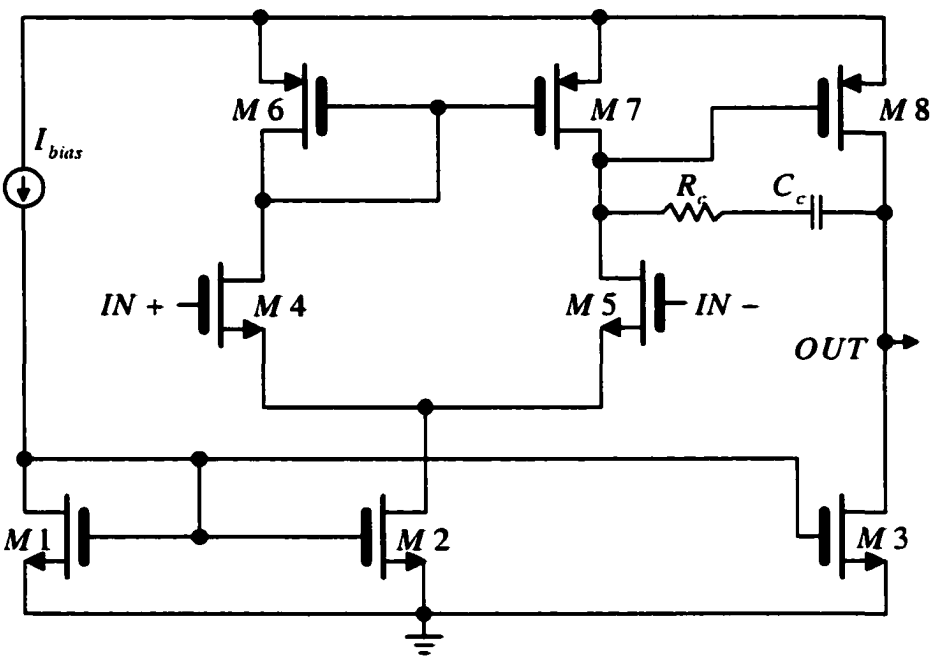


Figure 2.7 Two-stage op amp

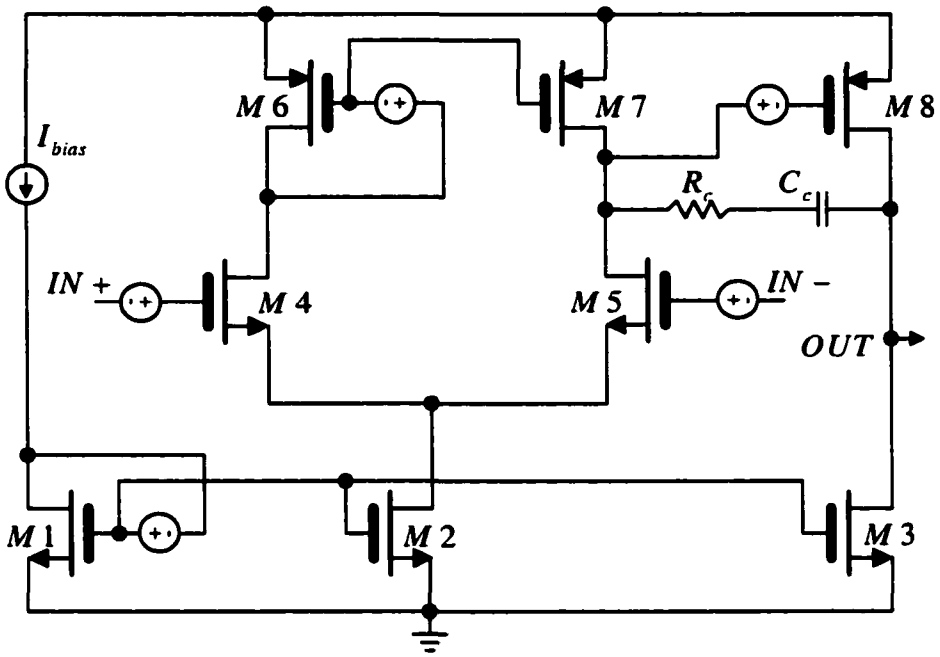


Figure 2.8 Low voltage op amp core

Table 2.3 Transistor sizes of the op amp core

	W/L ratio		W/L ratio		W/L ratio
M1/M2	120μ/1.8μ	M3	180μ/1.8μ	M4/M5	84μ/1.8μ
M6/M7	60μ/1.8μ	M8	180μ/1.8μ	Current	50μA (total, simulation)

2.5 Auxiliary Circuits

In order to realize the switch capacitor voltage sources, a set of supporting circuits are needed. Shown in Figure 2.9 is the block diagram of the auxiliary circuits and the op amp core. The oscillator generates a clock and the clock is then divided by a D flip flop (DFF) based frequency divider with a ratio of 16:1. To reduce the switching noise, a pulse generator is used to generate a short pulse over a long period for charging C_1. Clocks ϕ_1 and ϕ_2 are generated by a non-overlapping clock generator. The switches used in our design are all NMOS transistors. In order to fully turn on the switches, a clock booster circuit is designed to

boost the high voltage level of the clocks to about 2VDD to 3VDD. Finally, a very simple bias voltage generator is designed to provide the bias voltage V_{dc} to charge the capacitors. Nominally, this voltage is about 300mV below the threshold voltage, which gives the virtual transistor a 300mV effective threshold voltage. Each part of the auxiliary circuits will now be discussed.

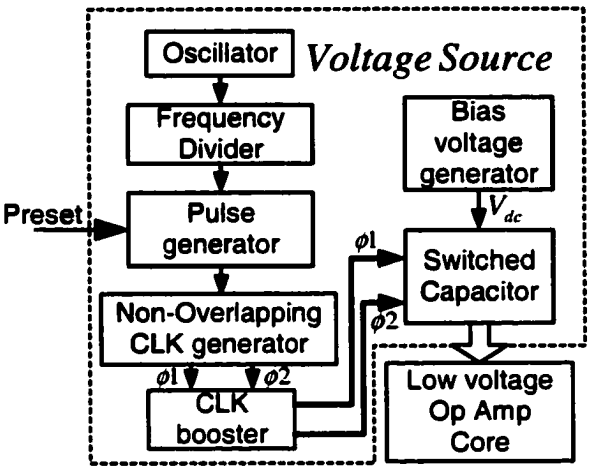


Figure 2.9 Block diagram of the auxiliary circuits with the op amp core

2.5.1 Oscillator

The oscillator of Figure 2.10 was used in our design. It is a 7-stage ring oscillator biased with a power supply of VDD. Each stage in the ring oscillator is just an inverter. The reason for using this simple structure is that the jitter and timing of the switching clocks are not a concern at all. This structure was designed to have all stages operating in the sub-threshold or weak inversion region.

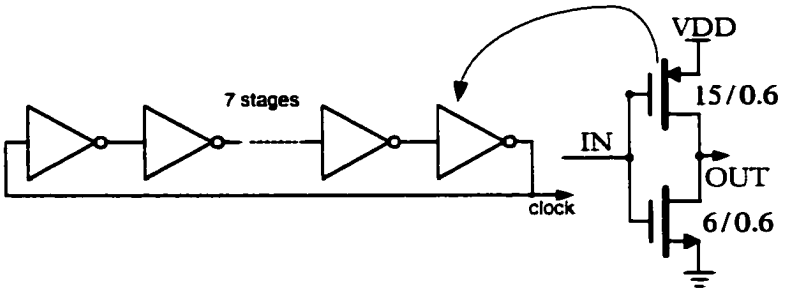


Figure 2.10 Schematic of the oscillator

2.5.2 Frequency Divider

The frequency divider is shown in Figure 2.11. It is implemented by 4 stages of DFF-type, divide-by-two circuit. The dividing ratio is 16:1. The use of frequency divider is to lower the clock frequency more effectively than adding more delay stages in the oscillator.

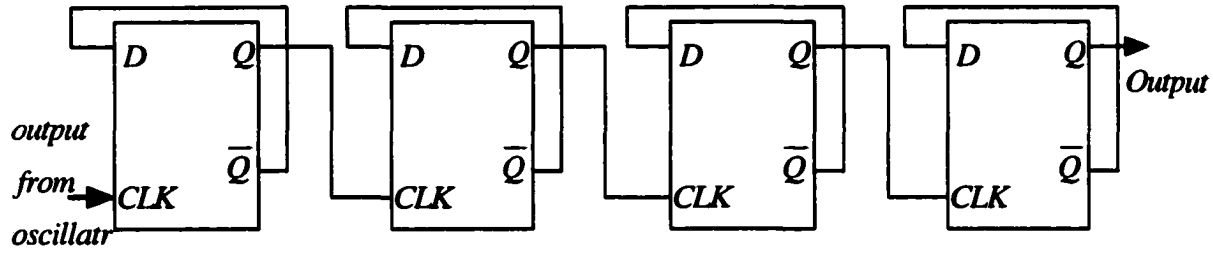


Figure 2.11 Schematic of the frequency divider

The schematic of the DFF is shown in Figure 2.12. It is a dynamic DFF circuits without feedback. Whenever there is rising edge on "CLK", the outputs will follow the input level at "D" and be able to hold this value even if "CLK" goes low or the level of "D" changes. It is a simple and highly efficient circuit and sufficient to implement the frequency divider.

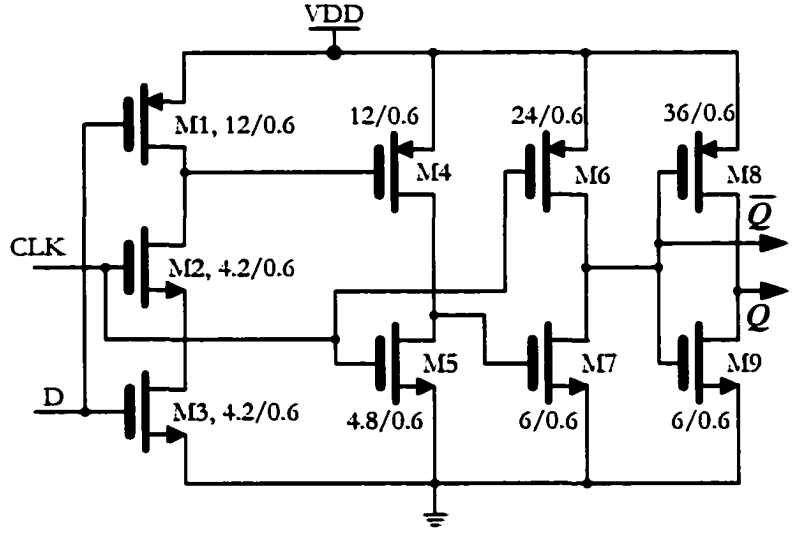


Figure 2.12 Schematic of the D flip flop used in frequency divider

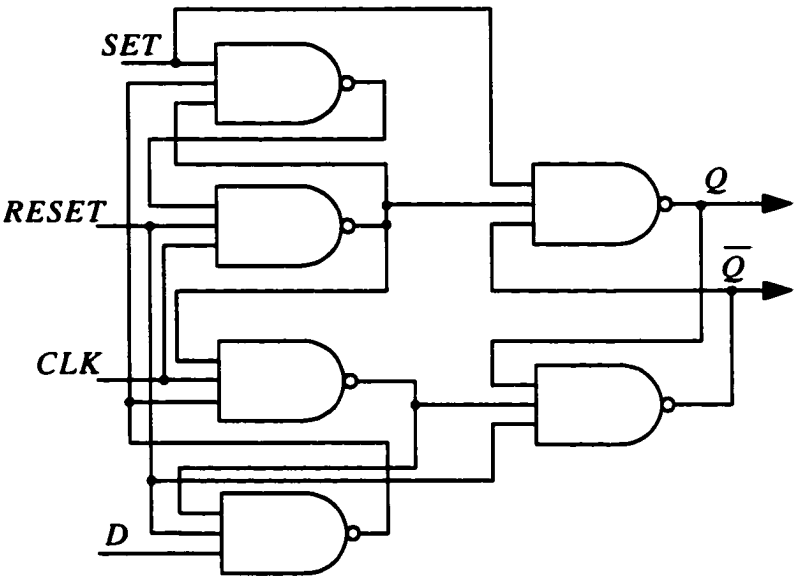


Figure 2.14 Schematic of the D flip flop used in pulse generator

The schematic of the DFF used in the pulse generator is shown in Figure 2.14. It is a master-slave type flip flop with SET and RESET. This structure is the same as that used in realizing DFFs in 74 series standard logic circuit families.

2.5.4 Non-overlapping Clock Generator

The clocks ϕ_1 and ϕ_2 of figure 2.6 must be non-overlapping to ensure the correct charging operation, that is, ϕ_1 has to be off before ϕ_2 can be on at the beginning of charging and ϕ_2 has to be off before ϕ_1 can be on at the end of charging.

A non-overlapping clock generator introduced in [2.18] was used in our design. This is shown in Figure 2.15. The delay blocks are used to ensure that the clocks remain non-overlapping. They are implemented by series of inverters (2 inverters in delay1, 4 inverters in delay2). This non-overlapping clock generator is widely used in the design of switched capacitor circuits.

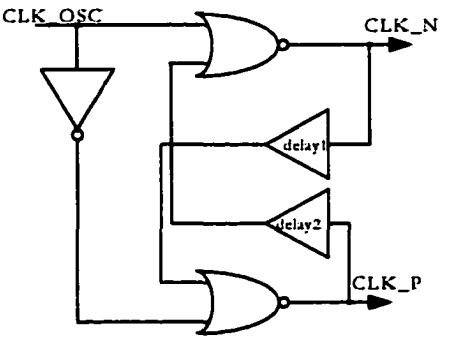
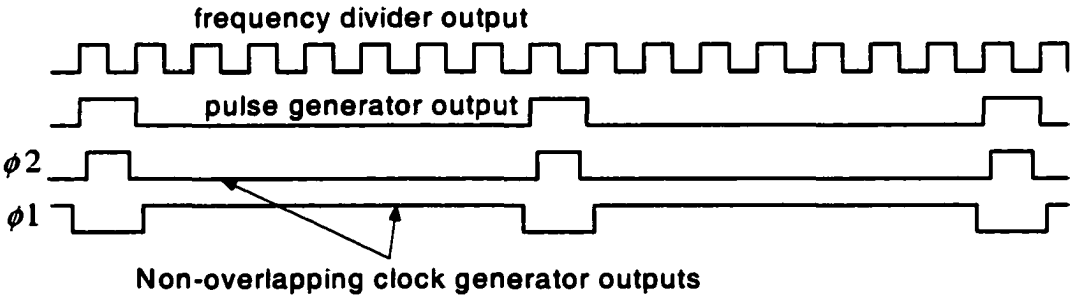
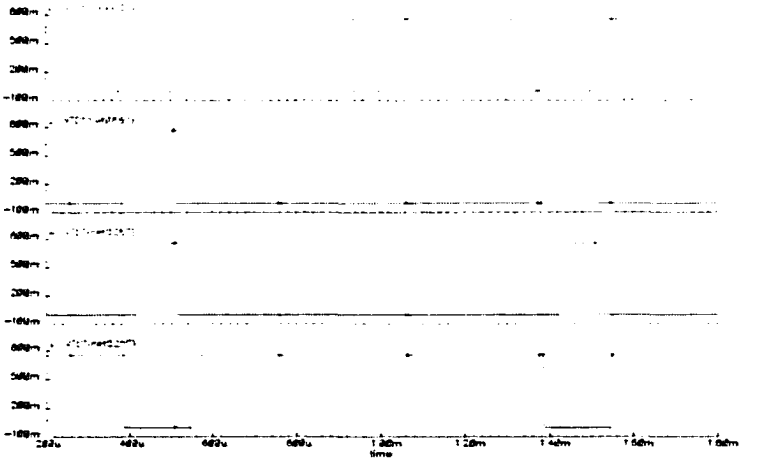


Figure 2.15 Schematic of the non-overlapping clock generator

Shown in Figure 2.16(a) is a timing diagram of the outputs of frequency divider, pulse generator and the non-overlapping clock generator. It gives a better understanding how those signals are related. Shown in Figure 2.16(b) are the simulation waveforms corresponding to Figure 2.16(a).



(a)



(b)

Figure 2.16 (a) Timing relationships of the clocks; (b) corresponding simulation results

2.5.5 Clock Booster Circuits

A stage of clock booster circuit is shown in Figure 2.17(a). When the input is low, M3 and M5 are on (sub-threshold or weak inversion) and M4 is off. Capacitor C is charged to VDD and the output is zero. When the input becomes high, M3, M5 are off, M4 is on. The voltage level at the gate of M3 becomes VDD and the voltage level at the drain of M3 becomes 2VDD. The voltage at the drain of M3 is transferred through M4 to the output. When the voltage at the drain of M3 becomes higher than VDD, however, the charge will start to leak through M3 because the drain voltage of M3 becomes higher than its source voltage. For this reason, the size of M3 is very small in our design. Although the leakage is small, in reality, the swing of the clock can only be boosted to about 1.5VDD–1.8VDD in one stage. In our design as shown in figure 2.17(b), two stages are used to boost the high clock level from VDD to about 2VDD to 3VDD. Shown in Figure 2.17(c) is the simulation results of the clock booster. Upper two waveforms are non-overlapping clock inputs. The bottom two waveforms are boosted non-overlapping outputs.

2.5.6 Bias Voltage Generator

The bias voltage generator we used is shown in Figure 2.18. It is a very simple structure. The output voltage is well defined, and the bias voltage generator can provide sufficient current under very low supply voltages. A relatively high-value resistor is used to alleviate the dependency of the output to the changes of the power supply, that is, the output should keep relatively constant without tracking the change of the VDD, which gives a relatively constant effective threshold voltage for the virtual transistors. Only two of bias voltage generators are required to supply the current for charging voltage sources for NMOS and PMOS. The sizing of the transistors is determined by both the output voltage and the charging current requirements.

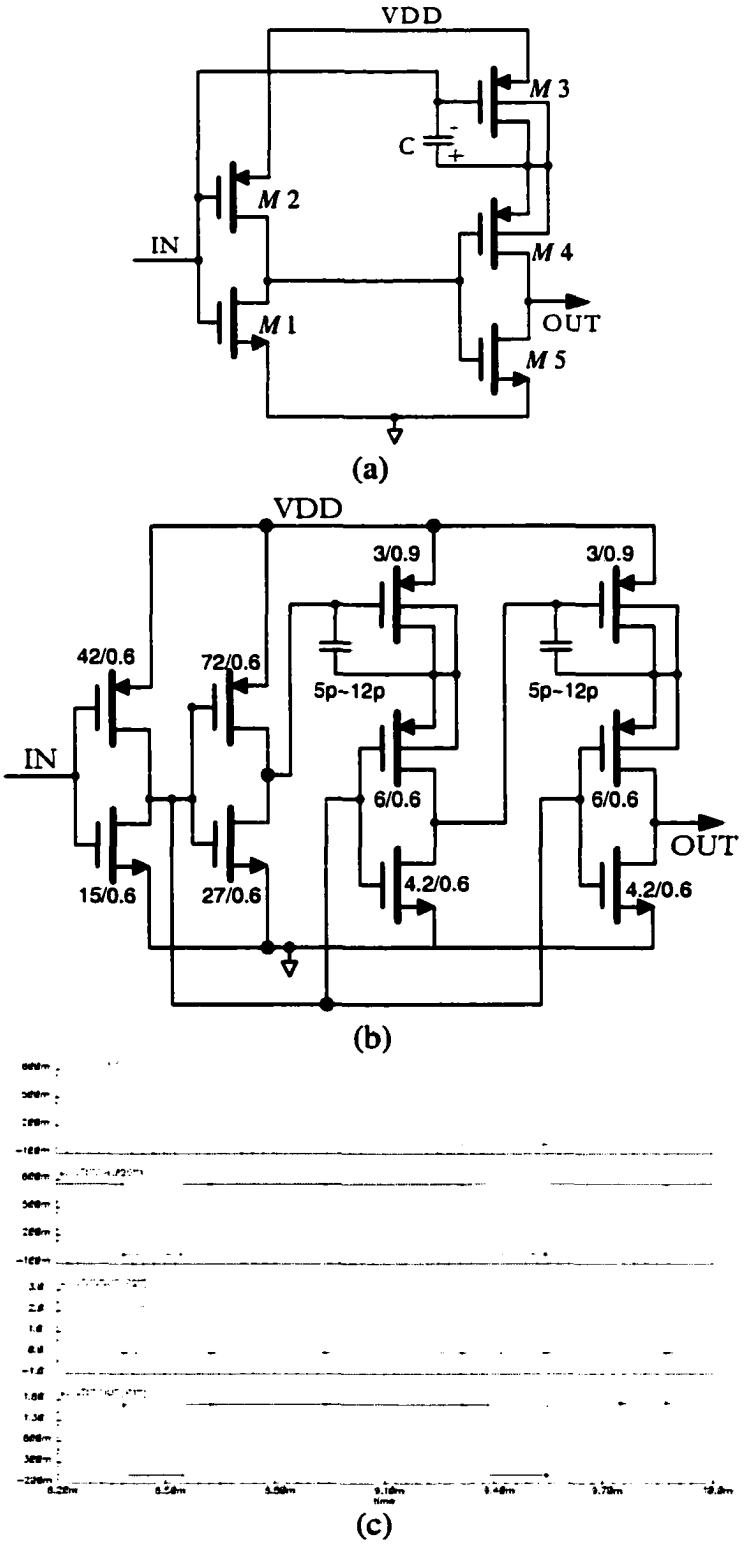


Figure 2.17 (a) one stage of clock booster; (b) actual implementation of the clock booster; (c) simulation results of the clock booster

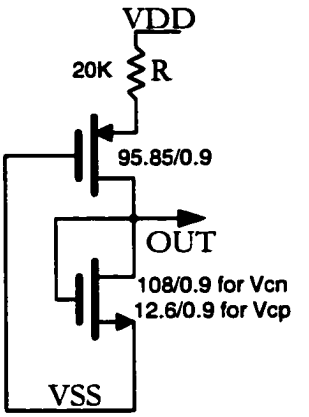


Figure 2.18 Schematic of the bias voltage generator

2.6 Other Design Issues

In our design, the projected process is the AMI 0.5μm CMOS process with a threshold voltage about 0.8V. The maximum gate-source voltage in the circuits is below 1.6V which is much smaller than the nominal 5V supply for the process. If it were used for a design in, for example, a 0.13μm process with threshold voltage of 0.4V, then the maximum gate-source voltage will not be higher than 0.8V which will not exceed the nominal supply for the process of about 1.5V. So using this design technique in any CMOS process, the actual gate-source voltages of all the transistors are always far lower than the nominal power supply. So there is no stressing to the gate oxide with this approach.

All the transistors in the auxiliary circuits work at either sub-threshold region or weak inversion region depending on the supply voltages. On the contrary, all the transistors in the low voltage op amp core work in the strong inversion region. This property makes it possible for the low voltage op amp to have comparable high frequency performance to that of the regular op amp.

Because most of the auxiliary circuits are digital and their operating frequencies are very low, they consume a very small amount of current. From our experimental measurements, the auxiliary circuits draw only 2.1μA (4% percent of total current) at 0.75V

and 2.3μA (3.4% of total current) at 0.8V. Most of this currents flow through the bias voltage generators. Furthermore, the same auxiliary circuits can be used for building larger low voltage circuits. The current consumption of the auxiliary circuits will remain at a similar level because a wider ϕ_2 can be designed for charging more capacitors.

2.7 Chip Layout

The chip micrograph is shown in Figure 2.19. In the upper right corner are the layouts of the normal 5V op amp and the low voltage op amp. In the left lower corner are the auxiliary circuits. Most of the active area is occupied by capacitors.

Several techniques were used in the chip layout in order to get optimal performance. The capacitors are poly-poly capacitors. The bottom plates of the capacitors were carefully placed to avoid any possible complications caused by bottom plate capacitance which is usually 10%-20% of the total capacitance. Refer to Figure 2.8, the bottom plates of the capacitors used for M4 and M5 are connected to the inputs acting as input capacitance. The bottom plates of the capacitors used for M6 and M7 are connected to the drain of M4 which lowered the pole frequency associated with this node. It does not affect the frequency response of the op amp because the dominant pole is still at the output node of the first stage. The bottom plates of the capacitors used for M8 are connected to the drain of M5. Its effect is enhancing the compensation a little bit. Finally, the bottom plate of the compensation capacitor is connected to the output node acting as load capacitance. There is no matching requirement for the capacitors. The capacitors were not laid out for matching performance.

Digital circuits and analog circuits are separated in the layout as far as possible in order to lower the switching noise injected into the analog circuits. Bias voltage generators are imbedded in the capacitor array and surrounded by guard rings in order to give clean outputs.

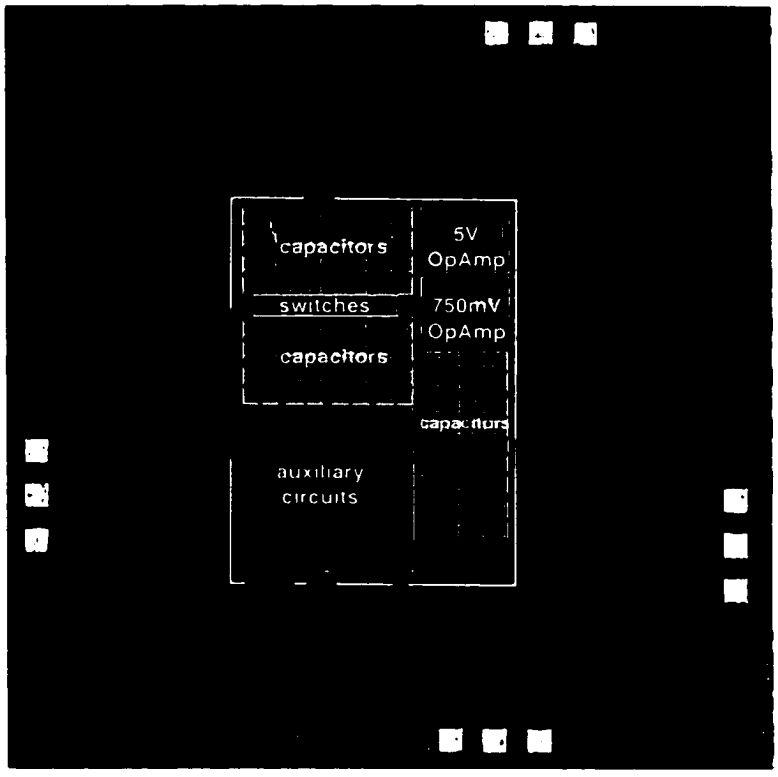


Figure 2.19 Die micrograph

The current mirrors in the op amp, M1/M2/M3, the transistor pairs M4/M5, M6/M7 are laid out using common-centroid technique in order to enhance its matching characteristics.

2.8 Experimental Results

The prototype of this work was fabricated in the AMI 0.5μm CMOS process with $V_{tn} \approx 0.8V$ and $V_{tp} \approx -0.9V$ by MOSIS. The active area is 560μm×760μm including the 5V op amp. The testing circuits were built on a breadboard.

Experimental results of the low voltage op amp are summarized in Table 2.4. It works with supply voltage as low as 750mV which is about $0.9V_t$. It also is able to work with a supply voltage up to 1V. This supply range is determined by both the effective threshold voltages and the V_{dsat} requirements. Slew rate performance is very close to that of a normal

op amp with similar current level. The input common-mode range can start from close to zero because of the voltage source we added to the gate of the input pair. During measurements, we found out that the charging time of the capacitor C_1 in Figure 2.6 can be as low as 0.4mS for a 3.2mS period while still maintain the performance.

Table 2.4 Summary of the experimental results of the low-voltage op amp

	@750mV	@800mV	@900mV	@1V
Slew Rate	3.1V/μS	3.8 V/μS	5 V/μS	6.36 V/μS
GBW	3.2MHz	3.7MHz	3.9MHz	4.2MHz
DC gain	62dB	64dB	64.6dB	64dB
Input offset voltage	*2.04mV (3σ value)	N/A	N/A	N/A
Input common mode range	0.1V-0.58V	0.07V-0.64V	0.02V-0.76V	0V-0.89V
Output swing for linear operation	0.31V-0.58V	0.27V-0.67V	0.15V-0.78V	0.1V-0.82V
PSRR at DC	82dB	N/A	N/A	N/A
CMRR at DC	56dB	N/A	N/A	N/A
Total power consumption	38.3μW (4% by auxiliary Circuits)	53.6μW (3.4% by auxiliary circuits)	81μW (2.7% by auxiliary circuits)	106μW (2.4% by auxiliary circuits)
Technology	AMI 0.5μm CMOS, double poly, triple metal			
Active area	560μm ×760μm			
Package	DIP28			

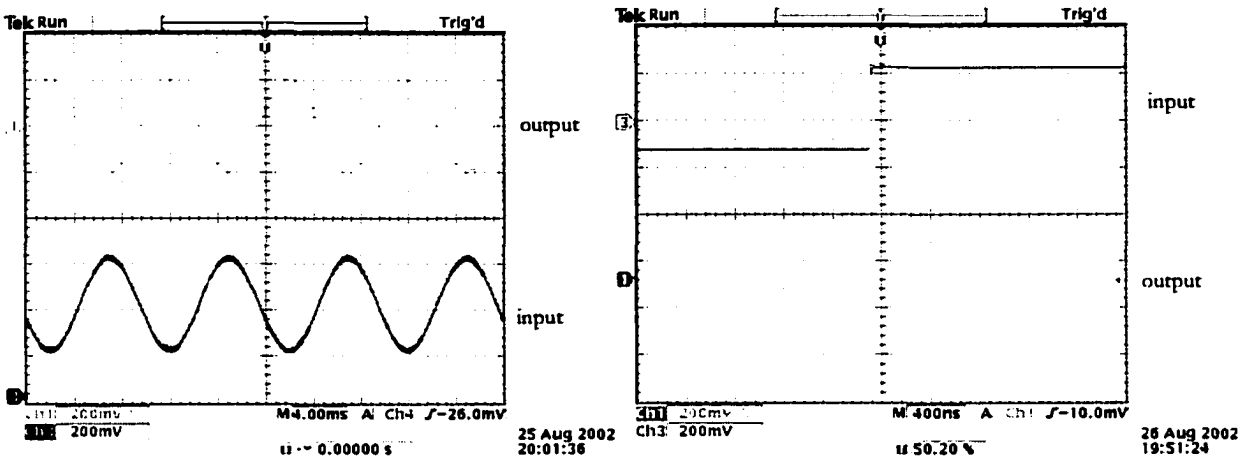
* Input offset voltages of 15 samples. Maximum value is 3.7mV and minimum value is 1.1mV. Standard deviation is 0.68mV.

Figure 2.20(a) shows the input and output waveform for an inverting gain feedback configuration with a gain of one. External resistors of value 10K were used to form the feedback network. Figure 2.20(b) shows the unity gain step response. In the design phase, we under-estimated the loading capacitance of the op amp. In the testing setup, the loading capacitance was about 15pF and that made the phase margin of the op amp to be around 40 degrees. That is why the overshooting appears in the step response. Table 2.5 shows the comparison of the experimental results of the low voltage and the 5V op amp. We observed that, with similar current consumption, GBW of the low voltage op amp degrades by less than 7% percent from what is achievable with the high voltage op amp when operating with a signal 0.8V supply. The DC gain of the low voltage op amp is about 20dB lower than the normal op amp. This is due to insufficient output impedance because of the lowered V_{DSAT} for the MOSFETs in the low voltage op amp. The power dissipation of the low voltage op amp when operating with the 750mV supply is only 11% of that of the high voltage op amp with comparable dynamic performance.

We tested 15 samples for the unity gain step response. All of them gave very similar results like what is shown in Figure 2.20(b). This result shows the robustness of this design technique. After testing the chip, we adjusted the simulation setup in Cadence to reflect the real testing situation. We added power supply models, pad frame models, package models and larger load capacitance. The experimental results suggested that the simulation results were pretty accurate. Shown in Table 2.6 is the comparison of the simulation results (after testing) and the experimental results at the 750mV power supply.

Table 2.5 Comparison of the normal and the low voltage op amps

	DC gain	Current	Power	GBW	Slew Rate
LV op amp@750mV	62dB	51μA	38.3μW	3.2MHz	3.1V/μS
LV op amp@800mV	64dB	67μA	53.6μW	3.7MHz	3.8V/μS
5V op amp	84dB	70μA	350μW	4MHz	3.9V/μS



(a) (b)

Figure 2.20 Oscilloscope captures of (a) Inverting configuration with gain of one; (b) step response of unity-gain configuration

Table 2.6 Comparison of the simulation and experimental results

	Simulation results	Testing results
DC gain	64.8dB	62dB
GBW	3.34MHz	3.2MHz
Slew Rate	3V/μS	3.1V/μS
PSRR at DC	86dB	82dB
CMRR at DC	56dB	56dB

2.9 Conclusion

A threshold voltage tuning technique for designing very low voltage analog circuits was introduced. To validate this technique, a low voltage op amp was designed and tested. This low voltage op amp used only the standard transistors available in any CMOS process and is able to work at a supply voltage LOWER than the threshold voltage (~ $0.9V_t$). All key transistors in the op amp core work in the strong inversion region despite the extremely-low supply voltage. It maintains comparable performance to that of a traditional high voltage

design operating at the same current levels while it greatly reduces the power consumption. To our knowledge, our design is the first implementation of a strong inversion op amp that works at a supply lower than the threshold voltage in standard CMOS process and the first very low voltage op amp that maintains dynamic performance comparable to that of op amps requiring much larger supply voltages.

References

[2.1] B. Razavi, "Design of Analog CMOS Integrated Circuits", McGraw-Hill Company, 2001

[2.2] A. L. Coban, P.E. Allen, "A low-voltage CMOS op amp with rail-to-rail constant-g_m input stage and high-gain output stage", 1995 IEEE International Symposium on Circuits and Systems, vol.2, pp. 1548 -1551, 1995

[2.3] G. Ferri, W. Sansen, "A 1.3 V op/amp in standard 0.7μm CMOS with constant g_m and rail-to-rail input and output stages", Digest of Technical Papers of Solid-State Circuits Conference, pp. 382 -383, 478, 1996

[2.4] G. Giustolisi, G. Palmisano, G. Palumbo, T. Segreto, "1.2V CMOS op-amp with a dynamically biased output stage", IEEE J. Solid-State Circuits, Vol. 35, pp. 632-636, April 2000

[2.5] E. K. F. Lee, "Low-Voltage Opamp Design and Differential Difference Amplifier Design Using Linear Transconductor with Resistor Input", IEEE Transactions on Circuits and Systems – II: Analog and Digital Signal Processing, Vol. 47, NO. 8, pp. 776-778, August 2000

[2.6] S. Sakurai, M. Ismail, "Robust Design of Rail-to-Rail CMOS Operational Amplifiers for a Low Power Supply Voltage", IEEE Journal of Solid-State Circuits, Vol. 31, NO. 2, pp. 146-156, February 1996

[2.7] G. N. Lu, G. Sou, "1.3V single-stage CMOS opamp", IEE Electronics Letters, Vol.34, pp. 2073-2074, Oct. 29, 1998

[2.8] G. Palmisano, G. Palumbo, R. Salerno, "A 1.5-V High Drive Capability CMOS Op-Amp", IEEE Journal of Solid-State Circuits, Vol. 34, NO. 2, pp. 248-252, February 1999

[2.9] R. Hogervorst, R. J. Wiegerink, P. A.L. de Jong, J. Fonderie, R. F. Wassennaar, J. H. Huijsing, "CMOS low-voltage operational amplifiers with constant-g_m rail-to-rail input stage", IEEE International Symposium on Circuits and Systems, Vol. 6 , pp. 2876-2879, 1992

[2.10] P. E. Allen, B. J. Blalock, G. A. Rincon, "A 1V CMOS op amp using bulk-driven MOSFETs", Digest of Technical Papers of IEEE International Solid-State Circuits Conference, pp. 192 -193, 1995

[2.11] K. Lasanen, E. Raisanen-Ruotsalainen, J. Kostamovaara, "A 1-V 5μW CMOS-Opamp with bulk-driven input transistors", Proceedings of the 43rd IEEE Midwest Symposium on Circuits and Systems, Vol. 3, pp. 1038-1041, 2000

[2.12] J. Ramirez-Angulo, R. G. Carvajal, J. Tombs, A. Torralba, "Low-voltage CMOS op-amp with rail-to-rail input and output signal swing for continuous-time signal processing using multiple-input floating-gate transistors", IEEE Transaction on Circuits and Systems - II: Analog and Digital Signal Processing, Vol. 48, pp. 111-116, Jan 2001

[2.13] T. Stockstad, H. Yoshizawa, "0.9V 0.5μA rail-to-rail CMOS operational amplifier", IEEE Journal of Solid-State Circuits, Vol. 37, pp. 286-292, March 2002

[2.14] C.-G. Yu, R. L. Geiger, "Very low voltage operational amplifiers using floating gate MOS transistor", IEEE International Symposium on Circuits and Systems, Vol. 2, pp. 1152-1155, May 1993

[2.15] E. Sanchez-Sinencio, "Low voltage analog circuit design techniques: a tutorial", IEEE Circuits and Systems Dallas workshop 2000, March 27, 2000

[2.16] J. Zhou, "500mV low voltage operational amplifier design", M.S. thesis, Iowa State University, 1997

[2.17] Y. Tang, R. L. Geiger, "A 0.6V Ultra-low voltage Operational Amplifier", IEEE International Symposium on Circuits and Systems, Phoenix, May 2002

[2.18] D. A. Johns, K. Martin, "Analog Integrated Circuit Design", John Wiley & Sons, Inc. 1997

CHAPTER 3

A HIGH-SPEED PHASE-LOCKED LOOP WITH NON-SEQUENTIAL LINEAR PHASE DETECTOR FOR DATA RECOVERY

3.1 Clock/Data Recovery and Phase-Locked Loop

Clock and data recovery (CDR) is a critical function in high-speed transceivers. Such transceivers serve in many applications, including optical communications, backplane routing, and chip-to-chip interconnects. The data received in these systems are both asynchronous and noisy, requiring that a clock be extracted to allow synchronous operation. Furthermore, the data must be "retimed" so that the jitter accumulated during transmission is reduced. CDR circuits must satisfy stringent specifications defined by communication standards thus posing difficult challenges to system and circuit designers.

At gigahertz data rates, CDR circuits are often implemented by expensive GaAs, SiGe, bipolar or BiCMOS processes. With the shrinking of gate length, deep sub-micron CMOS technology can also achieve fast operation which makes CMOS implementation of gigahertz transceivers possible. Designers face major challenges to take full advantage of the high speed capability of the sub-micron technology while still maintain the correct operation of the CDR circuits.

The task of CDR is often realized by using Phase-Locked Loops (PLL). A typical CDR system is shown in Figure 3.1. The Phase Detector (PD) is used to compare the phase difference of the data and the clock generated by the local voltage-controlled oscillator (VCO). The feedback loop is used to adjust the frequency (thus phase) of the clock until the clock has the same phase as the data. Ideally, the recovered clock is then used to sample the data at the center of each bit period. Because the center of the bit period has the best

possibility of being sampled correctly, the re-generated data will have much lower bit error rate and thus much wider eye openings if viewed as eye diagram.

The PD is a key component of the PLL. The performance of CDR circuits critically depends on the characteristics of the PD. With existing circuit implementations, the PD is often the bottleneck that limits the data rates that can be achieved by the PLL.

The PD can be categorized into two types. One is used in PLLs that lock to a reference clock signal. Many PDs can perform this function. Included in the group are Gilbert multipliers, XOR gates and RS latches etc. The other can be used in PLLs that lock to random Non-Return-to-Zero (NRZ) data and recover the clock signal which is embedded in the data stream. Since the spectrum of the NRZ data has reduced energy at the data rate, this makes the task of data recovery more difficult and places more severe restrictions on the performance of the PD. Often it requires a nonlinear operation at the front end of the PD circuit to generate more energy at its data rate.

With the booming of telecommunication applications in the late 90's, significant progress has been made on designing high speed CMOS PDs. Many novel configurations and design techniques have emerged. In the next section we will give a review of the existing PD structures.

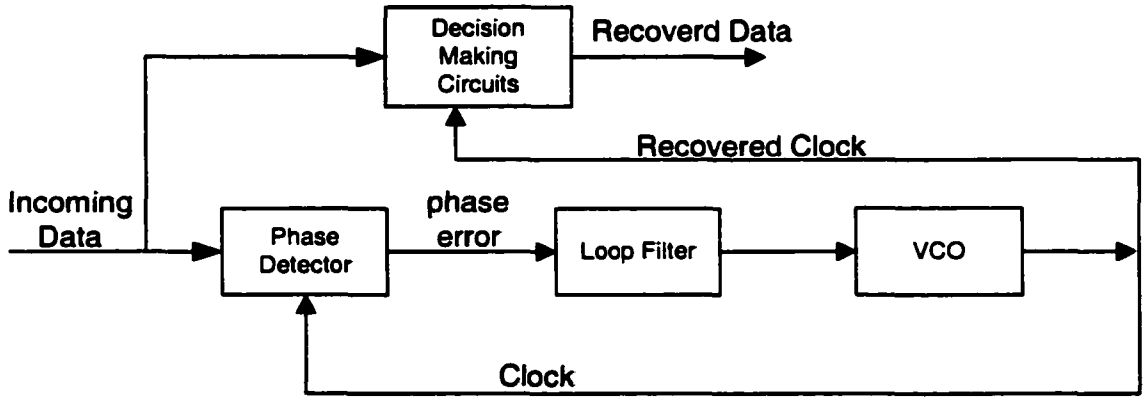


Figure 3.1 Block diagram of a typical data recovery system

3.2 Phase Detector Review

A brief review of the existing PDs will be given in this section. Basic operation for each PD and its advantages, as well as shortcomings, will be summarized. Details about how those PDs can be used as phase detector will not be repeated here. For those who are interested, some details can be found in the references.

3.2.1 Phase Detectors for Clock Recovery

A. Gilbert Cell Phase Detector

The schematic of the CMOS Gilbert cell [3.1] is shown in Figure 3.2. The output of the Gilbert cell can be expressed as:

$$I_{out} = \frac{uCox}{2} V_X \left[\sqrt{\left(\sqrt{\frac{I_{ss}}{k} - \frac{V_Y^2}{2}} + \frac{V_Y}{\sqrt{2}} \right)^2 - V_X^2} - \sqrt{\left(\sqrt{\frac{I_{ss}}{k} - \frac{V_Y^2}{2}} - \frac{V_Y}{\sqrt{2}} \right)^2 - V_X^2} \right] \quad (3.1)$$

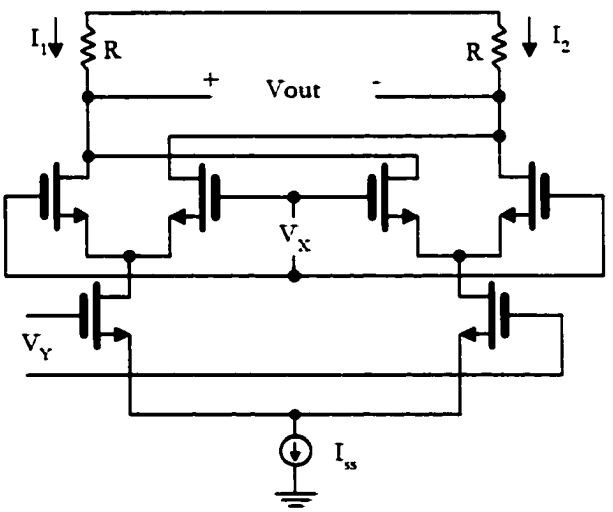


Figure 3.2 Schematic of Gilbert cell

The output current, therefore, has a nonlinear relationship with V_X and V_Y. When V_X and V_Y are very small, the output can be approximated by

$$I_{out} = |I_7 - I_8| = \sqrt{2} k V_X V_Y \quad (3.2)$$

We conclude that when signals of small amplitude are applied to the inputs of the cell, it behaves as an analog multiplier. If the phase difference of the inputs is in the vicinity of 90°, the average value of the output is linearity proportional to the phase difference.

The advantage of the Gilbert cell as a PD is its high speed compared to other structures. However, it suffers from a severe disadvantage. Its gain depends on the amplitude of the inputs. It also consumes static current which is not desired.

The Gilbert cell is seldom used in modern digital data communication systems.

B. XOR Phase Detector

The principle of the XOR gate used as PD [3.2] is shown in Figure 3.3. As the phase difference between the inputs "A" and "B" deviates from 90°, the output duty cycle departs from 50% resulting in an average output that is proportional to the phase error.

The Gilbert cell can actually be used as an XOR gate if the amplitude of the inputs are large. The advantage of the XOR gate as a PD is that it has a low sensitivity to the noise. Unfortunately its performance will be greatly impaired by asymmetric inputs (different duty cycle).

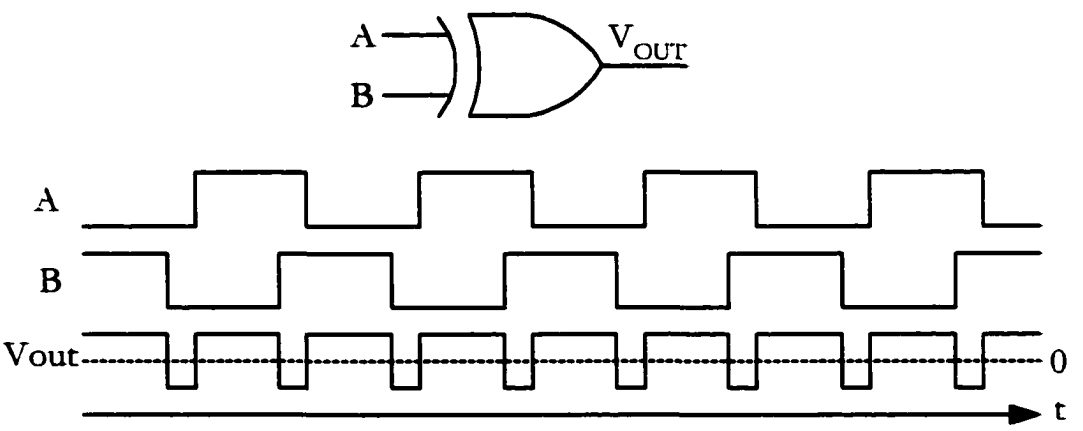


Figure 3.3 Operation of the XOR as a phase detector

C. *Two-state Phase Detector*

The name of "two-state" PD [3.2] comes from the fact that this kind of PD has two operation states. Shown in Figure 3.4 are the two-state PD and its state transition diagram. The high and low output indicate the two states.

The principles of the operation of the two-state PD are: the rising edge on signal R will make the output Q=1 while the rising edge on signal S will make the output Q=0. The advantages of this PD are the independence of the average value of its output on the duty cycle of the inputs and the improved acquisition range. Some drawbacks of this structure are that they are more sensitive to noise compared to the XOR and they may make the PLL lock to the harmonics (false lock).

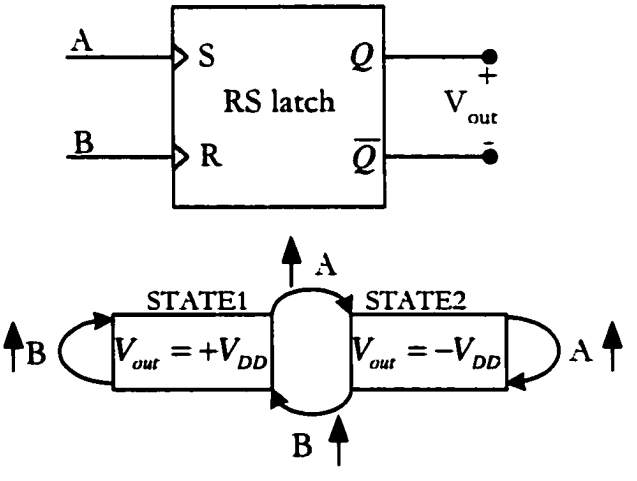


Figure 3.4 Two-state phase detector and its state transition diagram

D. *Three-state Phase Detector*

The three-state PD [3.2] [3.3] is similar to the two-state PD. The schematic of one possible implementation of the three-state PD and its state transition diagram are shown in Figure 3.5. It employs two edge-triggered resettable D flip-flops with their D inputs connected to VDD (logic HIGH). Signals A and B act as the clock input of the two flip-flops. The state of the PD is determined by the output of the two D-flip flops. The three states are:

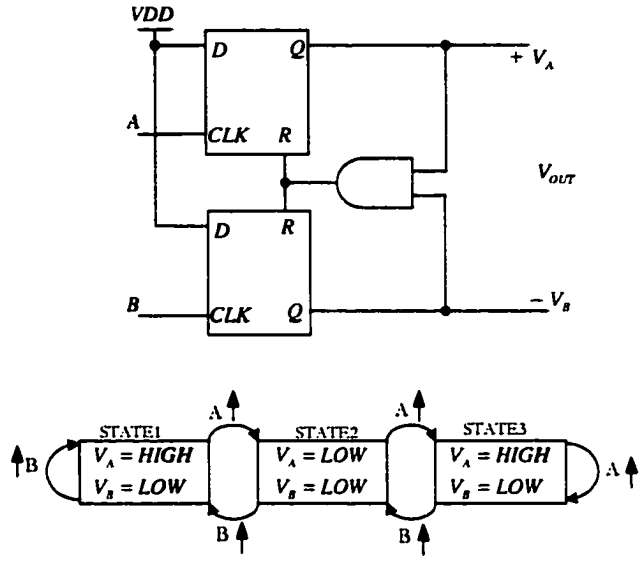


Figure 3.5 Three-state phase detector and its transition diagram

$(V_A = High, V_B = Low)$, $(V_A = Low, V_B = Low)$ and $(V_A = Low, V_B = High)$

V_A and V_B can not be high at the same time because this will reset the DFFs. The state transition diagram in Figure 3.5 clearly shows the operation principles of the three-state PD.

The performance of the three-state PD is better than that of the XOR and the two-state PD because it can also detect frequency difference. From its state transitions operation, if $\omega_A > \omega_B$, there are only one or no rising edges of V_A between two adjacent V_B rising edges. So the PD will stay at state 2 or 3, it can not reach state 1. The output V_{out} will remain positive. If $\omega_B > \omega_A$, there are only one or no rising edges of V_B between two adjacent V_A rising edges. So the PD will stay at state 2 or 1, it can not reach state 3. The output V_{out} will remain negative. This is a great aid in acquiring lock when the two frequencies are initially different.

This PD is edge-triggered and it is not sensitive to the duty cycle of the inputs. But it is very sensitive to the loss of transitions in the inputs which means it is not suitable for data recovery.

3.2.2 Phase Detectors for Data Recovery

A. Hogge Phase Detector

The Hogge PD [3.4] and its variations [3.5] [3.6] are probably the most widely-used PD for data recovery. The schematic of Hogge PD is shown in Figure 3.6(a). It uses two D flip-flops and two XOR gates. Complementary clocks are used to drive the two DFFs. The operation of the Hogge PD is shown in Figure 3.6(b). The signal "DOWN" is used as a reference to "UP". The width of pulse appeared on "DOWN" will always be half the period of "CLOCK". Figure 3.6(b) shows the situation when PLL is locked. We see that the pulse widths on "UP" and "DOWN" are equal. The output of the PD is the duty cycle differences between "UP" and "DOWN". Using this PD to drive a charge pump, the output of the charge pump will not change when the phase difference between data and clock is 0.

Consider the situation when "CLOCK" is leading "DATA" (phase leading), the pulse widths on "UP" will be shorter while the pulse width on "DOWN" will not change. Similarly, when "CLOCK" is lagging "DATA" (phase lagging), the pulse widths on "UP" will be wider while the pulse width on "DOWN" will not change. Therefore, the output of the PD will change negatively or positively according to the phase difference at the input.

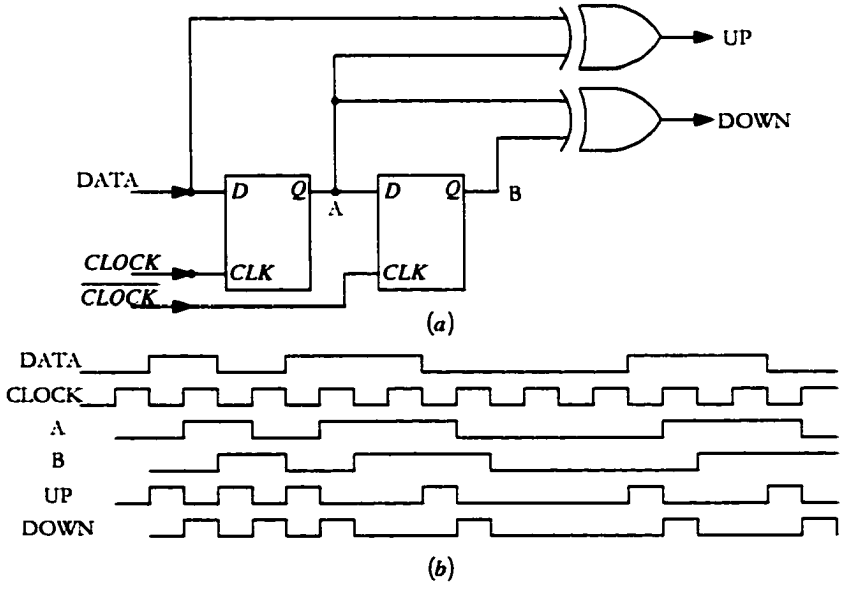


Figure 3.6 Hogge phase detector (a) schematic; (b) its operation

The Hogge topology is a linear PD generating a small average as the phase difference approaches zero and generating an output that is linearity proportional to the phase difference during normal operation. This linear behavior is desirable, particularly so when contrasted to that of the bang-bang PD as we will discuss in the next section.

B. Bang-Bang Phase Detector

Bang-bang PD [3.7] [3.8] [3.9] [3.10] refers to a group of PDs that only have two output states. This is in contrast to the linear PDs that have an output that is either proportional to the phase difference or at least varies continuously with phase difference. All the PDs we discuss so far fall into the linear this category including the Hogge PD.

The simplest bang-bang PD is just a D flip-flop. Its structure and transfer characteristic are shown in Figure 3.7. Circuit of Figure 3.7(a) operates as follows. Upon turn-on, the DFF multiplies the data by the VCO output, generating a beat that drives the VCO frequency toward the input bit rate. If the initial difference between the VCO frequency and the data rate is sufficiently small, the loop locks, establishing a well-defined phase relationship between "DATA" and "CLOCK". In fact, with the bang-bang characteristic provided by the DFF PD, the data edges settle around the zero-crossing points of the clock. Even for a slight phase error, the PD generates a large output.

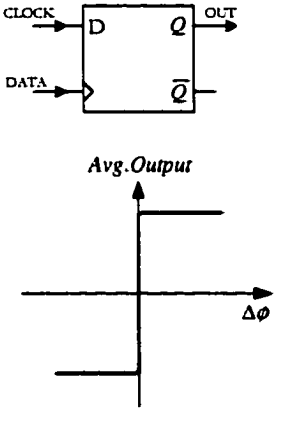


Figure 3.7 Simplest Bang-Bang phase detector and its transfer characteristic

This PD has a lot of drawbacks and has limited applications. More sophisticated bang-bang PDs have been proposed over the last decade. All of them bear a similar transfer characteristic to that shown in Figure 3.7(b). Both the advantages and the drawbacks of the bang-bang PD are related to its two-state output. Because of the simplicity of the output, it can employ very simple circuit to achieve high speed. On the other hand, the two-state output of the bang-bang PD will create a significant ripple on the control line of voltage-controlled oscillators when the PLL is in lock which causes high jitter. Usually additional circuits are needed to suppress the ripple on the control line.

C. Phase Detector with half-rate clock

PDs with half-rate clock input [3.11] [3.12] [3.13] have become a hot topic recently where data rates of the transceiver design moves into the gigabit/s range. At very high speeds, it may be very difficult to design oscillators that provide an adequate tuning range with reasonable jitter. For this reason, PLL circuits may sense the input random data at full rate but utilize a VCO running at half the input rate. Such PLL topologies require a PD that provides a valid output while sensing a full-rate random data stream and a half-rate clock.

One example [3.13] of this type of PD is shown in Figure 3.8(a). The circuit consists of four latches and two XOR gates. The data is applied to the inputs of two sets of cascaded latches. Each cascaded latches constitutes a flip-flop that retimes the data.

The operation of the PD can be described using the waveforms depicted in Figure 3.8(b). The basic unit employed in the circuit is a latch whose output carries information about the zero crossings of both the data and the clock. The output of each latch tracks its input for half a clock period and holds the value for the other half, yielding the waveforms shown in Figure 3.8(b) for points X_1 and X_2. The two waveforms differ because their corresponding latches operate on opposite clock edges. Produced as $X_1 \oplus X_2$, the "*Error*" signal is equal to ZERO for the portion of time that identical bits of X_1 and X_2 overlap, and

equal to the XOR of two consecutive bits for the rest. In other words, "*Error*" is equal to ONE only if a data transition has occurred.

The random nature of the data and the periodic behavior of the clock make the average value of "*Error*" pattern dependent. For this reason, a reference signal must also be generated whose average conveys this dependence. The two waveforms Y_1 and Y_2 contain the samples of the data at the rising and falling edges of the clock. Thus, $Y_1 \oplus Y_2$ contains pulses as wide as half the clock period for every data transition, serving as the reference signal. The amplitude of "*Error*" must be scaled up by a factor of two with respect to Reference so that the difference between their averages drops to zero when clock transitions are in the center of the data eye.

This structure is very similar to Hogge PD. Actually it is an interleaved Hogge type PD modified to make it able to work with a half-rate clock. The speed potentials and limitations of the Hogge PD and this PD are the same.

There are several other types of PDs such as sample and hold PD [3.14] and Alexander PD [3.15], which will not be discussed here in details.

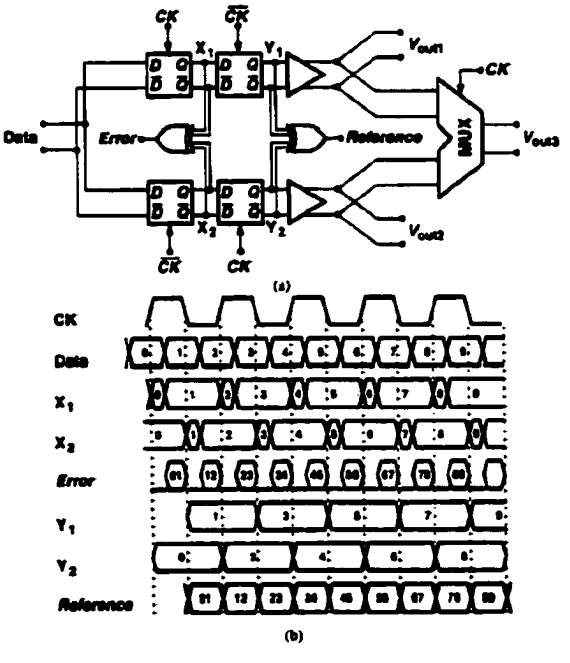


Figure 3.8 Phase detector with half-rate clock (a) Schematic; (b) its operation

3.3 A New Phase Detector for High Speed Data Recovery

All the PDs used for data recovery that we discussed so far are based on state-machines, that is, they use flip-flops to memorize the state of the PD to determine the phase difference between the random data and the clock. This approach has its drawbacks. Take the Hogge PD, the most popular one, for example. Typical of the state-based PDs in this class, the performance of the Hogge PD deteriorates rapidly at higher frequencies. These performance limitations are due mainly to the inadequate settling performance of the flip-flop used to form the state machine.

In order to overcome the inability of the Hogge PD to work at higher speeds, a new PD was introduced [3.16]. It can be used in PLLs designed to recover high-frequency clocks embedded in pseudo-random NRZ data streams. The simple architecture and the elimination of the state machine contribute to the improved high-frequency performance of this circuit.

In contrast to existing PDs that use a single-phase clock and multi-phased data signals, the new PD uses multi-phase clock signals and the actual data sequence to achieve simplicity and high speed operation. A general structure of the proposed PD is shown in Figure 3.9. It is applicable for all kinds of PLL designs. Multi-phase clock signals ("CLK_d1" and "CLK_d2") are generated by delaying the clock signal "CLK". When the PLL acquires lock, "CLK_d1" will phase-lock to "Data_d1". The operation details of this PD will be explained later.

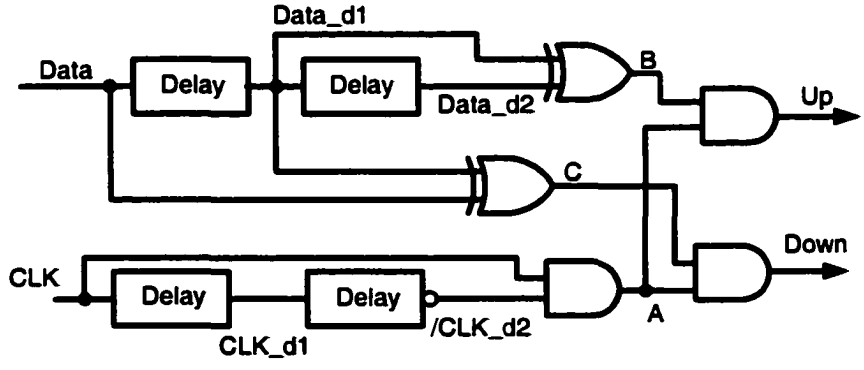


Figure 3.9 General structure of the proposed phase detector

In PLLs using ring-oscillator type VCOs, the multi-phase clock signals are inherently available and two delay stages for generating the delayed clock signals are not necessary thus simplifying the structure of the PD. For example, the PD used with a 3-stage ring oscillator VCO is shown in Figure 3.10. The two signals extracted from the VCO labeled "CLK_lead" and "/CLK_lag" are the leading and inverting lagging signals of the Clock ("CLK") signal. Comparing Figure 3.9 and Figure 3.10, "CLK_d1" is analogous to "CLK" of Figure 3.10, "CLK" of Figure 3.9 is analogous to "CLK_lead" and "/CLK_d2" is analogous to "/CLK_lag". For either the structure of Figure 3.9 or Figure 3.10, two data delay cells and the two XOR gates are used to detect the edges of transitions in the random input data.

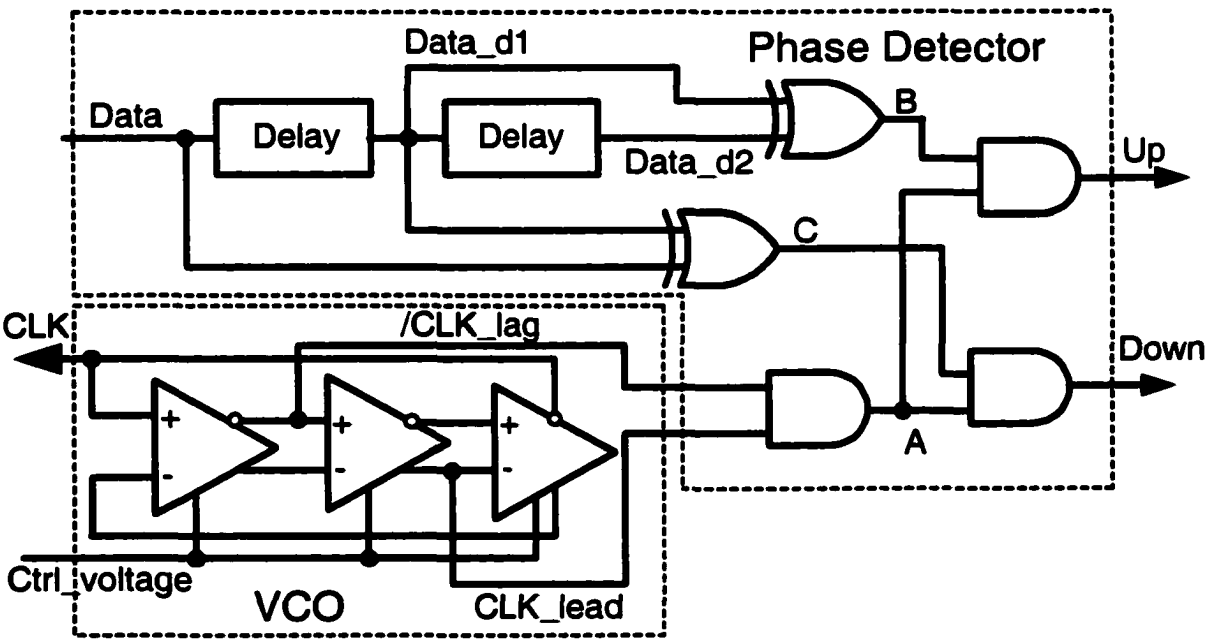


Figure 3.10 Phase detector structure used with ring oscillator with odd-number stages

Figure 3.11 shows the timing diagram (for circuits shown in Figure 3.10) for a segment of random input data when the PLL is in lock. The circuit aligns the rising edges of "CLK" with the middle of signal "A" independent of the data at the input. The falling edges of "C" and the rising edges of "B" are aligned at the dotted line, which, when the PLL is in

lock, are also aligned with the middle of the signal "A". The "Up" and "Down" signals are generated by using "B" and "C" to partition the "A" signal into equal width segments at each data transition. Therefore, the "Up" and "Down" signals have the same duty cycles when in lock and the output of the loop filter, which filters the difference in the duty cycles of the "Up" and "Down" signals, will not be driven up or down. The "Up" and "Down" signals are only generated whenever there are transitions in the incoming data stream. This property provides the ability to handle random NRZ data. Note that when the PLL is in lock, "CLK" phase-locks to "Data_d1" instead of "Data".

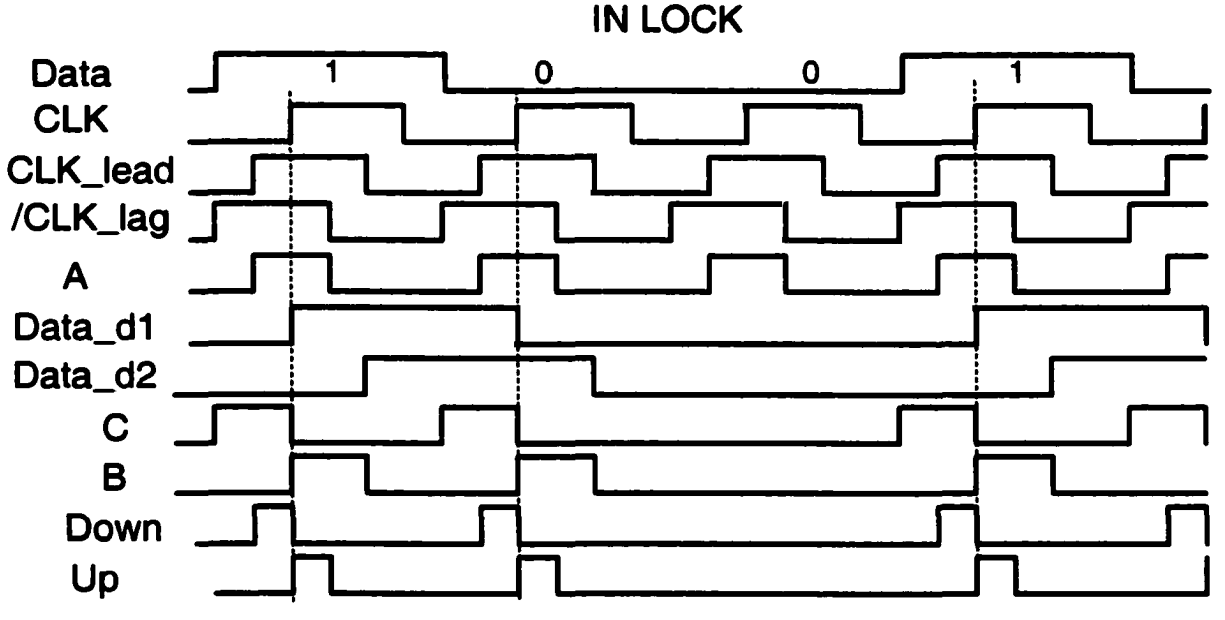


Figure 3.11 Operation principle of the proposed phase detector (in lock)

Figure 3.12 shows the situation when "Data" is leading and lagging the "CLK". When "CLK" leads "Data" as depicted in Figure 3.12(a), the falling edge of "C" and the rising edge of "B" no longer align in the middle of "A". Instead, they will move to the right of the pulse "A". Thus, the width of the "Up" will decrease and the width of "Down" will increase, which, in turn, will bring down the frequency of "CLK" through PLL and eventually make

the PLL back to lock. When "CLK" lags "Data" shown in 3.12(b), opposite change will happen for "Up" and "Down" which will increase the frequency of the "CLK".

The accuracy requirements for the delay time of the delay cells in the PD are lax. Proper operation of the proposed PD will be achieved provided the delay cell satisfies the inequality:

$$\max\left(\frac{1}{2}T_0, \frac{1}{2}(T - T_0)\right) < T_{delay} < \min\left(T - \frac{1}{2}T_0, \frac{1}{2}(T + T_0)\right) \tag{3.3}$$

where T is the period of the signal "CLK", "T_0" is the pulse width of the signal "A", and "T_{delay}" is the delay time of the delay cell. From the timing diagram, we can see the delay time mismatch of the two delay cells will not cause any problem for the correct operation of the PD.

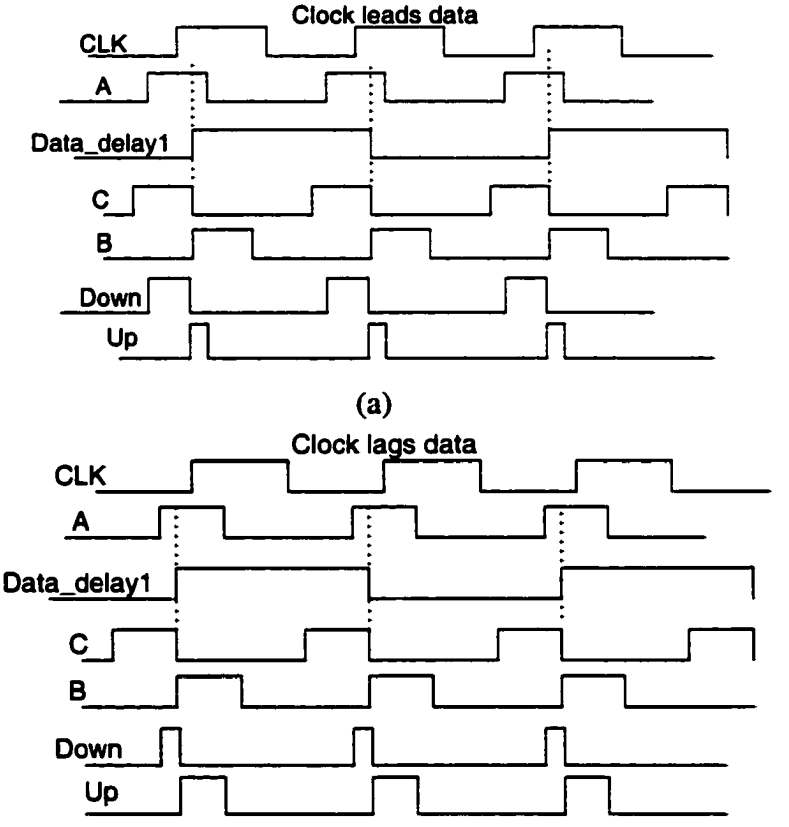


Figure 3.12 Timing diagram when (a) clock leads data; (b) clock lags data

For any VCO with odd number (≥ 3) of stages, the "CLK_lead" signal can come from the non-inverting output of the stage immediately preceding the clock output stage and the "/CLK_lag" signal can come from the inverting output of the immediately following stage. Other signals can also be used for "CLK_lead" and "/CLK_lag" when there are more than 3 delay stages depending on the VCO design.

When the VCO has an even number (≥ 4) of stages, the PD structure becomes even simpler. For example, Figure 3.13 depicts a 4 stage VCO. In this case, we can eliminate the AND gate that is used to generated the signal "A" since the signal "A" can be directly extracted from the VCO. For more than 4 delay stages in the VCO, the number and position of the cross-overs must be considered when extracting the "A" signal.

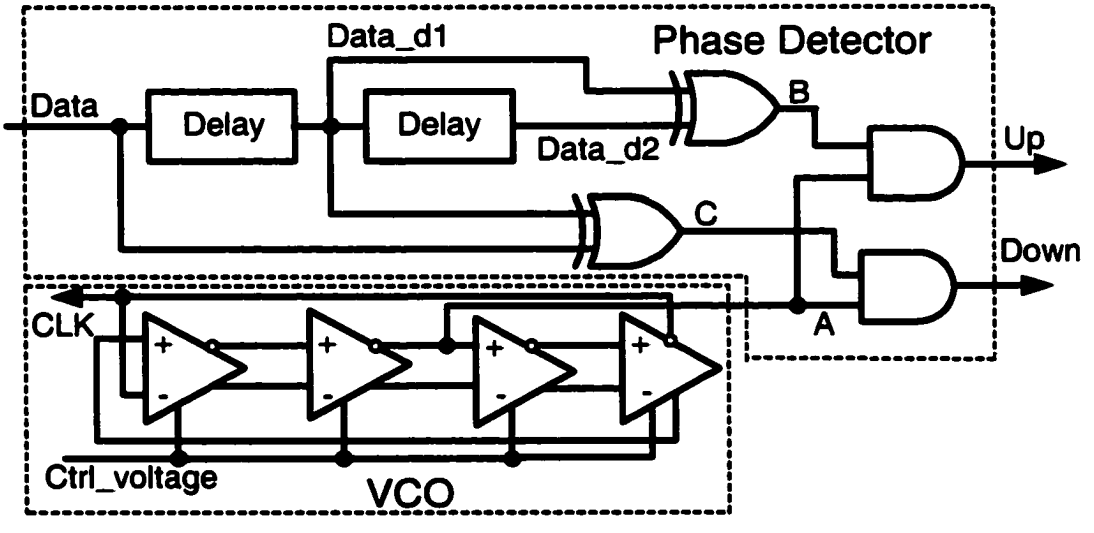


Figure 3.13 Phase detector structure used with ring oscillator with even-number stages

The transfer characteristic of the PD for a typical T_{delay} is shown in Figure 3.14. This sinusoid-like relationship is typical of a linear PD. Corresponding to different delay times, the shape of the curve will change modestly, but its functionality as a useful PD will be maintained.

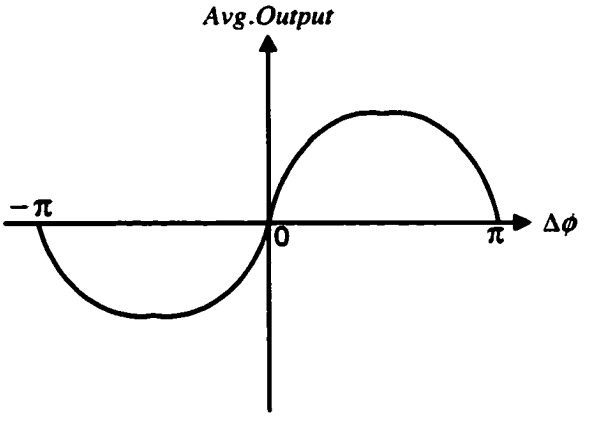


Figure 3.14 Typical transfer characteristic of the proposed phase detector

3.4 Analysis of the speed of new PD and Hogge PD

In our realization of the new PD, we used NAND and NOR gates to implement the function of AND gates because complementary signals are available.

The major factors that affect the speed of these two PDs are the rise time, fall time and the propagation delay of their components. Actually the propagation delay T_{delay} as defined in Figure 3.15 contains rising (falling) time information. So it is fair to use the propagation delay of the gates to make a simple relative speed comparison of the Hogge PD and our new PD.

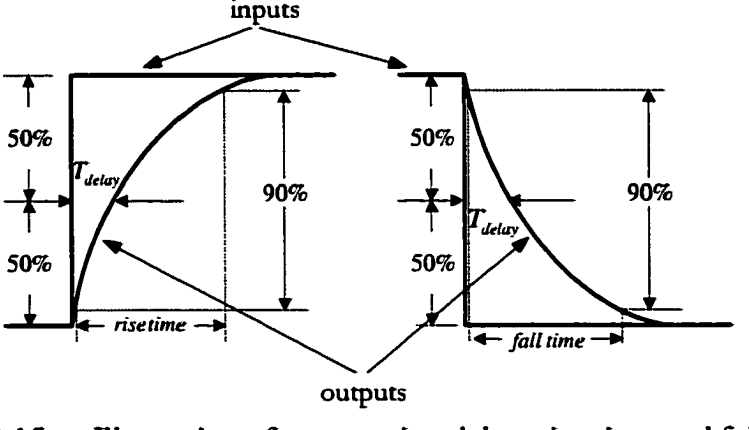


Figure 3.15 Illustration of propagation delay, rise time and fall time

The propagation delay of a DFF, an NAND (NOR) gate and an XOR gate can be obtained from circuit simulations. To make the speed comparisons, we will first characterize the propagation delays of the basic gates.

One of the fastest DFF realization is shown in Figure 3.16(a) [3.20]. It is also used as the frequency divider in the de-embedding circuits for testing our PLL design. Transistors are sized to achieve high speed. The implementations of the NAND and NOR gates are static CMOS logics as shown in Figure 3.16(b). They have very similar speed performances. The pass-transistor XOR gate shown in Figure 3.16(c) is used to achieve high speed.

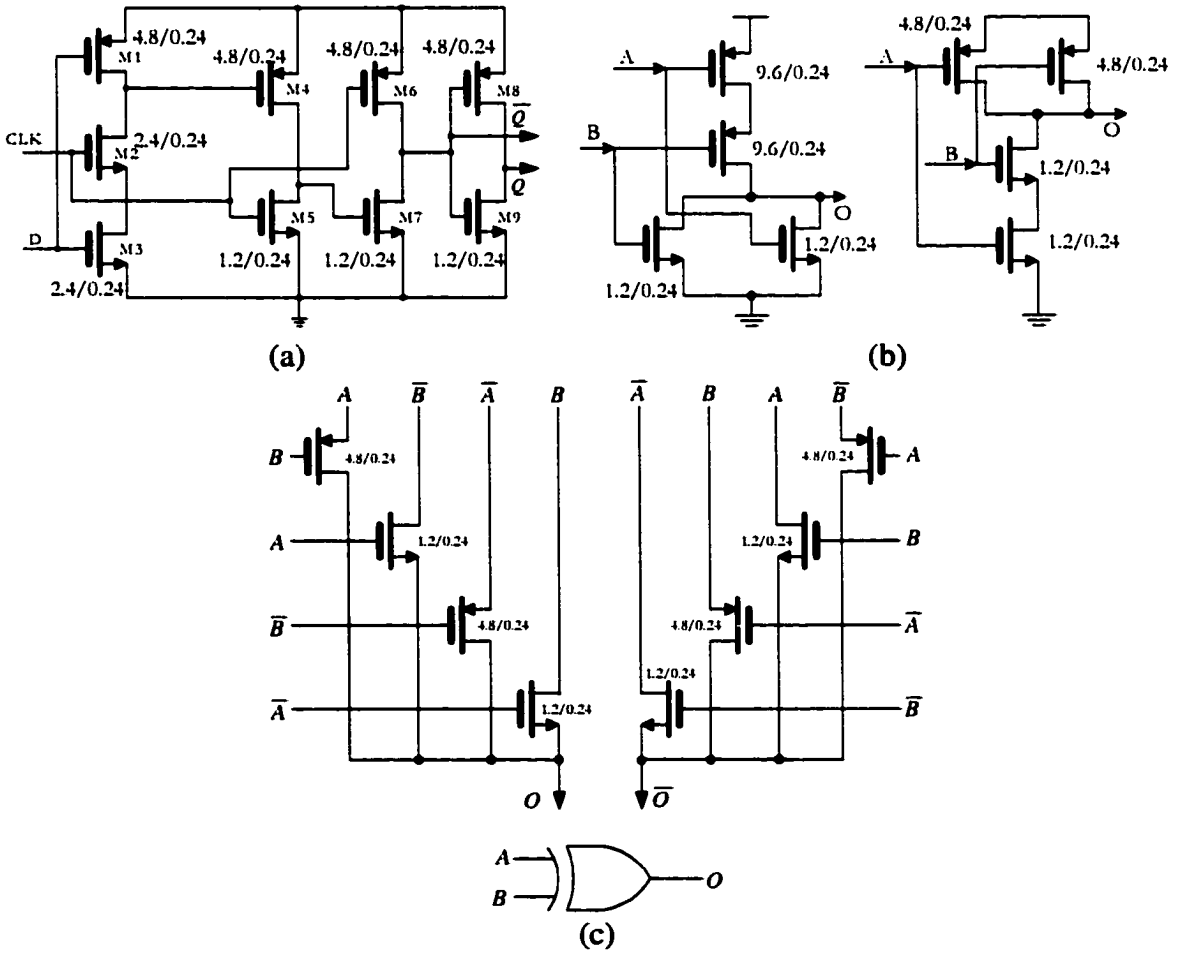


Figure 3.16 (a) 9-transistor dynamic DFF; (b) NOR and NAND gates; (c) Pass-transistor XOR gate

Using TSMC 0.25μm process device models and the HSPICE simulator, their propagation delays are summarized in Table 3.1. The rising and falling time for the input signals are both 10ps. The loads for these circuits are 20fp.

We see that the NAND (NOR) and XOR gates have great advantage of much smaller propagation delay.

Table 3.1 Speed performance of the DFF and NAND gate

	DFF	NAND	NOR	XOR
Propagation	105ps (rise)	48ps (rise)	43ps (rise)	40ps (rise)
delay	121ps (fall)	60ps (fall)	59ps (fall)	39ps (fall)

Our first comparison of the speeds of new PD and Hogge PD will be based on the propagation delay of DFF and NAND gate and assuming everything else in the PDs are ideal.

For Hogge PD, it normal operation is shown in Figure 3.17(a). As shown in Figure 3.17(b), when there is a delay in the DFF, the "A" and "B" will be delayed and the width of "DOWN" signal will be the same as in ideal case while the "UP" signal is wider. This will cause a static phase offset when the PLL is in lock. We can also see if the delay is bigger than half of the clock period as shown in Figure 3.17(c), the problem is very serious. The falling edges the "CLOCK" are sampling the wrong places of signal "A" and the operation of the PD is totally screwed up. We can conclude that the highest clock frequency that Hogge PD can handle is given by:

$$f = \frac{1}{2T_{delay}} \tag{3.4}$$

where T_{delay} is the propagation delay of the DFF.

This means that in the TSMC 0.25μm process, the upper limit set by the propagation delay of the DFFs for the Hogge PD is about $1/(2\times \sim 113ps) \approx 4.4GHz$.

For our new PD, the delays of the NAND and NOR gates do NOT affect the operation of the PD at all as shown in Figure 3.18. They only appear as delays in the outputs.

Thus, the propagation delay of the DFFs will cause a static phase offset and set the upper bound for the clock frequency that the Hogge PD can handle. The propagation delays of the NAND and NOR gates in new PD do not affect its operation.

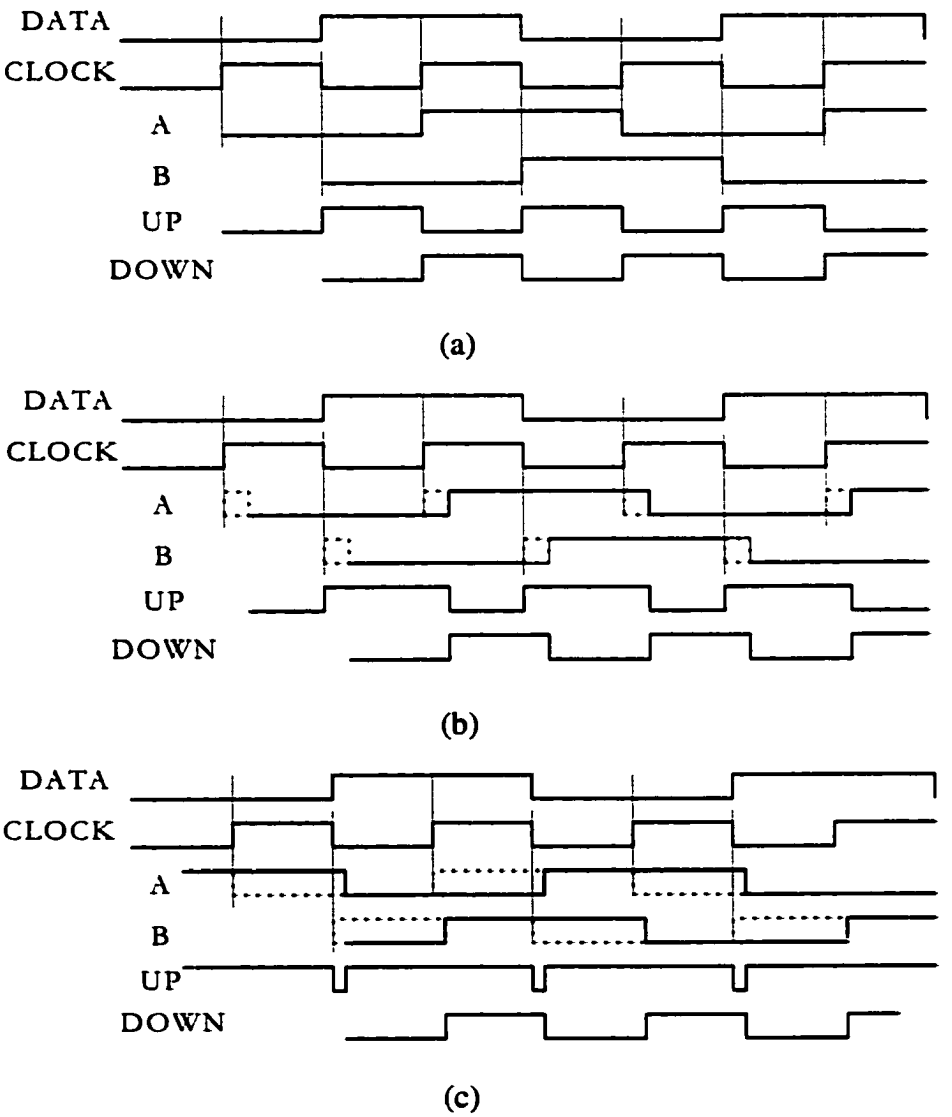


Figure 3.17. Hogge PD operation with DFF propagation delays (a) ideal case without delay; (b) delay is smaller than half clock period; (c) delay is larger than half clock period

Secondly, let us see the effects of the propagation delay of the XOR gates to the speeds of the Hogge and new PD.

As shown in Figure 3.19(a), the delay of the XOR will shift the output of the Hogge PD. It does not affect its speed. The timing diagram for new PD with XOR delay is shown in Figure 3.19(b). The XOR delay will cause a static phase offset. But unlike the case of DFF delay for Hogge PD, the XOR delay does NOT limit the upper bound of the input clock frequency. This is a benefit of the pure combinational logic operation instead of sequential logic operation.

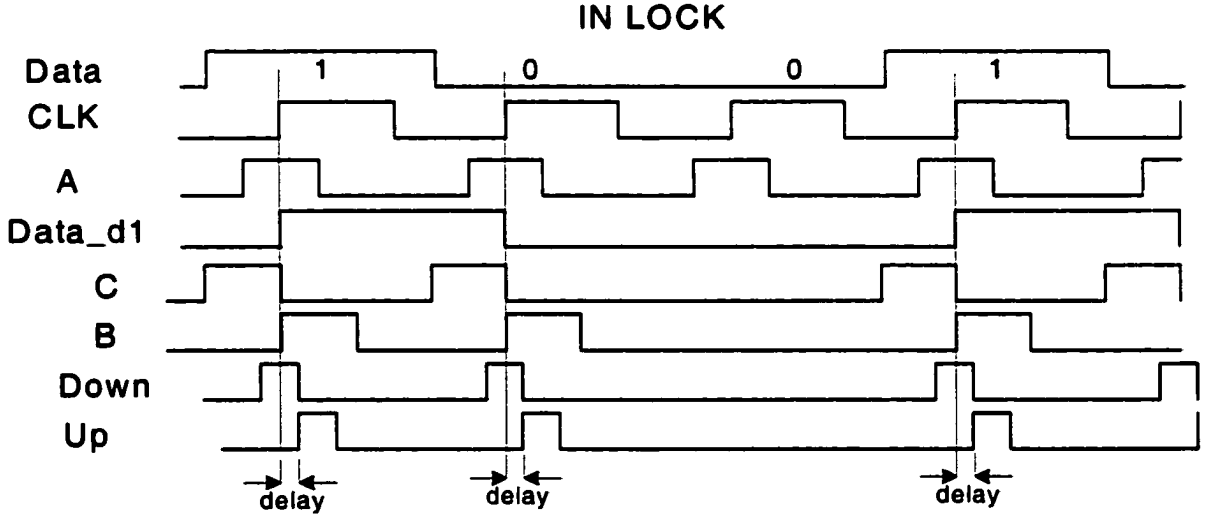


Figure 3.18. Effect of NAND (NOR) delay to new PD

Overall, the speed of the Hogge PD is limited by the DFF delay and the speed of the new PD does not limited by either NAND (NOR) delay or XOR delay. This property show great speed advantage of new PD over Hogge PD.

Having all the information above, we can now evaluate the speed of the new PD and the Hogge PD from a different point of view. That is to compare the static phase offset that is introduced by propagation delays. This offset will result in a higher bit error rate in the data recovery system because the clock sampling edge is not in the center of the data period. Usually a certain static phase offset budget is set to ensure an acceptable bit error rate.

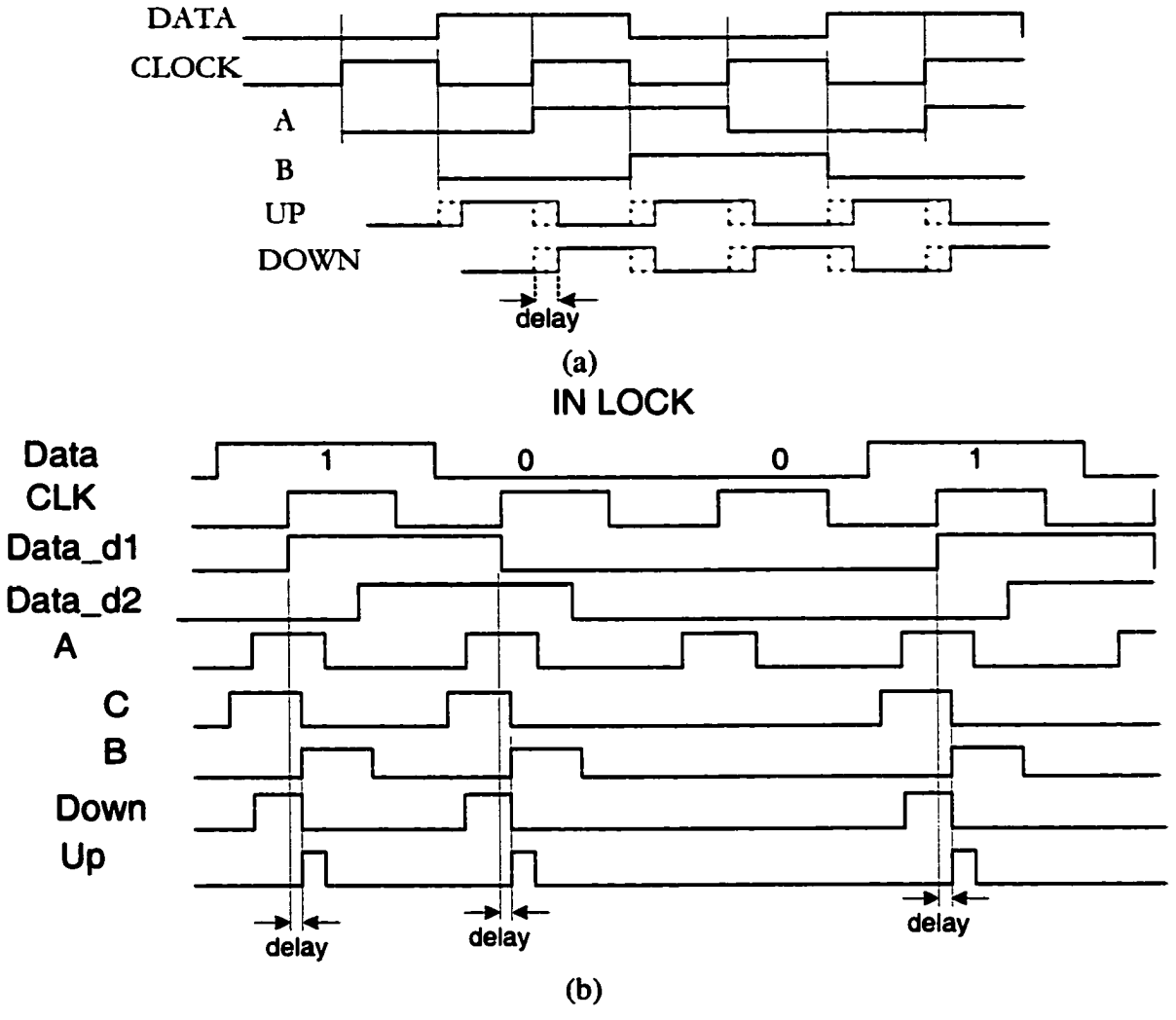


Figure 3.19. (a) Effect of XOR delay to Hogge PD; (b) effect of XOR delay to new PD

From the analysis above, we know the DFF delay and the XOR delay will cause a static phase offset in the Hogge PD and our new PD respectively. For both PDs, the amount of static phase offset is given by

$\frac{T_{delay}}{T_{CLK}} \times 360^{\circ}$ where T_{CLK} is the clock period

Using the parameter we got from simulation, $T_{delay\_DFF} \approx 113ps$ and $T_{delay\_XOR} \approx 40ps$ Figure 3.20 shows the static phase offset of the Hogge PD and the new PD versus input clock frequency when the clock is locked to data. The slopes of the two lines are proportional to

their propagation delays. By setting the phase offset budget at any point will result close to 3 times higher operating frequency for the new PD than for the Hogge PD. As shown in Figure 3.20, the line of the Hogge PD is cut off because of inability to operate at any higher speed. Basically, this graph shows the advantage of the much smaller XOR propagation delay over the DFF propagation delay.

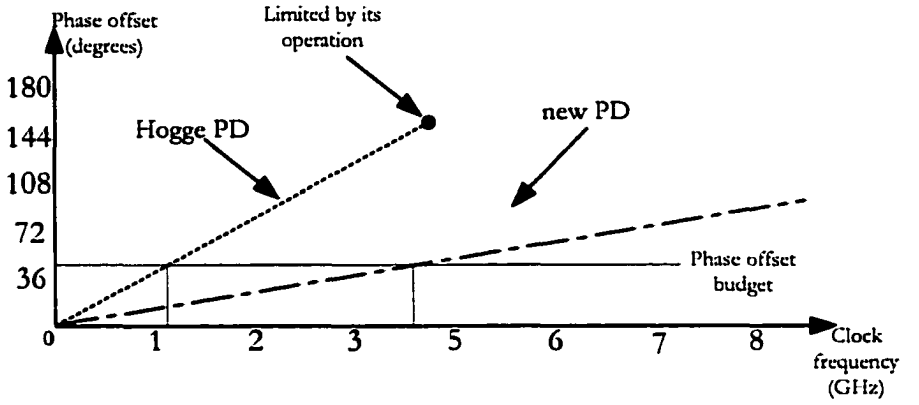


Figure 3.20 Clock frequency vs. phase offset plot for Hogge PD and new PD

The preceding analysis is not intended to indicate the ACTUAL speed for Hogge and our new PD in TSMC 0.25μm CMOS process. It is intended to demonstrate the relative speed advantage of the new PD.

Actually the motivation for us to propose the new PD was that we found the Hogge PD was not able to work as fast as we need. In one of our PLL design projects, we spent more than 6 weeks trying to design a Hogge PD using fully-differential current-steering logic to achieve a 2GHz operating frequency in the HP 0.35μm CMOS process. We were unable to meet our goal over process and temperature variations and were only able to reliably achieve about 1.5GHz performance. In contrast, we spent less than one week to design our new PD to work at 2GHz over the corners and temperatures using static and pass transistor logic in the same process.

3.5 Phase-Locked Loop Implementation

A high speed PLL using the proposed PD was designed. The architecture of the PLL is shown in Figure 3.21. It is a typical charge pump PLL structure. A high resolution current-steering charge pump follows the PD to convert the outputs, "Up" and "Down", to a control voltage referenced to positive power supply. A second-order passive loop filter is used. The VCO has a bias generator and four delay stages.

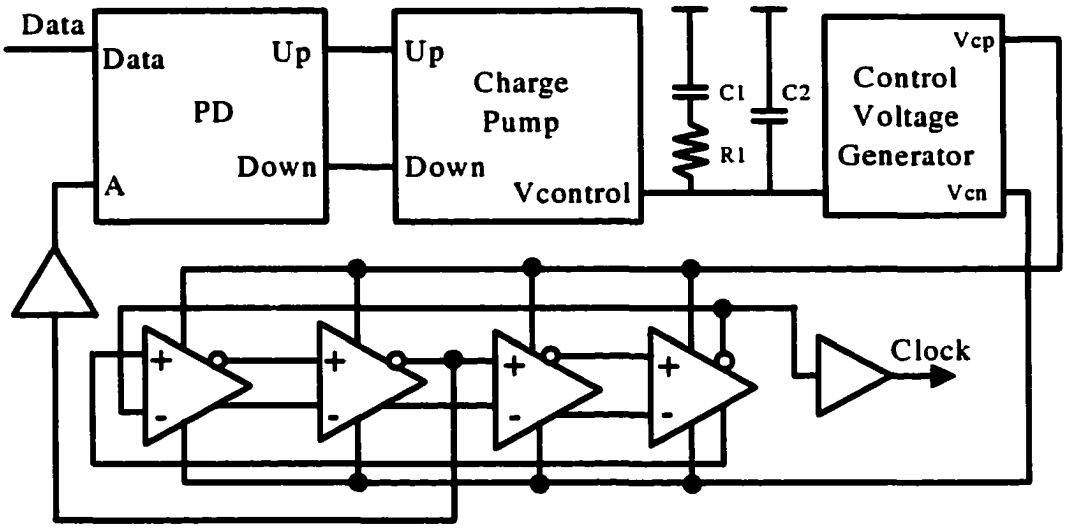


Figure 3.21 Structure of the phase-locked loop

Choosing different parameters for the design will greatly affect the loop performance of the PLL, especially the locking characteristics, stability and jitter. Assume the charge pump bias current is I_p, the transfer function of the loop filter is $F(s)$, and the gain of the VCO is K_{VCO}. The input and the output of the PLL are represented as θ_{in} and θ_{out}. A standard small signal analysis [3.17] shows that the small-signal transfer function of the PLL, θ_{out}/θ_{in} is given by:

$$\frac{\theta_{out}}{\theta_{in}} = \frac{K_{VCO} I_p F(s)}{2\pi s + K_{VCO} I_p F(s)} \tag{3.5}$$

which shows the standard low pass characteristics. The loop will reject high-frequency phase noise from the input and reject low-frequency phase noise from the VCO. Since the VCO is a major contributor to the jitter in the recovered data, to minimize the impact of the VCO phase noise, we need to make the bandwidth of the PLL large. This will not only suppress the phase noise of the VCO, but also increase the tracking speed of the PLL.

The gain of the VCO, K_{VCO}, is about 1.2GHz/V in our design. The bias current of the charge pump is about $I_p \approx 50\mu A$. Loop filter components value of C1=40p, C2=10p and R1=10K are selected. With these design values, the loop bandwidth is about 6MHz and the phase margin of the loop is around 65 degrees.

3.5.1 Phase Detector

The structure of the PD is the same as that shown in Figure 3.13. The delay cells are simply implemented by a 3 stages of inverters as shown in Figure 3.22. Since the delay time requirement for them is not very critical, it is easy to control the delay time over the temperature range and the process corners.

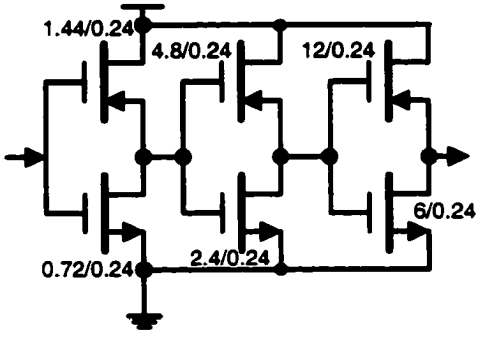


Figure 3.22 Implementation of the delay cells in phase detector

The initial goal of designing this PD is to eliminate the sequential logic circuits that are difficult to operate at high speeds. The proposed PD consists of only combinational logic.

It has the ability to operate at a higher speed than sequential logic circuits. A good choice of architectures for the XOR gates and AND gates is, however, crucial to achieve the high-speed operation in the proposed PD.

Several styles of CMOS logic can be considered. One is classic complementary CMOS logic which is built from NMOS pull-down and PMOS pull-up logic networks. Simple gates, such as NAND/NOR can be realized very efficiently with only a few transistors and a few circuit nodes. Other gates, such as XOR and AND gates, require more complex circuit realizations.

Another choice is pass-transistor logic. Pass-transistor logic XOR gates are very simple and can operate at very high speeds. However, special care must be taken to circumvent the swing degradation problems which are of concern in pass-transistor logic.

Several styles of pass-transistor logic are available including Complementary Pass-transistor Logic (CPL), Swing Restored Pass-transistor Logic (SRPL), Double Pass-transistor Logic (DPL), and Single-Rail Pass-transistor logic (LEAP). We used DPL [3.19] as the structure for XOR gates because of its speed advantage in our simulation. In DPL style, both NMOS and PMOS logic networks are used in parallel. This provides full swing on the output signals (i.e., no level restoration circuitry is needed), and circuit robustness is therefore high.

The schematic is shown in Figure 3.23. All the signals in the PD are complementary. It is easier to have complementary signals to drive the charge pump and complementary signals are also helpful in minimizing the switching noise injected into the substrate.

Because that all the signals are complementary, the AND gates in Figure 3.13 are able to be implemented by standard static CMOS NAND gates and NOR gates as shown in Figure 3.24.

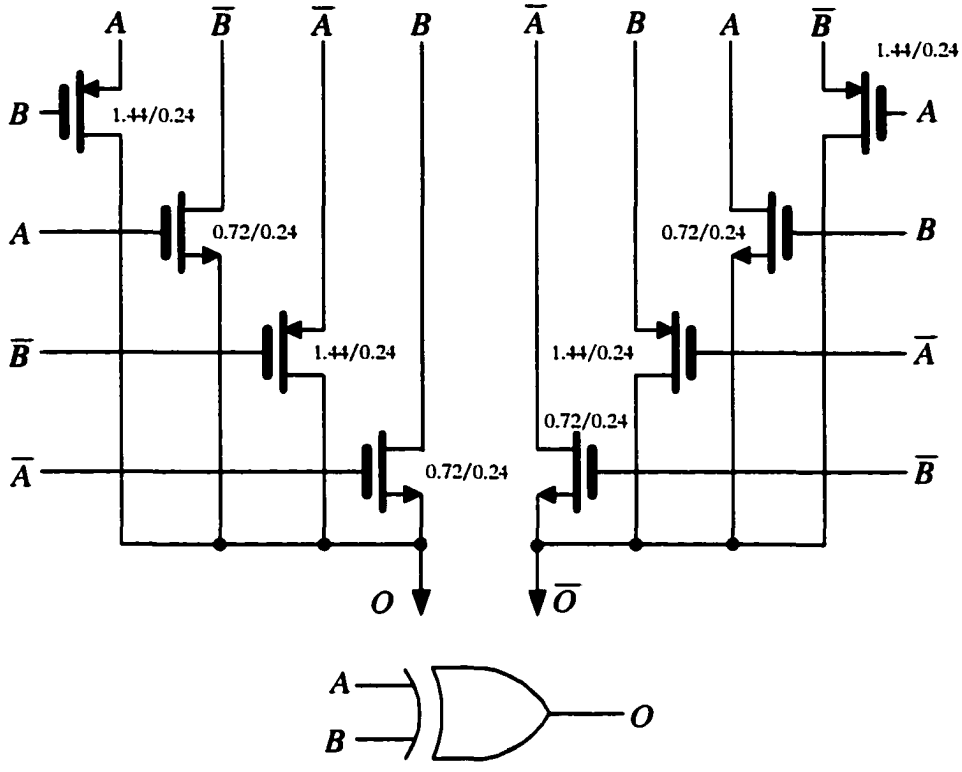


Figure 3.23 Schematic of the DPL-style XOR gate

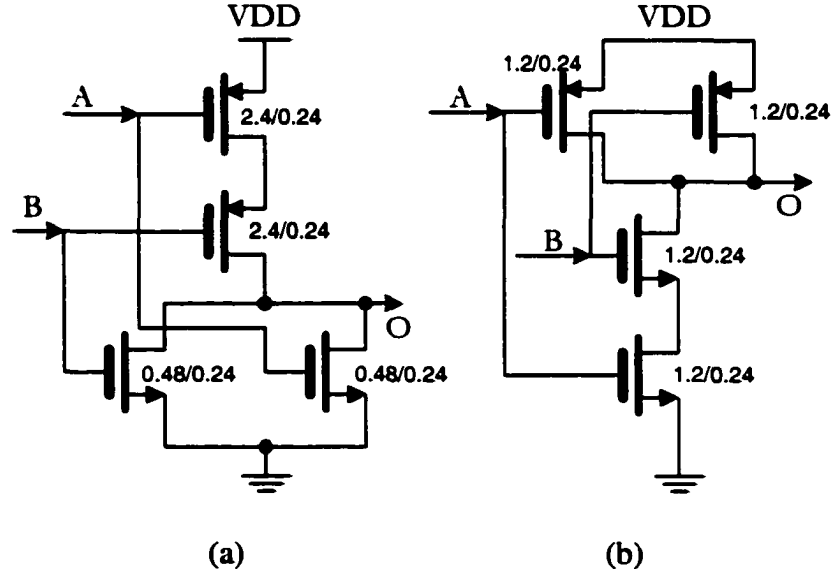


Figure 3.24 Schematics of (a) NOR gate; (b) NAND gate

The circuits were first designed and simulated in HP 0.35μm CMOS process. The clock frequency is 2GHz. All schematic and anticipated layout parasitics are included. Additional 10fF capacitors were added at each connection nodes to model the interconnection capacitance. The simulation covers the temperature range from 0°C to 100°C and all process corners. The input data stream is a 1GHz signal with 50% duty cycle which represents the NRZ data pattern "0101010 ....." (not random). Using HSPICE simulator and level 49 BSIM3 device models, the transfer characteristics of the proposed PD is shown in Figure 3.25. One of the three curves is the PD transfer characteristic at room temperature with the normal model. The other two are under extreme conditions; specifically 100°C at the slow process corner and 0°C at the fast process corner.

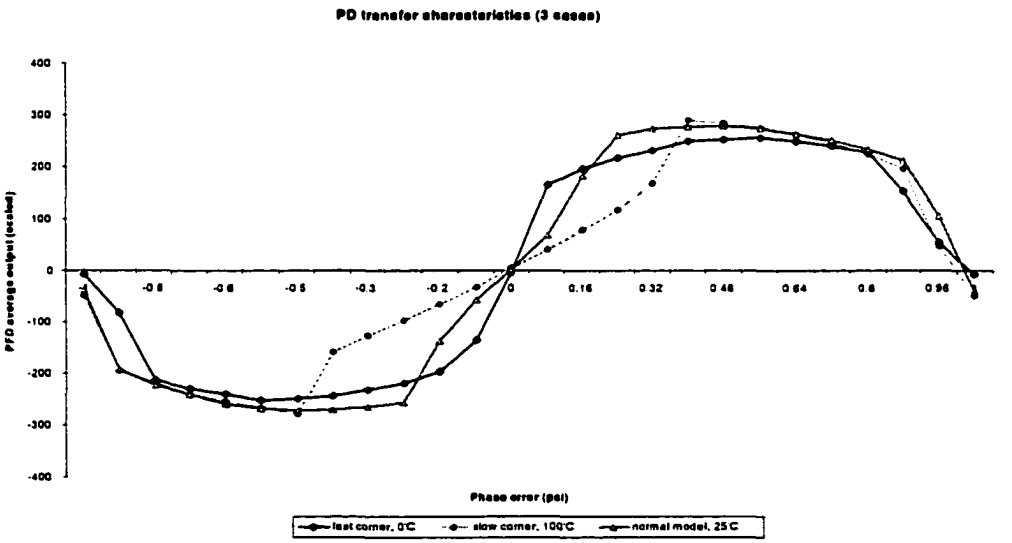


Figure 3.25 Simulated transfer characteristics of the phase detector

The PD operates correctly under all the conditions. From these simulations, it is apparent that very high gain is achieved around the zero phase error point. The "Dead zone" which is inherent in many phase detectors is absent.

The simulation results with random input data are shown in Figure 3.26. The performance of the PD with two patterns of NRZ data was tested. One pattern is series of

"1"s with one "0" ("11110"); another is series of "0"s with one "1" (00001). Results show that they all have zero output at zero degree phase shift and the PD gain is reasonably constant.

A snapshot of the output waveforms of the PD is shown in Figure 3.27. Both "Up" and "Down" signals are complementary. The pulse widths of them determined by distances between the cross points of the waveforms.

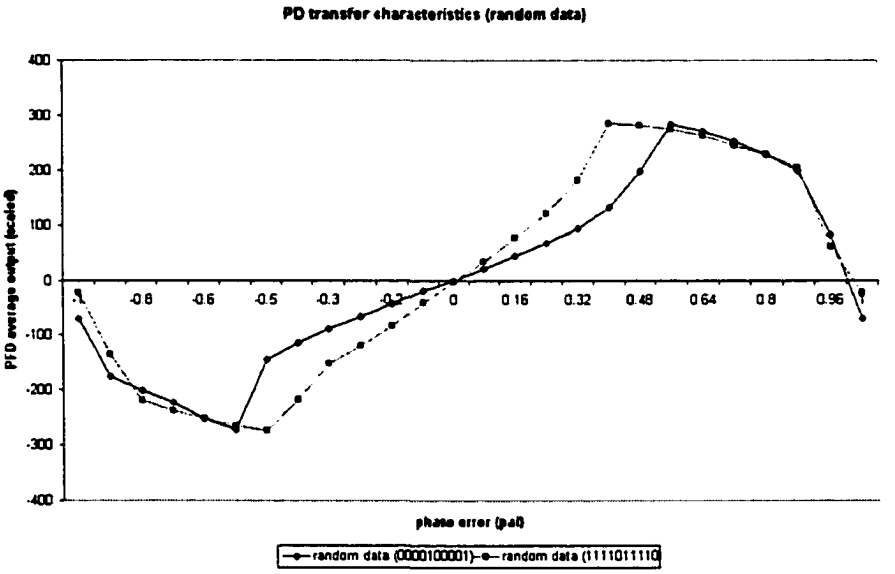


Figure 3.26 Simulated transfer characteristics of the phase detector for random data

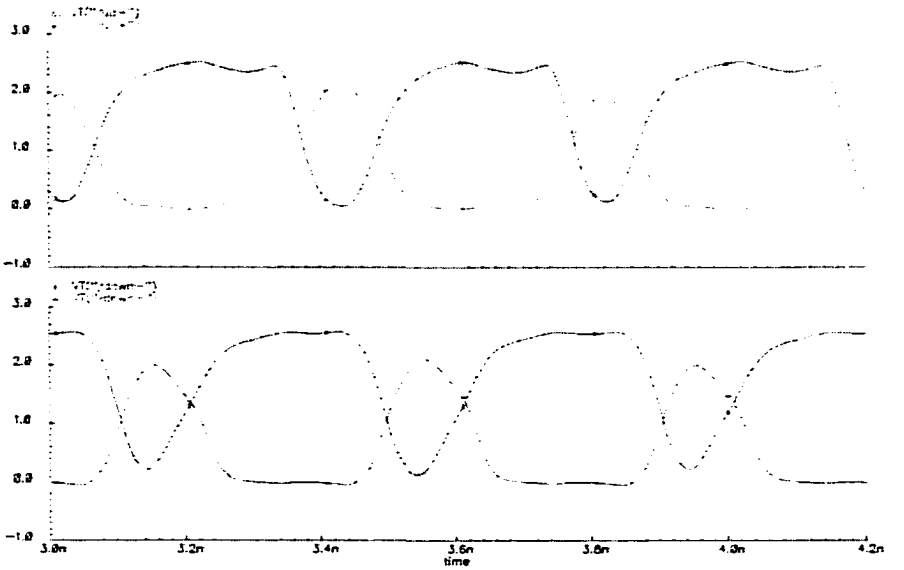


Figure 3.27 Output waveforms of the PD ("Up" and "Down" signals)

3.5.2 Charge Pump and Loop Filter

In order to achieve high resolution, we chose a current-switching type charge pump. The schematic of the charge pump, together with the loop filter, is shown in Figure 3.28. The "Up" and "Down" inputs are driven by the complementary outputs of the phase detector.

Several considerations of the charge pump design deserve mention:

- To maximize the speed of the charge pump, the bias currents should never be cut off during the operations
- Relatively large transistors are used to minimize the effects of mismatch
- Properly set the bias voltages so that the switching transistors would operate at active region instead of triode region when they are turned on

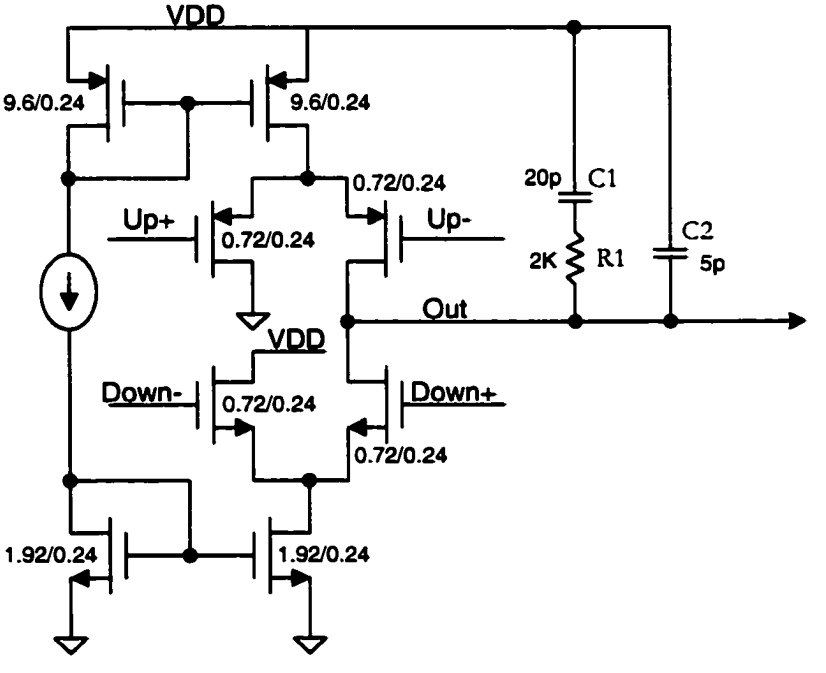


Figure 3.28 Schematic of the current steering charge pump and the loop filter

A simple passive second-order loop filter was used. It reduces the ripple which is inherently present in second-order loops at the control voltage node. The values of C1, C2 and R1 are carefully chosen in order to maintain an adequate phase margin in the third-order

loop and minimize the control voltage ripple. The whole PLL is a third-order loop. But its behavior can be approximated as a second-order loop if $C_1 >> C_2$. [3.17]

3.5.3 Control Voltage Generator and Voltage-Controlled Oscillator

The VCO is a 4-stage ring oscillator based structure. One of the delay stages in VCO and the control voltage generator are shown in Figure 3.29. Both of them are based on the topologies presented in [3.18].

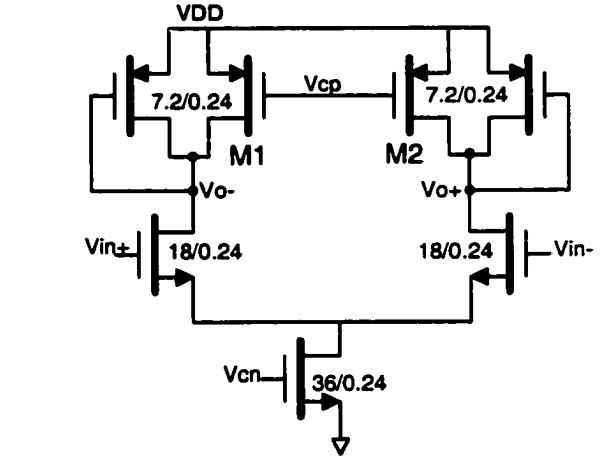
The delay stage shown in Figure 3.29(a) is a transitional fully-differential delay stage design except the symmetric resistive loads. The symmetric loads consist of a diode-connected PMOS device in shunt with an equally sized biased PMOS device. The effective resistance of the load elements changes almost linearly with the change of PMOS bias voltage Vcp [3.18]. Thus the delay time will change linearly with Vcp. It not only provides good control over delay time, but also leads to high dynamic supply noise rejection.

Two control voltages Vcp and Vcn are generated from Vcontrol by the control voltage generator which is shown in Figure 3.29(b). It consists of a replica circuit of the VCO delay stage and a single-stage operational amplifier. It establishes a current that is held constant and independent of power supply by adjusting Vcn so that Vcp is equal to Vcontrol, which greatly helps the power noise rejection performance of the VCO. The main function of this generator is to continuously adjust the bias currents for delay stages providing a tuning range wide enough to compensate for the temperature and process variations.

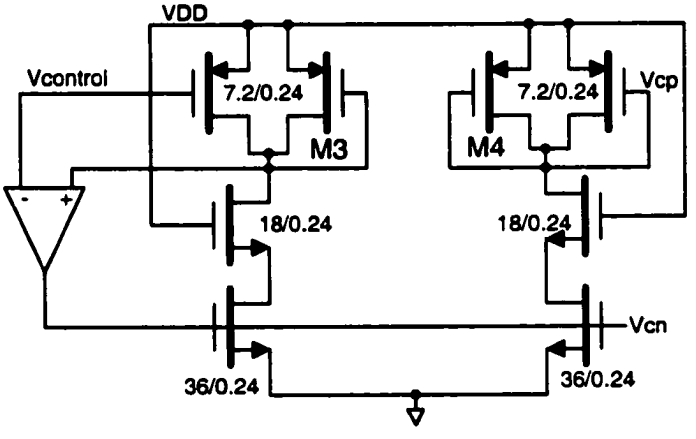
Shown in Figure 3.30 is the simulation waveform of the VCO differential outputs.

3.5.4 Other Circuits

Except the circuits that we have already discussed, there are other circuits implemented in the PLL.



(a)



(b)

Figure 3.29 Schematics of (a) delay stage; (b) control voltage generator

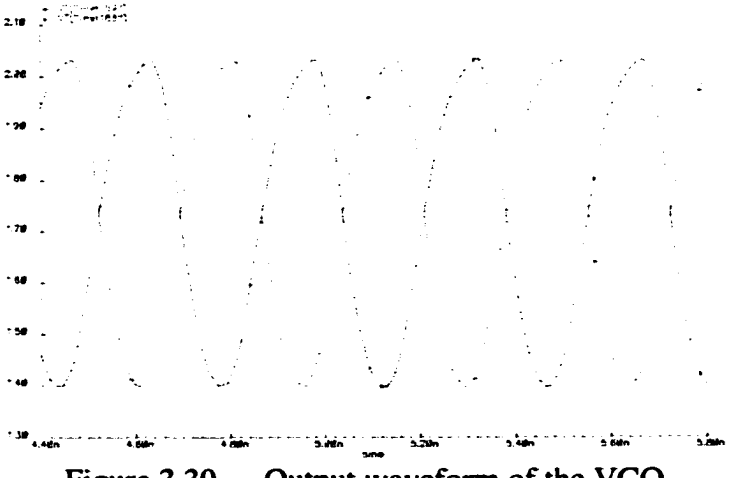


Figure 3.30 Output waveform of the VCO

Preset circuit, which is used to preset the initial control voltage, is shown in Figure 3.31(a). The input inverter is used to isolate the "Preset" control signal. M1 and M2 are sized so that the output is higher than 1.5V when "Preset" signal is active.

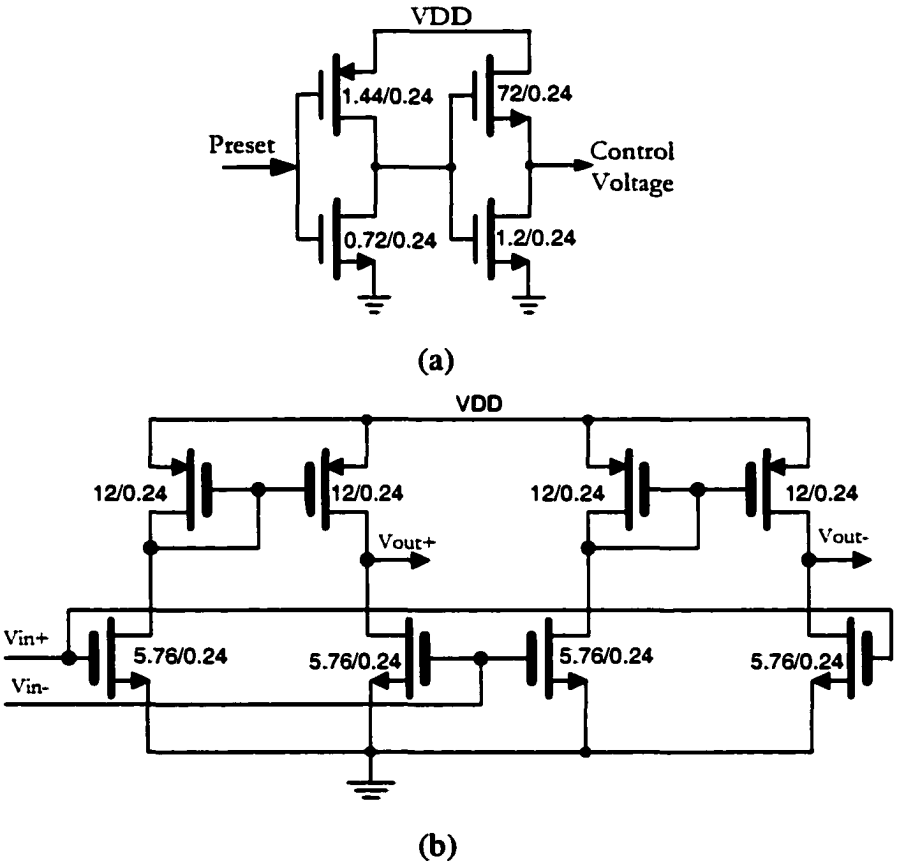


Figure 3.31 Schematics of (a) preset circuit; (b) output buffer

Because the speed of this PLL is very high, we designed frequency divider circuits to lower the frequency of the clock in order to have an easier measurement in case the high frequency clock may not be good enough for testing. The frequency divider is a DFF based divider very similar to the one shown in Figure 2.11 except we have 5 DFFs in the PLL. So the dividing ratio is 32:1. The structure of the DFF is the same as the one shown in Figure 2.12.

Output buffers are designed to drive the load capacitance. It is for testing purposes. The schematic of one stage is shown in Figure 3.31(b). It consists of two opposite rail-to-rail NMOS differential amplifiers. Several stages of such circuits were used in series for buffering the clock output and the divided clock output.

3.6 Simulation Results

Using HSPICE simulator and TSMC 0.25μm CMOS process device models, the PLL successfully locks to a pseudo-random input data with a data rate of 2.5Gbit/s under normal and extreme cases.

All the schematic parasitics were included in the simulation. Input data was distorted by passing it through a cascaded string of inverters before going into the PLL. Initial conditions were set to the control voltage.

Figure 3.32 shows the locking characteristics of the control voltages under 3 situations. The simulations show the locks were successful acquired.

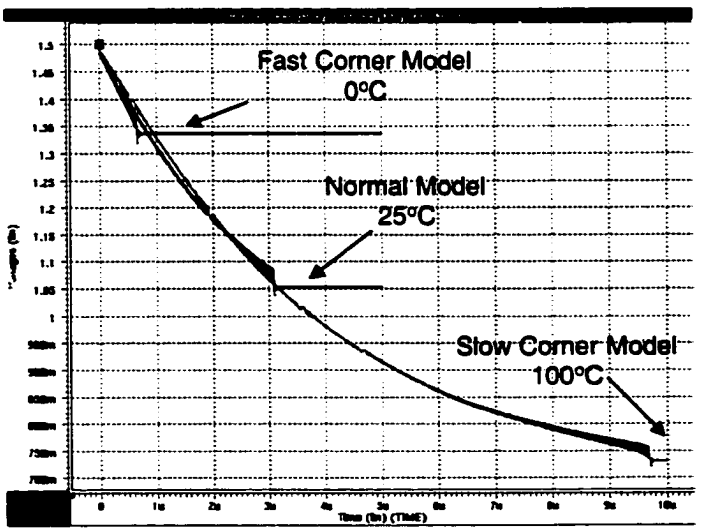


Figure 3.32 Simulated acquisition processes of the phase-locked loop

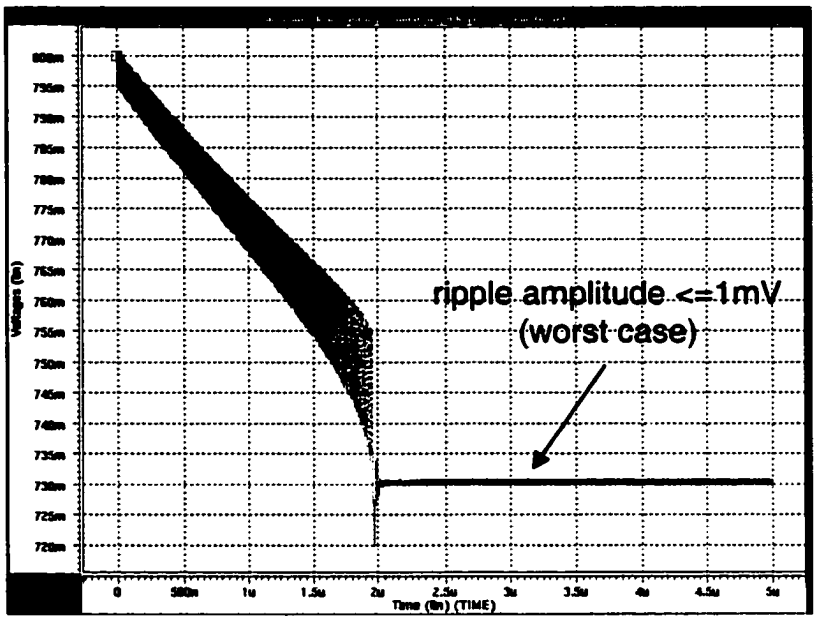


Figure 3.33 Details of the control voltage ripple when PLL acquires lock

Figure 3.33 shows the locking transient with the worst ripple amplitude, which is the case of slow corner model at 100°C. We can see the ripple amplitude is still very small (<=1mV). Though this simulation didn't consider any noise sources such as those associated with the power supply, ground or any associated digital circuits, it still demonstrated the low noise level from the system design view. In the real world, such noises would be minimized by using techniques such as fully-differential circuits.

Additional simulation results showed the locking range of the PLL is 1.5GHz-2.7GHz. The lower bound was obtained at fast model corner at 0°C and the upper bound occurred at the slow model corner at 100°C. The power dissipation is about 40mW (nominal) under 2.5V power supply for PLL core which was quite low.

3.7 Chip Layout

It is extremely important to have a well-considered layout, especially for high speed circuits. The floor-plan for this PLL is shown in Figure 3.34. PD is separated from VCO and

the output buffers as far as possible to the corners of the layout. For sensitive analog circuits such as PD, charge pump and loop filter, they all are surrounded by double guard rings in order to isolate the substrate noise. Furthermore, the metal line for control voltage is shielded by two parallel metal lines connected to VDD at each side. Also, a massive area of substrate contacts are put between the output buffers the analog circuits.

The capacitors in the loop filter are implemented by NMOS gate capacitor. Resistor is implemented by available high-resistance N-diffusion ploy resistor which has the lowest variation and temperature coefficient in this process.

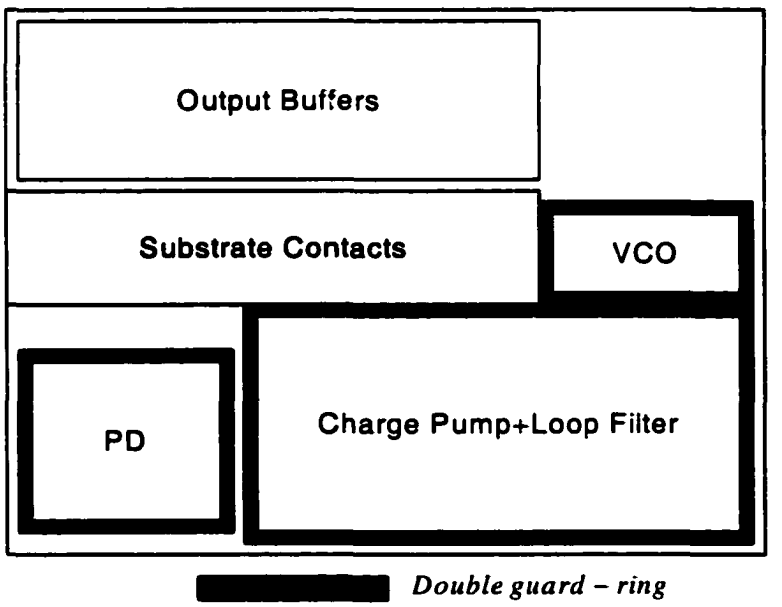


Figure 3.34 Floor planning of the phase-locked loop layout

The supply nets for PLL and output buffers are separated to minimize the noise injection. On-chip power supply decoupling was implemented by NMOS gate capacitors. A very simple low-ESD pad frame is used to lower the parasitics and achieve high speed. On-chip 50Ω terminations were also implemented.

3.8 Experimental results

The prototype was fabricated in TSMC 0.25μm CMOS process. The chip micrograph is shown in Figure 3.35. On the left upper corner is the PLL. The package chosen for the PLL is LCC52 with gold lead. Critical signals for the PLL were connected through the package with the shortest electrical path to minimize the inductance which is the biggest concern in high speed circuits.

A printed-circuit board with four copper layers was designed for testing PLL. The upper layer is used for routing high speed signals. Special care was taken to make sure there are no parallel signal traces. The width of the copper line was calculated and designed for the data rate we were shooting for. The second layer is ground layer. The third layer is used for power supply. The bottom layer is used for routing low speed signals.

The board under testing is shown in Figure 3.36(a). A LCC52 socket with gold lead was used to connect the PLL chip to the board. At the initial testing, we found that the on-chip termination resistors were far larger than 50Ω. We decided to use off-chip terminations. The off-chip terminations and the decoupling capacitors are shown in Figure 3.36(b) which is the bottom side of the board.

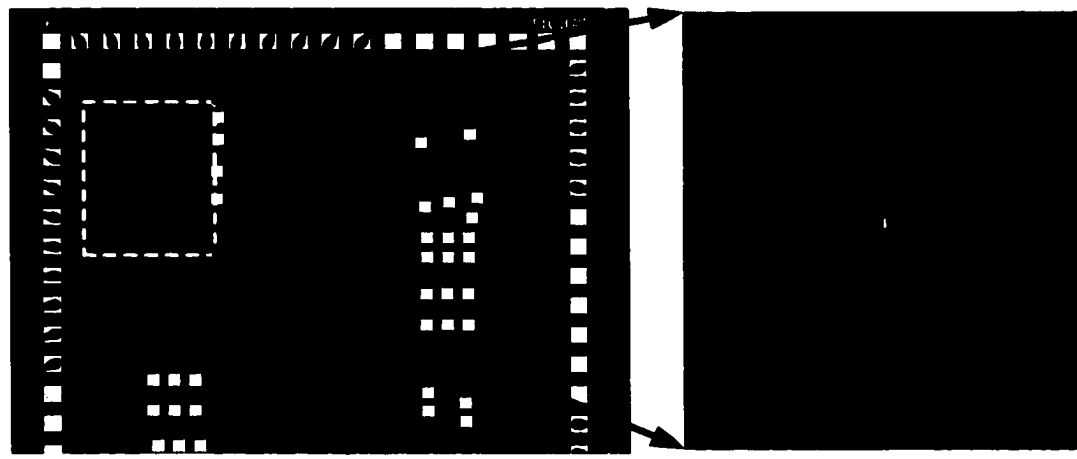
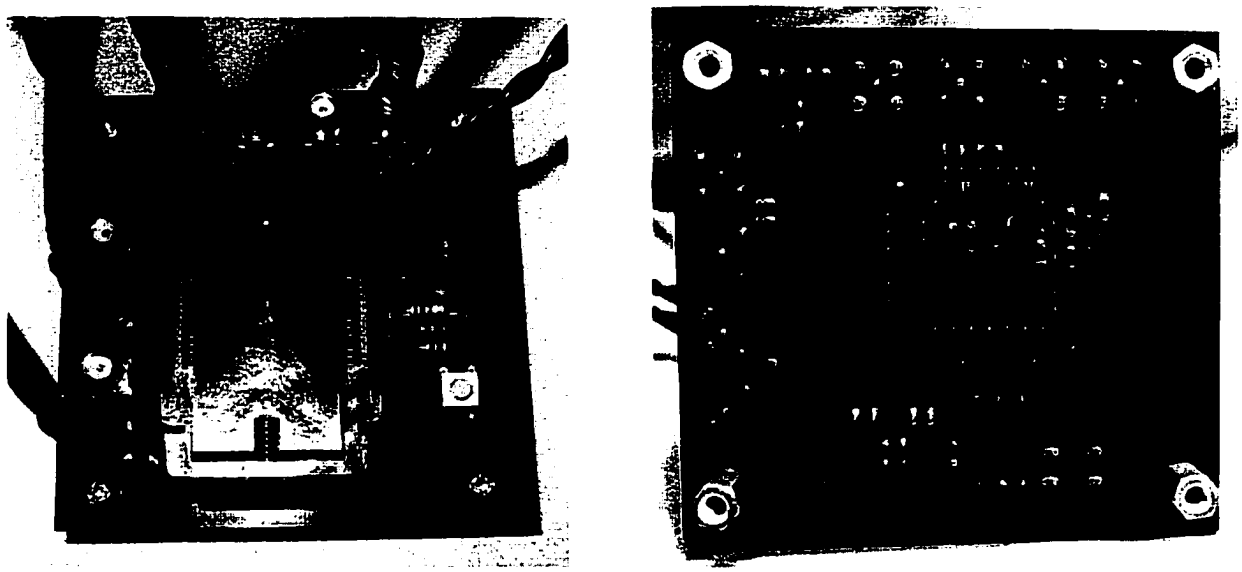


Figure 3.35 Chip micrograph



(a) (b)

Figure 3.36 Photos of the (a) prototype under testing; (b) off-chip termination and decoupling (bottom side)

During the testing, we found that the output clock was not stable enough to measure the frequency. At the position where the frequency-divided signal has transitions, there are spikes appeared on the output clock. The reason for this phenomenon is that the output buffer for frequency-divided signal is too strong and whenever there is transition, a large amount of current is flowing through the power supply causing a large power supply noise that affects the output clock. Furthermore, the large switching current also injects noise to the substrate causing the same problem on the output clock.

Fortunately, the frequency-divided clock is stable and its frequency can be reliable be measured. To verify the locking condition of the PLL, we used the frequency-divided clock to observe if it keeps tracking the changing frequency of the input reference.

The PLL successfully locks to both 1.05GHz clock (equivalent to 2.1Gb/s data sequence with "010101010...." Pattern) and 2.1Gbit/s $2^{23}-1$ PRBS data sequence. The performance of the PLL is summarized in Table 3.2.

Table 3.2 Summary of the experimental results

Power supply	2.3V
Locking range	1.88GHz-2.1GHz
Tracking range	1.65GHz-2.3GHz
Power consumption	Total 54mA, PLL core ~10mA*
Technology	TSMC 0.25μm CMOS
Active area	$400\mu m \times 290\mu m$
Package	LLC52

* Due to the limitation of the design, the current for PLL core can not be directly measured. It is estimated by using the current ratio between PLL core and the output buffer in the simulation.

Some waveform captures are shown in Figure 3.37. The upper waveform is the input clock (the frequency shown is the half of the data rate). The lower waveform is the divided output clock. The three captures show the tracking properties of the PLL at input data rate at 1.8Gbit/s, 2Gbit/s and 2.2Gbit/s. The frequency of the divided clock tracks the input successfully.

3.9 Conclusion

A new linear non-sequential PD structure that is capable of operating at very high speeds was introduced. Using this PD, a 2.3V CMOS PLL for data and clock recovery was implemented. Experimental results show that it can operate at the 2.1Gbit/s data rate in TSMC 0.25μm CMOS processes. It successfully verifies the functionality and the speed capability of the PLL. It is among the fasted full-rate PDs for data recovery.

Secondly, a half-rate PD based on the idea of this PD may be viable. It would further improve the speed of the PD (up to two times higher).

References:

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

[3.2] D. H. Wolaver, "Phase-Locked Loop Circuit Design", Prentice Hall Inc. 1991

[3.3] S.-O. Jeon, T.-S. Cheung, W.-Y. Choi, "Phase/frequency Detectors for High-speed PLL Applications", IEE Electronics Letters, Vol. 34, pp. 2120-2121, October 1998

[3.4] C. R. Hogge, "A Self-Correcting Clock Recovery Circuit," IEEE Journal of Lightwave Technology, vol. LT-3, pp. 1312-1314, December 1985

[3.5] Y. M. Greshishchev, P. Schvan, "SiGe Clock and Data Recovery IC with Linear-Type PLL for 10Gb/s SONET Application", IEEE Journal of Solid-State Circuits, Vol. 35, NO. 9, pp. 1353-1359, September 2000

[3.6] A. Hati, M. Ghosh, B. C. Sarkar, "Phase Detector for Data-Clock Recovery Circuit", IEE Electronics Letters, Vol. 38, pp. 161-163, February 2002

[3.7] M. Ramezani, C. A. T. Salama, "An Improved Bang-Bang Phase Detector for Clock and Data Recovery Applications", IEEE International Symposium of Circuits and Systems, Vol. 1, pp. 715-718, 2001

[3.8] M. Rau, T. Oberst, R. Lares, A. Rothermel, R. Schweer, N. Menoux, "Clock/Data Recovery PLL Using Half-Frequency Clock", IEEE Journal of Solid-State Circuits, Vol. 32, NO. 7, pp. 1156-1159, July 1997

[3.9] M. Wurzer, J. Bock, H. Knapp, W. Zirwas, F. Schumann, A. Felder, "A 40-Gb/s Integrated Clock and Data Recovery Circuit in a 50-GHz fT Silicon Bipolar Technology", IEEE Journal of Solid-State Circuits, Vol. 34, NO. 9, pp. 1320-1324, September 1999

[3.10] B. Lai, R. C. Walker, "A Monolithic 622Mb/s Clock Extraction Data Retiming Circuit", Digest of Technical Papers of IEEE International Solid-State Circuits Conference, pp. 144-145, 306, 1991

[3.11] P. Larsson, "An Offset-Cancelled CMOS Clock-Recovery/Demux with a Half-rate Linear Phase Detector for 2.5Gb/s Optical Communication", Digest of Technical Papers of IEEE International Solid-State Circuits Conference, pp. 74-75, 434, 2001

[3.12] X. Maillard, F. Devisch, M. Kuijk, "A 900-Mb/s CMOS Data Recovery DLL Using Half-Frequency Clock", IEEE Journal of Solid-State Circuits, Vol. 37, NO. 6, pp. 711-715, June 2002

[3.13] J. Savoj, B. Razavi, "A 10-Gb/s CMOS Clock and Data Recovery Circuit with a Half-Rate Linear Phase Detector", IEEE Journal of Solid-State Circuits, Vol. 36, NO. 5, pp. 761-767, May 2001

[3.14] S. B. Anand, B. Razavi, "A CMOS Clock Recovery Circuit for 2.5Gb/s NRZ Data", IEEE Journal of Solid-State Circuits, Vol. 36, NO. 3, pp. 432-439, March 2001

[3.15] J.D.H. Alexander, "Clock Recovery From Random Binary Signals", IEE Electronic Letters, vol. 11, pp. 541-542, October 1975

[3.16] Y. Tang, R. L. Geiger, "A Non-sequential Phase Detector for PLL-based High-Speed Data/Clock Recovery", Proceedings of Midwest Symposium of Circuits and Systems, 2000

[3.17] F. M. Gardner, "Charge-Pump Phase-Lock Loops," IEEE Transaction of Communications, Vol. COM-28, pp. 1849-1858, November 1980

[3.18] J. G. Maneatis, "Low-Jitter Process-Independent DLL and PLL Based on Self-Biased Techniques", IEEE Journal of Solid-State Circuits, Vol. 31, pp. 1723-1732, November 1996

[3.19] R. Zimmermann, W. Fichtner, "Low-Power Logic Styles: CMOS Versus Pass-Transistor Logic," IEEE Journal of Solid-State Circuits, Vol. 32 NO. 7, pp. 1079-1090, July 1997

[3.20] R. Rogenmoser, N. Felber, Q. Huang, W. Fichtner, "A 1.16GHz dual modulus 1.2μm CMOS prescaler", Proceeding of IEEE 1993 Custom Integrated Circuits Conference, pp.27.6.1- 4, May 1993

CHAPTER 4

TRANSIENT BIT ERROR RATE ANALYSIS OF DATA RECOVERY SYSTEMS USING JITTER MODELS

4.1 Introduction

We introduced a high speed PLL design for data recovery in chapter 3. In this chapter, we will present a method of evaluating the performance of the data recovery system by analyzing the transient bit error rate (BER) of the recovered data.

The performance of the data recovery system is usually characterized by the BER of the recovered data. The BER of these links is dominantly determined by the characteristics of the data recovery system. High-speed low-power phase-locked loops are an integral part of data/clock recovery system. Although the performance of the PLL after it is in lock is reasonably well understood, its performance during lock acquisition has received minimal attention in the literature but is also of concern since this determines how long it will take for a PLL to attain an acceptable BER.

The BER is determined by the jitter of the incoming data and the jitter performance of the PLL. In this chapter, we develop the relationship between the BER of the recovered data and the jitter of the incoming data both when the PLL is in lock and when the PLL is acquiring lock.

Research on PLLs has been ongoing for decades and the term "lock" is widely used to indicate the PLL is in a special "steady state" mode of operation. But until now, a rigorous definition of "lock" has not been presented in the literature. It is generally assumed that a PLL is in lock when the output of the loop filter stabilizes and that one just "knows" when the PLL is in lock but, in reality, the control voltage for the VCO comes from a loop filter that generally has an infinite impulse response and, as such, only asymptotically approaches

a steady-state value or a steady-state average value. In this work, a practical criterion for determining if the PLL is in "lock" will be developed and this "lock" condition will be contingent upon establishing a given BER level of performance.

To determine if a data recovery system is working correctly, we usually establish a maximum acceptable value for the BER, denoted in this work as B_{acc}. If we assume that the frequency of the incoming data does not change for time $t > t_0$, then if the BER of the recovered data satisfies the relationship $BER < B_{acc}$ for time all $t > t_1$, where $t_1 > t_o$, then we say the PLL is in "lock" for $t > t_1$. If t_1 is the minimum value of t for which the BER satisfies the inequality $BER(t_1) \leq B_{acc}$, then we say the PLL acquires lock at time t_1.
In the following sections, we will analysis the BER of the recovered data based on the acquisition behavior of the PLL and the jitter model.

4.2 Acquisition Behavior of the Phase-Locked Loop

The acquisition behavior of the PLL can be studied most conveniently by considering the response of the loop to an initial phase error or a frequency error. Consider the common second-order PLL shown in Figure 4.1.

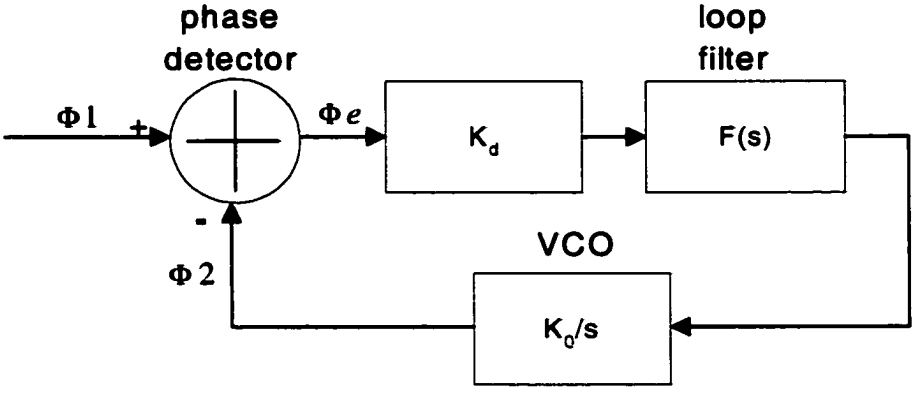


Figure 4.1 Model for 2nd order phase-locked loop

Assume that the phase detector is linear. The transfer function of the second-order PLL is given by [4.1] [4.2] [4.3]

$$H(s)=\frac{\Phi_2(s)}{\Phi_1(s)}=\frac{K_0K_dF(s)}{s+K_0K_dF(s)} \tag{4.1}$$

The phase error transfer function is given by

$$H(s)=\frac{\Phi_e(s)}{\Phi_1(s)}=\frac{s}{s+K_0K_dF(s)} \tag{4.2}$$

where K_d is the phase detector gain, K_o is the VCO gain and the $F(s)$ is the transfer function of the loop filter.

The acquisition process of the PLL is classified into two distinct types, lock-in process and pull-in process.

Assuming initially the PLL is in lock, the lock-in process is the re-acquisition process during which the output of the phase detector will only sweep once within its output range before the PLL returns to lock. Pull-in is the re-acquisition process during which the output of the phase detector will sweep within its output range more than one time before the PLL returns to lock. The pull-in process is more complicated and takes much longer time than the lock-in process and it is a highly nonlinear process. The typical control voltage response of the lock-in process and the pull-in process are illustrated in Figure 4.2.

4.3 Jitter and its Model

Jitter is the deviation from the ideal timing of an event. It is composed of both deterministic and random (Gaussian) components. [4.4]

The deterministic jitter is the jitter with a non-Gaussian probability density function. It is always bounded in amplitude and has specific causes. Deterministic jitter is characterized by its bounded, peak-to-peak value.

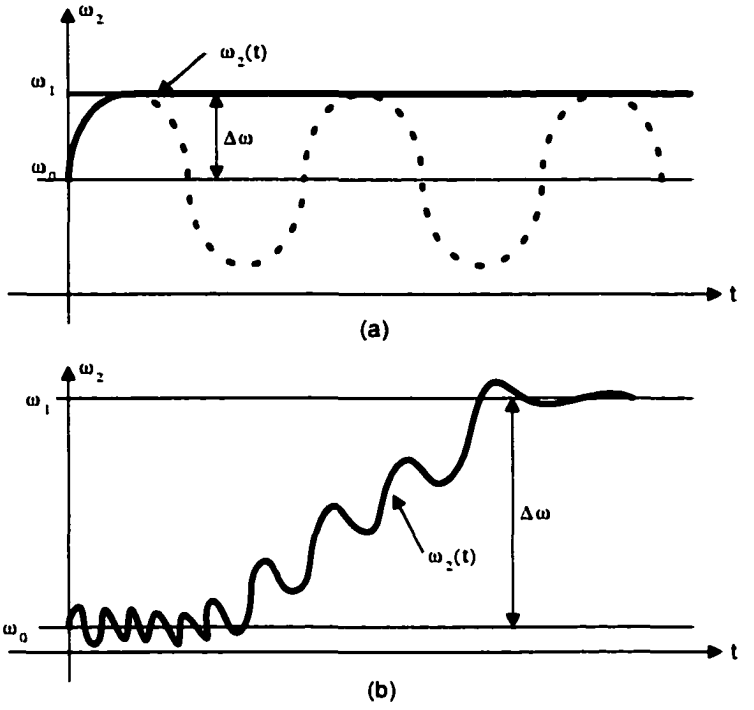


Figure 4.2 Acquisition processes of (a) lock-in process; (b) pull-in process

Random jitter is the jitter that is characterized by a Gaussian distribution. It is defined to be the peak-to-peak value which is given to be 14 times the standard deviation of the Gaussian distribution for a BER of 10^{-12}.

In the following, we will define the jitter models which will generate plots of eye closure vs. BER with various amount of random and deterministic jitter components.

The error probability is defined as

$$P(Q)=\frac{1}{2}\left(1-erf\left(\frac{Q}{\sqrt{2}}\right)\right) \tag{4.3}$$

where $erf(\ )$ is the error function which is given by

$$erf(x)=\frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt \tag{4.4}$$

and Q, the average signal to noise ratio, is defined as

$$Q = \frac{1}{2}\left(\frac{V}{\sigma}\right) \tag{4.5}$$

where V is the peak to peak signal amplitude and σ is the root-mean-square noise. To arrive at this expression, it is assumed that the noise has a Gaussian probability density function with zero mean.

4.3.1 Effects of Random Jitter

Let QT_1 be the ratio of the eye opening to the amount of random jitter at an eye crossing, i.e.

$$QT_1(T_0, t, \sigma) = \frac{1}{2}\left(\frac{T_0}{\sigma}\right) \tag{4.6}$$

To include the effect of sampling time, QT_1 can be rewritten as

$$QT_1(T_0, t, \sigma) = \frac{1}{2}\left(\frac{T_0 + |t|}{\sigma}\right) \tag{4.7}$$

where t is a dummy variable that defines the offset of the sampling instant from the eye crossing. When t=0, a worst-case BER is obtained, i.e. t=0 defines the position of the eye crossing.

If the decision threshold is made at the eye crossing, then the eye opening is essentially, zero, i.e. $T_0 = 0$.

Following the analysis in the signal domain, the BER in the time domain is defined as

$$PT_1(T_0, t, \sigma) = \frac{1}{2}\left(1 - erf\left(\frac{QT_1(T_0, t, \sigma)}{\sqrt{2}}\right)\right) \tag{4.8}$$

In order to study the eye closure, let us define the position of the second eye crossing. The second eye crossing would have similar characteristics as the first one and occurs a bit period away, i.e.,

$$QT_2(T_0, t, \sigma) = QT_1(T_0, t - |T|, \sigma) \tag{4.9}$$

Similarly,

$$PT_2(T_0,t,\sigma)=\frac{1}{2}\left(1-erf\left(\frac{QT_2(T_0,t,\sigma)}{\sqrt{2}}\right)\right) \tag{4.10}$$

The BER now is given by

$$P(T_0,t,\sigma)=PT_1(T_0,t,\sigma)+PT_2(T_0,t,\sigma) \tag{4.11}$$

4.3.2 Effects of Deterministic Jitter

Deterministic jitter (DJ) is caused by varying patterns or duty cycle creating predominant spectral components or DC baseline drift in the transmitted signal. DJ reduces the eye width and can be assumed to have larger amplitude than random jitter. To account for DJ, both QT_1 and QT_2 can be written as

$$QT_1(T_0,t,\sigma,DJ)=\frac{(T_0-DJ)+|t|}{2\sigma} \tag{4.12}$$

Similarly,

$$QT_2(T_0,t,\sigma,DJ)=QT_1(T_0,t-|T|,\sigma,DJ) \tag{4.13}$$

The BER is now

$$PT_1(T_0,t,\sigma,DJ)=\frac{1}{2}\left(1-erf\left(\frac{QT_1(T_0,t,\sigma,DJ)}{\sqrt{2}}\right)\right) \tag{4.14}$$

and

$$PT_2(T_0,t,\sigma,DJ)=\frac{1}{2}\left(1-erf\left(\frac{QT_2(T_0,t,\sigma,DJ)}{\sqrt{2}}\right)\right) \tag{4.15}$$

The total probability over the window of interest is therefore

$$P(T_0,t,\sigma,DJ)=PT_1(T_0,t,\sigma,DJ)+PT_2(T_0,t,\sigma,DJ) \tag{4.16}$$

4.3.3 Total Jitter Model

A complete jitter model due to the total jitter can be obtained by combining the random jitter model and the deterministic jitter model together.

$$BER(RJ,DJ)=P(T_0,t,\sigma)+P(T_0,t,\sigma,DJ) \tag{4.17}$$

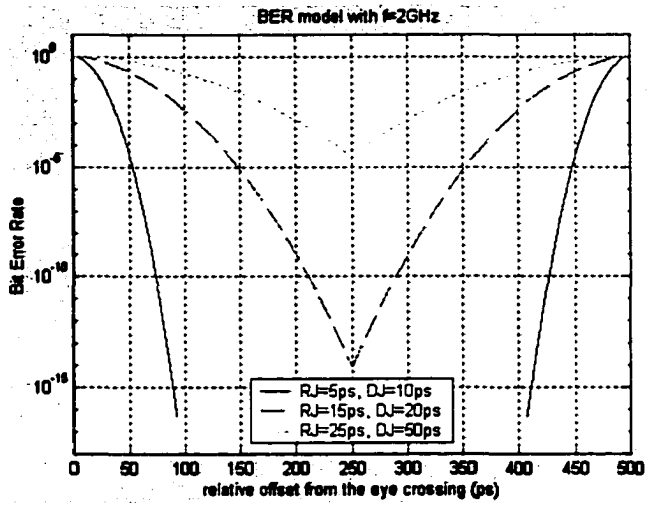


Figure 4.4 BER with different RJ and DJ combinations

The effects of different combination of the random jitter and deterministic jitter to BER are shown in Figure 4.4.

4.4 BER Analysis of the Data Recovery System

Once we have obtained the jitter models and the acquisition behavior of the PLL. We can use them to calculate the transient BER during acquisition or even at anytime.

As before, we can get an optimal BER when the clock samples the data at the middle between the two eye crossings. This is also the principle for the decision-making circuits. In the decision-making circuits, the incoming data should be re-sampled by the recovered clock with sampling edges at the middle of a bit period.

We assume that initially the PLL is in lock with reference signal with frequency ω_0. At t=0, there is a frequency step $\Delta\omega$ applied to the reference signal. After the step, the angular frequency of the reference becomes $\omega_1(t)=\omega_0+\Delta\omega\, u(t)$; the phase of the reference signal $\phi_1(t)$ is the integral over the frequency variation $\Delta\omega$. So that $\phi_1(t)=\Delta\omega t$.

From the transient response of the VCO, specifically from the output $\omega_2(t)$, we can get the phase of the VCO output $\phi_2(t)$ which is given by

1. Phase detector is a sinusoid phase detector,

$$U_d = K_0 \sin(\Delta\phi),\ K_0 = 3 \tag{4.20}$$

2. Loop filter is a passive loop filter

3. Initial PLL locking frequency = 2GHz, frequency step=20MHz

4. VCO gain $K_0 = 3.5 \times 10^8\ radians/V$

Using above parameters for the PLL and combining with the jitter model with deterministic jitter=20pS and random jitter=15pS, the following simulation results were obtained.

Figure 4.6(a) shows the transient response of the phase detector and loop filter outputs. The output of the phase detector swept the output range many times until it merged with the output of the loop filter. It's a pull-in process because of the large frequency step.

Figure 4.6(b) shows the corresponding BER during the acquisition. At the early stages of the acquisition, the BER changes dramatically with a large range. During this time, the BER is unacceptably large. When the PLL approaches lock, the BER drops steeply. Figure 4.6(c) shows the magnified BER response. From this figure, we see that after about t=6.242 μS, the BER dropped below 1×10^{-12} .

This example does not include jitter of the PLL because the transient analysis of the acquisition process is ideal. However, it can provide an easy and quick method to approximately evaluate a data recovery system.

This example shows an application of this method on early (behavioral level design) design stage. It is also applicable for after-design verification. When the data recovery system design is completed, combining the transient simulation results and the jitter models, we can get results that are very close to the real world.

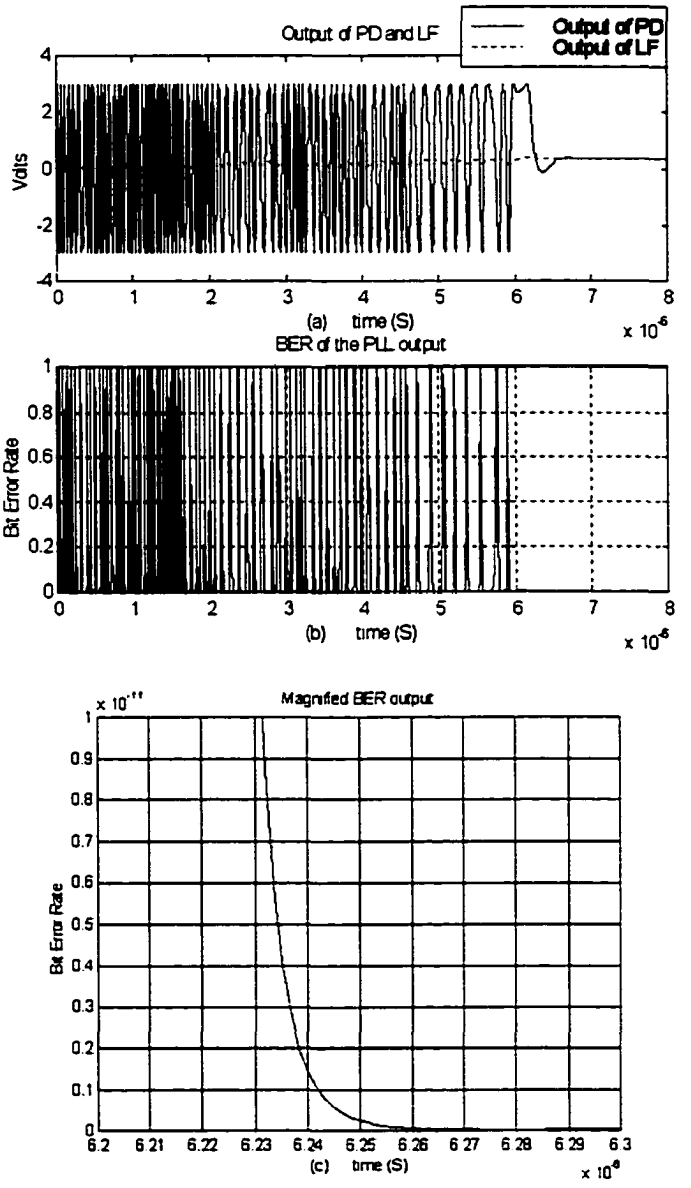


Figure 4.6 Simulated results of (a) transient response of the PLL; (b) transient BER; (c) magnified transient BER

4.6 Conclusion

For a data recovery system, the BER can be calculated by using the jitter model and the transient response of the PLL. This makes it possible to predict when the data recovery system will enter lock.

References:

[4.1] Roland E. Best, "Phase-locked loops: design, simulation, and applications", 3rd edition, McGraw Hill, 1997

[4.2] Dan H. Wolaver, "Phase-locked loop circuit design", Prentice Hall, 1991.

[4.3] Floyd Martin Gardner, "Phaselock Techniques", 2nd edition, John Wiley & Sons, Inc. 1979

[4.4] "Information Technology-Fiber Channel - Methodologies for Jitter Specification", Working Draft, T11.2 / Project 1230/ Rev 2.0, February 1998

CHAPTER 5

A HIGH PRECISION HIGHLY LINEAR VARIABLE GAIN AMPLIFIER

5.1 Background

Variable gain amplifier (VGA) finds a very wide range of applications where Automatic Gain Control (AGC) is needed, such as hearing aids, imaging and wireless communications. In such applications, the signal strength varies over a large range. VGA is used for either controlling the transmission signal power or adjusting the received signal amplitude. In order to let system work under such situations, a feedback loop is usually required to implement AGC. VGA plays a key part in this loop. Usually a highly linear VGA is needed to maintain good system linearity. The linearity of the VGA is almost entirely determined by its amplifier design. A highly linear amplifier design is crucial for the linearity of the VGA.

There are two types of VGAs. One is a discrete gain-step type with a digital control signal, and the other is a continuously variable gain type which is controlled by an analog gain-control signal.

Sophisticated analog design usually realized using expensive BiCMOS, SiGe processes. Large-scale integration of a mixed-signal system or SoC in deep submicron process can only be achieved when analog circuits are also implemented with ultra-short channel devices in CMOS. This project is trying to implement a digitally controlled high precision highly-linear CMOS variable gain amplifier with the performance that was only achievable in more expensive processes before. The projected process is 0.25μm standard CMOS process. The power supply voltage is 3.3V.

Some design challenges and specifications in this design are:

1. Precise gain step of 6.02dB
2. Gain range: -6dB - +36dB
3. Maintain enough bandwidth (>250MHz)
4. The third harmonic distortion for input signal at 160MHz should be no worse than $-55dB@V_{opp}=1V$.
5. Less than $2.5nV/\sqrt{Hz}$ input-referred thermal noise at maximum gain
6. Better than 12dB noise figure at the maximum gain assuming 200Ω source impedance.
7. Fixed differential input impedance of 200Ω and differential output impedance of 600Ω.

The toughest specification to achieve is the linearity requirement. Our linearity performance target in simulation is $-60dB@V_{opp}=1V$ which is higher than the requirement. This is because the measurement results of linearity usually would be worse than simulation because of the mismatch and the process variation. The linearity performance of the VGA is almost entirely determined by the linearity performance of the amplifier section of the VGA.

5.2 Structure of the VGA

The first natural choice for designing highly linear amplifier is negative feedback amplifier configuration. Extensive investigation has been done to evaluate the linearity and bandwidth performance of the negative feedback amplifier configurations. The idea is to build a high open loop DC gain amplifier and connect it as feedback configurations. The benefits are precise gain control and better linearity. The inherent problem would be that the limited speed in CMOS compared to BiCMOS or bipolar technologies may make this approach not viable. After extensive investigation, unfortunately, the projected CMOS process doesn't have the luxury of extra bandwidth to play with negative feedback. So our focus shifted to the low gain high bandwidth open loop amplifier structure.

Because of the tight specification on gain step accuracy, traditional transconductance adjusting approach can not meet the requirement even with the help of tuning circuits. The final structure of the VGA as shown in Figure 5.1 is an R-2R ladder plus fixed-gain amplifier structure.

The inputs are AC coupled to an R-2R ladder which is controlled by digital control inputs (MSB, ISB and LSB) to attenuate the input. The output from the ladder is then feed into the fixed-gain open loop amplifier. If the resistors in R-2R ladder can be well laid out to minimize the mismatch, the gain step of the VGA is totally determined by the ladder which can be very accurate.

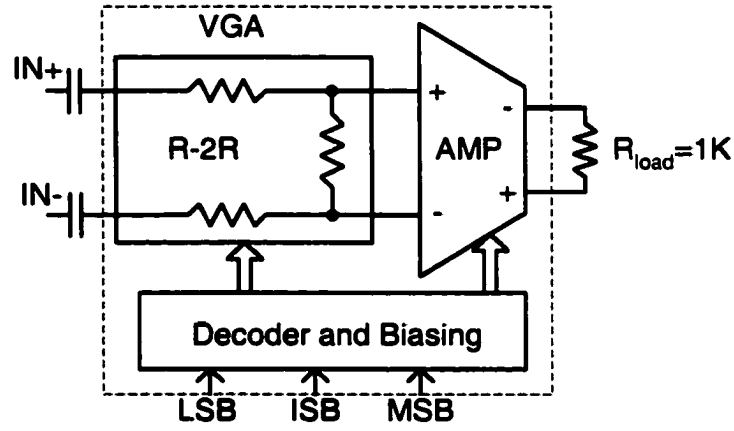


Figure 5.1 Structure of the VGA

We chose to use a two-stage open loop configuration for this amplifier design which is shown in Figure 5.2. The first stage is a transconductance stage that converts the input small signal voltage to current. The second stage is a simple current mirror which drives a resistive load of 300 ohms. This approach can also meet the requirement of 600 ohms differential output impedance easily. The total gain of this structure is given by

$$A = g_m \cdot K \cdot R_L \tag{5.1}$$

where g_m is the transconductance of the first stage, K is the mirror gain, R_L is the load resistance.

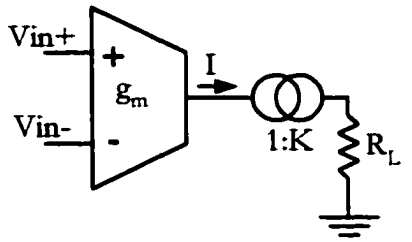


Figure 5.2 Two stage amplifier configuration

Several issues need to be solved in the open loop situation. First, the gain of the amplifier needs to be stabilized within a certain range (±3dB). Though the gain requirements for the amplifier is not very tight, it should not vary too much over process variations and temperatures, which is usually the case if the gain of the amplifier relies on the transconductance of the devices. Secondly, the amplifier needs to be linearized because the MOS transistors are inherently not quite linear especially in modern deep sub-micron process that is prone to short-channel effect and deviates from the classic square-law equation.

5.3 Linearization Schemes Review

In this section, we will review several widely-used linearization schemes for the transconductor design. In the following analysis we will consider perfectly quadratic $i-v$ characteristic for the MOS transistors in the saturation region and the channel length modulation effect will be neglected for simplicity. First, let us study the most basic structure - simple differential pair. It is shown in Figure 5.3(a). Assuming $v_i = v_{in+} - v_{in-}$ and $i_o = i_1 - i_2$. It has a transfer characteristic given by

$$i_o = \sqrt{2\beta I_{M1,2}} \cdot v_i \cdot \sqrt{1 - \frac{v_i^2}{4(V_{GS} - V_T)^2}} \qquad (5.2)$$

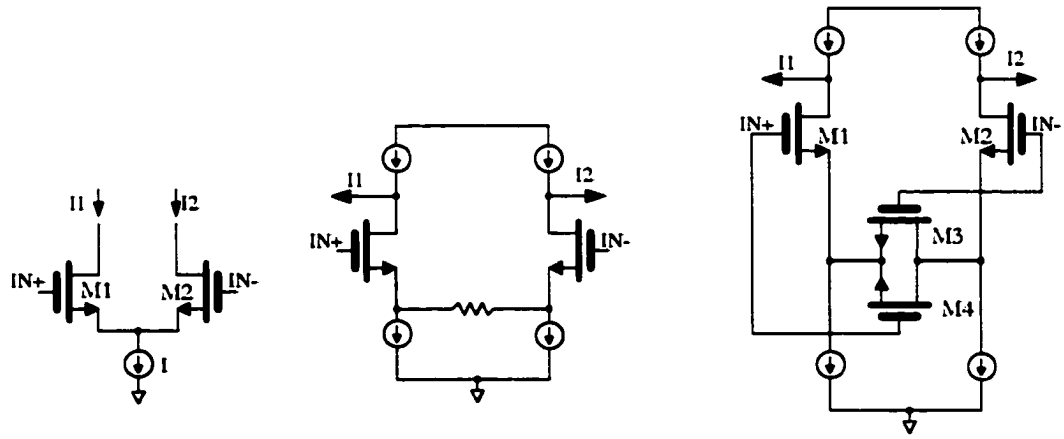
Even without considering channel length modulation, the $i-v$ characteristic is not linear. A better linearity can be get for a larger excess bias $V_{GS} - V_T$. In our simulation, we were able to get -51dB third harmonic distortion with $V_{opp} = 1V$.

5.3.1 Source Degeneration Scheme

One of the simplest topologies to linearize the transfer characteristic of the MOS transconductor is the one with source degeneration using resistors and depicted in Figure 5.3(b). The disadvantage of this configuration is the large resistor value needed to achieve a wide linear input range. By replacing the degeneration resistors with two MOS transistors operating in the triode region, the circuit in Figure 5.3(c) is obtained [5.1] [5.2] [5.3] [5.4] [5.5]. Considering perfectly matched transistors, and neglecting the body and channel length modulation effects, the transfer characteristic of this transconductor is given by

$$i_o = \frac{\sqrt{2\beta_1 I_{M1,2}}}{\alpha} v_i \sqrt{1 - \frac{\beta_1 v_i^2}{\alpha^2 I_{M1,2}}} \tag{5.3}$$

$$\text{where } \alpha = 1 + \frac{\beta_1}{4\beta_3} \tag{5.4}$$



(a) (b) (c)

Figure 5.3 (a) simple differential pair; (b) source degeneration linearization; (c) source degeneration using two triode transistors

Usually, the nonlinear term under the square root can be made much smaller than unity and improved linearity and larger input dynamic range can be obtained. The circuit has bandwidth and noise performances comparable to the simple differential pair.

This linearization scheme was proposed more than 10 years ago. To test its performance at sub-micron CMOS process, we built the circuits in Figure 5.3(c) and were able to get -58dB third harmonic distortion at $V_{opp} = 1V$. Still, it is not good enough for us.

5.3.2 Constant Drain-Source Voltage Scheme

Another linearization scheme is using "constant drain-source voltages". Recall the model for transistors in triode region is given by

$$I_D = \mu C_{ox}\left(\frac{W}{L}\right)\left[(V_{GS} - V_T)V_{DS} - \frac{V_{DS}^2}{2}\right] \tag{5.5}$$

We see here that if V_{DS} is kept constant, the drain current is linear with respect to gate-source voltage. Several implementations based on constant drain-source voltage were presented before [5.9] [5.10]. One possible implementation is shown in Figure 5.4.

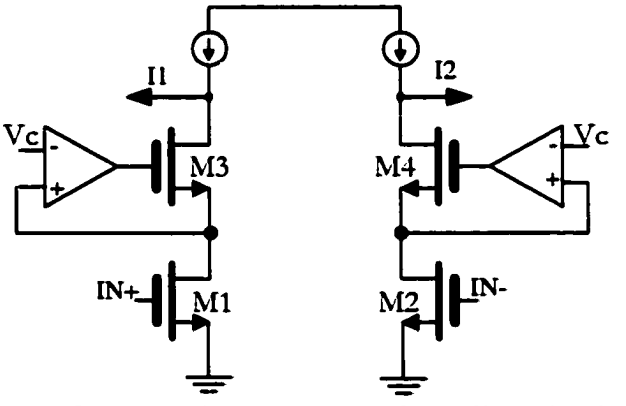


Figure 5.4 Constant drain-source voltage linearization scheme

M1 and M2 are placed in the triode region and their drain-source voltages are set equal to Vc through the use of M3, M4 and two extra amplifiers. The transconductance of this structure is given by

$$g_m = \mu C_{ox}\left(\frac{W}{L}\right)_1 V_{DS} \tag{5.6}$$

Note that the transconductance is proportional to the drain-source voltage. A major disadvantage of this structure is the limited bandwidth.

5.3.3 Constant Sum of Gate-Source Voltage Scheme

Similar strategy named "constant sum of gate-source voltage" is also widely used [5.11] [5.12]. The difference is that in this strategy, transistors are working in saturation

region in stead of triode region. Assuming two transistors are operating in saturation region as shown in Figure 5.5(a), the output differential current is given by

$$(I_1 - I_2) = \beta(V_{GS1} + V_{GS2} - 2V_T)(V_{GS1} - V_{GS2}) \tag{5.7}$$

We see that the output differential current is linear if the sum of the two gate-source voltages remains constant. One important point is that, although the differential current is linear, the individual drain currents are not linear. Thus, if the subtraction between currents has some error, some distortion will occur even if perfect square-law devices are obtainable.

There were a variety of ways to make the sum of the gate-source voltages remain constant when applying an input signal. One of them, as depicted in Figure 5.5(b), is to use differential pair with floating voltage sources. Writing a voltage equation around the loop, we have

$$V_{GS1} - (V_x + V_t) + V_{GS2} - (V_x + V_t) = 0 \tag{5.8}$$

thus, $$V_{GS1} + V_{GS2} = 2(V_x + V_t) \tag{5.9}$$

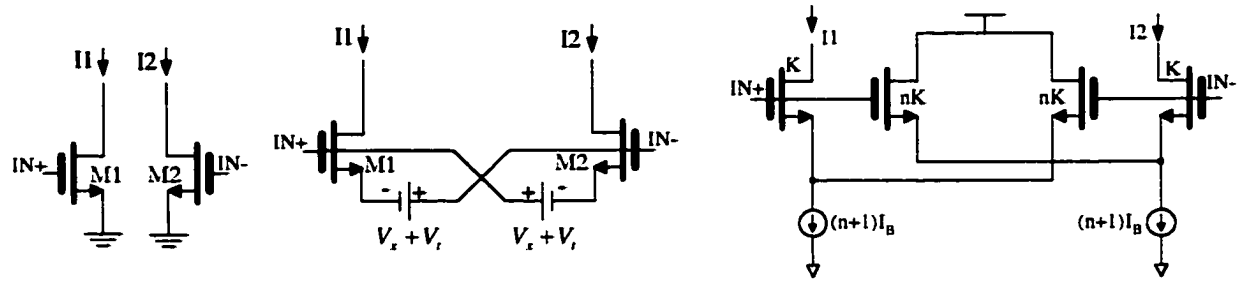
As a result, this circuit maintains a constant sum of gate-source voltages even if the applied differential signal is not balanced. Also, we can find the differential output current is given by

$$I_1 - I_2 = 4\beta V_x V_i \tag{5.10}$$

A simple way to realize the floating voltage sources of Figure 5.5(b) is to use two source followers, as shown in Figure 5.5(c) [5.12]. The transistors labeled nK are n times larger than the other two transistors. They act as source followers when n is large. The transconductance of this structure is given by

$$g_m = \frac{n}{n+1} 4\sqrt{KI_B} \tag{5.11}$$

A major disadvantage of this structure is a large amount of quiescent current pass through two source followers. To test its performance at sub-micron CMOS process, we built this transconductor and were able to get -57dB third harmonic distortion at $V_{opp} = 1V$.



(a) (b) (c)

Figure 5.5 Constant sum of gate-source voltage linearization scheme

5.3.4 Bias-offset Cross-Coupled Differential Pairs

Another approach to realize a transconductor with active transistors is to use two cross-coupled differential pairs where input into one pair is intentionally voltage offset [5.13] [5.14] [5.15]. One example of this approach is shown in Figure 5.6. MOS transistors M1-M4 and M5-M8 have the same dimensions and operate in saturation region. Because of the same current flow through M5 and M7 (M6 and M8), their gate-source voltages will be the same. Thus the inputs will be voltage-shifted to be applied to M3 and M4 and this voltage shift can be controlled by biasing voltage V_B.

Applying square law of the MOS transistor, we have

$$I_1 = I_{d1} + I_{d4} = \beta(V_P - V_T)^2 + \beta(V_N - V_B - V_T)^2 \tag{5.12}$$

$$I_2 = I_{d2} + I_{d3} = \beta(V_N - V_T)^2 + \beta(V_P - V_B - V_T)^2 \tag{5.13}$$

Thus, the differential output current

$$I_o = (I_1 - I_2) = 2\beta V_B(V_P - V_N) \tag{5.14}$$

Which yields

$$g_m = 2\beta V_i \tag{5.15}$$

We see that the differential current is linear with respect to differential input voltage as expected.

There are still some other schemes proposed to linearize the transconductor [5.16] [5.17] [5.18] whose details will not be discussed here.

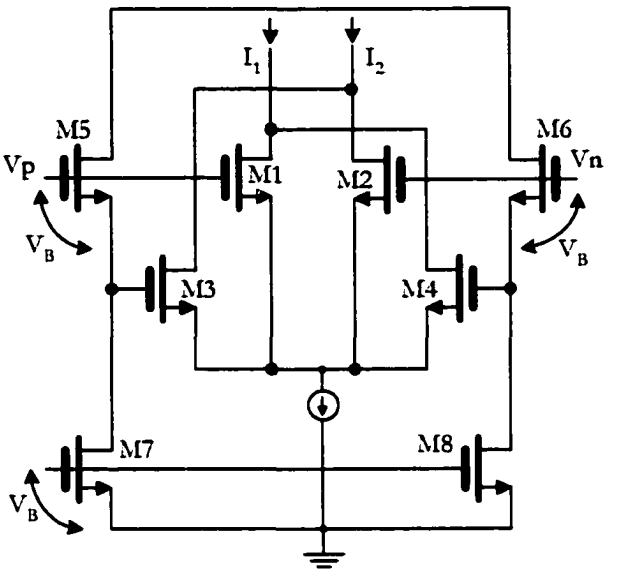


Figure 5.6 Bias-offset cross-coupled differential pairs linearization scheme

These linearization schemes can bring us moderate improvements. The ideas behind them were all based on the square-law model of the MOS devices. Compared to long-channel processes at the time when they were published, the improvements that can be obtained from those schemes in short-channel processes are limited. This is due to the large deviation from the classic square-law model for the short-channel devices. Because of the short-channel effects, it's not a good idea to use active devices to linearize the transconductor in order to achieve high linearity. Discussed in next section is a linearization scheme that doesn't depend on the MOS devices.

5.4 Open Loop Amplifier with Linearized Transconductor

For all the transconductors that we discussed in the previous section, none of them was accepted for our amplifier design. Some structure's linearity performances are not good enough for our application. Others have the problem of gain stability because their transconductance depend on process parameters.

As mentioned before, our amplifier is a two stage structure. The first stage is a linearized transconductance stage. The second stage is a simple current mirror. The linearized

transconductor structure we used is similar to those used in [5.6] [5.7] [5.8]. In [5.8], authors used a "floating linear resistor" formed by triode region transistors to linearize the transconductor. Because the resistance value in our design is small, we replaced the floating linear resistor with a real resistor in order to get better linearity. Its structure is shown in Figure 5.7.

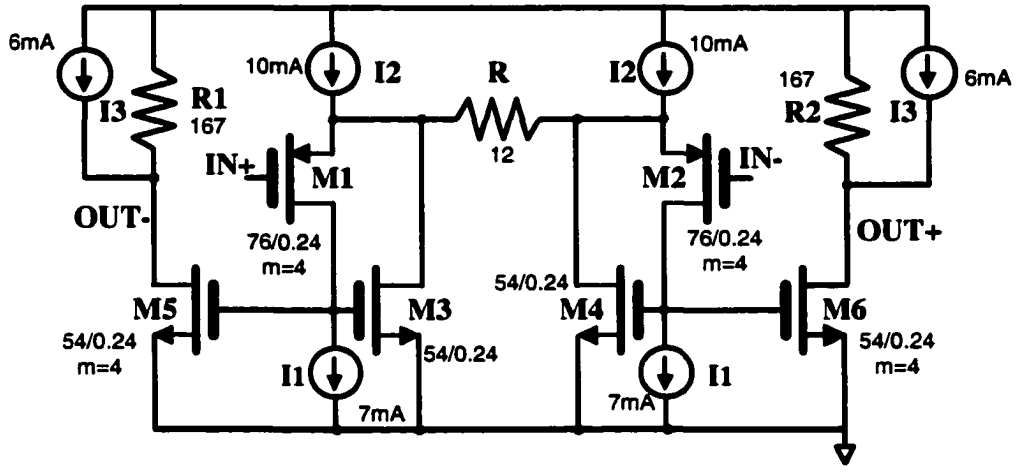


Figure 5.7 Linearized amplifier used in VGA

The transconductance stage includes M1, M2, R and the current sources I1 and I2, while the current mirror stage is consist of M3-M6 to drive resistive loads.

Current I1 is forced flew through M1 and M2 at any time that keeps the constant V_{GS} for both M1 and M2. Thus the p-channel devices will serve as voltage followers buffering the input small signal across the resistor R. The small-signal current will then flow through M3 and M4 and be mirrored to the output to drive the loads R1. Theoretically, the transconductor stage is very linear and it doesn't rely on the square law of the transistors.

This structure also takes care of the gain stability problem for the open loop amplifiers. The transconductance of the first stage is simply $g_m = 1/R$. The gain of the amplifier is given by:

$$A = \frac{R1 // Rx}{R} M \tag{5.16}$$

where M is the mirror gain, Rx is the external load.

The gain of the amplifier is determined by the mirror gain and the ratio of the two resistances. This property greatly enhances the gain stability. Actually, the gain of the amplifier will change because the sheet resistance of the integrated resistors varies while the external resistive load keeps constant. Simulation results considering this effect will be given later in this chapter which shows an acceptable performance.

Because the output impedance of the amplifier is fixed in order to interface with the loads, the only two design variables that we can control are R and M. A combination of them must be carefully chosen to ensure the low distortion and ease the realization of the resistors.

To get the best linearity out of the current mirror, the output common mode voltage was chosen to be around 1.1V. Because the output impedance of our design is fixed, the quiescent current level in the output transistor would be limited in a small range. This limited quiescent current in the output stage can't sustain large current swing while still maintain the required linearity. Two current sources were added to the output devices to increase the quiescent current in M5 and M6. These additional currents provide an additional 4dB better linearity.

The input transistors were chosen to use PMOS devices. This is based on several considerations. First, its body can be connected to the source in the projected N-WELL CMOS process. It eliminates the body effect and improves the linearity. Secondly, PMOS devices are less noisy than NMOS devices. Finally, to complement the design, NMOS current mirrors can be used that have better frequency response than PMOS current mirrors with NMOS input stage.

Current mirrors contribute part of the overall nonlinearity. The intuitive thoughts to improve the linearity of the current mirrors were to use cascode current mirrors. Hope the additional cascoded devices can shield the drains of M5 and M6 from large voltage swings. But several investigations revealed that this approach has little to do with the linearity of the

amplifier. We have two observations about the linearity performance that apply for both simple and cascode current mirror configurations under small-feature size processes and BSIM3 models:

(1) Couplings from the output (drains of M5 and M6) back to the transconductance stage (through gates of M3 and M4) have a major impact on the linearity of the gm stage at high frequencies. At low frequencies, the linearity of the signal current of the gm stage is keeping constant and very high. It starts to get worse when the input frequencies are higher than 100MHz.

(2) Small drain voltage swings at the output devices (M5 and M6) do not necessary give a better linearity compared to larger swings. The linearity is more depend on the harmonies between two drain voltages, i.e. it would be more linear to have a constant large difference between two drain voltages than a variable small difference between them.

The speed of this amplifier is moderate, the dominate pole appears at the gates of M4 and M6. In order to meet the bandwidth requirement, about 40mA current is pumped into the amplifier.

5.5 R-2R ladder

The R-2R ladder is used in series with the amplifier in order to have a very accurate gain step control. The fully-differential R-2R ladder structure is shown in Figure 5.8. All the switches are implemented by NMOS transistors. They are controlled by the output of the digital decoder. According to the digital control codes MSB, ISB and LSB, one tap of the ladder will be selected. The attenuation ratios of the output with respect to input are shown in the figure. The common mode voltage of the output is set to 1/3 of the supply voltage. It is also the input common mode voltage for the amplifier. Each resistor is 50 ohm.

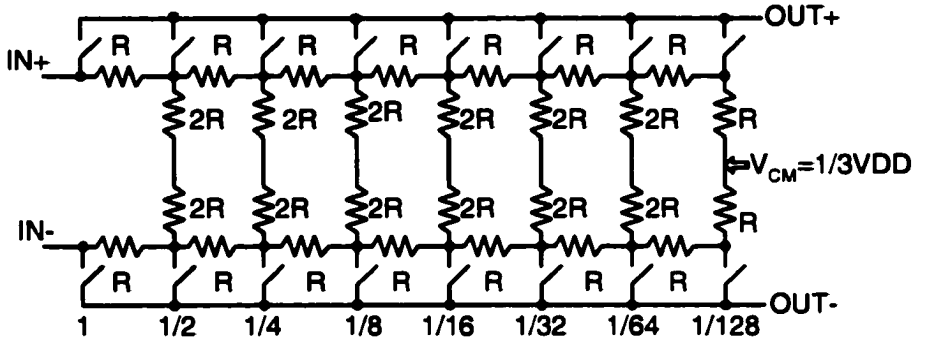


Figure 5.8 Structure of the R-2R ladder

5.6 Digital Circuits

Digital control of the gain of VGA is accomplished by a 3-bit parallel gain control input, a data valid signal to latch the data. If the data is not latched, the VGA continuously updates its gain setting.

The digital circuits shown in Figure 5.9 are basically a 3-to-8 decoder with latches and buffers. The buffers, inverters, AND gates and the latches are all implemented in standard digital circuits using thick gate-oxide transistors.

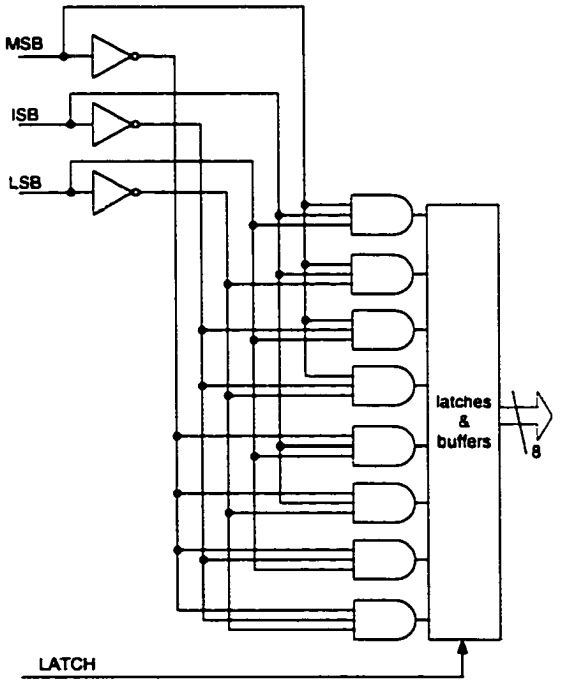


Figure 5.9 Schematic of the decoder

5.7 Biasing Circuits

The biasing circuits provide bias for amplifier, R-2R ladder. It is shown in Figure 5.10. M3-M6 and resistor R form a constant g_m, supply insensitive bias circuits. Two outputs BiasP and BiasP provide the biasing voltages for N-type and P-type current sources. This circuit has two stable quiescent states – normal working state and the state with all the currents are zero and all the transistors are off. M2 is used as start-up circuit to prevent circuit in zero-current working status. When the currents are zero, M2 forces M4 to conduct current and increase the gate voltage of the M3. Thus bring the circuits out of the zero-current state. When circuits work at normal state, M2 is off because of the higher source voltage.

M2, M7 and the inverter are used for power down function. When PowerDown input signal is high, BiasN is brought down and BiasP is brought up so that the amplifier will not function and the current consumption will reduced significantly.

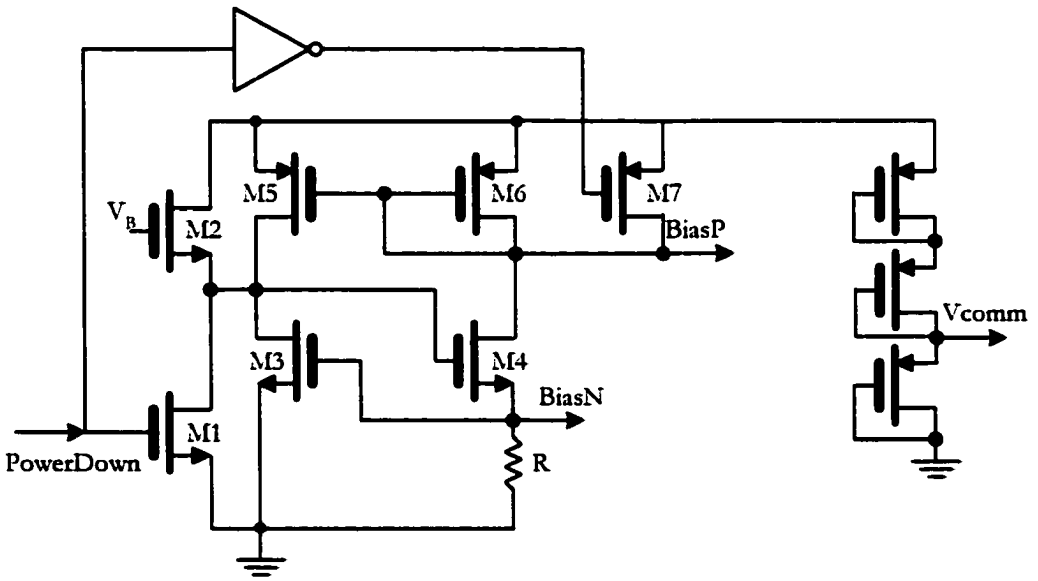
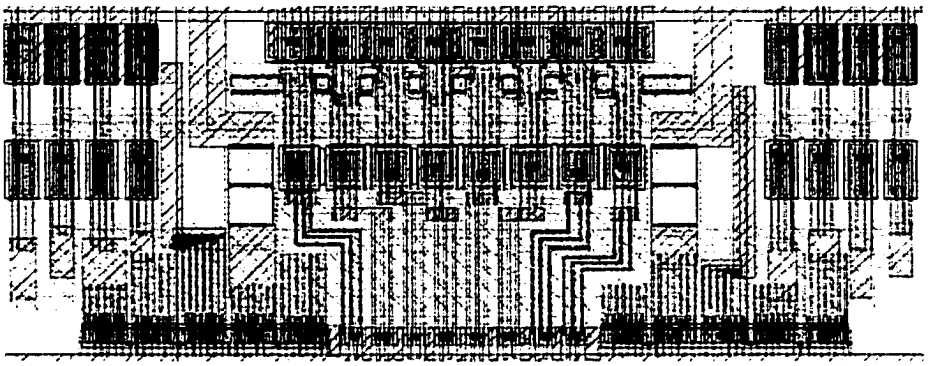


Figure 5.10 Biasing circuits for the VGA

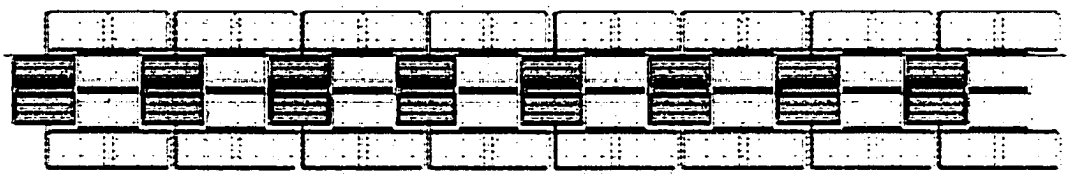
5.8 Chip Layout

The layout of the amplifier is shown in Figure 5.11(a). Inter-digitizing and symmetric layout techniques were used for better matching performance. Because of the large current, 60% of the area was consumed by metal interconnections.

The layout of the R-2R ladder is shown in Figure 5.11(b). It is a somewhat straight-forward layout. The resistors were laid out using single-unit resistor cells, vertically symmetric. The NMOS switches were surrounded by the resisters and between each ladder stages.



(a)



(b)

Figure 5.11 Layout of (a) amplifier; (b) R-2R ladder

5.9 Simulation Results

This project was designed in a standard CMOS 0.25μm process. It was simulated using HSPICE simulator and BSIM3, level 49 models with package and power supply models at all-transistor level. The accurate simulation option was switched ON in order to get good approximation for the expected measurement results.

The AC response of the amplifier for different gain settings at room temperature and normal device models is shown in Figure 5.12. The gain step is almost exactly 6.02dB

because there is no mismatch in schematic simulation. The 3dB bandwidth of the amplifier is about 300MHz. Some margin on bandwidth in our design would bring more confidence in the future testing. The AC performance under different process corners and different supply voltages are also summarized in Table 5.1

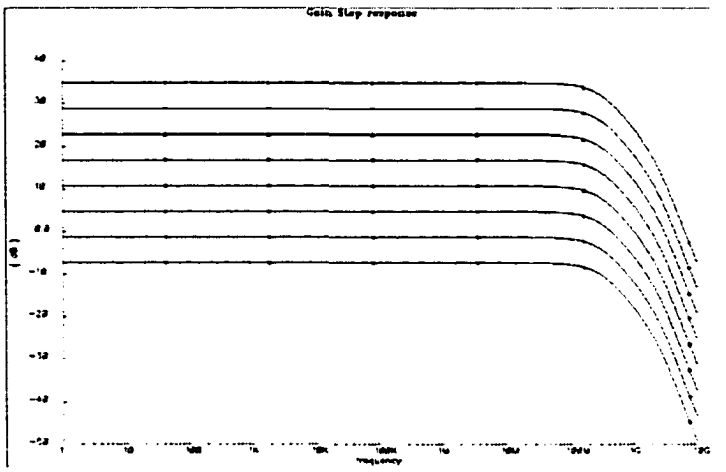


Figure 5.12 AC response of the VGA for different gain settings

Table 5.1 VGA AC performance

	3.3V		3V		3.6V	
	Gain	3dB BW	Gain	3dB BW	Gain	3dB BW
Normal	34.8dB	294MHz	34.5dB	293MHz	35.1dB	297MHz
Fast	34.7dB	321MHz	34.4dB	318MHz	34.8dB	325MHz
Slow	34.8dB	274MHz	34.2dB	274MHz	35.2dB	275MHz

The simulation results for the linearity of the amplifier under different process corners are shown in Table 5.2. This measurement was done at the input signal frequency f=160MHz. Some margin was left to meet the requirements for the linearity because the transistor mismatch was not taken into account. It's not surprising to see the worse real measurement results.

Table 5.2 Linearity@Vo_pp=1V (HD3), room temperature

	3.3V		3V		3.6V	
	Linearity	current	Linearity	current	Linearity	current
Normal	-62.8dB	46.3mA	-59.7dB	45mA	-62.9dB	47.4mA
Fast	-62.6dB	42.8mA	-61dB	41.6mA	-62.2dB	44mA
Slow	-61.8dB	50mA	-55dB	48.7mA	63.5dB	51.2mA

According to the data sheet of the projected process, its sheet resistance for poly resistor varies about ±20%. Shown in Figure 5.13 and Table 5.3 are the gain response and the linearity performance considering the resistance variation, i.e. all the integrated resistors vary their resistance while the external load keeps constant. The gain variation is controlled within ±1dB and the linearity also meets the requirements.

Table 5.3. Effects of the sheet resistance variation

	Linearity	Gain
Rmax	-62.3dB	35dB
Rmin	-61.6dB	34.4dB

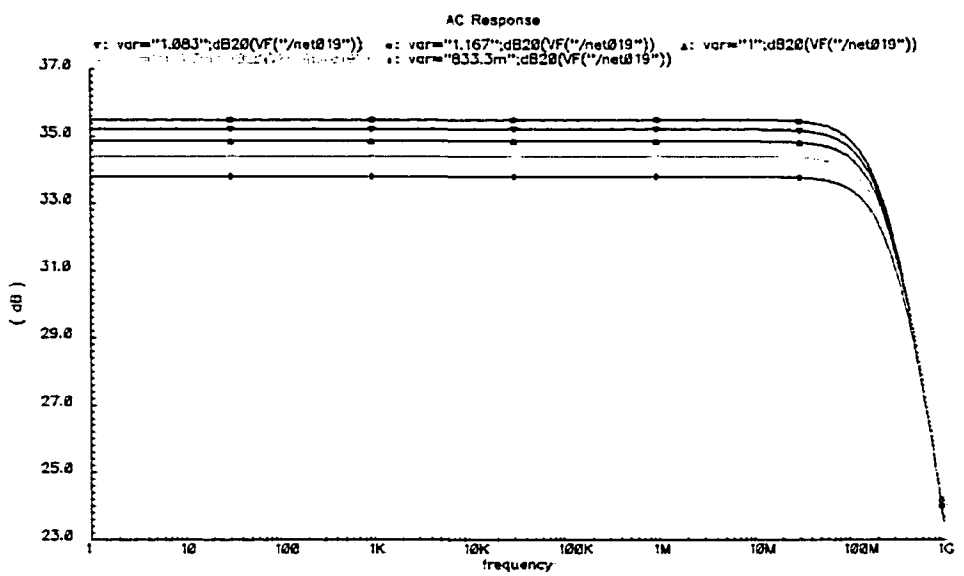


Figure 5.13 Gain response considering resistance variations

The simulated input equivalent noise for the amplifier is $1.44nV/\sqrt{Hz}$ at maximum gain. The noise performance for all the gain settings is depicted in Figure 5.14. Those values are the average value over the frequency range of 100MHz to 500MHz which covers the entire frequency range of our interest. The noise figures for all the gain settings assuming 200 ohms source impedance is shown in Figure 5.15.

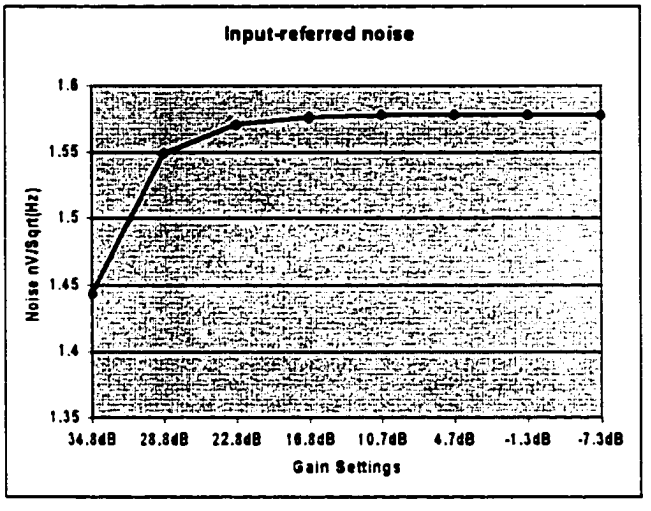


Figure 5.14 Simulated noise performance of the VGA

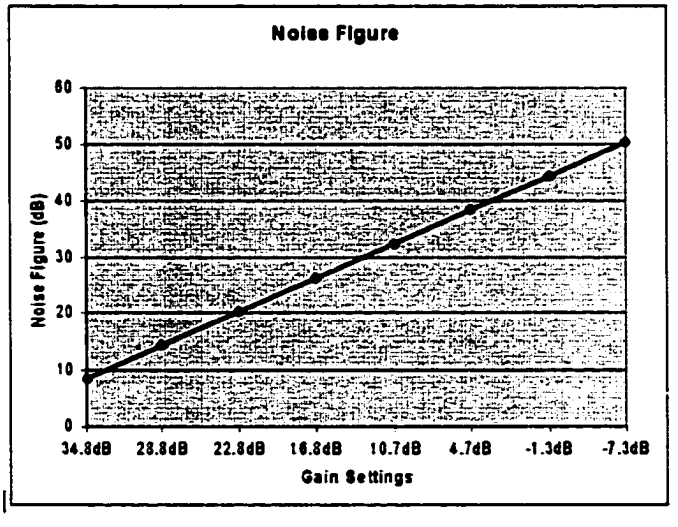


Figure 5.15 Simulated noise figure of the VGA

The temperature effect was also simulated and is shown in Table 5.4. The amplifier design keeps the acceptable performance on linearity and AC characteristics.

Table 5.4 Temperature effects on amplifier design

	Linearity	Gain	BW
-40C	-64dB	35.9dB	303MHz
85C	-60.8dB	33.7dB	297MHz

5.10 Conclusion

Dedicated analog function can also be realized in deep sub-micron CMOS process. It not only provides acceptable performance, but also cost effective. In this work, we demonstrated the design of a CMOS VGA with precise gain step and high linearity. An open loop linearized amplifier was used in this VGA to meet the requirements of bandwidth and linearity at the same time. This amplifier structure can also be used as a low distortion building block for very wide applications.

References

[5.1] F. Krummenacher, N. Joehl, "A 4-MHz CMOS Continuous-Time Filter with On-Chip Automatic Tuning", IEEE Journal of Solid-State Circuits, Vol. 23, NO. 3, pp. 750-758, June 1988

[5.2] U. Chilakapati, T. S. Fiez, A. Eshraghi, "A CMOS Transconductor With 80-dB SFDR up to 10 MHz", IEEE Journal of Solid-State Circuits, Vol. 37, NO.3, pp. 365-370, March 2002

[5.3] G. Bollati, S. Marchese, M. Demicheli, R. Castello, "An Eighth-Order CMOS Low-Pass Filter with 30-120 MHz Tuning Range and Programmable Boost", IEEE Journal of Solif-State Circuits, Vol. 36, NO. 7, pp. 1056-1066, July 2001

[5.4] V. Gopinathan, M. Tarsia, D. Choi, "Design Considerations and Implementation of a Programmable High-Frequency Continuous-Time Filter and Variable-Gain Amplifier in Submicrometer CMOS", IEEE Journal of Solid-State Circuits, Vol. 34, NO. 12, pp. 1698-1707, December 1999

[5.5] K.-C. Kuo, A. Leuciuc, "A Linear MOS Transconductor Using Source Degeneration and Adaptive Biasing", IEEE Transactions on Circuits and Systems-II: Analog and Digital Signal Processing, Vol. 48, NO. 10, October 2001

[5.6] A. Leuciuc, Y. Zhang, "A Highly Linear Low-Voltage MOS Transconductor", IEEE International Symposium of Circuits and Systems, Vol. 3, pp. 735-738, 2002

[5.7] D. R. Welland, S. M. Phillip, et al. "A Digital Read/Write Channel with EEPR4 Detection", Digest of Technical Papers of IEEE International Solid-State Circuits Conference, pp. 276 -277, 1994

[5.8] T. Kwan, K. Martin, "An Adaptive Analog Continuous-Time CMOS Biquadratic Filter", IEEE Journal of Solid-State Circuits, Vol. 26, NO. 6, pp. 859-867, June 1991

[5.9] T. Yamaji, N. Kanou, T. Itakura, "A Temperature-Stable CMOS Variable-Gain Amplifier With 80-dB Linearly Controlled Gain Range", IEEE Journal of Solid-State Circuits, Vol. 37, NO. 5, pp. 553-558, May 2002

[5.10] S. L. Wong, "Novel Drain-Biased Transconductance Building Blocks for Continuous-Time Filter Applications", IEE Electronics Letters, Vol.25, Issue. 2, pp. 100-101, January 1989

[5.11] G. Wilson, P. K. Chan, "Low-Distortion CMOS Transconductor", IEE Electronics Letters, Vol. 26 Issue. 11, pp. 720-722, May 1990

[5.12] A Nedungadi, T. R. Viswanathan, "Design of Linear CMOS Transconductance Elements", IEEE Journal of Solid-State Circuits, Vol. 31, pp. 891-894, October 1984

[5.13] Z. Czarnul, N. Fujii, "Highly-Linear Transconductor Cell Realized by Double MOS Transistor Differential Pairs", IEE Electronics Letters, Vol.26 Issue.21, pp.1819-1821, October 1990

[5.14] Z. Wang, "Novel Linearization Technique for Implementing Large-Signal MOS Tunable Transconductor", IEE Electronics Letters, Vol. 26, Issue. 2, pp. 138-139, January 1990

[5.15] Z. Wang, W. Guggenguhl, "A Voltage-Controllable Linear MOS Transconductor Using Bias Offset Technique", IEEE Journal of Solid-State Circuits, Vol. 25, NO. 1, pp. 315-317, February 1990

[5.16] S. Szczepanski, A Wyszynski, R. Schaumann, "Highly Linear Voltage-Controlled CMOS Transconductors", IEEE Transactions on Circuits and Systems – I: Fundamental Theory and Applications, Vol. 40, NO. 4, pp. 258-262, April 1993

[5.17] C. S. Kim, Y. H. Kim, S. B. Park, "New CMOS Linear Transconductor", IEE Electronic Letters, Vol. 28, Issue. 21, pp. 1962-1964, October 1992

[5.18] F. Munoz, A. Torralba, R. G. Carvajal, J. Tombs, J. Ramirez-Angulo, "Floating-Gate-Based Tunable CMOS Low-Voltage Linear Transconductor and Its Application to HF Gm-C Filter Design", IEEE Transactions on Circuits and Systems – II: Analog and Digital Signal Processing, Vol. 48, NO. 1, January 2001

CHAPTER 6

EFFECTS OF OPEN-LOOP NONLINEARITY ON LINEARITY OF FEEDBACK AMPLIFIERS

We described a highly linear amplifier design in Chapter 5. In this chapter, we will present a quantitative analysis of how the negative feedback would impact on the nonlinearity of the feedback amplifier with respect to the open loop amplifier nonlinearities (OLN). It will give a better understanding of the relationship between negative feedback and the linearity.

6.1 Introduction

Nonlinearity is a major nonideality of an amplifier circuit. It can be depicted as a nonlinear input/output characteristic [6.1] as shown in Figure 6.1. Usually when the input signal is small, the output has a reasonable linear relationship to the input. But with an increase of the input level, the output typically exhibits an increase in nonlinearity as depicted in the figure.

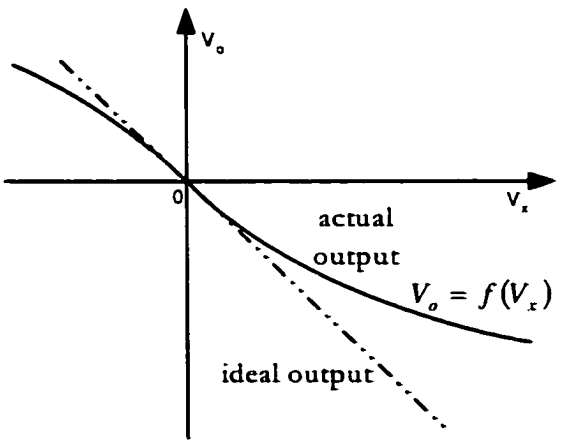


Figure 6.1 Nonlinearity in the amplifier

The nonlinearity of a circuit can also be considered as the "variation" of the slope (gain) in the input/output characteristics as a function of operating point. It means that a given incremental change at the input results in different incremental changes at the output depending on the quiescent input level.

Several techniques have been used to improve linearity. One of the most widely used linearization strategies is using negative feedback and the linearization properties associated with negative feedback were one of the major reason feedback concepts were developed. It is well known that another property of feedback circuits is gain desensitization. Since nonlinearity can be viewed as a variation of the small-signal gain with the input level, negative feedback techniques also decrease the variation.

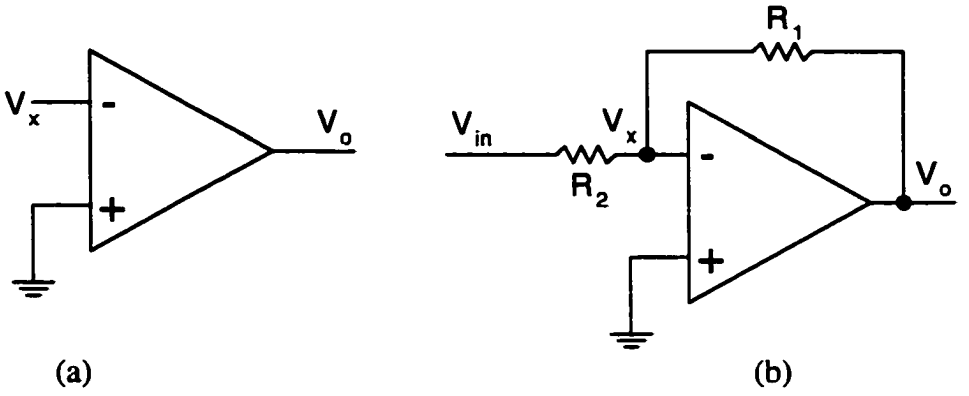
While the general effect of negative feedback on linearity is well known, little research has been done from a quantitative viewpoint on how much nonlinearity can be reduced through feedback. The issue of what effect feedback will have on different order harmonics that contribute to the nonlinearity in the open loop amplifier has also not received much attention. The work presented in this chapter provides a quantitative assessment of how several nonlinearity properties of the open loop amplifier affect feedback amplifiers.

6.2 Definition and Quantization of the Nonlinearity

In order to make a meaningful and fair comparison of the nonlinearities under different circumstances, a rigorous definition of the nonlinearity that is suitable for both open loop and feedback structures are needed.

Consider the open loop amplifier shown in Figure 6.2(a). The input-output relationship is $V_o = f(V_x)$, where it will be assumed that $f(V_x)$ can be approximated by a desired first-order term and two undesired nonlinear terms. Thus, $f(V_x)$ can be expressed as

$$f(V_x) = -AV_x + BV_x^2 + CV_x^3, \quad A, B, C > 0, V_x > 0 \tag{6.1}$$



(a) (b)

Figure 6.2 (a) open loop amplifier; (b) feedback amplifier

This equation characterizes the open loop transfer characteristics of the amplifier in the fourth quadrant. Its characteristic in the second quadrant is similar as depicted in Figure 6.1. This expression includes the second and the third harmonic distortions that generally dominate the nonlinearity in most open loop amplifier.

Assume that the transfer characteristic of the open loop amplifier is as shown in Figure 6.1 with the solid line that shows an increase in nonlinearity as the input amplitude increases. When negative feedback is applied, the gain of the feedback amplifier is usually decreased and considerably less distortion is experienced.

Ideally, the amplifier should have a linear input-output relationship of $V_o = -AV_x$. This ideal linear relationship corresponds to the tangent line through the origin with a slope of $k = f'(V_x)\big|_{V_x=0} = -A$. This ideal output is shown in Figure 6.1 as the dotted line.

For the feedback amplifier shown in Figure 6.2(b), it follows that:

$$V_x = \frac{R_1}{R_1 + R_2} V_o + \frac{R_2}{R_1 + R_2} V_{in} \tag{6.2}$$

The feedback gain (amount of feedback applied) is usually defined as:

$$\beta = \frac{R_1}{R_1 + R_2} \tag{6.3}$$

It follows that $V_x = \beta V_o + (1-\beta) V_{in}$ (6.4)

Combining equation (6.4) with (6.1), we can obtain an exact input output relationship $V_o = g(V_{in})$ for the feedback amplifier. The closed-loop form of the expression for $g(V_{in})$ is

unwieldy, even in the presence of only second-order and third-order nonlinearities. This function can be solved with the help of the MATLAB symbolic toolbox. The solution is too complicated to show here because of the existence of the third order harmonics in the solution.

Again, the ideal output of the feedback amplifier is defined as the tangent line that passes through the origin with a slope of

$$k = g'(V_{in})\Big|_{V_{in}=0} = -\frac{1-\beta}{\beta+1/A} \tag{6.5}$$

The nonlinearity for any specific input is defined to be the deviation of the actual output from the ideal output at the given input. With this definition, each input to an amplifier has its own nonlinearity value. What we are interested here is to see the effect of feedback on linearity. We need to choose a reference point where nonlinearities are investigated.

The nonliearity of an amplifier is usually closely associated with the output level. In what follows nonlinearity will be compared not at a certain input level but at a fixed ideal output level that is within our range of interest. We will base our comparisons on the output level rather than the input level because the gain of the feedback amplifier varies a lot with β. The quantization of the nonlinearity is shown in Figure 6.3. For comparison purposes, the nonlinearies of the feedback amplifier will be compared at the ideal output level of $V_o = -1$ which corresponds to the input level of V_{-1}. The actual output for input V_{-1} is V_{ao} because of the nonlinearity. The nonlinearity, expressed in percentage, can be expressed as:

$$Nonlinearity(\%) = 100 \times (1 + V_{ao}) \tag{6.6}$$

It should be mentioned here that more simulations based on different reference points yield similar results and the same conclusions.

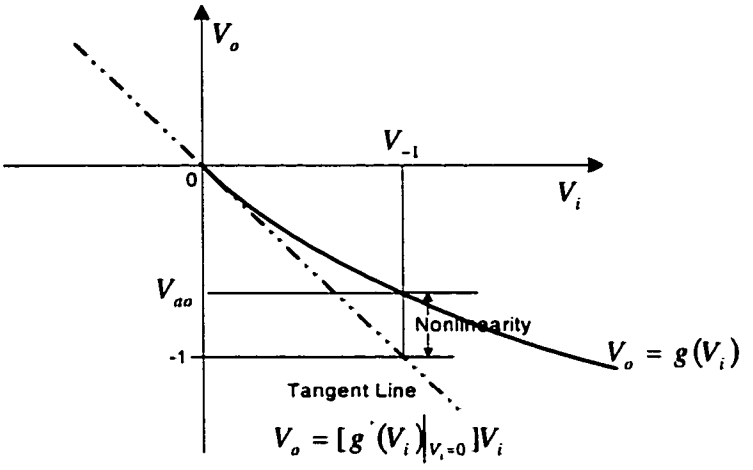


Figure 6.3 Quantization of the nonlinearity

6.3 Effects of the Feedback on Nonlinearity

6.3.1 Effects of the Feedback Factor on CLN

It is well known that with deeper negative feedback (larger β), more nonlinearity can be reduced. But no quantitative analysis has been done to resolve the relationship between feedback factor and the nonlinearity. The following investigation will look at how the amount of nonlinearity is related to the feedback factor β.

A typical feedback system is shown in Figure 6.4. The gain of the amplifier can be expressed as:

$$A_f = \frac{V_o}{V_{in}} = \frac{A}{1 + A\beta}. \tag{6.7}$$

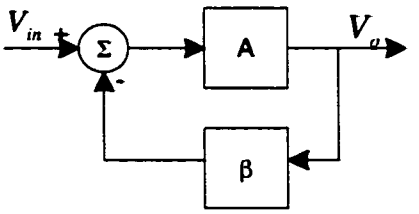


Figure 6.4 Negative feedback system

For the original open loop amplifier shown in Figure 6.2(b), we assume DC gain A=1000; total nonlinearity at the ideal output level $V_o = -1$ is $OLN = 10\%$. Depending on the percentage combination of the second and the third harmonics that constitute the nonlinearity, the coefficients B and C in equation (6.1) can be determined accordingly.

Special care must be take to guarantee the monotonically of the input-output relationship of the original amplifier within the range of our interest so that the solutions of the feedback amplifier equation are real. This was done by limiting the amount of the nonlinearity in the open loop amplifier in the calculations.

The nonlinearities measured at the ideal output level of $V_o = -1$ in several feedback amplifiers are shown in Figure 6.5. X axis shows the inversion of the feedback factor, i.e. $1/\beta$. Y axis shows the percentage of the nonlinearities in the feedback amplifier. Two cases are shown in Figure 6.5. One is that 100% of the OLN is due to the 2nd order harmonic, the other is that 100% of the OLN is due to the 3rd harmonic.

In both cases, the amount of nonlinearity is linearly proportional to the inverse of the feedback factor. We conclude that the amount of Closed-Loop Nonlinearity (CLN)

$$CLN = \frac{k}{\beta} + C \tag{6.8}$$

where k and C are constants and only determined by open loop amplifier characteristics.

In our calculations, the amount of OLN was fixed. Therefore, another conclusion can also be drawn,

$$\frac{CLN}{OLN} = \frac{m}{\beta} + D \tag{6.9}$$

where m and D are constants and only determined by open loop amplifier characteristics.

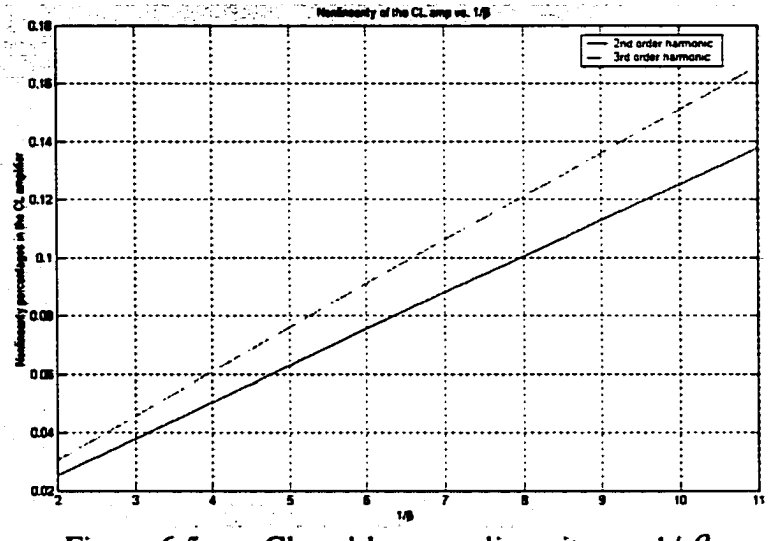


Figure 6.5 Closed-loop nonlinearity vs. $1/\beta$

6.3.2 Effects of the Open Loop Gain on CLN

Under that same assumption (except the open loop gain) as in previous section, the effects of open loop gain on nonlinearities in the feedback amplifier were investigated. As shown in Figure 6.6, open loop gain was swept from 1000 to 10000. Their corresponding CLN were calculated. The relationship between open loop gain and the amount of CLN is not linear, either for 2nd or 3rd order harmonics, or for different amount of OLN.

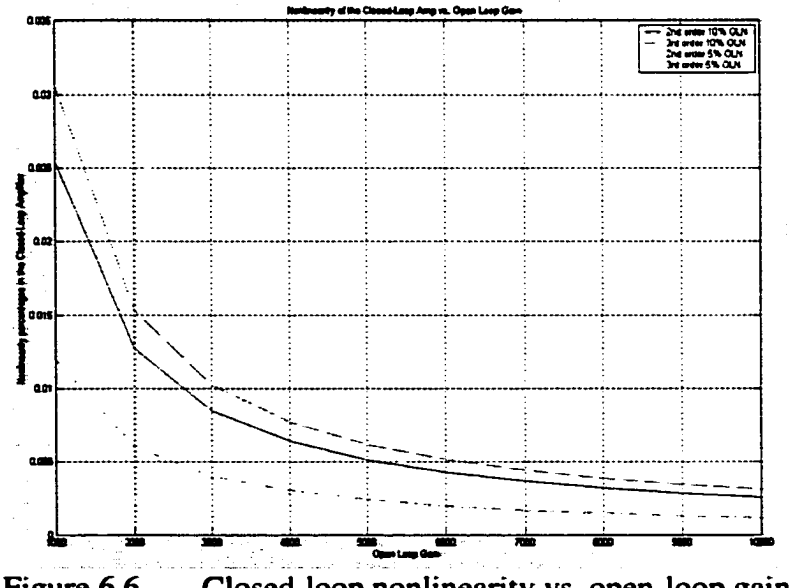


Figure 6.6 Closed-loop nonlinearity vs. open-loop gain

CLN drops dramatically when open loop gain start to increase. After open loop gain becomes larger than 3000-4000, the decrease of the CLN is much less. This property suggests that the effect of high open loop gain on reducing CLN is limited.

6.3.3 Effects of the Amount of OLN on CLN

Another interesting topic would be whether different amount of OLN would be suppressed linearly upon feedback. As shown in Figure 6.7, different amount of OLN were tested with feedback factor of 0.5 and open loop gain A=1000. For both 2nd and 3rd order harmonics, the suppressions of the nonlinearity through feedback were not linear.

The increase of the CLN becomes faster with the increase of the OLN. This trend is more obvious for the 3rd harmonic. This property suggests the importance of limiting the OLN in design a low-distortion amplifier.

6.3.4 Effects of Different Harmonics on CLN

Modern integrated circuits design is often based on fully differential structure in order to eliminate even order harmonics. It would an interest to investigate if different order harmonic behaves differently in the feedback amplifier.

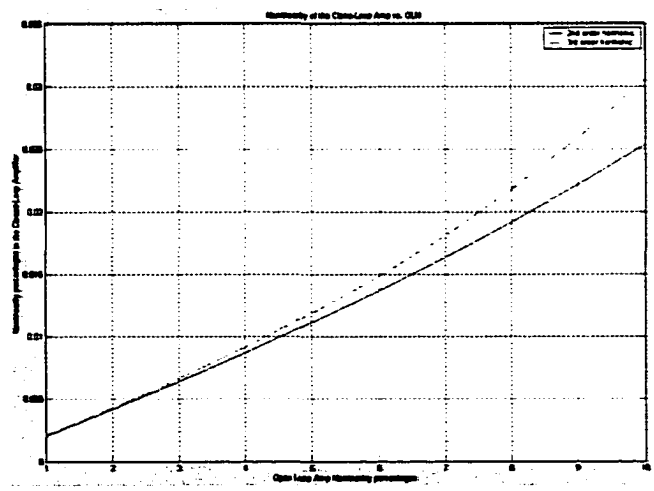


Figure 6.7 Amount of closed-loop nonlinearity vs. amount of open loop nonlinearity

It is already shown in Figure 6.5 that for the second and the third order harmonics, the same amount of nonlinearity in open loop amplifier will result in different amount of nonlinearity in the feedback amplifier. The third order harmonic will result a higher amount of nonlinearity in the feedback amplifier.

More investigations were done to see how different combination of the second and the third harmonics in open loop amplifiers would affect the nonlinearity in the feedback amplifier. As the surface plot shown in Figure 6.8, CLN were calculated with different percentages of 2nd and 3rd harmonics in the open loop amplifier.

It is very clear that for any percentage combination of the second and the third order harmonics, the amount of CLN is still linearly proportional to the inverse of the feedback factor. For a certain feedback factor, the amount of the CLN changes linearly with the percentage of the 2nd and the 3rd harmonics.

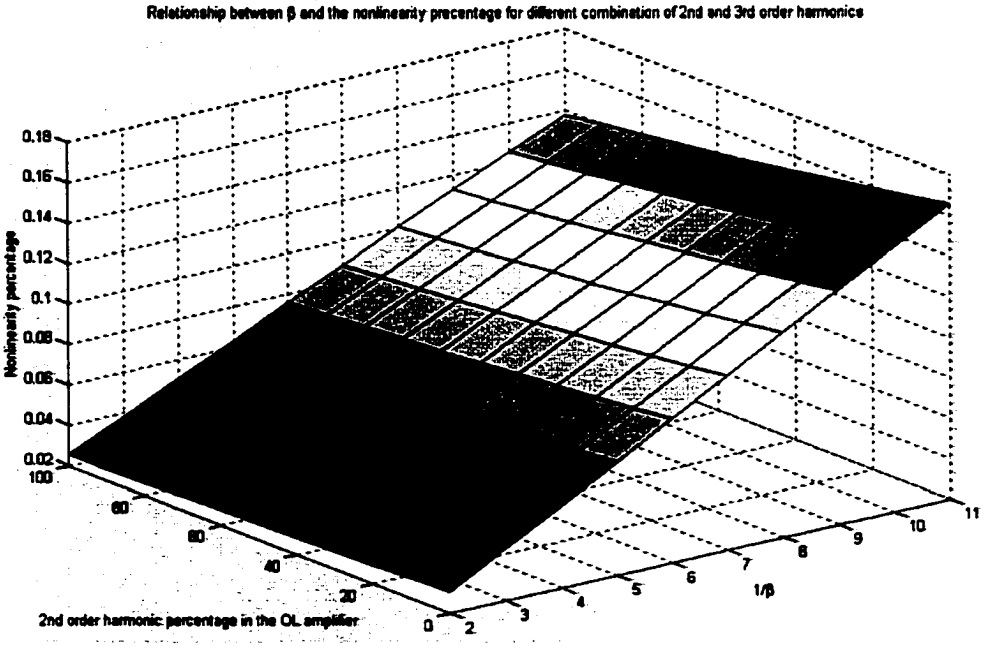


Figure 6.8 Nonlinearity vs. 1/β vs. different percentages of the second and the third harmonics

6.4 Conclusion

This chapter presented an analysis of the nonlinearity in feedback amplifiers. A new way to quantize the amount of nonlinearity was proposed. Using this method, a general-purpose negative feedback amplifier was analyzed for its nonlinearity under several different situations. We observed that the nonlinearity in the feedback amplifier is linearly proportional to $1/\beta$ and lower order harmonic nonlinearity will be reduced more through feedback. Results also show that the effect of high open loop gain on reducing nonlinearities through feedback is limited.

References

[6.1] Behzad Razavi, "Chapter 13: Nonlinearity and Mismatch", "Design of Analog CMOS Integrated Circuits", Preview Edition, McGraw Hill, 2000.

[6.2] Paul R. Gray, Robert G. Meyer, "Chapter 8: Feedback", "Analysis and Design of Analog Integrated Circuits", third edition, John Wiley & Sons, Inc. 1993

[6.3] Kenneth R. Laker, Willy M.C. Sansen, "Chapter 3: Feedback and Sensitivity in Analog Integrated Circuits", "Design of Analog Integrated Circuits and Systems", McGraw-Hill, Inc. 1994