1

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS
STANDARD REFERENCE MATERIAL 1010a
(ANSI and ISO TEST CHART No. 2)

National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file - Votre référence

Our file - Notre référence

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

The Study and Development of a BiCMOS Operational Amplifier

by

Fred G. Black, B.Eng.

A thesis submitted to the Faculty of Graduate Studies and Research in partial
fulfillment of the requirements for the degree of

Master of Engineering

**Department of Electronics
Carleton University
Ottawa, Ontario**

April 9, 1993

**©copyright
1993, Fred G. Black**

National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-315-92974-X

Canada

Thesis contains black & white
photographs &/or explanatory tables
which when microfilmed may lose their
significance. The hardcopy of the
thesis is available upon request from
Carleton University Library.

The undersigned recommend to the Faculty of Graduate Studies
and Research the acceptance of the thesis:

**"The Study and Development of a BiCMOS
Operational Amplifier"**

submitted by Fred G. Black, B.Eng. in partial fulfillment
of the requirements of the degree of Master of Engineering.

A handwritten signature in black ink, reading "M A Copeland", written over a horizontal line.

**Prof. M. A. Copeland
Thesis Supervisor**

A handwritten signature in black ink, reading "J S Wight", written over a horizontal line.

**Prof. J. S. Wight
Chairman
Department of Electronics**

Abstract

This thesis reports on work on possible uses of recently available BiCMOS integrated circuit technology in order to realize high frequency operational amplifiers. An investigation was carried out on the relative merits of bipolar and MOS devices when considered for use in the realization of the various circuit functions required. The investigation was begun early in the development of a particular BiCMOS process and the results have been documented here for future reference. An operational amplifier suitable for use in high frequency switched capacitor circuits has been designed in this BiCMOS technology.

Acknowledgements

I would to thank my program supervisor, Dr. M. A. Copeland for his patience, support and guidance in the preparation of this thesis.

I would also like to thank Mr. D. Lynch and Mr. T. Kozelj of Bell-Northern Research, Ottawa and Mr. R. A. Hadaway of Northern Telecom Electronics, Ottawa for their assistance and for arranging for the use of IC design facilities.

I would also like to acknowledge the financial support during the course of this work from Bell-Northern Research, Ottawa, the Natural Sciences and Engineering Research Council of Canada (NSERC), the Telecommunications Research Institute of Ontario (TRIO) and Micronet.

Table of Contents

Acceptance Sheet	ii
Abstract	iii
Acknowledgements	iv
Table of Contents	v
List of Tables	ix
List of Figures	x
List of Symbols and Terms	xiv
1 Introduction	1
1.1 Thesis Objective	1
1.2 Thesis Motivation	1
1.3 Thesis Format	3
2 CMOS, Bipolar and BiCMOS Integrated Circuit Processes	5
2.1 Comparison of MOS and Bipolar Transistors	5
2.1.1 Bipolar Transistor Design Equations	6
2.1.2 MOS Transistor Design Equations	10
2.1.3 Comparison of Bipolar and MOS Transistors	14
2.2 MOS and Bipolar Integrated Circuit Processes	25
2.2.1 Bipolar Integrated Circuit Processes	26
2.2.2 CMOS Integrated Circuit Processes	28
2.2.3 BiCMOS Integrated Circuit Processes	30
2.3 A Generic BiCMOS Process	32

2.3.1	Device Complement	33
2.3.2	Summary of Generic BiCMOS Process	41
2.4	Summary	42
3	A Consideration of Operational Amplifier Topologies	43
3.1	Operational Amplifiers	43
3.1.1	Operational Amplifier Non-Idealities	45
3.2	CMOS Operational Amplifiers	49
3.2.1	CMOS Two Stage Operational Amplifier	49
3.2.2	Techniques for Performance Improvement In CMOS Operational Amplifiers	55
3.2.3	CMOS Folded Cascode Operational Amplifier	56
3.3	Bipolar Operational Amplifiers	60
3.3.1	Simplified Two Stage Bipolar Operational Amplifier	60
3.3.2	A Practical Two Stage General Purpose Bipolar Operational Amplifier	61
3.3.3	Input Current Cancellation	63
3.3.4	Differences Between Bipolar and CMOS Operational Amplifiers	66
3.4	Summary	66
4	Topology Considerations for a BiCMOS Operational Amplifier	68
4.1	Options for a BiCMOS Operational Amplifier Topology	68
4.1.1	Devices Available	69
4.1.2	Operational Amplifier Design Options	69
4.2	Choice of Current Source	70
4.2.1	The MOS Current Mirror	73
4.2.2	The Bipolar Current Mirror	75
4.2.3	Cascaded Transistor Current Mirrors.	76
4.2.4	Summary of Current Mirrors.	79
4.3	Differential Input Circuits	79
4.3.1	The MOS Differential Pair	80

4.3.2	The Bipolar Differential Pair	84
4.3.3	The Bipolar Differential Pair with Emitter Degeneration	89
4.3.4	The Darlington Differential Pair	94
4.3.5	The Darlington Differential Pair With Emitter Degeneration .	100
4.3.6	The Bipolar Differential Pair Driven by MOS Source Followers	102
4.3.7	Comparison of Differential Pairs	107
4.3.8	Summary of Input Pairs	116
4.4	Circuit Options for Active Loads	117
4.4.1	Use of Cascaded Current Mirrors as Active Loads	118
4.4.2	Summary of Active Loads	122
4.5	Choice of Output Stage	122
4.5.1	Common Source and Common Emitter Amplifiers	123
4.5.2	Common Emitter Amplifier with Emitter Follower Input . . .	124
4.5.3	Single Stage Operational Amplifier	125
4.5.4	Summary of Output Stages	126
4.6	Summary	127
5	Design and Simulation	128
5.1	Towards a Proposed Topology for a BiCMOS Operational Amplifier .	128
5.1.1	Choice of Differential Pair	129
5.1.2	Choice of Current Source	131
5.1.3	Choice of Output Stage	131
5.1.4	Proposed Topology	135
5.1.5	Summary	137
5.2	Simulation Conditions	138
5.3	A BiCMOS Operational Amplifier with NMOS Differential Pair	139
5.3.1	Design Process	141
5.3.2	Biasing Scheme	143
5.3.3	Performance	144
5.4	Test and Measurement	165

5.4.1	Test Circuits for BiCMOS Operational Amplifier with NMOS Differential Pair	165
5.4.2	Open Loop Test Structure	167
5.4.3	Closed Loop Test Structure	169
5.4.4	Emitter Follower Test Structure	170
5.4.5	Measurements	171
5.5	Additional Topologies	171
5.5.1	A BiCMOS Operational Amplifier with NPN Differential Pair and Input Current Cancellation	172
5.5.2	A BiCMOS Operational Amplifier with Darlington Differential Pair and Emitter Degeneration	174
5.5.3	BiCMOS Operational Amplifiers with NPN Differential Pair and PMOS Source Follower Inputs	176
5.6	Summary	179
6	Conclusions and Future Work	181
6.1	Summary and Conclusions	181
6.2	Recommendations	185
A	SPICE Models for Hypothetical BiCMOS Process	188
B	Layout of Opamp and Test Structures	190
	References	195

List of Tables

2.1	Typical Bipolar Process Flow and Masks	27
2.2	Typical CMOS Process Flow and Masks	29
4.1	Difference Mode Gains of Differential Input Stages	108
4.2	Relative Difference Mode Gains of Differential Input Stages	110
4.3	Common Mode Range of Differential Input Stages	112
4.4	Input Impedance Expressions for Differential Input Stages	112
4.5	Approximate Input Impedance of Differential Input Stages	113

List of Figures

2.1	NPN Bipolar Transistor and Small Signal Model	6
2.2	1 x 3 um NPN Bipolar Transistor Characteristics	7
2.3	1 x 3 um NPN Bipolar Transistor Characteristics	8
2.4	MOS Field Effect Transistor and Small Signal Model	10
2.5	1 x 3 um NMOS Transistor Characteristics	12
2.6	1 x 3 um NPN Transistor	14
2.7	1 x 3 um NMOS Transistor	15
2.8	Transconductance vs Current for Several Transistors	16
2.9	Comparison of Output Conductances	18
2.10	Input Stage With Emitter Degeneration	19
2.11	Darlington Connected Input Stage	23
2.12	1x3 um NMOS Transistor Characteristics	24
2.13	1x3 um PMOS Transistor Characteristics	25
2.14	Vertical NPN Transistor	34
2.15	Polysilicon Emitter Structure	35
2.16	Trench Isolation Structure	36
2.17	Lateral PNP Transistor	38
2.18	N Channel MOS Transistor	39
2.19	P Channel MOS Transistor	40
3.1	Ideal Operational Amplifier and Equivalent Circuit	44
3.2	Inverting Amplifier Configuration	44
3.3	CMOS Pole Splitting Operational Amplifier	50

3.4	MOS Transistor Cascode	56
3.5	CMOS Operational Amplifier With Cascoded Differential Stage	57
3.6	CMOS Folded Cascode Operational Amplifier	58
3.7	Simplified Two Stage Bipolar Operational Amplifier	61
3.8	711 Monolithic Operational Amplifier	62
3.9	Input Current Cancellation Circuit 1	64
3.10	Input Current Cancellation Circuit 2	65
4.1	Generic Differential Input Stage with Difference Mode Small Signal Model	71
4.2	Simple MOS Current Mirror	74
4.3	Simple Bipolar Current Mirror	75
4.4	Cascoded MOS Current Mirror	77
4.5	MOS Differential Pair	80
4.6	Small Signal Half Circuits of the MOS Differential Pair	81
4.7	Bipolar Differential Pair	85
4.8	Small Signal Half Circuits for the Bipolar Differential Pair	86
4.9	Bipolar Differential Pair With Emitter Degeneration	89
4.10	Small Signal Half Circuits of the Bipolar Differential Pair With Emitter Degeneration	90
4.11	Darlington Differential Pair	95
4.12	Darlington Differential Pair Small Signal Half Circuits	96
4.13	Modified Darlington Differential Pair	98
4.14	Darlington Differential Pair With Emitter Degeneration	100
4.15	Small Signal Half Circuit of the Darlington Differential Pair With Emitter Degeneration	101
4.16	NPN Differential Pair With PMOS Source Follower Inputs	103
4.17	MOS Source Follower with Small Signal Model	104
4.18	Small Signal Half Circuits of NPN Differential Pair With PMOS Source Follower Inputs	105
4.19	Capacitor Discharge Through a Current Source	114

4.20	NPN Cascode Current Mirror	119
4.21	Wilson Current Mirror	120
4.22	Common Source and Common Emitter Amplifiers	123
4.23	Common Emitter Amplifier With Emitter Follower Input	125
4.24	Emitter Follower - Common Emitter Amplifier Small Signal Model . .	126
5.1	Common Emitter Amplifier With Emitter Follower Input	132
5.2	Proposed NMOS Input BiCMOS Operational Amplifier Topology . .	131
5.3	Proposed PMOS Input BiCMOS Operational Amplifier Topology . .	135
5.4	BiCMOS Operational Amplifier with NMOS Differential Pair	139
5.5	NMOS Input Operational Amplifier Layout Plot	140
5.6	Biasing Scheme for BiCMOS Op Amp with NMOS Differential Pair .	144
5.7	NMOS Input Opamp Frequency Response	146
5.8	NMOS Input Opamp DC Common Mode Response (Closed Loop) . .	152
5.9	MOS Differential Pair	154
5.10	MOS Differential Pair with Active Load	155
5.11	NMOS Input Opamp Common Mode and Power Supply Rejection Ratio	158
5.12	NMOS Input Opamp DC transfer Function	160
5.13	NMOS Input Opamp Open Loop Transient Response	161
5.14	NMOS Input Opamp Closed Loop Transient Response	162
5.15	NMOS Input Opamp Input Referred Noise	164
5.16	Open Loop Test Structure for BiCMOS Op Amp with NMOS Differ- ential Pair	166
5.17	Low Frequency Test Structure	167
5.18	Closed Loop Test Structure for BiCMOS Op Amp with NMOS Differ- ential Pair	170
5.19	Emitter Follower Test Structure	171
5.20	BiCMOS Operational Amplifier with Bipolar Inputs and Input Current Cancellation Circuitry	172
5.21	BiCMOS Operational Amplifier with Darlington Inputs and Emitter Degeneration	175

5.22 BiCMOS Operational Amplifier with Bipolar Differential Pair and PMOS Source Follower Inputs	177
5.23 BiCMOS Operational Amplifier with Source Follower Inputs and Emit- ter Degeneration	178
B.1 NMOS Input Operational Amplifier	191
B.2 Closed Loop Test Structure	192
B.3 Open Loop Test Structure	193
B.4 Emitter Follower Test Structure	194

List of Symbols and Terms

A_v	Voltage gain
A_{dm}	Difference mode voltage gain
A_{cm}	Common mode voltage gain
β	Low frequency current gain of a bipolar transistor. Also known as H_{FE} .
BiCMOS	An integrated circuit process in which both bipolar and MOS devices are available.
BNR	Bell Northern Research
CAD	Computer Aided Design
Cadence	CAD framework and IC layout editor
CMOS	Complementary Metal Oxide Semiconductor
CMRR	Common Mode Rejection Ratio
codec	COder - DECoder. A mixed analog/digital integrated circuit.
dB	decibel. A measure of relative voltage or power.
ESD	Electro-Static Discharge
f_c	Transconductance frequency. A figure of merit for MOS transistors.
f_T	Frequency at which high frequency current gain drops to unity. A figure of merit for bipolar transistors.
g_m	Transconductance of a transistor
g_o	Output conductance of a transistor
hspice	Circuit Simulator
Hz	Hertz. A measure of frequency.
I_C	Collector current of a bipolar transistor
I_D	Drain current of a MOS transistor
$\frac{kT}{q}$	Thermal Voltage, often referred to as V_T .
MHz	MegaHertz

MOS	Metal Oxide Semiconductor
NMOS	N channel MOS field effect transistor
NPN	A type of bipolar transistor having an N type emitter and collector and a P type base.
NTE	Northern Telecom Electronics
opamp	operational amplifier
pF	Pico Farad. A measure of capacitance.
phase margin	A measure of stability of a feedback system. It is defined as the difference between the phase shift of an opamp and - 180 degrees. It must be positive to ensure stability.
PMOS	P channel MOS field effect transistor
PNP	A type of bipolar transistor having a P type emitter and collector and an N type base.
PSRR	Power Supply Rejection Ratio
R_i	Input resistance, usually of a gain or buffering stage.
r_o	Output resistance of a transistor. It is the reciprocal of g_o .
SPICE	Circuit simulator
S_r	Slew rate
μA	microampere
UGBW	Unity gain bandwidth. The frequency at which the gain of an opamp drops to unity.
μm	micron
V_A	Early Voltage. Proportional to the output resistance of a bipolar transistor.
VAF	Alternative notation for V_A
V_{BE}	Voltage from base to emitter of a bipolar transistor
V_{CE}	Voltage from collector to emitter of a bipolar transistor
V_{DS}	Voltage from drain to source of a MOS transistor
V_{GS}	Voltage from gate to source of a MOS transistor
V_T	Threshold Voltage of a MOS transistor

V_{dd}	Positive power supply
V_{ss}	Negative power supply
ω_u	Unity gain bandwidth in radians per second.

Chapter 1

Introduction

1.1 Thesis Objective

The objective of this thesis is to explore the use of a recently available BiCMOS technology in the realization of improved operational amplifiers for use in monolithic analog signal processing on integrated circuits. A detailed comparison will be made between MOS and bipolar transistors in order to identify the characteristics of each that are advantageous for the design of analog circuits in general and operational amplifiers in particular. Bipolar, MOS and BiCMOS circuit techniques for operational amplifier design will be explored with emphasis on achieving increased bandwidth in operational amplifiers suitable for use in switched capacitor circuits.

1.2 Thesis Motivation

Historically, bipolar and CMOS integrated circuit technologies have evolved separately and have been motivated by different concerns and applications. Generally, bipolar technology has been preferred where speed is of primary concern and has typically been used to provide analog subsystems such as component opamps, rather than complete analog signal processing systems on a single integrated circuit. CMOS technology has been preferred where circuit density and the ability to combine low power digital circuitry and complete analog signal processing systems on the same

chip are of concern. An example of such a chip is the codec (COder - DECoder), used to interface between analog telephone sets and a digital switch.

Advancement of technology and market demand has created the need to operate integrated circuits at increasingly higher speeds. Bipolar integrated circuit technologies have not often been a desirable choice for use by the telecommunications industry due to cost and power consumption considerations and the need to perform analog signal processing in the sampled data domain. However the operating frequencies of telecommunications circuits are rapidly approaching or exceeding the capabilities of current CMOS processes. One solution is to adopt a BiCMOS process, a combination of bipolar and CMOS on the same chip. With such a process, bipolar circuitry can be used where needed to allow increased speed of operation while retaining the use of CMOS where needed for its digital circuit densities and other benefits.

The primary building block for monolithic analog system design is the operational amplifier. Operational amplifiers find application as voltage amplifiers, signal buffers and in active filtering, both continuous and sampled data [6]. As was stated above, there is a continuing trend to operate systems at increasing speeds, thus necessitating higher bandwidth operational amplifiers. The use of bipolar transistors and subcircuits where advantageous allows one possible option to increase the bandwidth over available CMOS operational amplifiers. One major objective of this thesis is to investigate the new possibility of using combinations of bipolar and CMOS transistors and subcircuits to advantage in operational amplifier design.

As will be described in chapters 2 and 3, bipolar and MOS transistors each have different advantages and disadvantages when considered for use in the implementation of the various circuit functions required by an operational amplifier. Over the years, fairly complete sets of analog subcircuits have evolved in each of the technologies. As a result of the different properties of bipolar and CMOS transistors, these sets of subcircuits have somewhat different properties and we would think these differences might be exploited to advantage if both CMOS and bipolar devices are available on the same chip. As quick examples, MOS current mirrors are not affected by β error of bipolar current mirrors; however, a bipolar transistor makes a much better

current source as its output impedance (shunt) is higher and the voltage required to keep the device in active operation is lower. In the case of differential input stages, bipolar transistors provide greater gain and potentially greater bandwidth, but at the price of low input impedance. MOS differential pairs provide effectively infinite input impedance but more moderate gain. The correct choice depends on the particular application. The ability to combine subcircuits from each technology or create new circuit blocks employing both device types potentially allows the designer an opportunity to improve the performance of both analog and digital circuits. The desire to have the ability to use each type of device to attempt to overcome the limitations of the other device in circuits is one motivation to develop a BiCMOS technology.

1.3 Thesis Format

This thesis is on a study of operational amplifier design techniques as they apply to BiCMOS technology with an emphasis on high frequency design, and with a consideration of how special combinations of the two types of circuits, now possible, might lead to useful innovation. This study was begun early in the development of the particular BiCMOS process studied and documents the capabilities of the available active devices in an analog design context. Thus this thesis serves as an exploratory document for future researchers.

Introductory material including the objective, motivation and format of the thesis is presented in Chapter 1. In Chapter 2, typical bipolar and CMOS transistors and the integrated circuit processes that fabricate them will be described, first separately and then as a combined process. In Chapter 3, topologies of existing CMOS and bipolar operational amplifiers will be presented and discussed. In Chapter 4, bipolar and MOS implementations of various circuit blocks that are combined to realize operational amplifiers will be discussed. A BiCMOS operational amplifier topology is then proposed in Chapter 5. The design and performance of an operational amplifier based on this topology are then discussed and simulation results are presented.

Finally, conclusions and suggestions for further work are given in Chapter 6.

Chapter 2

CMOS, Bipolar and BiCMOS Integrated Circuit Processes

In this chapter, the operation of bipolar and MOS transistors will be discussed and design equations for quantities of interest to analog design will be presented. This allows the two types of device to be compared with emphasis on qualities of concern in analog design.

The traditional bipolar and CMOS integrated circuit processes will be described with emphasis on available features and the advantages and drawbacks of each with respect to analog circuit design. The motivation for the use of a BiCMOS process for analog applications will be considered, and an example of such a process will be described. The devices available in this process are representative of what is available in state of the art separate bipolar and CMOS processes and will be used for comparison between the two types of process.

2.1 Comparison of MOS and Bipolar Transistors

The basic design equations for MOS and bipolar transistors will be presented and discussed below. Comparison between the results for bipolar transistors and MOS transistors will then be made. Reference will be made to the results in other chapters.

2.1.1 Bipolar Transistor Design Equations

The schematic symbol and an equivalent small signal model for forward active operation of an NPN bipolar transistor are shown in figure 2.1. The small signal model shown is often referred to as the *hybrid π* model [7]. The bipolar transistor has

Figure 2.1: NPN Bipolar Transistor and Small Signal Model

four modes of operation: cutoff, saturation, forward active and reverse active. The modes of cutoff and saturation are normally useful only for switching applications, while the reverse active mode offers poor performance compared to forward active. As the forward active mode of operation is the only mode useful in most types of analog design, all subsequent discussion will refer to this mode.

The collector current of a bipolar transistor in forward active operation may be related either to the base current, or to the base-emitter voltage. Thus the collector current may be given by:

$$I_C = I_s \exp \frac{qV_{BE}}{kT} \quad (2.1)$$

$$I_C = \beta I_B \quad (2.2)$$

β refers to the low frequency current gain of the device. For typical processes, β is around 100 and I_s has a value of 10^{-18} amperes per square micron of emitter area.

Transconductance can be defined as the ratio of the change in the output current of a circuit element to the change in the input voltage of that element. Thus the transconductance of the bipolar transistor is:

$$g_m = \frac{dI_C}{dV_{BE}} = \frac{qI_C}{kT} \quad (2.3)$$

It can be seen from equation 2.3 that the transconductance of the transistor is a function only of the current through the device and the temperature.

The output resistance of a bipolar transistor is due to modulation of the effective base width, also known as the Early effect. As a result of this effect, the current voltage characteristics have a finite slope in the forward active region. This can be seen in fig. 2.2 or fig. 2.3. These characteristics have been obtained by HSPICE

Figure 2.2: 1 x 3 um NPN Bipolar Transistor Characteristics

simulation of the NPN bipolar devices available in the process that will be described

Figure 2.3: 1 x 3 um NPN Bipolar Transistor Characteristics

in Section 2.3. If these slopes are extended, they will intersect at a point on the voltage axis. This point defines the Early voltage V_A . Equation 2.1 may be modified to account for the Early effect thus giving:

$$I_C = I_s \left(1 + \frac{V_{CE}}{V_A} \right) \exp \frac{qV_{BE}}{kT} \quad (2.4)$$

By differentiating equation 2.4 with respect to the collector emitter voltage, the output conductance g_o of the transistor can be found to be:

$$g_o = \frac{dI_C}{dV_{CE}} = \frac{I_C}{V_A} \quad (2.5)$$

There are three parasitic resistances included in this small signal model of the bipolar transistor. The base spreading resistance exists because the active base and the electrical contact to the base are electrically separate. Typical values for r_b are on the order of a few hundred Ohms and are smaller for large transistors, or transistors

with appropriate base contact design. The collector series resistance is due to the fact that current must flow through silicon from the effective collector to the collector contact. Typical values of r_{cs} are also on the order of a few hundred Ohms. The emitter series resistance is due to the bulk silicon between the emitter-base junction and the emitter contact. Due to the high doping in this area, it is much smaller than r_b or r_{cs} , and typically has a value on the order of 10 ohms. In most applications not involving high current, its effect is small and is usually ignored.

r_e is the dynamic resistance of the base-emitter junction. It is a function of the bias current and the β of the transistor and may be expressed as

$$r_e = \frac{\beta}{g_m} \quad (2.6)$$

The capacitances in the small signal model will now be described. C_μ is the collector-base junction capacitance. It is a function of doping, collector-base voltage and the area of the base. It can be approximated by

$$C_\mu \approx A_B \sqrt{\frac{q\epsilon N_C}{V_{CB}}} \quad (2.7)$$

where A_B = area of collector-base junction

N_C = doping concentration of collector

The effective capacitance associated with the base-emitter junction is represented by C_π and is composed of the emitter diffusion capacitance and the junction capacitance of the base-emitter junction. The value of C_π is approximately given by equation 2.8. Note that the term f_T will be described presently.

$$C_\pi \approx \frac{g_m}{2\pi f_T} + C_\mu \quad (2.8)$$

A popular figure of merit for the high frequency capabilities of bipolar transistors is the charge carrier transit-time cutoff frequency or f_T . This is the frequency at which the common emitter short circuit current gain drops to unity. It can be expressed as

$$f_T = \frac{1}{2\pi\tau_{ec}} \quad (2.9)$$

where τ_{ec} is the average time for a charge carrier to transit from the emitter to the collector. Typical silicon bipolar transistors have f_T 's in the range of 1 to 10 GHz.

2.1.2 MOS Transistor Design Equations

Figure 2.4 below shows the schematic symbols for both n and p channel MOS field effect transistors as will be used in this thesis. The small circle on the gate of the symbol indicates a PMOS transistor. A small signal model of an n channel MOS transistor operating in the saturation region is also shown in figure 2.4. MOS tran-

Figure 2.4: MOS Field Effect Transistor and Small Signal Model

sistors have three regions of operation: saturation, linear or triode, and subthreshold. The region of primary interest in amplifier design is the saturation region and is in many ways similar to the forward active mode of a bipolar transistor. The linear region is useful when it is desired for an MOS transistor to act as a resistor. The subthreshold region can be considered to be essentially in cutoff, however a small exponential current does flow in this region and its properties have been exploited in novel circuits [12].

In the triode region, the drain current is given by

$$I_D = K_P \frac{W}{L} (V_{GS} - V_T - \frac{V_{DS}}{2}) V_{DS} \quad (2.10)$$

where $K_P = \mu_n C_{ox}$

$V_T =$ the threshold voltage of the device.

It is often useful to employ a MOS transistor operating in the triode region as a

resistor. If it is assumed that $V_{DS} \ll (V_{GS} - V_T)$ then the resistance between the source and drain terminals is given by

$$R_{DS} = \frac{V_{DS}}{I_D} = \frac{1}{K_P \frac{W}{L} (V_{GS} - V_T)} \quad (2.11)$$

The most often used region of operation for MOS transistors used in analog circuit design is the saturation region. Saturation occurs when the drain voltage increases to the point that the channel depth at the drain decreases to zero. This is known as pinch off. If the drain voltage is increase further, the pinch off point will move toward the source, however, the drain current will not change (in the ideal case) because the voltage at the pinch off point will remain at V_{Dsat} and the number of carriers flowing from the source to the pinch off point will be unchanged. Pinch off occurs when

$$V_{DS} \geq V_{GS} - V_T \quad (2.12)$$

thus giving the condition for saturation. Ignoring output resistance, the drain current for a MOS transistor operating in the saturation region is given by

$$I_D = \frac{K_P}{2} \frac{W}{L} (V_{GS} - V_T)^2 \quad (2.13)$$

When the transistor operates in saturation and the drain-source voltage is increased, the pinch off point moves towards the source, thus shortening the channel length and increasing the drain current slightly. This effect is known as channel length modulation. The actual drain current may be approximated by including the factor $(1 + \lambda V_{DS})$ in the equation 2.13 giving

$$I_D = \frac{K_P}{2} \frac{W}{L} (V_{GS} - V_T)^2 (1 + \lambda V_{DS}) \quad (2.14)$$

The term λ is often referred to as the slope factor and is generally proportional to $1/L$. The output conductance of a MOS transistor can be expressed as a function of the drain bias current and the slope factor. The relation is simply

$$g_o \approx \lambda I_D \quad (2.15)$$

As the slope factor increases with decreasing channel length, the output resistance of the MOS transistor decreases with decreasing channel length. The effect of the slope factor can be seen in the MOS transistor characteristics in fig. 2.5.

Figure 2.5: 1 x 3 μm NMOS Transistor Characteristics

In a manner similar to that used for the bipolar transistor, the transconductance of a MOS transistor can be determined by differentiating equation 2.14 with respect to V_{GS} . There are two popular forms of expressing this result, which are shown in equation 2.16 and equation 2.17 below.

$$g_m = K_P \frac{W}{L} (V_{GS} - V_T)(1 + \lambda V_{DS}) \quad (2.16)$$

$$g_m = \sqrt{2K_P(W/L)I_D(1 + \lambda V_{DS})} \quad (2.17)$$

The capacitances shown in the small signal model in saturation will now be discussed. C_{GD} is the capacitance due to the overlap of the gate and the drain diffusion. Its value can be approximated as

$$C_{GD} \approx WL_{ov}C_{ox} \quad (2.18)$$

where C_{ox} is the gate oxide capacitance per unit area and L_{ov} is the overlap between the gate and the drain and source diffusions, both of which are assumed equal. C_{GS} is composed of the capacitance due to the overlap of the gate and the source diffusion, and the capacitance between the gate and the unpinched region of the channel. An often used simplification for a MOS transistor operated in the saturation region is to consider the latter capacitance to be 2/3 of the oxide capacitance between the gate and the surface of the substrate [8]. C_{GS} is then given by

$$C_{GS} = WC_{ox}(L_{ov} + \frac{2}{3}L_{eff}) \quad (2.19)$$

where C_{ox} and L_{ov} are as above and L_{eff} is the effective channel length, ie

$$L = 2L_{ov} + L_{eff} \quad (2.20)$$

The capacitances C_{SB} and C_{DB} are largely due to the junction capacitance between the source and drain diffusions and the substrate. Being junction capacitances, they are voltage dependent.

For comparison, both between different MOS transistors and between MOS and bipolar transistors, it is useful to define a figure of merit for MOS transistors similar to the f_T used for bipolar transistors. The definition of the bipolar f_T involves current gain and is not really applicable to MOS transistors due to the fact that MOS transistors operate as voltage to current amplifiers. A similar figure of merit is known as the transconductance frequency [7]. This is given as

$$f_c = \frac{g_m}{2\pi C_{in}} \quad (2.21)$$

where C_{in} is the input capacitance to the device and is the sum of C_{GS} and C_{GD} . From equations 2.18 and 2.19

$$C_{in} = C_{GS} + C_{GD} = WC_{ox}(2L_{ov} + \frac{2}{3}L_{eff}) \quad (2.22)$$

It can be seen from equation 2.22 that C_{in} is proportional to the area of the gate, ie

$$C_{in} \propto C_{ox}LW \quad (2.23)$$

By combining equations 2.21, 2.23 and 2.16, f_c can be expressed as

$$f_c \propto \frac{\mu(V_{GS} - V_T)}{2\pi L^2} \quad (2.21)$$

From equation 2.24, it can be seen that to optimize the f_c it is necessary to use a short channel length.

2.1.3 Comparison of Bipolar and MOS Transistors

Various similarities and differences between bipolar and MOS transistors will be explored in this section. In many of the cases, it will be appropriate to use examples. The devices used for the examples will be those available in the generic process described in Section 2.3 below. The bipolar example device will be an NPN transistor with an emitter area of 1 by 3 microns. The MOS example device will be an NMOS transistor with a 1 by 3 micron gate. Plan views of these two example devices are presented in fig. 2.6 and fig. 2.7. Where a complementary device is needed, a PMOS

Figure 2.6: 1 x 3 um NPN Transistor

device of the same size as the NMOS device will be used. Unless otherwise specified, the example devices will be operated at a current of $20 \mu A$.

Figure 2.7: 1 x 3 μm NMOS Transistor

Area and Transconductance

Bipolar transistors have traditionally been considered to be well suited to applications involving high frequency or requiring the driving of large loads. MOS transistors have traditionally been favored in less demanding applications where power and cost are of greater concern. The transconductance of the two types of device are given in equations 2.3 and 2.17. It can be easily seen that the transconductance of a bipolar transistor depends only on the collector current and the temperature. Ignoring the effects of channel width modulation, the transconductance of a MOS transistor depends on the drain current and the W/L ratio. This gives the designer of MOS circuits the advantage of conveniently being able to vary the transconductance and current independently by varying the size. The transconductance of a MOS transistor is affected by the temperature, but this is a second order effect. As the transconductance of a bipolar transistor is directly related to the ambient temperature, bipolar designers must take additional care that their circuits will develop adequate transconductance at high temperatures, both ambient and those due to current dissipation in the device. The relationship between the transconductance of a transistor and the current flowing through it was simulated for a number of MOS transistors and an NPN bipolar transistor using the SPICE parameters in Appendix A. The results for currents of $20\ \mu\text{A}$ to $200\ \mu\text{A}$ have been plotted in fig. 2.8. The models used do

Figure 2.8: Transconductance vs Current for Several Transistors

not accurately model the subthreshold behavior of the MOS transistors and thus, the transconductance at very low current levels has not been plotted. It can be seen from fig. 2.8 that at typical current levels used in analog circuits, the MOS transistor can only approach the transconductance of the bipolar transistor by using large W/L ratios. In addition to the costs of increased silicon area, large transistors are undesirable for high frequency application as much of the capacitance associated with the device is proportional to the area consumed by the device. Compared to MOS transistors, bipolar transistors are quite compact and as such have relatively small intrinsic and parasitic capacitances. It is desirable to have small capacitances as this will increase the frequency of non-dominant poles, thus limiting their effect on the stability of the final circuit. Where frequency of operation is concerned, bipolar transistors have the advantages of high transconductance and small capacitances when compared to MOS

devices.

Current Handling Ability

Current handling ability is of concern in analog integrated circuit design, both for the need to drive large loads as might be encountered off chip, and for the need to slew load capacitances rapidly. Bipolar transistors can operate at considerably higher current densities, thus they can handle larger currents than MOS transistors in a reasonable amount of space. As was mentioned above, excessively large transistors can introduce parasitic poles at low frequencies due to the large capacitances. This serves to reduce the stability of the circuit in which the device is used, thus such large transistors are undesirable.

Output Conductance and Voltage Gain

In high frequency analog design, it is necessary to utilize circuits having large voltage gains. To achieve a large voltage gain, it is necessary to have devices available with low small-signal output conductance. Low output conductance is also desirable when designing current sources or current mirrors. The output conductance of a $1 \times 3 \mu\text{m}$ NPN transistor and $3/1 \mu\text{m}$ NMOS and PMOS transistors from a hypothetical process that will be described in Section 2.3 has been plotted in fig. 2.9 below. It can be seen from fig. 2.9 that the output conductance of MOS transistors is somewhat worse than that of bipolar transistors. It was mentioned in Section 2.1.2, the output conductance of a MOS transistor increases as the channel length is decreased. Thus it is possible to improve the output conductance of MOS transistors by increasing the channel length. This approach has drawbacks as the silicon area use and capacitances will be substantially increased. In most applications, frequency of operation is of greater concern than output conductance or gain, thus the use of short channel MOS transistors is normally preferred.

Figure 2.9: Comparison of Output Conductances

Transconductance and Current Level

As has been mentioned above, the transconductance of a bipolar transistor depends primarily on the current through the device and the temperature of the device. Of these factors, the designer only has control over the current. This is a drawback as it is often necessary to operate opamp stages at high current levels in order to have adequate slew rate. In many cases, the resulting gain will be sufficient to allow the unity gain bandwidth to increase to the point that parasitic poles affect the stability of the circuit. This is not a problem in MOS design as it is possible to choose the current level and the transconductance of a transistor independently by changing the (W/L) ratio of the transistor. In bipolar design, it is necessary to resort to emitter degeneration to limit the gain of a circuit. An example of an input stage employing emitter degeneration is shown in fig. 2.10. Unfortunately, there are a number

Figure 2.10: Input Stage With Emitter Degeneration

of drawbacks to the use of emitter degeneration. As resistors are used for emitter degeneration, it is necessary that resistors of an appropriate size range be available in the process. This usually involves one or more additional masks, thus increasing the cost of the process. The tolerance on the value of most readily available integrated circuit resistors is quite large, requiring some circuit performance to be traded off against the need for the circuit to work under conditions of extreme processing.

Noise

There are several sources of noise that have been observed in electronic devices. Of these noise mechanisms, there are three that are of interest in the design of analog integrated circuits. The noise mechanisms of interest are shot noise, thermal noise and flicker noise. Each of these mechanisms and their applicability to MOS and bipolar transistors will be discussed below.

Shot noise occurs whenever charged particles cross a potential barrier as occurs in the semiconductor junctions of a bipolar transistor. Noise arises due to the fact that the passage of an electron or hole is a random event dependent on the kinetic energy of the carrier. Thus small fluctuations in the total current may be observed. The mean square of the shot noise current of a semiconductor junction may be shown to

be [5]

$$i_{nT}^2 = 2qI\Delta f \quad (2.25)$$

Shot noise may be observed in both the base and collector currents of a bipolar transistor. Current conduction in MOS transistors is due to drift and not carriers crossing semiconductor junctions, thus shot noise is not observed in MOS transistors. In order to compare the magnitude of bipolar shot noise to thermal noise in MOS transistors, it is convenient to express the shot noise of the collector current as a noise voltage applied to the base of the bipolar transistor. Thus the 'base referred' shot noise of the collector current is given by

$$v_{nT}^2 \approx (i_{nT}/g_m)^2 = \frac{2kT\Delta f}{g_m} \quad (2.26)$$

Thermal noise is due to random thermal motion of carriers within a conductor having finite resistance. Since there are free carriers in the active area of a MOS transistor, and there are parasitic resistances in both devices, there will be thermal noise present. The mean square of the thermal noise voltage for a resistor may be shown to be on the order of

$$v_{nT}^2 = 4kTR\Delta f \quad (2.27)$$

As current conduction in MOS transistors involve free carriers in a conductor having finite resistance, equation 2.27 can be applied with a suitable equivalent for R . In the case of a MOS transistor, R is given by the incremental channel resistance [8] and is approximately

$$R \approx \frac{2}{3g_m}$$

thus the mean square of the "gate referred" thermal noise voltage is approximately

$$v_{nT}^2 = \frac{8}{3} \frac{kT\Delta f}{g_m} \quad (2.28)$$

Thermal noise is also generated in the source and drain resistances of a MOS transistor and in the base, collector and emitter parasitic resistances of a bipolar transistor.

It can be seen in equations 2.26 and 2.28 that the input referred shot noise for both bipolar and thermal noise for MOS transistors is in inverse proportion to the transconductance of the transistor. As has been discussed above, bipolar transistors typically operate with much higher transconductances than MOS transistors and thus have much lower levels of shot noise than MOS transistors have of thermal noise.

Flicker noise is due to electron energy states or traps that exist at the Si/SiO₂ interface at the surface of the semiconductor or at crystal defects in the bulk. These states trap and release electrons resulting in noise. This process is relatively slow, thus the noise energy is concentrated at low frequencies and tends to display a $1/f$ relationship, and is often referred to as $1/f$ noise. As MOS transistors are surface devices, it follows that they will exhibit high levels of flicker noise due to the electron energy states at the Si/SiO₂ interface. For typical processes, flicker noise becomes dominant at frequencies below approximately 1 kHz. The mean square of the noise voltage due to flicker noise of a MOS transistor may be approximated as [8]

$$v_{nf}^2 = \frac{K}{C_{ox}WL} \frac{\Delta f}{f} \quad (2.29)$$

It can be clearly seen from equation 2.29 that the level of flicker noise is related to the gate area of the transistor and may be minimized by using a long channel length thus increasing the gate area. The tradeoffs in doing this have been described above. Flicker noise may be observed in bipolar transistors due to crystal defects in the semiconductor. As defect levels are low in modern processes, the noise energy due to flicker noise in bipolar transistors is small and is generally ignored.

It can be concluded from the above that for devices considered in isolation, bipolar transistors have more desirable noise characteristics due to the absence of significant flicker noise and the lower level of shot noise as compared to thermal noise in MOS transistors. In applications where a transistor is driven by a high impedance signal source, it is possible that a MOS transistor will offer better noise performance. This is because the high source impedance will add thermal noise to the base input of the bipolar transistor which is a low impedance node. In a MOS transistor, the gate is a high impedance node and will be less affected. When designing amplifiers bipolar transistors have the additional advantage that the high gain available in an input

stage allows the effects of subsequent stages to be considerably reduced. For the reasons outlined above, bipolar transistors have many advantages where circuit noise is of concern.

Input Impedance

One serious limitation of bipolar transistors is the fact that they have both a low small signal input impedance and the requirement for a DC bias current on the input. This makes them unsuitable for switched capacitor circuits as such circuits depend on samples of the signal stored on capacitors which would be discharged by the DC base current. Any significant discharge of the capacitors during the sampling period will introduce errors into the output signal. MOS transistors have an effectively infinite input impedance, thus the capacitors will not discharge significantly during a clock period. The low input impedance of bipolar transistors is also of concern in continuous circuits as a bipolar transistor operating at high current levels may undesirably load down the previous stage. The input impedance of a bipolar transistor may be improved by the use of emitter degeneration or a Darlington connected device. The Darlington configuration shown in fig 2.11 will also reduce the DC bias current drawn from the input or previous stage. The drawbacks of the Darlington connection include additional poles, reduction of the voltage gain by at least half and the necessity to maintain a base-emitter voltage double that required by a single transistor.

Emitter degeneration was discussed previously as a technique to limit the transconductance of a bipolar transistor operating at a given current level. When viewed at the base, the emitter resistance appears to be multiplied by the β of the transistor, thus increasing the input impedance. An example of an input stage employing emitter degeneration has been shown in fig. 2.10. The drawbacks of emitter degeneration include the fact that for practical amounts of degeneration, the input impedance is still low, and the gain is considerably reduced. Other considerations involved in the use of emitter degeneration have been discussed previously in this section.

Much of the DC bias current drawn by a bipolar input stage can be supplied by a suitable current mirroring network, however there will still be a non-zero bias current

Figure 2.11: Darlington Connected Input Stage

that must be supplied from the previous stage, and this technique does nothing to improve the input impedance. The low input impedance and the requirement for DC bias current of most bipolar opamps renders them unsuitable for sampled data applications.

Complementary Devices

When choosing between bipolar and MOS transistors for a particular application, it is necessary to consider the requirement for complementary devices. For reasons that will be discussed in the next section, most bipolar processes do not offer a complementary device, or the complementary bipolar device is considerably inferior to the NPN transistor. In contrast, all CMOS processes offer a complementary device, and the characteristics of PMOS and NMOS transistors are acceptably similar in most processes.

It can be seen in fig. 2.8 above that the transconductance available from a PMOS transistor is less, but of the same order as the transconductance available from an NMOS transistor of the same size operating at the same current. The transistor characteristics for an NMOS transistor and a PMOS transistor are shown in fig. 2.12

Figure 2.12: 1x3 μ m NMOS Transistor Characteristics

and fig. 2.13, with both figures plotted to the same voltage and current scales. It can be seen that the current levels obtained for a given gate-source voltage from a PMOS transistor are roughly one third that obtainable from an identically sized NMOS transistor. This is not ideal, however it is somewhat better than can be expected in bipolar processes that do not provide a complementary vertical PNP transistor. Due to manufacturing difficulties, it is frequently the case that such a complementary vertical PNP transistor is not available.

One may also compare the output resistances or conductances of the two complementary MOS devices. The output conductance of the two MOS devices, along with that of an NPN transistor, was plotted in fig. 2.9. It can be seen from fig. 2.9 that the output conductances of the PMOS transistor is better than that of the NMOS transistor, however neither is as good as that of the NPN bipolar transistor.

Figure 2.13: 1x3 μm PMOS Transistor Characteristics

2.2 MOS and Bipolar Integrated Circuit Processes

Typical MOS and bipolar integrated circuit processes will be described. Emphasis will be on device complement and process steps as the comparison between technologies is primarily centred on the comparison between types of device, which has been covered in some depth above. It should be noted that many of the processing steps and techniques used in the two technologies are similar. Processing steps and considerations for a BiCMOS technology will then be discussed.

2.2.1 Bipolar Integrated Circuit Processes

For reasons discussed above, bipolar processes are well suited to applications requiring high frequency operation, high gain, low noise, or the driving of large loads.

Device Complement

The primary device in most bipolar processes is the vertical NPN transistor. Other devices available include lateral PNP transistors and Schottky clamped vertical NPN transistors. The latter of these are used mainly in digital circuits. The devices found in typical modern bipolar processes are very similar to the bipolar devices in the generic BiCMOS process to be described below. The reader is referred to Section 2.3 for a more detailed description. One serious limitation of bipolar technology from the point of view of analog design is the lack of a true complementary device in most processes. Lateral PNP transistors are available in most modern bipolar processes, but typically have gains and f_T 's considerably less than that of the NPN transistors available. High quality complementary vertical PNP transistors are possible, however they are often omitted as they require additional process steps and thus are costly.

Linear resistors are commonly available in bipolar processes for use in biasing and emitter degeneration. These resistors may be diffused, polysilicon or precision resistors of various materials that are often trimmed after fabrication. Capacitors are usually available for use in compensation of stability of amplifiers and realization of continuous filters. These capacitors are usually of the Metal-Insulator-Semiconductor variety.

Process Flow

The process sequence and masks for a typical modern bipolar process are shown in Table 2.1. This process features two levels of metal and requires a total of 12 photolithographic masks. Table 2.1 describes the processing steps necessary to fabricate a vertical NPN transistor. By suitable design of the masks, it is possible to form most other devices in the process using these same masks and process steps. High

Mask Number	Process Step
1	Define and Implant N+ Underlay Underlay Drive and Anneal
2	Define and Implant P+ Isolation Isolation Drive and Anneal Deposition of Epitaxial Layer
3	Define Sunken Field Oxide Regions
4	Define N+ Collector Sinker and Implant
5	Define and Implant Intrinsic Base
6	Define and Implant Extrinsic Base
7	Define and Implant Emitter
8	Define and Etch Contact Windows Deposit Barrier Metal
9	Define First Level Metallization and Etch Deposit Intermetallic Dielectric
10	Define Vias and Etch
11	Define Second Level Metal and Etch Deposit Passivation
12	Define and Etch Pad Openings

Table 2.1: Typical Bipolar Process Flow and Masks

performance processes may have additional masks and steps in order to implement features such as precision resistors or advanced isolation schemes.

Advantages and Drawbacks

The advantages of bipolar technology include the ability to drive large capacitive loads, the ability to operate at high frequencies and the compactness of the layout for analog or high power circuitry. The drawbacks of bipolar technology include a low input impedance, thus precluding its use in sampled data applications, and the lack

of a complementary device in most cases.

2.2.2 CMOS Integrated Circuit Processes

Device Complement

The primary devices in CMOS integrated circuit design are n and p channel MOS field effect transistors. CMOS processes intended for analog applications usually provide a capacitor which is formed with two layers of polysilicon with a dielectric, usually oxide in between. This offers a capacitance which is independent of voltage and has good process control. Resistors are not normally needed in CMOS circuits and are generally not provided. Where resistors are needed, device wells or gate polysilicon are used; however the tolerances and parasitics on such resistors can be troublesome.

Process Flow

The process flow and masks required in a typical modern CMOS process are described in Table 2.2. The process features two levels of metallization and requires 12 photolithographic masks. Table 2.2 describes the processing steps necessary to fabricate all devices offered in a typical CMOS process. It should be noted that where linear capacitors are not required, mask # 5 and its associated steps may be omitted.

Advantages and Drawbacks of CMOS

There are three principal advantages of CMOS technology over bipolar technology for analog applications. MOS transistors have an effectively infinite input impedance which makes CMOS technology ideal for sampled data applications. CMOS processes have the advantages of offering a complementary device whose performance is reasonably well matched to the primary device. MOS transistors can be easily scaled, and the transconductance is variable independently of the current level. These qualities make CMOS the preferred technology for most monolithic analog systems. The

Mask Number	Process Step
1	Define and Implant N Well Grow Protective Oxide Over N Wells P Well Implant on Areas That Are Not N Wells
2	Define Device Wells and Grow Nitride Etch Areas That Are Not Device Wells
3	P Guard Etch and Implant (NMOS) Grow Field Oxide Strip Nitride Over Device Wells Grow Gate Oxide Threshold Adjust Implant Gate Poly Silicon Deposition, Doping and Anneal
4	Gate Poly Etch Capacitor Poly Deposition, Doping and Anneal
5	Capacitor Poly Etch
6	N+ Source/Drain Implant and Anneal
7	P+ Source/Drain Implant and Anneal
8	Define and Etch Contact Windows Deposit Barrier Metal
9	Define First Level Metallization and Etch Deposit Intermetallic Dielectric
10	Define Vias and Etch
11	Define Second Level Metal and Etch Deposit Passivation
12	Define and Etch Pad Openings

Table 2.2: Typical CMOS Process Flow and Masks

primary drawbacks of CMOS technology are the limited ability to drive large loads, relatively large parasitics and noise. These advantages and drawbacks have been discussed in depth in Section 2.1.3.

2.2.3 BiCMOS Integrated Circuit Processes

The term BiCMOS process refers to an integrated circuit process which has available at least some of the devices and features normally found in each of bipolar and CMOS processes. The set of features in any particular process is dependent upon the applications for which the process is intended and cost constraints.

There are a number of motivations for the development of a BiCMOS process. In the case of digital circuits, BiCMOS offers the ability to combine CMOS and ECL logic on a single chip, where critical paths are implemented in ECL, and the remainder of the circuitry is implemented in CMOS to reduce power consumption. BiCMOS also offers the opportunity to use a combination of CMOS, TTL and ECL I/O on a single chip. In the area of mixed analog/digital integrated circuits, BiCMOS technology has the potential to allow high quality analog components such as amplifiers, references and analog/digital converters to be implemented using bipolar circuitry on the same chip as CMOS digital control circuitry. This has the potential to allow single chip systems where in the past, multi chip systems were required, thus reducing the total system cost [11].

There are three major approaches to designing a BiCMOS process. Each of the strategies, its features and possible motivations will be described below.

Minimum Cost BiCMOS Process

A minimum cost BiCMOS process consists of a CMOS process to which NPN transistors are added with the minimum possible number of additional process steps, and hence, the minimum cost. Typically a single mask and implant are added to form the bases of NPN transistors, whose collector is an N well otherwise used for PMOS devices, and whose emitter is an n type diffusion as used for an NMOS source or drain.

This approach severely compromises NPN transistor performance for a number of reasons. As n-wells in most CMOS processes lack a buried layer, the collector resistance of the resulting NPN transistor will be quite large, and may very well limit the applications of such transistors to those that do not require the rapid slewing of a capacitance. The NPN transistor performance will further be hampered by the fact that the doping profile of a MOS drain diffusion will, in all likelihood, be far from optimal for an NPN emitter. The NPN transistors are self isolating due to the reverse bias junction between the n-well and the p-substrate, as is the case for MOS transistors in the same process.

This approach gives a BiCMOS process at minimum cost, however there are severe performance penalties to be paid that would likely limit the applications to those than make minimal demands on the bipolar transistors such as voltage references.

Analogue Compatible BiCMOS Process

This type of process is typically used in interface work where it is necessary to operate at higher voltages than the 5 Volts that is typical for digital circuits. Telecommunications applications where these requirements may exist include telephone line cards and integrated circuits that are required to drive transducers such as speakers directly. The higher operating voltage allows a wider dynamic range and increased noise margins. In order to allow operation at higher voltages, the MOS devices are designed with a longer minimum channel length than is the case in digital processes. As this severely limits the achievable logic density, it is desirable to limit the use of logic circuitry in a process such as this to that used to control the analog blocks or to interface to them. As this type of process is motivated by analog concerns, components such as double polysilicon capacitors, precision resistors, and good quality PNP transistors are frequently available. The PNP transistors may be any or all of lateral, substrate or complementary vertical. The number of masks required in such a circuit can vary widely, depending on the features and devices chosen. As circuit density is of less concern than in digital processes, self isolation is used in order to avoid the costs of more advanced isolation schemes. This type of process is motivated by somewhat

specialized concerns and is usually not intended to operate at high frequencies.

High Performance BiCMOS Process

High performance BiCMOS processes are used where performance is of greater concern than cost. The generic BiCMOS process to be described in Section 2.3 is an example of such a process. High performance processes feature both high performance NPN bipolar transistors and high performance MOS transistors. Depending on the motivating factors, the process may employ advanced isolation schemes, complementary devices and passive components. One possibility to accommodate both the need to have a device set suitable for the design of high quality analog circuitry and the need to minimize process cost and complexity for digital circuits is to make some of the process features optional as has been done by Northern Telecom [10]. Such a process is then suitable, both for the design of analog signal processing blocks, as well as the high speed digital circuitry required in the switching and transmission of voice and data in a modern telecommunications system. In order to satisfy the needs of high speed logic circuitry, most high performance BiCMOS processes are designed to operate at the industry standard 5 Volts. This does complicate analog design, however the problems are not insurmountable, and analog circuitry has been used in 5 Volt CMOS for some time.

2.3 A Generic BiCMOS Process

The process that will be used for purposes of explanation in this thesis is a 1 micron high performance analog BiCMOS process. Due to the proprietary nature of most BiCMOS processes in use at the time of writing, it is necessary to resort to the use of a hypothetical process. This process was used in course work [9] and is somewhat similar, at least in terms of device complement, to the process in which fabrication was attempted. The basic features and the device complement of this process will be discussed in the following paragraphs. The process was supplied as a set of SPICE parameters without specific design rules. These SPICE parameters

are included in Appendix A. Thus many of the features described below are typical of high performance processes and are assumed to be present. A copy of this set of SPICE parameters may be found in Appendix A. Following the discussion of process features, the performance of the various available devices will be discussed.

2.3.1 Device Complement

This process can best be described as a high performance 5 Volt analog/digital BiCMOS process. The minimum feature size is $1\text{ }\mu\text{m}$ for both NPN bipolar transistors and MOS transistors.

Vertical NPN Transistor

One of the primary motivations for a process such as this is the desire to produce digital circuits having high speed or high drive capability, while retaining the low power consumption and high circuit density given by CMOS technology. For this reason, a vertical NPN transistor is provided. A plan view and a cross section of this device are shown in fig. 2.14. The minimum emitter window size in this process is $1\text{ by }3\text{ }\mu\text{m}$ and this will be referred to as the unit transistor in later simulations. In addition to the structures necessary to make a simple NPN transistor, this device includes a number of features intended to maximize performance. These features are described below.

Reduction of Parasitic Resistances Where high current operation is of concern, it is desirable to minimize the collector series resistance of the transistor. This resistance is due to the electrical path through the collector diffusion that the carriers must traverse from the collector p-n junction to the collector contact at the surface. The resistance of a semiconductor is related to its doping level such that heavily doped material has much less resistance. There is an upper bound to the doping level that may be used in the collector diffusion due to the need to form the p type base within it as a subsequent fabrication step. One approach to reducing the collector series resistance is the use of a heavily doped buried layer, referred to as 'n underlay'

Figure 2.14: Vertical NPN Transistor

in fig. 2.14. This is formed by implanting a heavily doped N^+ diffusion in the surface of the semiconductor. An n-type epitaxial layer is then grown over the buried layer. The remainder of the transistor is formed in the epitaxial layer in the usual manner. The buried layer provides a lower resistance path for carriers moving parallel to the base, horizontally in the case of fig. 2.14. The collector series resistance may be further reduced by forming a heavily doped N^+ sinker or plug below the collector contact, thus providing a low resistance path from the contact to the buried layer. The vertical NPN transistor in the process uses both of these techniques in order to

minimize the collector series resistance.

Where high speed operation is of concern, it is important to minimize the series resistance between the contact to the extrinsic base and the intrinsic base. As the impurity concentration in the intrinsic base is determined by the desired junction properties, little can be done to reduce its bulk resistance. Reductions in the base series resistance may be achieved by the selection of a higher impurity density in the extrinsic base, thus lowering its bulk resistance and by silicidation of the surface of the extrinsic base. Silicidation involves metallization of the surface of diffusion or polysilicon with either Nickel or Titanium silicide which provides a low resistance path for current to flow along the surface, thus reducing the overall resistance of the diffusion. Silicidation also provides a lower resistance contact than is attainable with a diffused ohmic contact. In the case of fig. 2.14, the above techniques contribute to a reduction in the series resistance seen between the base contact and the intrinsic base.

Polysilicon Emitters A frequently employed technique to improve the performance of bipolar transistors is the use of polysilicon emitters. Such an emitter consists of a thin emitter diffusion with n-doped polysilicon deposited on top. A cross sectional view of such a structure is presented in fig. 2.15. This approach allows a

Figure 2.15: Polysilicon Emitter Structure

shallow emitter which will not be degraded by spikes from a metal contact as the

ohmic contact is made to the n polysilicon above the shallow emitter. The use of polysilicon emitters allows the choice of increased current gain or increased current handling capability as higher doping levels in the base may be used. There is a thin native oxide present at the interface between the single crystal emitter diffusion and the polysilicon. This thin oxide and dopant segregation to grain boundaries serve as a barrier to collector current, thus reducing the base current level. This gives the transistor better characteristics under high level injection, resulting in less of a decrease in the current gain at high current. Polysilicon emitters are assumed to be employed in this process.

Isolation In order to get maximum speed it is desirable to minimize the capacitances associated with the transistors. In order to maximize circuit density, and thus minimize cost, it is desirable to be able to locate individual transistors as close to each other as possible. The traditional technique is to surround the transistor with a p-type diffusion that extends through the epitaxial layer and reaches to the p-type substrate. This has the effect of surrounding the transistor with a reverse biased p-n junction. The drawbacks of this approach are relatively large sidewall capacitances due to the reverse biased junction and leakage currents which necessitate increased spacing between devices.

A technique that may be employed to improve isolation is trench isolation or dielectric isolation. A cross section of such a scheme is shown in fig. 2.16. The

Figure 2.16: Trench Isolation Structure

trench is formed by the use of an anisotropic etch. This results in a deep V shaped groove in the silicon. Oxide is grown on the walls and bottom of this groove and the trench is then filled with polysilicon. This type of isolation effectively eliminates the sidewall capacitance of the NPN collector, resulting in a considerably lower parasitic capacitance for the device. Isolation trenches are compact and have very low leakage currents, thus allowing a very dense layout. The process described here is assumed to employ trench isolation. The location of isolation trenches with respect to the active areas of an NPN transistor is shown in fig. 2.14.

PNP Transistors

A complementary, or vertical PNP transistor is not available in this process. This is due to the technical difficulties and cost of providing such a device, and to the fact that such devices are generally not used in high speed digital circuits. There is, however a lateral PNP transistor available as this can be fabricated using masks and process steps required for other devices. Plan and cross section views of this device are presented in fig. 2.17. This device has a number of drawbacks due to its structure. Since the active areas of the transistor are near to the surface, the lateral PNP transistor will be subject to many of the surface phenomena that affect MOS transistors. For high performance, a bipolar transistor requires a narrow base width. This is relatively easy to achieve in a vertical device as the base width is determined by ion implantation which can be well controlled. In the case of a lateral device, the base width is determined by photolithography. This has dual drawback of being less well controlled than an ion implanted layer, and the minimum reliable width is much greater than that available with a vertical device. The lateral PNP transistor is at a further disadvantage in typical processes in that the details of the processing steps from which it is formed are optimized for other devices and are likely less than optimal for the performance of the lateral PNP. Due to reasons discussed above, the lateral PNP transistor usually has low gain and low frequency response. Its use would generally be limited to low frequency applications where gain is not important, such as current sources. This device has not been employed in any of the circuits to be

Figure 2.17: Lateral PNP Transistor

discussed later in this thesis.

MOS Transistors

Both n and p channel MOS transistors are available in this process. Cross sections and plan views of the n and p channel MOS transistors are shown in fig. 2.18 and fig. 2.19.

The p channel MOS transistors feature a p+ polysilicon gate. This produces a more controlled p channel transistor than can be obtained using the n+ polysilicon

Figure 2.18: N Channel MOS Transistor

used for the n channel device gates. Both devices feature lightly doped drains [14] which are formed by introducing lightly doped regions between the channel and the source and drain diffusions. The lower doping concentration at the junction results in a wider depletion region and thus a reduction in the magnitude of the electric field at the drain pinchoff region. This lower electric field increases the breakdown voltage and reduces impact ionization and thus the injection of hot carriers, allowing the transistor to safely operate with a shorter channel length for a given voltage.

This process features twin wells which involves separate wells for each type of device in a thin epitaxial layer. This is convenient as the epi layer is necessary for the

Figure 2.19: P Channel MOS Transistor

fabrication of the vertical NPN transistors. As was mentioned above in the case of the NPN bipolar transistor, silicides are used to minimize the resistance of polysilicon and diffused layers. Due to the short channel length used, the process uses a very thin gate oxide in an attempt to maximize the apparent electrical length of the channel. It should be noted that many or all of the MOS transistor features mentioned above would also be found in a pure CMOS process of similar scale.

Other Active Devices

There are a few devices likely to be provided in a high performance BiCMOS process that are mostly of interest for the design of digital circuits. Schottky diodes have a lower forward bias voltage than junction diodes and are useful in clamping NPN transistors, thus limiting the degree of saturation of the transistor. Substrate diodes with the anode connected to the negative power supply are available and are useful in the design of overvoltage and undervoltage protection circuitry. It may be assumed that these devices are available in the hypothetical BiCMOS process however they are not normally useful in the design of amplifiers and will not be discussed further.

Passive Devices

This process has all of the passive device features available in a CMOS process of similar scale. These include linear capacitors (polysilicon - oxide - polysilicon) and resistors. The resistors may be any of diffused resistors, polysilicon resistors, thin film or other varieties. As details of the composition and performance of the resistors and capacitors in the hypothetical process are not available, for design purposes they will be assumed ideal.

2.3.2 Summary of Generic BiCMOS Process

The features available in this hypothetical process have been described. The process can be thought of as a 1 micron CMOS process as might be employed in telecommunications applications, with the addition of the bipolar devices mentioned above. This process has been used to generate the examples of bipolar and MOS transistor performance in Section 2.1.3, using the SPICE parameters presented in Appendix A. Thus the performance of the devices in this process is presented there. This process as described will be used for the design of a novel operational amplifier, and will also be used in the discussion of bipolar and CMOS devices in general.

2.4 Summary

In this chapter, the device characteristics and process flow for MOS field effect transistors and bipolar transistors were discussed. Small signal models and design equations including device current, transconductance, output resistance and frequency figures of merit were presented for both MOS and bipolar transistors. The performance of the two types of device were compared in several areas of interest including device size, current carrying ability, output resistance, transconductance, noise and input impedance.

The device complement and typical process flows for bipolar and MOS integrated circuit processes were discussed. Similarities between the two were noted and three possible approaches to a BiCMOS process were discussed.

The physical characteristics and features of a number of devices forming a hypothetical BiCMOS process were discussed. This process forms the basis of operational amplifier designs that will be presented in later chapters.

Chapter 3

A Consideration of Operational Amplifier Topologies

The design of operational amplifiers in both bipolar and CMOS technologies has been well covered in the literature [6, 18]. General considerations in operational amplifier design and some of the circuits more commonly in use will be discussed. The use of such circuits as part of monolithic analog systems will be discussed for each technology. Following this background discussion, a circuit topology suitable for use as a high bandwidth BiCMOS operational amplifier will be proposed. This design will be discussed in detail in the next chapter.

3.1 Operational Amplifiers

An operational amplifier is a voltage amplifier with differential inputs, single-ended output, high gain and bandwidth, high input impedance and low output impedance. An ideal opamp would have infinite voltage gain and bandwidth, infinite input impedance and zero output impedance. Obviously such a device does not exist, however an objective of operational amplifier design is to minimize the non-idealities of the circuit in the context of the intended application. The schematic symbol for an opamp and an equivalent circuit for an ideal opamp are shown in fig. 3.1.

The motivation for approximating an ideal opamp is to allow the performance of a

Figure 3.1: Ideal Operational Amplifier and Equivalent Circuit

circuit using the opamp to be determined by the components in the feedback circuit, which is the case if the opamp is sufficiently close to ideal. An example of this will now be presented. The equation describing the operation of the ideal opamp shown above in fig. 3.1 is

$$V_{out} = A(V_+ - V_-). \quad (3.1)$$

Now consider this same opamp in a feedback circuit such as the inverting amplifier configuration shown in fig. 3.2. As was stated above, an ideal operational amplifier

Figure 3.2: Inverting Amplifier Configuration

has infinite input impedance, ie. no current flows into either input. Thus the sum of the currents through the input and feedback resistors as shown in fig. 3.2 must be zero. This may be expressed as

$$\frac{V_{in} - V_-}{R_{in}} + \frac{V_{out} - V_-}{R_{fb}} = 0.$$

From equation 3.1, the output voltage of the circuit in this configuration is

$$V_{out} = -AV_-$$

The above two equations may be combined and manipulated to yield

$$A_V = \frac{V_{out}}{V_{in}} = -\frac{R_{fb}}{R_{in} + \frac{R_{in} + R_{fb}}{A}} \quad (3.2)$$

It can be seen from equation 3.2 that for large values of gain A , the gain of the inverting amplifier is simply the ratio of the feedback resistance to the input resistance.

$$A_V = \frac{V_{out}}{V_{in}} \approx -\frac{R_{fb}}{R_{in}}$$

3.1.1 Operational Amplifier Non-Idealities

There are a number of ways in which the performance of an actual operational amplifier may differ from the ideal. These will be briefly described below.

Input Offset Voltage Ideally, if the two input terminals of an opamp are shorted together, the output voltage should be zero. This is not likely to occur in an actual operational amplifier due to mismatches between devices, and causes intrinsic to the circuit such as the use of a bipolar current mirror to convert a differential signal to single-ended. It is usually more convenient to refer the offset to the input, thus the input offset voltage is that voltage that must be applied across the input terminals in order to reduce the output voltage to zero. MOS operational amplifiers typically have much higher input referred offset voltages than bipolar opamps. This is due both to the fact that bipolar opamps typically have much greater open loop gains and that matching between bipolar devices is generally better than matching between MOS devices. The higher open loop gain of bipolar operational amplifiers serves to minimizing the value of the offset voltage when it is referred back to the input.

Input Bias Current In bipolar opamps, it is necessary to supply a DC bias current to the inputs in order to keep the input transistors operating in the active region. The magnitude of this current depends on the current gain of the transistors used, the

current level at which they operate, and the use of Darlington inputs or input current cancellation schemes. Input bias current is not of concern with opamps employing MOS input devices.

It is also possible for there to be a mismatch between the current gains of the bipolar input transistors of an opamp. Thus there will be a difference in the bias current required by each input. This is only a concern with opamps employed in systems where the differential input properties are required. In virtual ground applications mismatch between the bias current levels is not of concern.

Input Impedance It is desirable to have a very high input impedance in an operational amplifier in order to minimize any loading effect that the input impedance might have on circuit elements external to the opamp. This can be a problem when considering the use of bipolar opamps due to the low input impedance of bipolar transistors. The effects of a non-infinite input impedance become less significant as the open loop voltage gain increases, thus it is often acceptable to use bipolar transistors as input devices due to the high gains they offer. As MOS transistors have an effectively infinite input impedance, input impedance is not a concern in opamps employing MOS input devices. Such opamps have the additional advantage of being usable in switched capacitor circuits.

Output Impedance All real operational amplifiers will have a non zero output impedance, usually mostly resistive, which is dependent upon the design of the output stage of the opamp. Large output resistances affect the speed at which an opamp may charge a capacitive load, which limits the frequency of operation of the system. Generally, opamps employing bipolar output stages have a somewhat lower output resistance than those employing MOS output stages. This is largely because bipolar opamps are typically designed to drive off chip resistive loads whereas MOS opamps are designed to drive on chip capacitive loads, rather than any inherent differences between bipolar and MOS transistors.

Input Common Mode Range Input common mode range refers to the range of input voltage that may be applied to an opamp without causing any of the gain stages inside to leave their high gain regions, or any of the current sources in the amplifier to operate outside of those regions of operation where their output current is approximately constant. In the case of bipolar technology, the devices in the gain stage may not be allowed to saturate, break down, or enter the cutoff region. In the case of MOS technology, all devices in the gain stages must remain in saturation. The input common mode range is also referred to as linear range, which can be described as range of input voltages over which the output voltage remains an approximately linear function of the input voltage.

Common Mode Rejection Ratio One of the properties of an opamp is the ability to reject common mode signals. Ideally, the common mode gain should be small, preferably zero. In real opamps, the common mode gain is finite, and the common mode rejection ratio may be defined as the ratio of the differential open loop gain to the common mode open loop gain. Common mode rejection is of concern in many analog systems as interference and noise will often manifest themselves as common mode signals on a pair of lines carrying the desired difference mode signal.

Power Supply Rejection Ratio In most opamp circuits, it is possible for noise signals present on the power supply busses to be propagated through an opamp to its outputs. This is particularly a problem in mixed analog and digital integrated circuits as the switching of logic gates can cause current spikes on the power supplies. The power supply rejection ratio is defined as the ratio of the open loop difference mode voltage gain to the voltage gain from the power supplies to the output. The power supply rejection ratio will likely be different for each of the power supplies.

Open Loop Voltage Gain Real opamps have a finite open loop voltage gain. High open loop gain is desirable as it improves the steady state accuracy of a feedback system using the opamp. For typical operational amplifiers, the open loop gain is on the order of 40 to 100 dB. MOS opamps tend to be at the lower end of this range

while bipolar opamps tend to be towards the higher end of this range. This is due largely to the greater transconductance of the bipolar differential pair used in the input stage.

Unity Gain Bandwidth A consequence of finite DC gain in an opamp and the need for single pole response is the fact that the gain of a real opamp decreases as the frequency of the signal is increased. The unity gain bandwidth of an opamp is defined as the frequency at which the open loop gain drops to unity. The values of unity gain bandwidth can range from a few MHz to about 100 MHz in currently available opamps.

Slew Rate The speed at which an opamp may change the voltage of a load capacitor is known as the slew rate. It is of concern since in switched capacitor circuits, it is necessary that the output of one stage reaches its final value and settles before it is sampled by the next stage. The slew rate is typically determined by the size of the load or compensation capacitor and the current level in the stage driving that capacitor. Normally the opamp is designed such that the slew rate will be limited by a specific stage. It is desirable to have the positive and negative going slew rates to be equal, however they may be different, depending on the design of the opamp.

Settling Time In switched capacitor circuits, whenever a new sample is applied to the opamp, it is a step input. As with any other feedback system, there is the possibility of some amount of ringing when the output reaches its final value, depending on the value of the damping coefficient of the loop. The value of the damping coefficient depends on the characteristics of both the opamp and the feedback network. The settling time is defined as the time taken for the output of an opamp to settle to within some specified fraction of its final value. Settling time is of concern as it is necessary for the output of a stage of a switched capacitor circuit to reach its final value and settle before it is sampled by subsequent stages, otherwise non-linear distortion will take place.

Dynamic Range Dynamic range is normally defined as the ratio of the highest signal level that may be passed by a circuit or system, and the lowest signal level that may be used. The lower limit on the signal level is due to the necessity of having the signal at the input greater than the input referred noise. The upper limit on the input signal amplitude is due to the need to avoid non-linear distortion.

Noise In opamps, noise is generated by all active devices in the circuit, as well as any passive resistors that may be present. Due to the manner in which noise sources cascade, devices closest to the input of the opamp are of the greatest concern with regards to noise. Similarly, devices not on the signal path usually have negligibly small contributions to the overall noise of the opamp. The noise generated in an opamp is normally referred back to the input.

3.2 CMOS Operational Amplifiers

The majority of CMOS operational amplifier designs in use can be considered to be adaptations of the pole splitter operational amplifier or the folded cascode operational amplifier. These circuits and considerations in their design will be discussed below.

3.2.1 CMOS Two Stage Operational Amplifier

The CMOS two stage operational amplifier is widely used, especially in the telecommunications industry for voice band purposes. This configuration usually employs pole splitting compensation and is shown in figure 3.3. Such a circuit is often referred to as a *CMOS Pole Splitter Operational Amplifier*. The circuit is most useful for feedback circuits driving small loads as would be the case for an on chip capacitive load. The circuit has the advantages of simplicity and a simple and well understood design procedure as described in [8]. This design procedure will be covered in detail as it will be shown later that it can be applied, at least in part, to the circuit that will form the basis of this thesis.

Figure 3.3: CMOS Pole Splitting Operational Amplifier

Design Procedure

There exists a straightforward design procedure for the pole splitting operational amplifier.[8] The design procedure achieves its simplicity by the use of a few assumptions and rules of thumb that reduce the number of variables that must be considered. These will be discussed as they become required in the discussion of the design procedure.

The required unity gain bandwidth of the amplifier is usually specified by the intended application. In the case of a switched capacitor circuit, a useful rule of thumb is:

$$\omega_o \geq \frac{15}{T} \quad (3.3)$$

where T is time available for the opamp to slew and settle and is typically one half of the clock period. The unity gain bandwidth of the opamp can be described, in terms

of the input transistor and compensation capacitor as:

$$\omega_o = \frac{g_{m1}}{C_c} \quad (3.4)$$

The factors to consider when the compensation capacitor is chosen include the load capacitance, desired input stage transconductance and frequency of operation. A good compromise among the requirements is to choose the compensation capacitor to have the same value as the maximum load capacitor. Having chosen the compensation capacitor, equation 3.4 can be solved to give the required transconductance of transistors M1 and M2.

The lowest frequency non dominant pole for this configuration is that of the output stage. If one assumes a dominant pole and considers only the effect of the first non dominant pole, the phase margin is approximately given by:

$$\phi_m \approx 90^\circ - \arctan \frac{\omega_o}{|s_{p2}|} \quad (3.5)$$

The choice of $s_{p2} = 2\omega_o$ will give a phase margin only slightly greater than 60° . To allow for the existence of other non-dominant poles, and any error on the location of the poles, a good choice is to choose $s_{p2} = 3\omega_o$, which should ensure an adequate phase margin. Analysis of the circuit will indicate that:

$$|s_{p2}| \approx \frac{g_{m6}}{C_L} \quad (3.6)$$

Having chosen the location of the non-dominant pole, equation 3.6 may be used to determine the required transconductance of transistor M6. The response of the circuit to step inputs is of particular interest in the case of switched capacitor applications. This large signal behavior occurs at the input stage whenever the difference in the input voltages is sufficient that one input device is effectively off and the other is conducting heavily. There are obviously two possible situations, but due to the current mirror compensating M3 and M4, they are essentially similar. In either case, the compensation capacitor is being charged or discharged by a current equal to the differential pair bias current. Thus it can be shown that the slew rate is:

$$S_r = \frac{dv_{out}}{dt} = \frac{I_o}{C_c} \quad (3.7)$$

Similarly, the slew rate of the output stage can be shown to be:

$$S_{r_o} = \frac{dv_{out}}{dt} = \frac{I_7}{C_L} \quad (3.8)$$

In order to ensure that the opamp has a predictable slew rate, it should be designed such that the slew rate is limited by the input stage since that stage drives a predictable load (C_c). This can be achieved by setting the slew rate of the output stage to be 2.5 times that of the input stage. By using this condition and re-arranging equation 3.7 and equation 3.8, the current levels in the differential pair and the output stage may be determined using:

$$I_o = S_r C_c \quad (3.9)$$

$$I_7 = 2.5 S_r C_c \quad (3.10)$$

Having determined the current levels and required transconductance of the input transistors and the output driver, equation 2.17 can be used to determine the required (W/L) ratios.

It can be shown that the systematic voltage offset can be minimized by choosing:

$$\frac{(W/L)_3}{(W/L)_6} = \frac{(W/L)_4}{(W/L)_6} = \frac{I_o}{2I_7} = \frac{1}{5} \quad (3.11)$$

In order to keep the (W/L) ratios accurate, transistors M3, M4, M6 should be chosen to have the same channel length. Similarly this applies to the current source transistors M5 and M7.

When designing the current source transistors M5 and M7 one must avoid choosing a size that requires a large gate to source voltage as the need to keep the transistors in saturation will reduce the common mode range of the input and the available output swing. The designer should also avoid making the devices too large as such a choice will result in excessive area consumption on the chip and increased parasitic capacitances. A reasonable compromise between these requirements is to choose $V_{GS} - V_T$ to be 0.5 Volts. Using this value with equation 2.14 and the currents given by equations 3.9 and 3.10 above, the required size of these transistors can be determined.

Due to the fact that the compensation capacitor passes high frequency signals in either direction, there exists an undesired high frequency feed forward path through it. This results in a zero in the amplifier response and necessitates the use of R_c . The location of the zero is given by:

$$|s_z| = \frac{1}{(R_c - 1/g_{m6})C_c} \quad (3.12)$$

The designer must now choose the location of the zero. Two common choices are to locate the zero at s_{p2} or to cancel the zero by choosing:

$$R_c = 1/g_{m6} \quad (3.13)$$

Performance

The previous paragraphs described a simple procedure to design the two stage CMOS operational amplifier with pole splitting compensation. The following paragraphs will describe aspects of the circuits performance not discussed above.

Assuming symmetry in the differential input stage, the voltage gain of the input stage can be shown to be:

$$A_{v1} = \frac{g_{m1}}{g_{o2} + g_{o4}} \quad (3.14)$$

Similarly, the voltage gain of the output stage can be shown to be:

$$A_{v2} = \frac{g_{m6}}{g_{o6} + g_{o7}} \quad (3.15)$$

Combining the two equations, the overall gain of the opamp is given by:

$$A_v = \frac{g_{m6}}{g_{o6} + g_{o7}} \cdot \frac{g_{m1}}{g_{o2} + g_{o4}} \quad (3.16)$$

The common mode rejection ratio of the opamp is approximately given by:

$$CMRR = \frac{g_{m1}g_{m3}}{g_{o5} + g_{o1}} \quad (3.17)$$

The above are approximate results from theory. Once a design has been done on paper, it is customary to simulate the circuit and evaluate the performance of the circuit in that manner.

Drawbacks of the Two Stage CMOS Operational Amplifier

The two stage CMOS operational amplifier described above is widely used and has the advantages of reasonably good performance and ease of design. However, it does have a number of drawbacks when one attempts to use it in more demanding applications. The voltage gain of a two stage CMOS operational amplifier has been presented in equation 3.16 above. It can be seen that the gain of the entire circuit is the same as that for two cascaded common source amplifiers. The designer is constrained by practical limits on the size of the transistors, current levels, and trade offs with other performance criteria. If one considers the generic BiCMOS process described previously, and assumes a W/L of 100, a current level of 100 μA and a slope factor λ of 0.03, the voltage gain of a common source amplifier with an active load is

$$\begin{aligned} A_v &= \frac{g_{md}}{g_{od} + g_{ol}} \\ &= \frac{\sqrt{2k_p(W/L)I}}{2\lambda I} \\ &= \frac{\sqrt{2(100)(100)(100)}}{2(.03)(100)} \\ &= 47dB \end{aligned}$$

Two such stages cascaded would give a voltage gain of 94 dB. Although this is not an absolute limit, it represents a practical limit for the voltage gain of the two stage CMOS operational amplifier. Neglecting concerns over bandwidth, high DC gain in an operational amplifier is desirable due to the fact that in a feedback circuit, the steady state error is inversely proportional to $1 + A$. Thus the steady state error in a feedback circuit can be minimized by the use of an opamp with high DC gain.

An additional concern of modern analog design is power supply rejection. It is common to have analog and digital circuitry located on the same chip, thus switching transients from the digital gates will appear on the power supply rails. The two stage CMOS operational amplifier is generally considered to have poor power supply rejection [16]. At moderate frequencies, the output driver transistor $m6$ is effectively diode connected by the compensation capacitor. This couples any power supply noise

through the diode to the output.

3.2.2 Techniques for Performance Improvement In CMOS Operational Amplifiers

The CMOS two stage operational amplifier described above has adequate performance for many applications, especially those that have been traditionally encountered in the telecommunications industry. The trend to operate systems at higher frequencies has generated the requirement for operational amplifiers that operate at higher gains and bandwidths, with short settling times.

One possible approach to increasing the gain of an operational amplifier is to add one or more additional gain stages. Typically such a stage would be a common source MOS transistor amplifier such as is used in the second stage of the CMOS two stage operational amplifier. This approach has the serious drawback of adding high impedance nodes and thus poles at relatively low frequencies, making compensation of the resulting amplifier difficult, even for operational amplifiers intended for high gain operation at only moderate frequencies. Thus the approach of adding additional gain stages is usually not desirable when attempting to produce a high bandwidth or a high gain operational amplifier.

Another approach that allows the designer to increase the gain of an operational amplifier is to increase the output resistance of the active load. This may be done easily by the use of a cascode load. Consider the cascode of two MOS transistors shown in fig. 3.4. It can be shown [1] that the output resistance of the cascode is

$$r_o = r_{o1} + (1 + g_{m2}r_{o1})r_{o2} \quad (3.18)$$

which amounts to a considerable multiplication of the output resistance of a single transistor. This circuit can be used to replace an active load in an operational amplifier, thus increasing the gain. An example of a modified two stage CMOS op amp employing cascodes in the differential stage is show in fig. 3.5. Another virtue of the cascode circuit is the fact that the impedance at the common node between the two transistors is a low impedance node, thus any resulting pole will be at a high

Figure 3.4: MOS Transistor Cascode

frequency and can usually be ignored.

The primary drawback of cascode loads is that in order to keep both devices in saturation, the voltage required across the cascode is greater than that required to keep a single transistor in saturation. This has the effect of reducing the voltage swing of the stage in which the cascodes are used. If cascodes are used in the input stage, it will reduce the common mode input range of the opamp, and possibly restrict it to applications where one of the inputs is grounded. If cascodes are employed in an output stage, it limits the output swing of the opamp. This problem is of concern due to the low power supply voltages under which most CMOS operational amplifiers work.

3.2.3 CMOS Folded Cascode Operational Amplifier

In more demanding applications, it is useful to have an operational amplifier that has both high gain and high bandwidth. Techniques for increasing the gain of an opamp and their benefits and drawbacks have been discussed above. For stability, an opamp must have a dominant pole response. In other words, the response of the entire circuit must approximate that of a single pole low pass filter to frequencies

Figure 3.5: CMOS Operational Amplifier With Cascoded Differential Stage

well beyond that at which the voltage gain drops to unity. In order to obtain a high bandwidth circuit, it is necessary to have both high DC gain, and no poles at frequencies low enough to affect stability. One of the difficulties with the circuit in fig. 3.5 is the limited common mode input range, due to the cascodes on the differential pair. An additional difficulty arises from the fact that a level shift is necessary between the differential pair and the output. This is performed by the common source output stage. The use of the output stage introduces a pole at relatively low frequencies, thus possibly compromising stability. It also requires the use of an internal compensation network. One possible improvement to this circuit is to use the folded cascode topology, shown in fig. 3.6, the operation of which will now be described.

Transistors M1, M2 and M3 form the input differential stage as in the CMOS two stage operational amplifier. Transistors M4 and M5 are current sources that each supply twice the quiescent current drawn by each leg of the differential pair

It should be noted that the current mirror can be replaced with two pairs of NMOS cascode transistors, with the lower pair controlled by some sort of level sensing circuit. Such a circuit can readily be adapted to fully differential operation which is frequently advantageous when implementing analog systems.

The folded cascode circuit has a number of advantages. The circuit is effectively a one stage amplifier, and thus no pole splitting compensation is needed. The bandwidth of the circuit is related to the load capacitor by

$$\omega_o = \frac{g_{m1}}{C_L} \quad (3.20)$$

in the same manner as the bandwidth of the two stage opamp is related to its compensation capacitor. Since the output driver is effectively the input device, there is no non-dominant pole due to an output driver and the load capacitance. Since the load capacitance does not contribute to a non-dominant pole, increasing the load capacitance increases the phase margin and stability. This property makes the folded cascode operational amplifier suitable for achieving wide closed loop bandwidths while driving large capacitive loads. The non dominant poles occur at the low impedance nodes in the circuit, which are nodes A, B and C on fig. 3.6 above. The non dominant poles are due to the transconductance of the PMOS cascode devices and the capacitance at nodes A and B due to the sources of the cascode transistors and the drains of the current source transistors. There is also a non-dominant pole due to the double to single ended conversion. In the case of the circuit shown in fig. 3.6, the pole is due to the transconductances of the transistors in the mirror, and the capacitance at node C. Strategies for the location of poles to ensure stability has been discussed with the CMOS two stage operational amplifier above. It is necessary to ensure that the poles due to the low impedance nodes are at a sufficiently high frequency, such that the folded cascode operational amplifier will have adequate phase margin when driving the smallest expected capacitive load.

An additional advantage of the folded cascode operational amplifier is that it possesses a high power supply rejection ratio. It was discussed above that the relatively poor power supply rejection of the two stage operational amplifier was due to the compensation capacitor coupling positive supply noise to the output. As the

folded cascode operational amplifier has no compensation capacitor, this mechanism for introducing power supply noise is eliminated.

The folded cascode operational amplifier is not without its drawbacks however. It uses a larger number of devices than the two stage opamp described above, and thus typically consumes more area on an integrated circuit. The need to supply an appropriate bias voltage to one or more pairs of cascode transistors requires a more complicated bias circuit than a simple two stage operational amplifier. In order to keep the cascode transistors operating in the saturation regime, it is necessary to maintain a certain voltage across them, which limits the available output swing of the circuit.

3.3 Bipolar Operational Amplifiers

Bipolar operational amplifiers have been in use for many years. They have the advantage of high gain, high bandwidth and output stages with higher drive capabilities. Due to the inability of bipolar opamps to work well in sampled data systems and the high power consumption of bipolar logic, mixed analog digital bipolar integrated circuits are much less common than is the case for CMOS. A large number of the bipolar operational amplifiers produced are in the form of discrete amplifiers used in printed circuit board systems.

3.3.1 Simplified Two Stage Bipolar Operational Amplifier

It is possible to design a bipolar operational amplifier with a topology very similar to that of the CMOS pole splitting operational amplifier. An example of such a circuit is shown in fig. 3.7. The topology of this circuit differs from that of fig. 3.3 in that the circuit is 'flipped' and the polarity of all devices reversed. This allows the use of an NPN as the amplifying transistor of the second stage. The combination of NPN transistors Q1 and Q2, and common base connected PNP transistors Q3 and Q4 forms the equivalent of a PNP differential pair while retaining the better performance offered by the NPN transistor in most processes. Compensation is achieved by a single pole

Figure 3.7: Simplified Two Stage Bipolar Operational Amplifier

splitting capacitor between the output of the second stage and the output of the input stage. As most bipolar opamps are designed to drive off chip resistive loads, to make a useful circuit, one would have to add some sort of output stage to fig. 3.7.

3.3.2 A Practical Two Stage General Purpose Bipolar Operational Amplifier

The classic example of a general purpose bipolar operational amplifier is the 741, which was first introduced in 1966. A schematic of the 741 is shown in fig. 3.8. The 741 can be thought of as a two stage opamp, with the addition of an output stage, biasing circuitry and protection circuitry. The input stage can be considered to be an inverted version of the input stage used in the CMOS two stage operational amplifier described previously. The differential pair is composed of transistors Q1 to Q4. The

Figure 3.8: 741 Monolithic Operational Amplifier

use of PNP cascode transistors Q3 and Q4 allows NPN transistors Q1 and Q2 to be used as input devices. This is desirable due to the fact that NPN transistors generally have greater gain and bandwidth than PNP transistors in the same process, as was discussed previously. The current mirror used to convert the differential signal to single ended is composed of transistors Q5 to Q7. The presence of Q7 minimizes the loading of the differential pair due to the finite β of the transistors, and hence reduces the offset of the opamp. If low offset is critical to the application, an external resistor network may be connected to the 'Off' terminals shown in fig. 3.8. The second stage of the opamp is formed by a Darlington driver composed of Q16 and Q17, and an active load Q13. By virtue of its split collector, Q13 also serves to bias the class B output stage which is driven by Q23 operating as an emitter follower. The opamp

is internally compensated by C_1 which is connected between the input and second stages in the same manner as the CMOS two stage opamp. The remainder of the circuit forms either a bias network, or the class B output stage, and output short circuit protection. As these features are generally not required on opamps intended for monolithic systems, their operation will not be described here.

3.3.3 Input Current Cancellation

One of the drawbacks of a bipolar input operational amplifier is the need to supply a constant base current to transistors of the differential pair. This is particularly of concern if it is desired to employ the opamp in sampled analog circuits, as the base current would draw charge from the sampling capacitor in between samples, thus introducing an error into the output signal. In some designs, circuitry has been added to supply a current to the bases of the transistors of the differential pair which is approximately that which is drawn by the transistors. Two examples of base current cancellation networks applied to NPN differential pairs are presented in fig. 3.9 and fig. 3.10.

In fig. 3.9, the collector current of transistors Q_1 and Q_2 is sampled by transistors Q_3 and Q_4 respectively. Assuming that the characteristics of all transistors are well matched, the base currents of Q_3 and Q_4 will be the same as those of Q_1 and Q_2 respectively. The base current of Q_3 is mirrored by the PNP current mirror formed by Q_7 and Q_8 and is then fed into the base of Q_1 . Similarly Q_5 and Q_6 mirror the base current of Q_4 into the base of Q_2 . If the β of the PNP transistors is large and the NPN transistors are well matched, the currents supplied by Q_8 and Q_6 should be the same as the base currents drawn by Q_1 and Q_2 . In most processes, the β of the PNP transistors will not be large and thus there will be some difference between the two currents which will be seen at the input to the opamp. The magnitude of the resulting input current will be less than that for the unmodified input stage. The input offset will likely be made worse by a base current compensation scheme such as this due to the fact that there are now four pairs of transistors that operate as matched pairs as opposed to a single pair of the uncompensated differential pair.

Figure 3.10: Input Current Cancellation Circuit 2

source or sink the net input current. In addition to not entirely eliminating the DC input current, the current cancellation networks of figs. 3.9 and 3.10 will cause the magnitude of the input referred offset of the opamp to increase due to mismatches in the base current compensation network.

The connection of additional active devices to the input of an amplifier results in the input referred noise of the amplifier increasing by the amount of noise supplied by the added devices. This means that the noise performance of an operational amplifier employing base current cancellation will be inferior to an opamp with a simple differential pair as its input stage.

3.3.4 Differences Between Bipolar and CMOS Operational Amplifiers

As was mentioned above, the majority of bipolar operational amplifiers are designed as general purpose devices for use in board level systems. This is due to the fact that bipolar integrated circuit processes were commercially available long before CMOS VLSI processes and the applications to which they are suited. It is also due to the relatively low input impedance of bipolar transistors which makes the resulting opamps unsuitable for use in sampled data systems. Thus bipolar operational amplifiers typically feature some sort of output stage in order to adequately drive off chip loads, whereas most CMOS operational amplifiers are intended for on chip systems and do not require such circuitry. Despite the historical uses of bipolar and CMOS operational amplifiers, the recent availability of BiCMOS processes offers an opportunity to exploit the advantages of each type of device in the design of an operational amplifier. One possible such circuit will be described in a later chapter.

3.4 Summary

In this chapter, general information about operational amplifiers was presented, including the characteristics of an ideal operational amplifier. Non-idealities of real operational amplifiers were discussed, including how these non-idealities affect each of bipolar and MOS opamps. The differences in end application between typical bipolar and MOS operational amplifiers and the resulting design differences between the two types of opamps were discussed.

The topology and operation of two popular designs of monolithic CMOS operational amplifiers was discussed with particular attention paid to the design details of the pole-splitter opamp which will be employed in a later chapter.

The topology and operation of a practical bipolar operational amplifier was discussed. Two techniques to reduce the magnitude of the DC bias current that must be supplied to a bipolar input operational amplifier were presented and discussed.

Aspects of the circuit topologies and design techniques presented in this chapter

will be employed and expanded upon in Chapter 4 where circuit “building blocks” that can be used to realize an operational amplifier will be discussed. The design of a complete BiCMOS opamp will then be presented in Chapter 5.

Chapter 4

Topology Considerations for a BiCMOS Operational Amplifier

Various examples of operational amplifiers from CMOS and bipolar technology have been discussed in Chapter 3. The properties of MOS and bipolar transistors, including small signal models and simple design equations and the set of devices that would likely be available in a BiCMOS process have been discussed in Chapter 2. It is now reasonable to consider possibilities for combining elements of the two technologies in a single opamp. The merits of possible alternatives for several sections of a typical opamp will be discussed. A topology for a BiCMOS operational amplifier to be designed in detail will then be presented in Chapter 5.

4.1 Options for a BiCMOS Operational Amplifier Topology

A number of operational amplifier topologies have been discussed in Chapter 3. Each of these amplifiers can be considered to be composed of a number of circuit building blocks. Examples of these circuit building blocks include differential pairs, current mirrors, current sources and single or compound transistor amplifiers. Any topology for a BiCMOS operational amplifier will also be composed of such circuit

building blocks. As these circuits exist in both bipolar and CMOS technologies, it is necessary to evaluate the benefits of each approach in order that a choice can be made.

4.1.1 Devices Available

The various types of devices available in possible BiCMOS processes and their performance have been discussed in Chapter 2. Of these devices, those of interest in analog design are MOS and NPN transistors, resistors and capacitors. One device that is not likely to be available is the PNP transistor due to the poor performance of the lateral PNP transistor and the high cost of the complementary vertical PNP transistor. For these reasons, the use of PNP transistors will not be considered. The device set used for the design of a BiCMOS operational amplifier will then consist of N and P channel MOS transistors, NPN bipolar transistors, double polysilicon capacitors and polysilicon resistors.

4.1.2 Operational Amplifier Design Options

The remainder of this chapter will consider the circuit choices that need to be made in order to realize a one or two stage BiCMOS operational amplifier. First, the selection of a current source for the differential pair and the implications of this choice on the common mode performance of the opamp will be discussed. A number of possible differential input circuits will then be discussed. Considerable emphasis will be placed on this discussion as it will be shown that many of the important parameters of an opamp depend heavily on the differential input circuit. Discussion will then shift to a consideration of the options for an active load to the differential pair. Finally, options for an output gain stage will be considered.

The performance of the circuit blocks presented will be predicted using the small signal models and design equations presented in Chapter 2. The small signal model and design equations for MOS transistors are those of the SPICE level 1 model, also known as the "square law" relations. The small signal model and design equations for

bipolar transistors are those of the hybrid π model [7], which is a simplified version of the model used in most simulators. These models were chosen for use in comparing different circuit blocks because they are simple enough to be used for hand analysis and are accurate enough for meaningful comparisons to be made. The performance of the actual design to be presented in Chapter 5 was determined by simulation using the more detailed models that are available in the simulator used, thus some small disagreements will be expected when these results are compared with predictions based on the small signal model equations.

Once the circuit possibilities for the various sections of an opamp have been presented, a topology for a BiCMOS operational amplifier can then be proposed. This will be done at the beginning of Chapter 5 and will be followed by a detailed design and simulation of an opamp based on that topology.

4.2 Choice of Current Source

A convenient way to realize a current source for biasing active devices on an integrated circuit in a manner that is relatively independent of voltage is to use a current mirror. The three areas of concern when selecting a current mirror are the degree of mismatch between the reference and output currents, the output impedance and the range of voltage over which the output current will remain reasonably constant. When selecting a current source to bias a differential pair or a simple amplifier, any mismatch between the reference and output currents is not of concern as it is generally predictable and is easy to design around. The output impedance and the output voltage range are of concern as they have a significant effect on the common mode behavior of the opamp. A considerable variety of current mirror circuits can be found in the literature, particularly since BiCMOS technology has become readily available. In the interest of brevity, discussion here will be limited to a small number of current mirror circuits that may possibly be of use. The discussion will concentrate on aspects of the circuits that affect the common mode performance of the differential pair. Other aspects of current mirror performance such as reference and output

current matching will also be considered.

The current source of the differential pair has a significant impact on for the common mode behavior of the opamp. Consider the generic differential input stage and small signal model shown in fig. 4.1. The small signal equivalent circuit shown

Figure 4.1: Generic Differential Input Stage with Difference Mode Small Signal Model

in the right half of fig 4.1 was obtained by applying the Bisection Theorem [7] to the left half of fig 4.1 and taking the small signal equivalent. In order to develop the small signal model shown in fig. 4.1, it was chosen to represent the transistors of the differential pair by a voltage controlled current source in parallel with an output resistance, as is the case with the simplest versions of both the bipolar and MOS small signal equivalent circuits. Analysis of the small signal circuit in fig. 4.1 gives a difference mode gain of

$$A_{dm} = -\frac{g_m R_L r_o}{R_L + r_o} \quad (4.1)$$

where double ended output is implied. Assuming the output resistance of the input transistors is large compared to the load resistances, equation 4.1 reduces to

$$A_{dm} \simeq -g_m R_L \quad (4.2)$$

which is the result often used in discrete circuit design [17]. When active loads are employed, r_o is not likely to be large compared to R_L , and thus equation 4.1 is more applicable. If the output resistance of the transistors of the differential pair is ignored, the common mode gain can be shown to be

$$A_{cm} \simeq -\frac{R_L}{2R_S} \quad (4.3)$$

It will later be shown that the effect of the output resistance of the transistors of the differential pair on the common mode gain is indeed small. Expressions for common mode gain, including the effect of transistor output conductance where appropriate, will be presented for selected differential pairs when they are presented later in this chapter.

From equations 4.2 and 4.3, the common mode rejection ratio of the differential pair is then approximately

$$CMRR = \frac{A_{dm}}{A_{cm}} \simeq 2g_m R_S \quad (4.4)$$

It is then obvious that in order to maximize the common mode rejection ratio, it is desirable to have a large R_S , which is the output resistance of the current source. It will later be shown that active loads to a differential pair incorporating current mirrors have different impedances to difference and common mode signals and thus also have a large impact on the common mode rejection of the differential pair. The common mode rejection ratio into a resistive load will still be presented for each of the differential pairs discussed in Section 4.3 as this is still a valid comparison between the relative merits of different configurations.

The choice of the current source also determines the lower limit of the common mode range of the opamp. For example, if a common mode voltage is applied to a differential pair and any mismatches between the two transistors are negligible, the current flowing in each side of the differential pair will not change. Since the current

flowing in each transistor is constant, the base-emitter voltage (bipolar) or the gate-source voltage (MOS) will remain constant. Thus the voltage between the input terminals and the common node of the differential pair will remain constant. Hence the voltage at the common node will follow any changes in the common mode input voltage. In the case of a negative going change in the common mode input, the voltage at the common node will decrease. With any real current source, when the voltage at the common node decreases sufficiently, the output transistor of the current source will leave saturation (MOS) or enter saturation (bipolar) and the output current source will begin to decrease substantially. The input voltage at which this occurs defines the lower limit of the common mode range of the opamp.

From the preceding discussion, it can be seen that a desirable current source will have a large output resistance and the ability to operate with a small output voltage. Obvious possibilities for the realization of a current source are the MOS current mirror and the bipolar current mirror. One might consider the use of cascode mirrors due to their very large output resistance, however it will be shown that this is not practical in a low voltage process.

4.2.1 The MOS Current Mirror

The MOS current mirror is shown in fig. 4.2. The connectivity of the circuit in fig. 4.2 is such that both transistors see the same gate-source voltage. If it is assumed that both transistors are in saturation, have the same geometry and that the effect of the output resistance of the transistors can be ignored, the drain currents of the two transistors will be equal. If the transistors do not share the same geometry, the ratio of the output current to the reference current will be the same as the ratio of the W/L 's of the two transistors. As no current flows into the gate of either transistor, the output current will be equal to the reference current. Thus the MOS current mirror does not have any systematic mismatch between the reference and output currents.

The output resistance of the MOS current mirror is simply the output resistance

Figure 4.2: Simple MOS Current Mirror

of the MOS transistor, which in the simplified (SPICE level 1) model is

$$r_o = \frac{1}{\lambda I_D}.$$

For minimum gate length transistors in the hypothetical process described in Chapter 2, the value of λ is between 0.07 and 0.13. For example, if a current level is chosen to be $100\ \mu\text{A}$, the output resistance will be between $78\ \text{k}\Omega$ and $138\ \text{k}\Omega$. It should be noted that the value of λ depends heavily on the channel length of the transistor and the current at which it operates. For transistors in a real process operating at high currents, λ may be significantly more than the values above, thus the output resistance in a real design may be substantially lower.

The minimum output voltage is determined by the fact that both transistors must remain in saturation for the assumption to apply that equal gate-source voltages will give rise to equal drain currents. The diode connected transistor into which the reference current flows will always be in saturation except when the reference current is zero, and thus the voltage at the output node must be sufficient to keep the output transistor in saturation. The boundary of saturation of a MOS transistor is given by equation 2.12 which is

$$V_{DS} = V_{GS} - V_T.$$

It is normal design practice as was discussed previously to choose $V_{GS} - V_T$ to be on the order of 0.5 Volts. Thus the minimum output voltage of the MOS current mirror will also be on the order of 0.5 Volts and will depend on the particulars of the design.

4.2.2 The Bipolar Current Mirror

The bipolar current mirror is shown in fig. 4.3. In a manner analogous to that

Figure 4.3: Simple Bipolar Current Mirror

of the MOS current mirror, the connectivity of the two transistors in fig. 4.3 is such that both see the same base-emitter voltage. Thus assuming that the effect of the output resistance of the transistors can be ignored, the two transistors will conduct the same collector current provided that both transistors remain in forward active operation and have the same geometry. However in the bipolar case, the transistors draw a non-zero base current and the output current will thus differ from the reference current by the amount of base current flowing into both transistors. Assuming that both transistors have the same large β , the ratio of the output current to the reference current may be expressed as

$$\frac{I_{out}}{I_{ref}} = \frac{1}{1 + \frac{2}{\beta}} \quad (4.5)$$

Thus the bipolar current mirror does have a systematic mismatch between the output currents. This is not a concern when the mirror is used to bias a differential pair as the β error can easily be accounted for by slightly increasing the reference current. This mismatch may be a concern if the current mirror is used to accomplish a double ended to single ended conversion, as this will introduce some asymmetry in the rising and falling edge behavior of the complete opamp.

As was the case for the MOS current mirror above, the output resistance and the minimum output voltage of the bipolar current mirror are determined by the output transistor. The output resistance of a bipolar transistor is

$$r_o = \frac{V_{AF}}{I_C}$$

where V_{AF} is the forward Early voltage, which for the bipolar transistors in the hypothetical process described in Chapter 2 is approximately 50 Volts. Thus at the current level used for the MOS example above, the output resistance of the bipolar current mirror would be approximately 500 k Ω . Assuming that V_{AF} and λ do not change, at a given current level the bipolar current source would have double the output resistance of the MOS current source. As described previously, the value for λ used in the example may be optimistic, thus the relative advantage of bipolar current sources may well be greater than suggested above.

In order to obtain a reasonably constant current from the bipolar current mirror, it is necessary to keep the output transistor out of deep saturation. Technically, the saturation regime of a bipolar transistor begins whenever the base-collector junction is forward biased (ie the collector-emitter voltage is less than the base emitter voltage), however in practice, an bipolar transistor continues to be a reasonably constant current source until the collector-emitter voltage decreases to approximately 0.2 to 0.3 Volts.

4.2.3 Cascaded Transistor Current Mirrors.

In some applications, it may be desirable to increase the output resistance of a current source beyond values that may be attained by either of the current mirrors

discussed above. To illustrate this technique and its advantages and drawbacks, the MOS cascode current mirror, shown in fig. 4.4, will be discussed. The output

Figure 4.4: Cascoded MOS Current Mirror

resistance of the MOS cascode current mirror can be shown to be

$$r_{out} = (g_{m2}r_{o2})r_{o1} \quad (4.6)$$

Under the conditions of the previous examples, the transconductance, g_m , of the transistors will be 400×10^{-6} and the output resistance of each transistor will be $250k\Omega$. Thus the output resistance of the cascode mirror will be $25M\Omega$. The drawback of the cascode mirror is that the voltage required to keep all of the transistors in saturation is a large fraction of a typical 5 Volt power supply, and the available signal swing at the output of the current mirror is small. Since all of the transistors carry the same current, and are the same size, the gate-source voltage V_{GS} of all transistors will be the same. Since transistors 3 and 4 are diode connected, their gate-source and

drain-source voltages are the same. Thus it can be seen that the voltage at node A will be $2V_{GS}$. It then be seen that the voltage at node D will be $2V_{GS} - V_{GS} = V_{GS}$. As in the case of the simple MOS current mirror, a good design choice is to make $V_{GS} - V_T$ approximately 0.5 Volts. In order to keep transistor 2 in saturation, the voltage between nodes C and D must be at least 0.5 Volts. The threshold voltage V_T of the hypothetical process is approximately 0.8 Volts. Thus the minimum voltage at node C may be calculated to be

$$\begin{aligned} V_{GS} &= (V_{GS} - V_T) + V_T \\ &= 0.5 + 0.8 \\ &= 1.3 \end{aligned}$$

$$\begin{aligned} V_C &= V_{GS} + (V_{GS} - V_T) \\ &= 1.3 + 0.5 \\ &= 1.8 \end{aligned}$$

If the bipolar analog to this circuit is analyzed, the results are similar, with a slightly lower minimum output voltage.

The lower limit of common mode range of a differential pair biased by a current source, such as has been discussed here, is obtained by adding the sum of the gate source or base-emitter voltage of the input transistors, and the voltage across the current source required to keep it in active operation, to the negative power supply potential. For example the lower limit of input common mode range of an NPN differential pair biased by the MOS cascode current mirror described above would be

$$V_{cm,min} = V_{SS} + V_{BE} + V_{CS,min} \quad (4.7)$$

$$= -2.5 + 0.7 + 1.8 \quad (4.8)$$

$$= 0V \quad (4.9)$$

assuming a 5 Volt process. The minimum input common mode voltage of 0V indicates that the resulting opamp would not be usable for signals centered about analog ground. It is now reasonable to conclude that the use of a cascode current mirror as a current source is not practical in a low voltage process. If the output of the

2

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS
STANDARD REFERENCE MATERIAL 1010a
(ANSI and ISO TEST CHART No. 2)

differential pair of an opamp can be appropriately level shifted, it may be possible to make use of a cascode current mirror as a differential to single ended conversion element. This possibility will be discussed in a later section.

4.2.4 Summary of Current Mirrors.

The output resistance and minimum output voltage of three varieties of current mirror have been discussed above. Due to the limitations that would be placed on the common mode range, any type of cascode current mirror is not normally practical for use as a current source for biasing a differential pair in a low voltage process. As will be discussed later, a cascode current mirror may be useful as an active load in applications where a large voltage swing is not necessary.

It has been shown above that the bipolar current mirror has both a higher output resistance and a lower minimum output voltage than a comparable MOS current mirror. As the reference current for biasing an opamp often comes from off-chip, or can be adjusted to account for β error, this effect is not normally of significant concern. Thus where the correct polarity of bipolar transistors are available, the bipolar current mirror is the preferred current source for the differential pair. In the case of the process used here, one should choose an NPN current source unless the differential pair used PMOS transistors.

4.3 Differential Input Circuits

All of the differential input circuits that will be considered here are variations on the simple differential pair. The MOS and bipolar differential pairs will be analyzed in detail. Possible additions to the differential pair will then be considered. These include emitter degeneration, Darlington inputs and source followers (common drain amplifiers) driving a bipolar differential pair. Expressions for the difference and common mode gains, assuming double ended output, will be derived for each differential input circuit using small signal equivalent circuits obtained by applying the bisection theorem [7]. Expressions for the common mode rejection ratio, common

mode range and where relevant, the input impedance will then be presented for each differential input circuit considered. Following the descriptions of each differential input circuit, these expressions will be compared between the different circuits.

4.3.1 The MOS Differential Pair

The MOS differential pair is shown in fig. 4.5. The advantages of this circuit are

Figure 4.5: MOS Differential Pair

its effectively infinite input impedance at low frequencies, and the ability to vary the transconductance of the stage independently of the current level. In the case of the process used here, it also has the advantage of being available in both polarities. Its drawbacks include relatively large parasitic capacitances, large area consumption and relatively low gain.

The difference and common mode gains of this circuit (for double ended output) can be calculated from the small signal half circuits shown in fig. 4.6. The difference

Figure 4.6: Small Signal Half Circuits of the MOS Differential Pair

mode gain can be calculated from fig. 4.6a as

$$A_{dm} = -\frac{g_m r_o R_L}{r_o + R_L} \quad (4.10)$$

The common mode gain can be calculated from fig. 4.6b as

$$A_{cm} = -\frac{g_m r_o R_L}{r_o + R_L + 2R_S(g_m r_o + 1)} \quad (4.11)$$

By noting that typically $g_m r_o \gg 1$ and $2g_m R_S \gg 1$, equation 4.11 reduces to

$$A_{cm} = -\frac{g_m r_o R_L}{R_L + 2g_m r_o R_S} \quad (4.12)$$

From equations 4.10 and 4.12, the common mode rejection ratio of the differential pair is

$$CMRR = \frac{A_{dm}}{A_{cm}} = \frac{R_L + 2g_m r_o R_S}{r_o + R_L} \quad (4.13)$$

It can usually be assumed that $2g_m r_o R_S \gg R_L$. Equation 4.13 then reduces to

$$CMRR = \frac{A_{dm}}{A_{cm}} = \frac{2g_m r_o R_S}{r_o + R_L} \quad (4.14)$$

In a complete opamp, the differential pair would likely be loaded with an active load incorporating a current mirror, thus the choice of load will have a significant impact on the common mode rejection of the opamp. As this can only be resolved in a specific design situation, the comparison of common mode rejection between different input stages will be done using simple resistive loads. It will later be shown that the common mode small signal gain of all of the differential pairs considered when loaded by resistors can be approximated as

$$A_{cm} \approx -\frac{R_L}{2R_S} \quad (4.15)$$

and assuming that all of the differential pairs being compared also use the same tail current source, the comparison of their common mode rejection ratio becomes the same as the comparison of their difference mode gains.

In order to illustrate the transconductance and common mode range of the MOS differential pair it is appropriate to make use of an example. Consider the case of a current level of $300 \mu A$ and a W/L of 100. Using the value of $K_P = 100 \times 10^{-6}$ and equation 2.17, the transconductance of the input transistors is

$$g_M = \sqrt{2K_P \frac{W}{L} I_D} \quad (4.16)$$

$$= 2.5 \times 10^{-3} \Omega^{-1} \quad (4.17)$$

This result will be referred to later when comparing the MOS differential pair to other differential pairs.

The simple "square law" equation for the drain current of a MOS transistor in saturation has been given previously in equation 2.13. This equation may be manipulated to give

$$V_{GS} = V_T + \sqrt{\frac{2I_D}{K_P(W/L)}} \quad (4.18)$$

The value of V_T will range from about 0.7 V to about 1.1 V, depending on the particular process used and the amount of substrate bias present. Using the same

values as used to calculate g_m above, and $V_T = 1.0V$, the value of V_{GS} may be calculated to be

$$V_{GS} = 1.25V$$

Once the value of V_{GS} is known, it is now possible to calculate the common mode range of the MOS differential pair. It was determined above that the NPN bipolar transistor makes a better current source than a MOS transistor, so the use of the NPN transistor for the current source will be assumed. As was discussed previously, the lower bound on the common mode range is determined by the input voltage at which the current source enters deep saturation (bipolar) or leaves saturation (MOS). Assuming a typical NPN transistor, this typically occurs when the collector-emitter voltage reaches a value between 0.2 and 0.3 Volts. It will be assumed that the onset of deep saturation of a bipolar transistor $V_{CE,sat}$ occurs at a collector-emitter voltage of 0.25 V. It can be seen from fig. 4.5 above that the minimum input voltage is the sum of the minimum voltage across the current source, and the gate-source voltage of the input transistor. Thus

$$\begin{aligned} V_{cm,min} &= V_{SS} + V_{GS} + V_{CE,sat} \\ &= V_{SS} + 1.25 + 0.25 \\ &= V_{SS} + 1.5V \end{aligned}$$

At lower common mode voltages than $V_{cm,min}$, the transconductance the differential pair will decrease rapidly due to the decreasing current from the NPN transistor, however this region of operation is not useful for a general purpose opamp.

The upper limit of common mode range is determined by the point at which one or both of the input transistors falls out of saturation. In the case of the two stage opamp, the resistive load is usually a current mirror, with the reference node being the output of M1. As the reference transistor of a current mirror is diode connected, it is always in saturation. If it is assumed that the load is a PMOS current mirror with the same size as the input transistors, the gate-source voltage of the load transistor can be calculated in a manner similar to that used for the input transistor to be

$$V_{GS,p} \approx 1.2V$$

Thus, at the current levels used in this example, the voltage across the load will be 1.2 volts. The condition of saturation for a MOS transistor is

$$V_{DS} \geq V_{GS} - V_T$$

and must be satisfied for the input transistor. By inspection of fig. 4.5, one can write an expression for the maximum input voltage using the preceding relations. The maximum input voltage is

$$\begin{aligned} V_{in,max} &= V_{DD} - V_{R_L} - V_{DS,min} + V_{GS} \\ &= V_{DD} - V_{R_L} - (V_{GS} - V_T) + V_{GS} \\ &= V_{DD} - V_{R_L} + V_T \\ &= V_{DD} - 1.2 + 1.0 \\ &= V_{DD} - 0.2V \end{aligned}$$

Thus the common mode range of a MOS differential pair with the current levels and device sizes used in the example here will be from $V_{SS} + 1.5$ Volts to $V_{DD} - 0.2$ Volts.

4.3.2 The Bipolar Differential Pair

The bipolar differential pair is shown in fig. 4.7. The advantages of this circuit are its high gain, small parasitic capacitances and small size. Its drawbacks include low input impedance, the requirement for a DC bias current at each input and the lack of ability to vary the transconductance of the stage independently of the current level.

Small signal equivalent half circuits of the bipolar differential pair are shown in fig. 4.8. In the interest of simplicity, the parasitic resistances (R_B , R_C and R_E) of the bipolar transistor have been omitted. The difference and common mode gains of the bipolar differential pair can be calculated from the small signal half circuits in the same manner as for the MOS differential pair. The difference mode gain is

$$A_{dm} = -\frac{g_m r_o R_L}{R_L + r_o} \quad (4.19)$$

Figure 4.7: Bipolar Differential Pair

which is the same result as that obtained for the MOS and generic differential pairs. The common mode gain is

$$A_{cm} = -\frac{R_L(g_m r_o + 1)}{(1 + \frac{1}{\beta})g_m r_o 2R_S + r_o + 2R_S}. \quad (4.20)$$

By applying the simplifications $(1 + \frac{1}{\beta}) \simeq 1$, $g_m r_o \gg 1$ and $g_m R_S \gg 1$, equation 4.20 reduces to

$$A_{cm} = -\frac{R_L}{2R_S} \quad (4.21)$$

Using equations 4.19 and 4.21, the common mode rejection ratio of the differential pair can be expressed as

$$CMRR = \frac{A_{dm}}{A_{cm}} \simeq \frac{2g_m r_o R_S}{r_o + R_L} \quad (4.22)$$

It can be seen that the expressions for the common mode rejection ratios of the bipolar differential pair (equation 4.22) and the MOS differential pair (equation 4.14) are

Figure 4.8: Small Signal Half Circuits for the Bipolar Differential Pair

identical. Thus the difference in common mode rejection ratio of the two differential pairs depends on the difference in the transconductance and output conductance of the two types of transistor. The effect of the output conductance of the differential pair transistors on the common mode rejection ratio can only be determined in a specific design situation as the output conductance interacts with the actual load, thus meaningful comparisons cannot be made at this point. If the output conductance is ignored, the common mode rejection ratio, assuming the same tail current source in all cases, is then simply proportional to the transconductance of the differential pair transistors.

In order to compare the transconductance and common mode range of the bipolar differential pair to that of its MOS counterpart, the same example will be used. Keeping the current level at $300 \mu A$, and ignoring the effect of the base and emitter

parasitic series resistances r_b and r_e , the transconductance of the input transistors is

$$g_M = \frac{I_C}{kT/q} \quad (1.23)$$

$$= 11.6 \times 10^{-3} \Omega^{-1} \quad (1.24)$$

This is considerably higher than the transconductance which was obtained with the MOS differential pair above, which was $2.5 \times 10^{-3} \Omega^{-1}$.

The equation for the collector current of a bipolar transistor in forward active operation has been given previously in equation 2.1. This equation may be manipulated to give

$$V_{BE} = \frac{kT}{q} \ln\left(\frac{I_C}{I_S}\right) \quad (1.25)$$

Using the current level in the example above, and the value of $I_S = 3 \times 10^{-18}$, the value of V_{BE} may be calculated to be

$$V_{BE} = 0.83V$$

It is now possible to calculate the common mode range of the bipolar differential pair. As before, the use of an NPN transistor having $V_{CE,sat} = 2.5V$ as the current source will be assumed. In a similar manner to the case of the MOS differential pair above,

$$\begin{aligned} V_{cm,min} &= V_{SS} + V_{BE} + V_{CE,sat} \\ &= V_{SS} + 0.83 + 0.25 \\ &= V_{SS} + 1.1V \end{aligned}$$

Once this value is reached, the differential pair will still operate, however the transconductance will be reduced due to the reduced current from the NPN transistor current source.

The upper limit of common mode range is determined by the point at which one or both of the input transistors enters deep saturation. Assuming the same load resistors as above, the voltage across the load resistors will be

$$V_{RL} \approx 1.2V$$

As was stated previously, a bipolar transistor will leave the forward active region and enter deep saturation when its collector-emitter voltage drops to a value typically between 0.2 and 0.3 Volts, which will be assumed to be 0.25 Volts for convenience. It can be seen from fig. 4.7 that the maximum input voltage is

$$\begin{aligned} V_{in,max} &= V_{DD} - V_{R_L} - V_{CE,sat} + V_{BE} \\ &= V_{DD} - 1.2 - 0.25 + .83 \\ &= V_{DD} - 0.6V \end{aligned}$$

Thus the common mode range of the bipolar differential pair will be from $V_{SS} + 1.1$ Volts to $V_{DD} - 0.6$ Volts. This is slightly better at the lower end and slightly worse at the higher end than the MOS differential pair. As the lower end of the common mode range is usually more limiting, the bipolar differential pair offers slightly better performance in this regard.

One of the obvious drawbacks of the bipolar differential pair is its low input resistance. Consider the bipolar differential pair shown in fig. 4.7 above. The resistance seen by a difference mode signal will simply be the sum of the input resistances of each transistor. Ignoring the effects of the parasitic resistances r_b and r_e , this is

$$R_{id} = 2r_{\pi} \quad (4.26)$$

The expression for the value of r_{π} was previously presented as equation 2.6. Using the example above and $\beta = 100$, the difference mode input resistance is

$$R_{id} = 2 \frac{\beta}{g_M} \quad (4.27)$$

$$= 2 \frac{100}{11.6 \times 10^{-3}} \quad (4.28)$$

$$\approx 17k\Omega \quad (4.29)$$

The value of the input resistance of $17k\Omega$ may be unacceptably low for many applications. In order to retain the advantages of the bipolar differential pair, the following subsections will investigate possible additions and modifications to this circuit to reduce its drawbacks.

4.3.3 The Bipolar Differential Pair with Emitter Degeneration

The bipolar differential pair with emitter degeneration is shown in fig. 4.9. The

Figure 4.9: Bipolar Differential Pair With Emitter Degeneration

normal motivations for employing emitter degeneration are to improve the input impedance, and to reduce the gain of the stage. This may be desired when designing a high speed opamp in which high current levels are required in order to achieve an adequate slew rate, and the bandwidth that would result from the transconductance at that current level is neither needed nor desired. One reason for not wanting the unneeded bandwidth is that stability could become a problem.

The difference mode and common mode small signal half circuits of the bipolar differential pair with emitter degeneration are shown in fig. 4.10, again with the parasitic series resistances R_B , R_C and R_E omitted for simplicity. It can be seen from fig. 4.10 that the difference mode small signal half circuit is the same as the

Figure 4.10: Small Signal Half Circuits of the Bipolar Differential Pair With Emitter Degeneration

common mode half circuit of the bipolar differential pair with $2R_S$ replaced by R_E . The difference mode gain can easily be written from equation 4.20 as

$$A_{dm} = -\frac{R_L(g_m r_o + 1)}{(1 + \frac{1}{\beta})g_m r_o R_E + r_o + R_E}. \quad (4.30)$$

Again using the approximations $g_m r_o \gg 1$ and $1 + \frac{1}{\beta} \simeq 1$ equation 4.30 can be reduced to

$$A_{dm} = -\frac{R_L g_m}{g_m R_E + 1}. \quad (4.31)$$

The common mode small signal half circuit shown in fig. 4.10b is the same as that for the bipolar differential pair (fig. 4.8b) with $2R_S$ replaced by $2R_S + R_E$. As typically $R_E \ll 2R_S$, the contribution of R_E to the common mode gain can be ignored, and the expression for the common mode gain will be the same as for the bipolar

differential pair and is

$$A_{cm} \simeq -\frac{R_L}{2R_S} \quad (4.32)$$

Thus the common mode rejection ratio is then

$$CMRR = \frac{A_d}{A_c} = \frac{2g_m R_S}{1 + g_m R_E} \quad (4.33)$$

It can be easily seen above that both the difference mode gain and the common mode rejection ratio are reduced by the factor of $1 + g_m R_E$. As was described above, it may be desirable to reduce the difference mode gain of the circuit, however the resulting reduction in the common mode rejection ratio is a drawback.

One of the frequent motivations for the use of emitter degeneration is the improvement in the input impedance of the resulting opamp. The difference mode input resistance may be shown to be

$$R_{id} = 2[r_\pi + (\beta + 1)R_E] \quad (4.34)$$

In the case of the current levels used in the differential pair example, the input resistance can be approximately doubled by the use of an emitter degeneration resistor having a value of 80 ohms. This will reduce the gain of the circuit to approximately half of that produced by the circuit without emitter degeneration.

The presence of the emitter degeneration resistance will reduce the lower limit of the common mode range by the amount of the voltage drop across the resistor. Assuming the same current level as the NPN differential pair and an emitter degeneration resistance of 200Ω , the voltage across the resistor will be

$$V_{R_E} = \frac{I}{R_E} = \frac{300 \times 10^{-6}}{200} = 0.06V \quad (4.35)$$

As the $0.06V$ across the emitter degeneration resistor represents only about 1% of the power supply, its effect should be small. The lower limit of the common mode range of the NPN differential pair with emitter degeneration is then

$$\begin{aligned} V_{cm,min} &= V_{SS} + V_{BE} + V_{CE,sat} + V_{R_E} \\ &= V_{SS} + 0.83 + 0.25 + 0.06 \\ &= V_{SS} + 1.1V \end{aligned}$$

which, when rounded to two figures does not differ from that of the non degenerated differential pair. The presence of emitter degeneration will not affect the upper limit of common mode range which is

$$\begin{aligned} V_{in,max} &= V_{DD} - V_{R_L} - V_{CE,sat} + V_{BE} \\ &= V_{DD} - 1.2 - 0.25 + .83 \\ &= V_{DD} - 0.6V \end{aligned}$$

The drawbacks of the use of emitter degeneration in a bipolar differential pair include the reduction in gain and common mode rejection discussed above. In addition there is the risk of increased offset due to mismatches between the emitter degeneration resistors. If large amounts of emitter degeneration are used, it is also possible that the voltage drop across the resistor will be large enough to noticeably reduce the lower limit of the common mode range. The use of emitter degeneration does not offer any improvement in the level of bias current required at the input of the opamp.

Use of Base Resistance Instead of Emitter Degeneration

It was suggested that an alternative to the use of emitter degeneration would be to add series resistance to the bases of the transistors of the differential pair. If the emitter degeneration resistance R_E in fig. 4.9 is replaced by a resistor R_B having the value $R_B = \beta R_E$ in series with the base of each of the differential pair transistors, and the operating current is unchanged, the difference mode gain and the input resistance will be approximately unchanged. There are at least three drawbacks to this approach.

One of the areas of concern in any symmetrical circuit is matching between the devices that form the two halves of the circuit. With emitter degeneration, it is possible to modify the topology of fig. 4.9 such that the emitter degeneration can be accomplished with a single resistor [7] while maintaining the same small signal equivalent circuit. If base resistors are used, any mismatch between them will affect the differential and common mode operation of the circuit.

If base series resistance is used instead of emitter degeneration, the small signal voltage gain seen across the internal shunt capacitance C_μ of the bipolar transistor will be larger by $\frac{r_\pi + R_B}{r_\pi}$ than would be the case with emitter degeneration. The Miller multiplied capacitance C_μ would form a pole with the external base resistance R_B which may cause problems.

Finally, each of the base resistors would be a noise source at the input to the opamp and would add directly to the input referred noise of the opamp. Expressions for the mean squared noise voltage of bipolar transistors and resistors were given in Chapter 2 and are

$$v_{nT}^2 \approx (i_{nT}/g_m)^2 = \frac{2kT\Delta f}{g_m} \quad (4.36)$$

for a the base referred noise of a bipolar transistor and

$$v_{nT}^2 = 4kTR\Delta f \quad (4.37)$$

for a resistor. Using the values $R_B = \beta R_E = (100)(200) = 20k\Omega$ and $g_{mi} = 11.5 \times 10^{-3}$ as before, and assuming a noise bandwidth of 150 MHz., the input referred noise contribution of the transistor is

$$v_{nT}^2 = \frac{2kT\Delta f}{g_{mi}} = \frac{2k(300)(150 \times 10^6)}{11.5 \times 10^{-3}} = 0.11 \text{ nV}/\sqrt{\text{Hz}} \quad (4.38)$$

and noise contribution of the resistor (which is input referred by connectivity) is

$$v_{nT}^2 = 4kTR\Delta f = 4k(300)(20000)(150 \times 10^6) = 49.7 \text{ nV}/\sqrt{\text{Hz}} \quad (4.39)$$

It can be readily seen from the above that the addition of large base resistors adds considerably to the input referred noise of the differential pair, however the case of an emitter degeneration resistor should be considered before any conclusions are drawn.

The small signal difference mode gain of the NPN differential pair with emitter degeneration was presented in equation 4.30 and rewritten in terms of conductances is

$$A_d = \frac{G_E g_{mi}}{G_L(g_{mi} + G_E)} \quad (4.40)$$

The small signal gain from the emitter degeneration resistor to the output of the differential pair can be calculated from the small signal equivalent circuit shown in fig. 4.10 to be

$$A_e = \frac{G_E g_{m1}}{G_L(g_{m1} + G_E) + g_o G_E} \quad (4.41)$$

By comparing equations 4.41 and 4.40, it can be seen that the small signal gain from the emitter degeneration resistor to the output of the differential pair can be, at most, the same as the difference mode gain. Thus the input referred value of the noise due to the emitter degeneration resistor will be no more than the noise at the resistor itself. Assuming this pessimistic case, the input referred noise due to the emitter degeneration resistor is

$$v_{nE}^2 = 4kTR\Delta f = 4k(300)(200)(150 \times 10^6) = 0.50 \text{ nV}/\sqrt{\text{Hz}} \quad (4.42)$$

It can be seen from equations 4.42 and 4.39 that the contribution to the input referred noise of the base resistor is much larger than from an emitter degeneration resistor sized to have the same effect on opamp performance. As the noise performance of an opamp depends primarily on the noise levels applied to the inputs and generated by the input pair, the use of base resistance is extremely undesirable from a noise point of view.

4.3.4 The Darlington Differential Pair

The Darlington differential pair is shown in fig. 4.11. The motivations for employing this configuration are to increase the input impedance of the differential pair, and to reduce the base current required by the stage. The small signal half circuit models of the Darlington differential pair are shown in fig. 4.12. It can be shown that the output resistance of the outer transistors has negligible effect on the overall gain of the differential pair. It can also be shown that the output resistance of the inner transistors has minimal effect on the common mode performance. Both of these and the parasitic series resistances R_B , R_C and R_E have been omitted from fig. 4.12 to minimize the complexity of the resulting equations. From fig. 4.12a, the difference

Figure 4.11: Darlington Differential Pair

mode gain of the Darlington differential pair may be shown to be

$$A_d = -\frac{g_{mo}(\beta + 1)}{g_{mi} + g_{mo}(\beta + 1)} \cdot \frac{g_{mi}r_{oi}R_L}{r_{oi} + R_L} \quad (4.43)$$

Equation 4.43 applies both to the Darlington differential pair shown in fig. 4.11 and to the modified Darlington differential pair shown in fig. 4.13, where the current level in the outer transistors is set by current sources rather than the base current of the inner transistors.

In the case of the unmodified Darlington differential pair shown in fig. 4.11, the current level of the outer transistors is set by the base current of the inner transistors and then

$$g_{mo} = \frac{g_{mi}}{\beta}$$

Figure 4.12: Darlington Differential Pair Small Signal Half Circuits

Equation 4.43 then reduces to

$$A_d = -\frac{g_{m1} r_{oi} R_L}{2(r_{oi} + R_L)} \quad (4.44)$$

which is half the gain of the ordinary bipolar differential pair. This result is reasonable as the difference mode input signal can be considered to be divided between the base-emitter drops of the two transistors on each side.

Similarly to the bipolar differential pair with emitter degeneration, the common mode gain depends primarily on the output resistance of the current source and on the load resistance. Thus it is expected that the common mode gain of this circuit will be the same as that of the bipolar differential pair. If the common mode small signal half circuit (fig. 4.12b) is analyzed, the common mode gain reduces to

$$A_c = -\frac{R_L}{2R_S} \quad (4.45)$$

as expected.

The common mode rejection ratio may then be determined to be

$$CMRR = -\frac{g_{mo}(\beta + 1)}{g_{mi} + g_{mo}(\beta + 1)} \cdot \frac{2g_{mi}r_{oi}R_S}{r_{oi} + R_L} \quad (4.16)$$

It can be easily seen above that the common mode rejection ratio when compared to the normal bipolar differential pair is reduced, due to the reduction in difference mode gain.

One of the frequent motivations for the use of Darlington inputs is the improvement in the input impedance of the resulting op amp. The difference mode input resistance may be shown to be

$$R_{id} = 2\beta_o \left[\frac{1}{g_{mo}} + \beta_i \left(\frac{1}{g_{mi}} + R_E \right) \right] \quad (4.17)$$

The above equation assumes the existence of an emitter degeneration resistance R_E . If no emitter degeneration is used, equation 4.17 reduces to

$$R_{id} = 2\beta_o \left[\frac{1}{g_{mo}} + \frac{\beta_i}{g_{mi}} \right] \quad (4.18)$$

Assuming the β of the outer transistors is reasonably large, it can be seen that the increase in input resistance is considerable. One drawback of the configuration shown in fig. 4.11 is that the performance and operating conditions are not easily predictable, particularly those affected mostly be the outer transistors. This is due to the fact that the outer transistors operate at a much lower current density, and thus will have somewhat different characteristics. The difference in current density is unavoidable as scaling the inner transistors such that they operate at the same current density as the outer transistors is impractical. In typical BiCMOS processes, the current density in a unit NPN transistor (ie. the smallest available,) operating at the current level used in previous examples ($300\mu A$) will offer a near maximum β and f_T near its peak. If the differential pair transistors are larger than this, the collector capacitance increases, lowering the frequency of a non-dominant pole. If the inner and outer transistors are of similar size, the current density in the outer transistors will be much lower than in the inner transistors as stated above. One method of alleviating this difficulty is to use the modified Darlington differential pair shown in fig. 4.13. The addition of current

Figure 4.13: Modified Darlington Differential Pair

sources connected to the emitters of the outer transistors allows those transistors to operate at a predictable current level. These current sources are typically chosen to operate at current levels between 5% and 10% of the tail current. As this current is somewhat higher than the current drawn by the base of the input transistors, it dominates the operating point of the outer devices. This approach has the drawback of increasing the base current required due to the fact that the outer transistors now operate at a higher current level. The differential pair shown in fig. 4.13 can also be considered to be an ordinary bipolar differential pair driven by a pair of emitter followers.

The drawbacks of the use of the Darlington connected differential pair include the reduction in gain and common mode rejection discussed above. In addition there is the risk of increased offset due to mismatch of the current gains of the outer transistors which operate at a very low current density. This may be improved somewhat by the

use of the modified Darlington differential pair discussed above. Probably the most serious drawback of the Darlington differential pair is the fact that between the input node and the common node of the differential pair, there are two base-emitter voltage drops that must be maintained. The lower limit of the common mode range of the Darlington differential pair may be expressed as

$$V_{cm,min} = V_{SS} + V_{BE,o} + V_{BE,i} + V_{CE,sat}$$

The value of $V_{BE,o}$ may be calculated in the same manner as was done for the V_{BE} of the NPN differential pair. Depending on the current level of any current sources connected to the emitters of the outer transistors, the value of $V_{BE,o}$ can range from 0.72 to 0.78. For purposes of this example, it will be assumed that

$$V_{BE,o} = 0.75V$$

Then

$$\begin{aligned} V_{cm,min} &= V_{SS} + V_{BE,o} + V_{BE,i} + V_{CE,sat} \\ &= V_{SS} + 0.75 + 0.83 + 0.25 + \\ &= V_{SS} + 1.8V \end{aligned}$$

In a 5 Volt process, this will result in an unacceptably small common mode range, making any operational amplifier using this input stage suitable only for applications where the amplitude of any common mode signal is small. The upper limit of the common mode range may be expressed as

$$\begin{aligned} V_{in,max} &= V_{DD} - V_{RL} - V_{CE,sat} + V_{BE,i} + V_{BE,o} \\ &= V_{DD} - 1.2 - 0.25 + 0.83 + 0.75 \\ &= V_{DD} + 0.1V \end{aligned}$$

As the upper limit of common mode range is greater than the positive supply, it is not of concern for opamps intended for monolithic applications.

4.3.5 The Darlington Differential Pair With Emitter Degeneration

It was stated previously that the motivation for employing a Darlington differential pair was to increase the input impedance of an opamp without abandoning bipolar inputs. It was also stated that the motivation for employing emitter degeneration was to reduce the gain of a differential pair at a given current level. This does reduce bandwidth but allows greater stability. It is not unreasonable to want to combine both characteristics in a single circuit, therefore an example of a Darlington differential pair with emitter degeneration is shown in fig. 4.14. This particular example has current

Figure 4.14: Darlington Differential Pair With Emitter Degeneration

sources to bias the outer transistors at a predictable current level.

The DC gain of this circuit can be determined by analyzing the small signal half circuits shown in fig 4.15. The output resistances of the transistors in fig. 4.15 are

Figure 4.15: Small Signal Half Circuit of the Darlington Differential Pair With Emitter Degeneration

unlikely to have a significant effect on the calculated gains, thus they have been omitted for simplicity, as have the series parasitic resistances R_H , R_C and R_E . From fig. 4.15a, the difference mode gain may be calculated to be

$$A_d = \frac{-g_{mi}R_L}{1 + g_{mi}R_E + \frac{g_{mi}}{\beta g_{mo}}} \quad (4.49)$$

In the special case where the current sources on the emitters of the outer transistors are not used, equation 4.49 reduces to

$$A_d = \frac{-g_{mi}R_L}{2 + g_{mi}R_E} \quad (4.50)$$

Through analysis of fig. 4.15b and the applications of the simplifications used in the calculations for the Darlington differential pair and the bipolar differential pair with

emitter degeneration, the common mode gain of this circuit may be expressed as

$$A_{cm} \simeq -\frac{R_L}{2R_S} \quad (4.51)$$

which is the same result as previous bipolar differential pairs. By combining equations 4.49 and 4.51 the common mode rejection ratio can be calculated as

$$CMRR = \frac{A_d}{A_{cm}} = \frac{2g_{mi}R_S}{1 + g_{mi}R_E + \frac{g_{mi}}{\beta g_{mo}}} \quad (4.52)$$

The input resistance for a Darlington differential pair was presented previously in equation 4.47, both for circuits with and without emitter degeneration.

As was the case for the NPN differential pair with emitter degeneration, the presence of emitter degeneration resistance will not significantly change the common mode range of this differential pair from that of the Darlington differential pair. Unfortunately, the lower limit of common mode range is sufficiently restrictive in a 5 Volt process that any opamp using this input stage could only be used in applications where the common mode input would be small.

4.3.6 The Bipolar Differential Pair Driven by MOS Source Followers

It was discussed above that the modified Darlington differential pair could be considered to be a normal bipolar differential pair with the inputs driven by emitter followers. The source follower is the MOS equivalent to the bipolar emitter follower. It is then reasonable to consider an input stage consisting of a NPN differential pair driven by MOS source followers. Such a circuit is shown in fig. 4.16. The obvious advantage of a circuit such as that shown in fig. 4.16 is that it offers the effectively infinite DC input impedance of the MOS transistor while retaining the high transconductance of the bipolar differential pair.

One of the advantages of using MOS source followers is that they are available with both N and P type devices. If an NMOS source follower is used, the common mode range will be unacceptably limited in the same manner as for the NPN Darlington differential pair, as the base-emitter voltage drop of the outer transistor is replaced

Figure 4.16: NPN Differential Pair With PMOS Source Follower Inputs

by the gate-source voltage of the MOS transistor. Indeed the problem would likely be much worse as the gate-source voltage of a typical MOS transistor is significantly higher than the base emitter voltage of a bipolar transistor. If a PMOS source follower is used, the gate-source voltage of the outer transistor will be in the opposite sense to the base-emitter voltage of the inner NPN transistors. This serves to reduce the lower limit of common mode input voltage which has been a limiting factor in the utility of the Darlington configurations discussed previously. For the above reasons, it was chosen to use PMOS source followers in fig. 4.16.

A source follower is a single transistor (plus current source) MOS level shift circuit that has a gain slightly less than unity and a wide bandwidth. An NMOS example of such a circuit including a small signal model is shown in fig. 4.17. From the small signal model shown in fig. 4.17, it can be shown that the gain of a MOS source follower

Figure 4.17: MOS Source Follower with Small Signal Model

is

$$A_V = \frac{g_m}{g_m + g_o + 1/R} \quad (4.53)$$

The result of equation 4.53 is always less than unity, and for practical transistor sizes and currents, the voltage gain is usually between 0.8 and 0.9.

The effect of using a source follower to drive a differential pair is to reduce the difference mode gain by the gain of the source follower. As common mode signals will be attenuated equally, there is no change in the common mode gain, and the common mode rejection ratio will be reduced due to the reduced difference mode gain.

One may also compute the difference and the common mode gains of this circuit from the small signal half circuits shown in fig. 4.18. As before, the parasitic series resistances R_B , R_C and R_E have been omitted for simplicity. From fig. 4.18a, the difference mode gain can be calculated to be

$$A_d = - \frac{g_{m1} g_{m2} r_{o1} R_L}{(R_L + r_{o1})(g_{o2} + \frac{g_{m1}}{\beta} + g_{m2})} \quad (4.54)$$

Figure 4.18: Small Signal Half Circuits of NPN Differential Pair With PMOS Source Follower Inputs

If the effects of r_{oi} and g_{oo} are considered to be negligible, equation 4.54 reduces to

$$A_d = -\frac{g_{mi}R_L}{1 + \frac{g_{mi}}{\beta g_{mo}}} \quad (4.55)$$

From fig. 4.18b, the common mode gain can be calculated to be

$$A_c = -\frac{g_{mi}r_{oi}R_L}{(R_L + r_{oi})((1 + g_{mi}2R_S) + \frac{g_{mi}}{\beta g_{mo}})} \quad (4.56)$$

By applying the approximation $g_{mi}R_S \gg 1$ and assuming the effect of r_{oi} to be negligible, equation 4.56 reduces to

$$A_c = -\frac{R_L}{2R_S + \frac{1}{\beta g_{mo}}} \quad (4.57)$$

By combining equations 4.55 and 4.57, the common mode rejection ratio is

$$CMRR = \frac{2R_S g_{mi} + \frac{g_{mi}}{\beta g_{mo}}}{1 + \frac{\beta g_{mo}}{g_{mi}}} \quad (4.58)$$

The lower limit of common mode range of the circuit of fig. 4.16 may be expressed as

$$V_{cm,min} = V_{SS} + V_{CE,sat} + V_{BE} - V_{GS} \quad (4.59)$$

Assuming that the source follower operates at a current of $50\mu A$ has $(W/L) = 100$ and all other conditions as were used to calculate the V_{GS} of the MOS differential pair, equation 4.18 can be used to calculate the V_{GS} of the source follower which is

$$\begin{aligned} V_{GS} &= V_T + \sqrt{\frac{2I_D}{K_P(W/L)}} \\ &= 1.0 + \sqrt{\frac{2(50 \times 10^{-6})}{(100 \times 10^{-6})(100)}} \\ &= 1.1V \end{aligned}$$

The lower limit of the common mode range is then

$$\begin{aligned} V_{cm,min} &= V_{SS} + V_{CE,sat} + V_{BE} - V_{GS} \\ &= V_{SS} + 0.25 + 0.83 - 1.10 \\ &= V_{SS} + 0V \end{aligned}$$

This is actually an improvement over any of the other differential pairs considered so far. As was mentioned previously, it is expected that the upper limit of common mode range will be more restrictive for this circuit. The upper limit of common mode range may be expressed as

$$\begin{aligned} V_{cm,max} &= V_{DD} - V_{R_L} - V_{CE,sat} + V_{BE} - V_{GS} \\ &= V_{DD} - 1.2 - 0.25 + 0.83 - 1.10 \\ &= V_{DD} - 1.7V \end{aligned}$$

As $1.7V$ is a large fraction of the positive supply ($+2.5V$), this input circuit would only be useful where the magnitude of any common mode inputs was known to be small.

4.3.7 Comparison of Differential Pairs

A number of potential differential input circuits have been presented in the preceding section, along with relations describing the performance of each. These relations will now be re-stated in a manner to allow easy comparison of the different circuits. As the objective here is to identify significant fundamental differences in the performance of the various circuits, the equations will be simplified as much as possible. For example the effects of the transistor output resistance (r_o) will be ignored as its effect is reasonable small, it affects all differential pairs and its inclusion would tend to mask the fundamental differences that are of interest. The comparison will focus on those areas which have a significant effect on important characteristics of a two stage operational amplifier designed around one of these input circuits.

DC Gain and Bandwidth

In order to compare the gain and bandwidth potential of the various differential pairs, it will be assumed that the topology of the operational amplifier used will be that of the CMOS two stage operational amplifier described in Section 3.2.1 with the current mirror load replaced by fixed resistors. The unity gain bandwidth of the two stage opamp was presented in equation 3.4 and is

$$\omega_o = \frac{g_{mi}}{C_c} \quad (4.60)$$

where g_{mi} is the transconductance of the differential pair. In the case of simple differential pairs, g_{mi} is simply the transconductance of the two transistors forming the differential pair. The transconductance of the more complex differential input circuits such as the Darlington differential pair has not been calculated. However by noting that the small signal DC gain of the generic differential pair which was previously presented as equation 4.2 is

$$A_{dm} \simeq -g_{mi}R_L \quad (4.61)$$

where the effect of the output resistance of the differential pair transistors was assumed to be negligible, the transconductance of the various differential pairs may

Differential Input Stage	Difference Mode DC Gain
MOS	$-g_{m1}R_L$
NPN	$-g_{m1}R_L$
NPN with Emitter Degeneration	$-\frac{g_{m1}R_L}{g_{m1}R_E+1}$
Darlington	$-\frac{g_{m1}R_L}{2}$
Darlington with Emitter Degeneration	$-\frac{g_{m1}R_L}{g_{m1}R_E+2}$
NPN with Source Follower Inputs	$-\frac{g_{m1}R_L}{1+\frac{g_{m1}}{g_{m2}}}$

Table 4.1: Difference Mode Gains of Differential Input Stages

be compared by comparing their small signal DC gains, which have been calculated previously. That is,

$$g_{m1} = \frac{|A_{dm}|}{R_L} \quad (4.62)$$

It can be seen from equation 4.62 that those differential pairs having the highest difference mode gain also have the highest transconductance and hence, will also offer the highest bandwidth. This is consistent with the practical requirement that an operational amplifier exhibit an approximately single pole response to beyond its unity gain bandwidth.

If the small signal DC gains of two differential pair circuits are compared, it can be seen that

$$\frac{A_{dm2}}{A_{dm1}} = \frac{g_{m2}R_L}{g_{m1}R_L} = \frac{g_{m2}}{g_{m1}} \quad (4.63)$$

thus the ratio of the transconductance of two differential pairs is equal to the ratio of the DC gains. Thus the transconductance and hence bandwidth potential of two differential pairs can be compared by comparing their small signal DC gains into a common load. The difference mode gain for all of the differential input circuits previously presented will now be restated in tabular form. In order to facilitate comparison, the effect of the output resistance of the transistors of the differential pairs will be ignored.

In order to compare the gains of the various circuits, it is useful to know reasonable values of some of the terms appearing in Table 4.1. The transconductance of the

MOS differential pair and the NPN differential pair for a typical current level were previously calculated and presented as equations 4.17 and 4.24 respectively. The transconductance of an example MOS differential pair using NMOS transistors having a W/L of 100 was

$$g_{mi,MOS100} = 2.5 \times 10^{-3} \Omega^{-1} \quad (4.64)$$

Provided that the parasitic capacitances do not become large enough to cause parasitic poles to be at low enough frequencies to affect stability, the transconductance of a MOS transistor in a given can be increased by increasing the W/L of the device. If a W/L of 500 is substituted into equation 4.17 with the same values for K_P and I_D as the above example, the transconductance will be

$$g_{mi,MOS500} = 5.5 \times 10^{-3} \Omega^{-1} \quad (4.65)$$

The transconductance of an example bipolar differential pair operating at the same current level was

$$g_{mi,Bipolar} = 11.5 \times 10^{-3} \Omega^{-1} \quad (4.66)$$

It should be noted that the transconductance of the MOS differential pair can be increased by increasing the W/L of the input transistors until the associated capacitances begin to compromise stability, however caution must be exercised to ensure that $(V_{GS} - V_T)$ is not too small.

A reasonable value of emitter resistance to be used in those differential pairs having emitter degeneration is a few hundred ohms. For purposes of illustration, 200 ohms will be chosen here. The term $g_{mi}R_E$ then evaluates to

$$g_{mi}R_E = (11.5 \times 10^{-3})(200) = 2.3 \quad (4.67)$$

To compare the differential input circuits incorporating source follower to the other circuits, it is necessary to make some educated guesses as to the size and current levels of the source follower transistors. It was shown previously that a MOS transistor operating at a current of $300\mu A$ and having a W/L of 100 would have a transconductance of $2.5 \times 10^{-3} \Omega^{-1}$. Assuming that the source follower would operate

Differential Input Stage	Relative DC Gain
MOS $W/L = 100$	$\frac{1}{4.5}$
MOS $W/L = 500$	$\frac{1}{2}$
NPN	1
NPN with Emitter Degeneration	$\frac{1}{3}$
Darlington	$\frac{1}{2}$
Darlington with Emitter Degeneration	$\frac{1}{4}$
NPN with Source Follower Inputs	$\frac{1}{1.5}$

Table 4.2: Relative Difference Mode Gains of Differential Input Stages

at a current level somewhere between 5% and 20% of the differential pair tail current and that the MOS transistors should not be too much larger than the $W/L = 100$ of the example, the transconductance of the source follower transistor (g_{mo} in Table 4.1) would likely range between $0.1 \times 10^{-3} \Omega^{-1}$ and $1 \times 10^{-3} \Omega^{-1}$. Assuming that β is approximately 100

$$\frac{g_{mi}}{\beta g_{mo}}, \max = \frac{11.5 \times 10^{-3}}{(100)(0.1 \times 10^{-3})} = 1.15 \quad (4.68)$$

and

$$\frac{g_{mi}}{\beta g_{mo}}, \max = \frac{11.5 \times 10^{-3}}{(100)(1 \times 10^{-3})} = 0.12 \quad (4.69)$$

By substitution of the preceding values into the equations in Table 4.1, the differential input circuits may be compared. It has been chosen to use two different transistor sizes to illustrate the performance of the MOS differential pair. The W/L of 500 represents quite a large transistor that will take up considerable space and will have large parasitic capacitances. The W/L of 100 represents a more typical size which is fairly large by the standards of lower frequency CMOS opamp designs. This has been done in Table 4.2, with the results rounded to assist comparison.

It was stated previously that the bandwidth of a two stage operational amplifier is proportional to the DC gain of the differential input stage. As no other areas of concern have been considered thus far, the NPN differential pair would appear to

be the best choice to realize the highest bandwidth opamp. This assertion may be modified as additional characteristics are considered.

Common Mode Rejection

The common mode rejection ratios of the various differential input stages were presented with each circuit. Rather than repeat them here, use will be made of the observation that the value of DC common mode gain for all of the circuits under consideration is approximately

$$A_{CM} \simeq -\frac{R_L}{2R_S}$$

The common mode rejection ratio of a differential pair is simply the ratio between the difference mode and common mode gains,

$$CMRR = \frac{A_{DM}}{A_{CM}}$$

As all of the differential input circuits under consideration have approximately the same common mode gain, the common mode rejection ratio is directly proportional to the DC gain. Thus the comparisons presented in Table 4.2 also apply to common mode rejection, and those input circuits offering the most potential bandwidth will also offer the highest common mode rejection ratio.

Common Mode Range

The approximate limits of common mode range have been calculated for each of the differential input stages in the previous sections. These results are summarized in Table 4.3. The input common mode range of the various circuits can easily be compared by inspection of Table 4.3 however the importance of common mode range depends greatly on the intended application of the opamp. In applications where a large input common mode range is not required such as virtual ground applications, it may be advantageous to choose an input stage that offers a narrower common mode range but superior characteristics in other areas. It will be assumed here that the power supplies are $\pm 2.5V$ relative to analog ground and that the lower and upper

Differential Input Stage	Common Mode Range	
	Min	Max
MOS	$V_{SS} + 1.5$	$V_{DD} - 0.2$
NPN	$V_{SS} + 1.1$	$V_{DD} - 0.6$
NPN with Emitter Degeneration	$V_{SS} + 1.1$	$V_{DD} - 0.6$
Darlington	$V_{SS} + 1.8$	$V_{DD} + 0.1$
Darlington with Emitter Degeneration	$V_{SS} + 1.9$	$V_{DD} + 0.1$
NPN with Source Follower Inputs	$V_{SS} - 0.1$	$V_{DD} - 1.8$

Table 4.3: Common Mode Range of Differential Input Stages

Differential Input Stage	Input Impedance
MOS	Large
NPN	$2r_\pi$
NPN with Emitter Degeneration	$2(r_\pi + (\beta + 1)R_E)$
Darlington	$2(r_{\pi_o} + \beta_o r_{\pi_i})$
Darlington with Emitter Degeneration	$2(r_{\pi_o} + \beta_o(r_{\pi_i} + \beta_i R_E))$
NPN with Source Follower Inputs	Large

Table 4.4: Input Impedance Expressions for Differential Input Stages

limits of common mode range are of equal concern. It can be noted from Table 4.3 that the allowable input common mode range is roughly the same for all of the Darlington and source follower input stages, and that all are somewhat inferior to the NPN or MOS differential pairs.

Input Impedance

The difference mode input impedance of each of the bipolar differential input stages has been previously presented with each circuit considered. The input impedance of those input stages where the input is a MOS gate has been assumed to be large. These results have been summarized in Table 4.4. Assuming that $\beta \simeq 100$ and using the previous current levels to calculate r_π , the expressions in Table 4.4 have

Differential Input Stage	Input Impedance
MOS	Large
NPN	17k Ω
NPN with Emitter Degeneration	58k Ω
Darlington	3M Ω
Darlington with Emitter Degeneration	8M Ω
NPN with Source Follower Inputs	Large

Table 4.5: Approximate Input Impedance of Differential Input Stages

been evaluated and are presented in Table 4.5. It is desirable for an operational amplifier to have high input impedance; however as most opamps are used in a feedback configuration, the importance of high input impedance becomes less. This can be illustrated by repeating the analysis of the inverting amplifier configuration shown in fig. 3.2 with a finite input impedance. The result of this analysis is

$$\frac{V_{out}}{V_{in}} = -\frac{Z_{fb}}{Z_{in}} \cdot \frac{1}{1 + \frac{1}{A}(1 + \frac{Z_{fb}}{Z_i} + \frac{Z_{fb}}{Z_{in}})} \quad (4.70)$$

where A is the open loop gain of the opamp, Z_i is the input impedance of the opamp and Z_{fb} and Z_{in} are the impedances of the feedback and input paths respectively. It can readily be seen from equation 4.70 that the effect of Z_i becomes small as the open loop gain A increases. Although the NPN differential pair has quite a low input impedance, it may still be useful on account of the high gain that it offers.

Switched Capacitor Considerations

Designs for switched capacitor circuits generally assume that the input of the operational amplifier is such that the amount of charge drained from from the capacitor during a clock period is negligible. This is true with MOS input operational amplifiers that have a very high input impedance and no DC input current. Bipolar input operational amplifiers have both relatively low input impedance and non zero DC input currents. It was stated above that the low input impedance (small signal) could be tolerated if the DC gain was sufficiently high, at least from a steady state

point of view. In steady state, the voltage at the inverting input to the opamp will be held at a constant or near constant value by the output of the opamp through the feedback network, assuming that the non-inverting input is grounded. Thus no large discharge of the sampling capacitor supplying the input signal should occur. During transients that occur when a sampled data circuit switches, the voltage at the input of the opamp can swing significantly, as the opamp may be effectively operating open loop until it begins to settle at the new steady state value. A substantial amount of charge may be drawn from the sampling capacitor during these transients because a low input impedance opamp undergoing a voltage excursion at its input will draw current which the feedback network may not be able to supply. This may affect the steady state value, even if the opamp has a high but finite open loop gain, and thus limit the utility of bipolar input opamps for switched capacitor circuits.

The base current drawn by a bipolar transistor also presents a problem. The DC base current can be reduced by the use of the Darlington configuration, or perhaps some sort of current cancelling arrangement as discussed in Chapter 3; however it cannot be eliminated. It should also be noted that the use of emitter degeneration, although increasing the small signal input impedance of the differential pair will not have any effect on the base current.

Figure 4.19: Capacitor Discharge Through a Current Source

The effect of a non zero input current of an operational amplifier used in a switched capacitor circuit will likely manifest itself as a downward level shift. This can be

illustrated by a capacitor discharging through a current source as shown in fig. 4.19. If the current flowing through the current source were zero, and assuming an ideal capacitor, the voltage across the capacitor would remain constant indefinitely. If the current is not zero, the capacitor will discharge through the current source. The voltage across the capacitor can be expressed as

$$V(t) = V(0) - \frac{It}{C}. \quad (4.71)$$

As the output voltage of a switched capacitor stage is sampled by the subsequent stage no earlier than the end of the half clock period in which the sample was taken, the value of t in equation 4.71 is usually half of the clock period. For example assuming a clock frequency of 1 MHz, one half of the clock period is

$$t = \frac{T}{2} = \frac{1}{2f} = \frac{1}{(2)(1 \times 10^6)} = 500ns.$$

Assuming that the current source represents a base current of $3\mu A$, the drop in the capacitor voltage is then

$$\Delta V = \frac{It}{C} = \frac{(3 \times 10^{-6})(500 \times 10^{-9})}{5 \times 10^{-12}} = 0.3V \quad (4.72)$$

It can be seen in equation 4.71 that the drop in voltage ΔV is inversely proportional to the duration of the current flow, thus the drop in voltage can be reduced by increasing the clock frequency. For example if the clock is increased to 10 MHz, the drop in voltage becomes only 0.03 V.

The drop in voltage across a sampling capacitor over the sampling period will result in a level shift seen by the next stage of the filter. If level shifts on the order of tenths of a Volt were allowed to accumulate through a succession of switched capacitor stages, the total level shift would quickly reach the limit of common mode range of the opamps. This makes it impractical to realize switched capacitor circuits using single ended bipolar input operational amplifiers. It may, however be possible to realize a fully differential switched capacitor circuit using double ended opamps. As the base current flows into both transistors of the differential pair, the level shift will be towards V_{SS} on both inputs. Thus the level shift becomes a common mode input and should be cancelled by the differential pair of the next stage.

In summary, the base current requirements of bipolar transistors present a major difficulty in the realization of switched capacitor circuits using operational amplifiers having bipolar differential pairs. This problem may be surmountable with the use of double ended opamps and an adequately high sampling frequency. The use of double ended opamps complicates the design and testing of the opamps, particularly the design of the bias circuit. The requirement to use a high sampling rate is less of a concern, but it is an undesirable restriction on the end use of an opamp. For these reasons, it is reasonable to conclude that the most appropriate input stage for an operational amplifier to be used in switched capacitor circuits is the MOS differential pair.

4.3.8 Summary of Input Pairs

The operation and various aspects of the performance of a number of differential input circuits that might be useful in the design of a BiCMOS operational amplifier have been presented in this section. One of the important requirements for a high bandwidth operational amplifier is that the input stage has high gain. The NPN differential pair offers the highest transconductance of the circuits considered at a given current level. The major drawbacks of the NPN differential pair are its low input impedance, its requirement for a DC bias base current and that its transconductance is not readily adjusted independently of the current level. In practical operational amplifiers, the current level in the differential pair is usually determined by the need to have an adequate slew rate when driving the compensation capacitor. This current level is usually sufficiently high that transconductance of a bipolar transistor will be too high, and the opamp will not be stable. The transconductance of a bipolar differential pair can be reduced at a given current level by the use of emitter degeneration; however the gain of the differential pair will be reduced to a level that is attainable by the MOS differential pair. It was stated above that it may be possible to accept the base current of a bipolar input opamp in a switched capacitor circuit by the use of double ended opamps and a sufficiently high sampling rate; however it would be preferable to avoid these limitations. The Darlington differential and the

NPN differential pair with PMOS source follower inputs offer improvements in the input impedance and input current, but are of limited utility in a 5 volt process due to their limited common mode range. For these reasons, it is reasonable to choose the MOS differential pair for use as an input stage in an operational amplifier intended for use in switched capacitor applications.

4.4 Circuit Options for Active Loads

In the examples presented previously, the differential pairs under consideration were loaded with passive resistors. This is not practical in a real opamp due to the following conflicting requirements. In order to develop adequate voltage gain, the small signal resistance of the load presented to the differential pair has to be fairly large. In order to avoid limiting the common mode range of the differential pair, the voltage drop across the load must be kept small. At a given current level, the need to limit the voltage drop across a resistor imposes a maximum value of its resistance. Values of a passive resistor small enough to ensure adequate common mode range would be too small to allow the differential pair to develop the voltage gain necessary for a high bandwidth opamp. The output conductance of both MOS and bipolar transistors is primarily a function of the current through the device, and much less so a function of the terminal voltages. Thus active loads can have high small signal resistances and hence provide more gain than a passive load having a similar voltage drop.

The issues of concern when considering the use of transistors as active loads are the output resistance of the load and the voltage range over which the load device will operate correctly. This is a similar problem to that of choosing a current source, and the discussion of devices and circuit configurations for current sources in Section 4.2 applies to active loads.

4.4.1 Use of Cascaded Current Mirrors as Active Loads

Cascaded current sources were considered impractical for use in simple input stages due to the fact that they reduce the common mode range of the differential pair by an unacceptable amount in a low voltage process. Through similar reasoning, one can conclude that a cascode load between the output of a differential pair and a power supply would unacceptably reduce the common mode range in the other direction. If the current output of the differential pair is "folded over" as is done in the folded cascode operational amplifier shown in fig. 3.6, it may be possible to use a cascode load, as the output voltage can then be centered about analog ground. This is particularly true if the input stage and cascode load is followed by a gain stage as the voltage swing required at the input to the gain stage will be small.

When using BiCMOS technology, there are a large number of circuit possibilities for use in current mirrors or active loads. Some of these that are of interest have been discussed in Section 4.2. Several others are discussed in the literature [13]. Two possibilities that are of interest and are not discussed in Section 4.2 are the NPN cascode current mirror shown in fig. 4.20 and the Wilson current mirror shown in fig. 4.21.

The operation of the NPN cascode current mirror shown in fig. 4.20 is as follows. Transistors 1 and 2 form a current mirror which causes the current in each side of the circuit to be approximately equal. Transistor 3 provides a suitable bias voltage for transistor 4 which acts as an impedance transformer. The reference and output currents will not be exactly equal due to the need to supply base current to the transistors from the reference current. Assuming that the transistors are well matched, and by making use of the relation

$$I_C = \beta I_B$$

the ratio of the output current to the reference current can be expressed as

$$\frac{I_{out}}{I_{ref}} = \frac{\beta^2}{\beta^2 + 4\beta + 2} \quad (4.73)$$

Noting that $\beta^2 \gg 2$, equation 4.73 can be reduced to

$$\frac{I_{out}}{I_{ref}} = \frac{\beta}{\beta + 4} \quad (4.74)$$

Figure 4.20: NPN Cascode Current Mirror

The ratio between the output and reference currents is not particularly important if the current mirror is used to load a single ended amplifier. This ratio is of concern if the current mirror is used as a combined load and double ended to single ended conversion for a differential pair as any asymmetry in the loads presented or the current levels will manifest itself as an asymmetric slew rate and as an offset voltage. As a large output voltage swing is not required here, the property of interest when selecting an active load is the output resistance. The output resistance for the NPN cascode can be shown to be [7]

$$r_{out} \approx \beta r_o \quad (4.75)$$

The principal drawback of the NPN cascode current mirror is the relatively large mismatch due to the base currents of the transistors. A variation of this circuit that reduces this difficulty at the expense of output resistance is the Wilson current mirror shown in fig. 4.21. The operation of this circuit is similar to that shown in fig. 4.20 except that the base current for transistors 1 and 2 comes from the leg carrying the output current. Intuitively, one would expect that this would reduce the degree of

Figure 4.21: Wilson Current Mirror

mismatch between the two currents. Analysis of the circuit in fig. 4.21 gives

$$\frac{I_{out}}{I_{ref}} = \frac{\beta^2 + 2\beta}{\beta^2 + 2\beta + 2} \quad (4.76)$$

As $\beta^2 \gg 2$, the ratio of the output current and the reference current is effectively unity. The output resistance of the Wilson current mirror can be shown to be [7]

$$r_{out} \approx \frac{\beta r_o}{2} \quad (4.77)$$

In double ended to single ended conversion, the Wilson current mirror would be preferred over the NPN cascode current mirror due to its significantly better current matching. In single ended loading applications, the cascode current mirror would be preferred as its 2 times higher output resistance would allow an additional 6 dB gain.

The MOS cascode current mirror was discussed previously and is shown in fig. 4.4. The output resistance of this circuit was presented as equation 4.6 and was

$$r_{out} = (g_{m2}r_{o2})r_{o1}.$$

As both the MOS cascode current mirror and the Wilson current mirror have little or no mismatch between the reference and output currents, it is useful to compare

their output resistances. Assuming a β of 100, and Early voltage of 50 and a current of $300\mu\text{A}$, equation 4.77 for the Wilson current mirror evaluates to

$$\begin{aligned} r_{\text{out}} &\approx \frac{\beta r_o}{2} \\ &= \frac{(\beta)(\text{VAF})}{(2)(1)} \\ &= \frac{(100)(50)}{(2)(300 \times 10^{-6})} \\ &= 8.3\text{M}\Omega \end{aligned}$$

Assuming the same current level, and that the (W/L) of the transistors in the MOS cascode current mirror is chosen such that $V_{GS} - V_T = 0.5\text{V}$, the transconductance of the transistors will be

$$g_m = 1.2 \times 10^{-3}$$

Assuming that the λ of the MOS transistors is 0.04, the output resistance r_o of each transistor will be

$$\begin{aligned} r_o &= \frac{1}{\lambda I} \\ &= \frac{1}{(0.04)(300 \times 10^{-6})} \\ &= 83.3\text{K}\Omega \end{aligned}$$

Equation 4.6 then evaluates to

$$\begin{aligned} r_{\text{out}} &= g_m r_o^2 \\ &= (1.2 \times 10^{-3})(83.3 \times 10^3)^2 \\ &= 8.3\text{M}\Omega \end{aligned}$$

which is the same result as for the Wilson current mirror. If the assumptions above are correct, there is no advantage to either the Wilson or the MOS cascode current mirrors in this application. In reality, the λ of 0.04 could only be achieved with a long channel MOS transistor. Thus the value of the output resistance of the MOS cascode current mirror represents the maximum reasonably achievable, while the value for the Wilson current mirror is fairly typical. As the current mismatch between the

reference and output currents of the Wilson current mirror is quite small, and as the Wilson current mirror will likely have at least as great an output resistance as a MOS cascode current mirror, the Wilson current mirror implemented with NPN bipolar transistors is the preferred choice for double ended to single ended conversion loads to a differential pair.

4.4.2 Summary of Active Loads

In situations where available voltage swing is a concern, it is appropriate to use a single transistor or simple current mirror as an active load. Compared to some form of cascode load, the primary drawback of this approach is the lower output resistance, particularly if it is a MOS device. The advantages are larger output swing, and simplicity, particularly in biasing. In situations where only a small voltage swing is necessary such as in a folded cascode stage that is followed by a gain stage, it is appropriate to use a cascode current mirror in order to exploit the additional small signal resistance available. If possible, this mirror should be realized using NPN bipolar transistors. If current matching is of concern, the load should be connected as a Wilson current mirror, otherwise it should be connected as a NPN cascode current mirror.

4.5 Choice of Output Stage

The choice of an output stage depends largely on the end application of the opamp. Component opamps are required to drive loads having relatively large capacitance, and are often required to source or sink significant currents. Such opamps typically require an output stage such as a totem pole, class A or AB amplifier, or an emitter or source follower. As these types of output stages generally have low voltage gains, it is usually necessary to include a gain stage between the differential pair and the output stage. The primary drawback of such a configuration is that opamps having three or more gain stages can be difficult to compensate for stability.

4.5.1 Common Source and Common Emitter Amplifiers

Figure 4.22: Common Source and Common Emitter Amplifiers

The applications that provided motivation for this thesis involve monolithic analog systems. Thus it is not necessary to drive off-chip loads and less rugged output stages may be employed. Two obvious candidates are the common source amplifier and the common emitter amplifier shown in fig 4.22. In each case, the NMOS or NPN amplifying transistor is loaded with a PMOS active load because there is no acceptable PNP current source available in the process. The small signal DC gain of either circuit can be shown to be

$$A_v = -\frac{g_m}{g_{on} + g_{op}} \quad (4.78)$$

where g_m and g_{on} are the transconductance and output conductance of the NPN or NMOS amplifying transistor and g_{op} is the output conductance of the PMOS

load transistor. In the case of the common emitter amplifier shown in fig. 4.22, equation 4.78 neglects the effects of emitter series resistance and collector spreading resistance.

A common source amplifier stage such as that shown in fig. 4.22 was used on the two stage CMOS operational amplifier discussed in Chapter 3. This circuit allows a reasonably large output voltage swing and can have a large gain. In a BiCMOS process, one has the choice of using either of the circuits shown in fig. 4.22. In Section 2.1.3, it was shown that for reasonably sized devices, a bipolar transistor has higher transconductance and lower output conductance than a MOS transistor. If equation 4.78 were evaluated for both cases shown in fig. 4.22, it could be easily seen that the common emitter output stage would have larger DC gain than the common source output stage operating at the same quiescent current. The major drawback of the common emitter output stage is that it will load down the differential pair more severely than the common source stage, due to the low input impedance of the NPN transistor. This would necessitate the use of an emitter follower before the output stage as shown in fig. 4.23 or some other buffering arrangement.

4.5.2 Common Emitter Amplifier with Emitter Follower Input

The drawback of the common emitter output stage is that the low input impedance of the NPN transistor will load down the previous stage. This effect may be reduced by the addition of an emitter follower, or eliminated by the addition of a source follower between the input stage and the common emitter stage. An example of a common emitter driven by an emitter follower is shown in fig. 4.23. As the emitter follower will likely operate at a lower current level, the DC loading on the previous stage will be less, and the input resistance of the common emitter stage will be multiplied by β . The voltage gain of fig. 4.23 can be calculated from the low frequency small signal model shown in fig. 4.24 to be

$$A_v = -\frac{g_{m1}}{(g_{o3} + g_{o4})} \cdot \frac{g_{m1}}{\frac{\beta+1}{\beta}g_{m1} + g_{o2} + g_{o1} + \frac{g_{m3}}{\beta}} \quad (4.79)$$

Figure 4.23: Common Emitter Amplifier With Emitter Follower Input.

By applying the approximations $\frac{\beta+1}{\beta} \simeq 1$ and $g_m + g_o \simeq g_m$, equation 4.79 reduces to

$$A_v = -\frac{g_{m1}g_{m3}}{(g_{o3} + g_{o4})(g_{m1} + g_{o2} + \frac{g_{m3}}{\beta})} \quad (4.80)$$

The input impedance of fig. 4.23 can also be calculated from the low frequency small signal model shown in fig. 4.24 and is

$$R_i = r_{\pi1} + \frac{\beta + 1}{g_{o1} + g_{o2} + g_{\pi3}} \quad (4.81)$$

4.5.3 Single Stage Operational Amplifier

The other output stage to be considered is none at all, as is the case with the folded cascode operational amplifier discussed in Chapter 3. The primary benefit of this configuration is that no compensation network is needed. The folded cascode operational amplifier is a single stage for compensation purposes with the necessary

Figure 4.24: Emitter Follower - Common Emitter Amplifier Small Signal Model

dominant pole being supplied by the load capacitance. The folded cascode offers improved power supply rejection as it lacks a pole splitting capacitor which is the route by which V_{dd} power supply noise appears at the output of the pole splitter operational amplifier. The drawbacks of this approach are that both the available voltage swing at the output of a folded cascode amplifier and the gain readily available in a single stage amplifier are limited. The limited voltage swing is a concern, particularly in a low voltage (5 Volt) process. The limited voltage gain available from a single stage is excessively limiting when the objective is to maximize the bandwidth of an operational amplifier.

4.5.4 Summary of Output Stages

The DC gain that must be developed by a high bandwidth operational amplifier would be difficult to achieve using a single stage such as the folded cascode opamp. The most reasonable choice for an output gain stage would be one based on the common emitter amplifier due to the high transconductance of the NPN bipolar transistor. In order to minimize the effects of the low input impedance of the common emitter amplifier on the preceding stage, it is necessary to buffer the input signal.

A suitable output stage incorporating both of these features is the common emitter amplifier with emitter follower input shown in fig. 4.23.

4.6 Summary

In this chapter, MOS and bipolar implementations of common analog "building block" circuits have been presented and analyzed as potential candidates to be exploited in the design of a BiCMOS operational amplifier. The NPN bipolar transistor has been chosen as the current source for biasing the differential pair due to its high output resistance and small size. The MOS differential pair has been chosen as the preferred input stage for an opamp intended for switched capacitor applications as it has the least severe drawbacks of the input stages considered. The common emitter amplifier has been chosen as the preferred output stage due to its high gain and compact size. In the following chapter, a topology for a BiCMOS operational amplifier will be proposed for further investigation, along with justification for the various design choices made. Detailed design and simulation of an opamp based on this topology will then follow.

Chapter 5

Design and Simulation

In Chapter 2, design equations for and performance of MOS and bipolar transistors were presented and compared. Circuits and topologies of CMOS and bipolar operational amplifiers were discussed in Chapter 3. Possible BiCMOS implementations of the various circuit blocks that are necessary to implement an operational amplifier were discussed in Chapter 4. A topology for a BiCMOS operational amplifier will now be proposed. A complete operational amplifier design based on this topology will then be presented along with calculated performance and simulation results. Following this, some design exercises that considered alternate topologies will be presented, eg; bipolar inputs instead of MOS.

5.1 Towards a Proposed Topology for a BiCMOS Operational Amplifier

In Chapter 4, MOS and bipolar implementations of common “building block” circuits, including current mirrors, differential pairs and single transistor amplifiers have been presented. The merits of these circuits have been discussed in the context of the various functions necessary to realize an operational amplifier. It is quite apparent that a large number of circuits could be proposed for an experimental BiCMOS operational amplifier, due to the large variety of constraints that could be imposed.

The primary constraint chosen here was the maximization of bandwidth, while still achieving high DC gain. The main topology will now be proposed as follows.

5.1.1 Choice of Differential Pair

The bandwidth of an operational amplifier employing a simple differential pair as its input stage is determined by transconductance of the differential pair and the size of the load or compensation capacitance. These are related by equation 3.4, presented previously. In the case of simple differential pairs such as those shown in fig. 4.5 and fig. 4.7, the transconductance of the differential pair is simply the transconductance of the input transistors. In the case of more complex differential pairs such as the Darlington differential pair, or differential pairs with emitter degeneration, the transconductance of the differential pair will depend on the transconductance of the inner transistors, and the effect of any other transistors and degeneration resistors on the gain of the differential pair. In Chapter 4, rather than calculate the transconductance of each differential pair, it was noted that the DC gain of a differential pair into a particular load is directly proportional to its transconductance, thus the transconductances and hence available bandwidth of different differential pairs, could be compared by comparing their DC gains into an identical load. It can be seen from equation 3.4 that in order to achieve high bandwidth, it is necessary to choose a differential input stage offering high transconductance, and to minimize the load capacitance.

The objective of this thesis is to develop a higher bandwidth operational amplifier, using BiCMOS design techniques, that is suitable for use in sampled data circuits, particularly those employing switched capacitor techniques. In order to achieve the highest possible gain and bandwidth, it would seem reasonable to choose NPN bipolar transistors for use in the differential pair, thus exploiting their high transconductance. This would also allow for an opportunity to experiment with the effects of bipolar inputs on sampled data circuits.

It was suggested in Chapter 4 that it may be possible to tolerate the base current requirements of a bipolar input opamp in a switched capacitor circuit, provided that

the circuit was clocked at the highest possible speed and was fully double ended. As the design of double ended operational amplifiers is somewhat more involved than single ended, particularly in the biasing section, it has been decided to avoid this option.

The desire to limit the design scope to that of a single ended operational amplifier that is readily usable in switched capacitor circuits effectively eliminates the use of a simple bipolar differential pair due to the low input impedance and the bias current in the base. It was discussed previously that techniques such as emitter degeneration and Darlington connected inputs could reduce these difficulties. Emitter degeneration will increase the input impedance of a bipolar differential pair, but will not make any improvement in the DC base current. Emitter degeneration also degrades the gain and bandwidth of the bipolar differential pair to a level that is attainable by MOS transistors. The Darlington configured input pair offers considerable improvement in both input impedance and base current. However this configuration has been shown in Chapter 4 to have a much narrower common mode range than the simple bipolar or MOS differential pairs, due to the number of devices that must be kept in forward active operation.

One of the differential input circuits considered consisted of an NPN differential pair driven by PMOS source followers. This would offer the advantage of infinite input impedance and potentially greater bandwidth than the MOS differential pair. Unfortunately, the common mode range of this configuration is narrow unless the source follower transistors are quite large. Thus it is not an appropriate choice.

The possible candidates for the differential input circuit of the proposed topology have been discussed above. It was decided to eliminate the Darlington and source follower input configurations due to their reduced common mode range. The NPN differential pair and the NPN differential pair with emitter degeneration were eliminated due to the constant base current that would discharge the capacitors in a switched capacitor circuit. The remaining choices are then the NMOS differential pair and the PMOS differential pair. The choice between an NMOS and a PMOS differential pair will have ramifications in the choices to be made for the remainder of the topology.

Two topologies will be proposed, one based on an NMOS differential pair and one based on a PMOS differential pair. A choice between the two will then be made.

5.1.2 Choice of Current Source

Circuits for use as a current source for biasing the differential pair were presented and discussed in Section 4.2. Cascoded current sources are not suitable for biasing a differential pair as the large voltage needed to keep the two devices in series operating as a current source is larger than for a single device and this would unacceptably reduce the input common mode range of the differential pair. If a P type differential pair is chosen, the only bias current source available is a PMOS transistor. If a N type differential pair is chosen, either a NMOS transistor or an NPN transistor could be used. The NPN transistor would be preferred, as a differential pair biased by an NPN transistor would have better common mode rejection and common mode range.

5.1.3 Choice of Output Stage

Possible circuits for use in the output stage of a BiCMOS operational amplifier were discussed in detail in Section 4.5. The advantages and drawbacks of these circuits will now be briefly summarized and a selection made. There are two obvious choices for the design of an output stage: a common emitter amplifier and a common source amplifier, both shown previously in fig. 4.22. Alternatively, there might be no output stage at all, with the output being taken directly as in the folded cascode operational amplifier, shown previously in fig. 3.6.

The option of no output stage (other than the cascode legs of a folded cascode topology) allows a simpler overall circuit. However the need to keep two series connected devices in active operation (in saturation if MOS) limits the available output voltage swing, particularly with a 5 Volt process. Also a folded cascode opamp will likely offer much lower DC gain than an opamp featuring some sort of cascode load and an output gain stage. Thus in order to maximize DC gain, a topology featuring some sort of cascode load and an output gain stage is preferred.

Figure 5.1: Common Emitter Amplifier With Emitter Follower Input

The DC gain of the common emitter and common source single transistor amplifiers shown in fig. 4.22 was presented previously as equation 4.78, and is proportional to both the transconductance and the output resistance of the transistor. As the transconductance and output resistance of bipolar transistors are considerably larger than those offered by MOS transistors of reasonable size, the use of an NPN common emitter amplifier offers the largest possible gain in the output stage.

The drawback of using an NPN transistor as an output driver is that it will draw a large base current due to its high collector current and low input impedance. The low input impedance can load the previous stage, thus severely degrading the gain of that stage. In order to minimize the impact on the previous stage of the opamp, an emitter follower will be used to drive the output transistor as previously shown in fig. 4.23 and repeated in fig. 5.1. This increases the impedance seen by the previous stage in the

same manner as in the Darlington input stage discussed previously. Compensation for the amplifier can be provided in the form of a pole splitting capacitor between the input stage and the output, in the same way as would be done when there was no impedance buffer between the input and output stages.

One constraint that both an NPN common emitter output stage and an NMOS common source output stage place on the remainder of the operational amplifier topology is that the signal applied to the output stage must be biased below analog ground, eg. as for Q3 in fig. 5.1. Such a bias below analog ground works compatibly if the accompanying input differential pair utilizes PMOS transistors. If it is desired to use NMOS transistors for the differential pair, the output of the differential pair will be biased above analog ground and it will be necessary to either use a PMOS common source output stage or to level shift the signal. The use of a PMOS common source output stage is undesirable because the available transconductance is less than that of both NPN and NMOS transistors, and thus maximum bandwidth is less likely to be achieved. Thus some sort of downward level shift circuit is required to allow the use of the NPN emitter follower - common emitter output stage.

All level shifting circuits require the use of complementary devices. In the BiCMOS process available, the only P type device is the PMOS transistor, which does not offer the same level of performance as the NPN transistor. However, it is necessary to use a PMOS device to accomplish level shifting, and hence it is desirable to choose a circuit that has minimal dependence on the performance of these PMOS devices. This can be accomplished by using two PMOS current sources as shown in fig. 5.2 (those biased by V_{B1}), as well as a pair of common gate connected PMOS transistors (those biased by V_{B2}). Thus we have a folded cascode operational amplifier. Under quiescent conditions, half of the current flowing through each of the PMOS current sources will flow through the differential pair, and half will flow through the common gate PMOS transistors and into the cascode load. Any change in the current flowing through the differential pair due to an input voltage will be folded over in to the cascode stage. This level shifts the signal to a potential that can be below analog ground, and an N type output driver can then be used. As the PMOS devices are used as current

sources and common gate stages, their performance is perhaps less bandwidth limiting than in circuits where they are used as gain elements, provided that it is possible to keep the parasitic pole at the source of the common gate stage (transistors biased by V_{B2} in fig. 5.2) at a frequency sufficiently higher than the unity gain bandwidth of the overall amplifier. This issue can only be resolved in a specific design situation.

Figure 5.2: Proposed NMOS Input BiCMOS Operational Amplifier Topology

A topology featuring an NMOS input differential pair with cascode folding and a common emitter output stage is shown in fig. 5.2. An alternate topology allowing the NPN common emitter output stage with a PMOS input differential pair and without

Figure 5.3: Proposed PMOS Input BiCMOS Operational Amplifier Topology

cascode folding is shown in fig. 5.3. The merits of these two choices will be discussed and a choice between them made for design and simulation below.

5.1.4 Proposed Topology

The circuit shown in fig. 5.2 features an NMOS differential pair loaded by a folded cascode stage as described above. The folding action of the PMOS current sources allows the use of the NPN common emitter output stage rather than a PMOS common source amplifier. An emitter follower is used as an impedance buffer between the

cascode load and the common emitter amplifier. It could be said that this topology is essentially a BiCMOS implementation of a folded cascode operational amplifier with the addition of an output stage. It can also be considered to have the topology of the two stage or pole splitter operational amplifier, with a folded cascode operational amplifier serving as the input stage.

The circuit shown in fig. 5.3 is a BiCMOS implementation of the two stage operational topology previously shown in fig. 3.3. In fig. 5.3, the circuit has been flipped to allow the use of a PMOS differential pair and an N type output device. The same emitter follower - common emitter output stage as used in fig. 5.2 was chosen as the output stage.

In both fig. 5.2 and fig. 5.3, a Wilson current mirror was used to load the differential pair and accomplish the double to single ended conversion. The Wilson current mirror offers better current matching than any other bipolar current mirror and can be realized using less area than a MOS cascode current mirror carrying a similar current. The output resistance is higher than can be obtained from MOS cascode current mirrors, although this is not likely to be important due to the low input impedance of the emitter follower - common emitter output stage. This current mirror is however well suited in terms of output voltage level compatibility with the emitter follower - common emitter output stage chosen. It can easily be seen from fig. 5.2 or fig. 5.3 that the voltage at the base of the emitter follower transistor will be "pinned" at $V_{SS} + 2V_{BE}$ by the base-emitter voltages of the common emitter output transistor and the emitter follower transistor. It can also be seen that the voltage at the corresponding point on the other leg of the cascode will also be "pinned" at $V_{SS} + 2V_{BE}$ by the base-emitter voltages of transistors in the Wilson current mirror. Under quiescent conditions, the common gate PMOS transistors of fig. 5.2 or the transistors of the PMOS differential pair in fig. 5.3 will see roughly the same drain-source voltage, thus helping to ensure that the cascode remains balanced.

Limitations on time and available silicon area precluded pursuing the design and layout of both topologies of fig. 5.2 and fig. 5.3. As both candidates are essentially two stage operational amplifiers, the design equations presented in Section 3.2.1 can

be useful in either case. An advantage of the PMOS input topology shown in fig. 5.3 over the NMOS input topology shown in fig. 5.2 is that the PMOS transistors can be formed in a well, eliminating the backgate bias, but at the expense of a greater input capacitance. This increased input capacitance is due to the larger gate area required for a given transconductance, and in the case of common mode inputs, the n-well to substrate capacitance. The well capacitance will reduce the common mode rejection at high frequencies. A second advantage of PMOS is that the power consumption of this circuit will be less than that of fig. 5.2 due to the absence of the cascode stage which consumes as much power as the differential pair. A drawback of the PMOS input topology is that the current source supplying the tail current to the differential pair must also be a PMOS transistor. Compared to the circuit of fig. 5.2, the lower output resistance of the PMOS current source will result in a reduced common mode rejection ratio, as described in Section 4.2. The larger minimum drain source voltage needed to keep the PMOS current source transistor in saturation may reduce the common mode range. The most compelling disadvantage of the circuit with a PMOS differential pair is due to the characteristics of the NMOS and PMOS transistors in the process. The transconductance parameter of the PMOS transistors is half or less than that of the NMOS transistors, due to the difference between electron and hole mobility. Thus adding the effect of a separate well, the transconductance of a PMOS transistor will be roughly 60% of the transconductance of an NMOS transistor having the same geometry. In order to maximize the bandwidth of an operational amplifier, the transconductance of the differential pair must also be maximum. For this reason, it was decided that the NMOS input operational amplifier topology shown in fig. 5.2 would be pursued.

5.1.5 Summary

A number of potential BiCMOS operational amplifier circuit components were presented and discussed in Chapter 4. It was then decided that a MOS differential pair should be employed as the input stage in the proposed topology for switched capacitor contexts. Two topologies were proposed, based on the use of NMOS and

PMOS differential pairs, and the NMOS differential pair topology shown in fig. 5.2 was selected. Details of the actual design and calculated performance will be presented in a subsequent section. Results from simulation will be presented and compared to performance predicted from the equations developed in Chapter 4. A measurement strategy will then be proposed to be used when fabricated test circuits are available. Finally a number of alternative topologies that were considered prior to the selection of the NMOS input topology of fig. 5.2 will be briefly presented.

5.2 Simulation Conditions

During the course of research for this thesis, several operational amplifier circuits have been designed and simulated in a number of different environments. Early simulations were performed using a set of SPICE parameters for a hypothetical 1 micron BiCMOS process used for course work [9]. These SPICE parameters are included in Appendix A. Later simulations were performed using a proprietary set of process parameters for the 0.8 micron BiCMOS process in which the circuits were designed and fabricated. Due to the proprietary nature of the process used, simulations of the opamp were performed only for nominal conditions and processing. The nominal temperature and power supply voltages were 27 degrees C and ± 2.5 Volts respectively. The simulations to be presented in Section 5.3 include the effect of parasitic capacitances which were extracted from the opamp layout using the extraction package of the CAD framework used in the opamp design.

The process used to design the following opamp can be described as an 0.8 μm analog and digital BiCMOS technology [10]. The devices used in the opamp design were N and P channel MOS field effect transistors, vertical NPN bipolar transistors, double polysilicon capacitors and polysilicon resistors. The substrate doping is P type, thus NMOS transistors are formed in the substrate, and PMOS transistors are formed in N doped wells. The NPN transistors are constrained to a library of standard layouts due to modeling concerns. These transistors are identified by the ratio of the emitter area of each transistor to the emitter area of the smallest transistor, for

example, a 5X transistor has five times the emitter area of a 1X transistor.

5.3 A BiCMOS Operational Amplifier with NMOS Differential Pair

The circuit that has been developed and submitted for fabrication is shown in fig. 5.4. A computer plot of the layout, on which the major structures of the opamp

Figure 5.4: BiCMOS Operational Amplifier with NMOS Differential Pair

have been identified, is shown in fig. 5.5 and also in Appendix B. This circuit is that of the proposed topology shown in fig. 5.2 with the addition of diode connected transistors M6 and Q11 to allow the use of current biasing rather than voltage biasing. The biasing scheme will be described in more detail presently. Major design decisions and assumptions will now be described.

Figure 5.5: NMOS Input Operational Amplifier Layout Plot

5.3.1 Design Process

The starting point of the design was the choice of reasonable values for the load and compensation capacitors and a reasonable value for the tail current of the differential pair. Based on typical capacitor values used in switched capacitor circuits, the values of the load and compensation capacitors were both chosen to be 5 pF. This was a relatively large capacitance but was chosen to make the amplifier general purpose. A smaller capacitance load would allow lower currents and hence reduce the power dissipation of the opamp as well as potentially increasing the bandwidth. The tail current (collector current of Q3) was chosen to be $600\mu A$, as a compromise between the desire to operate at high currents (higher transconductance) and the desire to keep device sizes from becoming excessive. Transistor Q3 was realized using six 1X transistors in parallel, each operating at a collector current of $100\mu A$. This approach was taken in order to maximize the accuracy of the the current ratios of Q3, Q9 and Q11, and to allow the bias current supplied to the collector of Q11 (a single 1X device) to be only $100\mu A$, thus reducing the power dissipated by the opamp. A drawback of this approach is that the collector capacitance of Q3 is much greater than would be the case for a single transistor of equal emitter area and thus the common mode rejection ratio will be slightly reduced. The silicon area consumed by an array of 1X transistors is also greater than would be consumed by a single larger transistor. The current density of $100\mu A$ in a 1X transistor, although somewhat less than that at which peak f_T will be reached, ensures that the transistor will still be operating at or near its maximum β , thus minimizing the β error of the current mirrors.

The bias current chosen for the PMOS current source transistors M7 and M8 was chosen to be the same as the tail current of $600\mu A$, as is usual for a folded cascode operational amplifier as described in Section 3.2.3. The bias collector current for the common emitter output stage was set at 2.5 times the differential pair tail current as described in the two stage opamp design procedure in Section 3.2.1, to bias its parasitic pole beyond the unity gain bandwidth and to ensure that the slew rate of the opamp will be limited by the input current and compensation capacitor rather than by the output stage and load. In order to minimize loading of the cascode stage,

the emitter follower in the output stage was biased at the lowest convenient current level ($100\mu A$) by using a 1X transistor (Q9) connected to the same current mirror as the tail current source transistor Q3.

The design objective of this opamp was simply to make use of a newly available BiCMOS process in order to attempt to obtain the maximum possible bandwidth of an opamp that would be suitable for switched capacitor applications. Thus it was decided to make the input transistors as large as practical and to use minimum gate length transistors to maximize the W/L ratio and hence, the transconductance of the input pair. This is allowable, as regards gain, because the cascode loading pins the drains of M1 and M2 at a relatively fixed voltage. Thus the ability to use shorter channel length inputs is another important advantage of the configuration shown in fig. 5.2 over that shown in fig. 5.3. The use of minimum channel length input transistors, along with an interdigitated layout as can be seen in fig. 5.5, minimizes the contribution of the transistors to the parasitic capacitances at the sources of cascode transistors M3 and M4. The capacitance at the sources of transistors M3 and M4 and the transconductance of these transistors form a non-dominant pole which may have a low enough frequency to affect the stability of the opamp. The final size of the input transistors was set after several simulation iterations to determine when the "point of diminishing returns" had been reached.

There are a number of drawbacks to the above approach and the large W/L ratios that were chosen. The large gate area means that the input capacitance of the opamp will be large (on the order of 0.7 pF) and may cause difficulties when the opamp is placed in a real switched capacitor system rather than being driven by a voltage source. The large W/L also means that the range of linear operation (voltage) of the opamp will be small and that the input transistors may be very close to the subthreshold region of operation.

When choosing transistor sizes for the PMOS current sources M7 and M8, it was desired to employ the guideline of $V_{GS} - V_T \geq 0.5V$, and to make this number on the high side to make the devices smaller and reduce their drain capacitances. This was possible because the output of the folded cascode stage does not need large voltage

swings, it being followed by another gain stage. It was then chosen to make transistors M3 and M4 the same size as transistors M7 and M8, allowing a compact interdigitated layout as can be seen in fig. 5.5 and reducing parasitic capacitances by sharing drains and sources, so as to reduce the effect of the parasitic pole at this point.

The transistors forming the Wilson current mirror (Q4, Q5, Q6 and Q7) were each made of two 1X transistors connected in parallel. This allowed the use of a common centroid layout arrangement (shown in fig. 5.5) which helps to minimize the effect of thermal gradients. This is of concern as the Wilson current mirror is located close to the output stage which dissipates substantial power. The output transistor Q10 was chosen to be a single larger device as neither matching nor thermal gradients were of concern here. This results in a current density for Q10 that gives a high β and is below that at which peak f_T is reached. It is desirable to operate at a lower current density than the peak f_T value as f_T decreases steeply beyond its peak value but decreases only gradually at lower current densities.

The final step of the design was to choose the value of the nulling resistor R_C . This was done by trial and error with the simulator until the value that maximized bandwidth while maintaining acceptable gain and phase margins was found.

5.3.2 Biasing Scheme

In order to eliminate the need to design a complex bias circuit, and to allow the bias conditions to be changed during testing, it was decided supply the bias currents from off chip. This allows investigation of how the bias conditions affect performance, and affords an opportunity to compensate at least in part for any design errors or process variation problems that may be uncovered during testing. The biasing scheme including external resistors and diode connected PMOS transistors on chip is illustrated in fig. 5.6. The current sources for the cascode legs and the output stage are biased by applying a current to diode connected transistor M6 through a potentiometer to V_{SS} . The current source for the differential pair and the emitter follower driving the output transistor are biased in a similar manner by supplying a current to Q11. The sizes of M6 and Q11 have been chosen such that the magnitude

Figure 5.6: Biasing Scheme for BiCMOS Op Amp with NMOS Differential Pair

of the two bias currents should be the same. Thus it should be possible to connect them together with a single resistor of suitable value rather than using separate resistors between each bias input and the power supply rail. The gates of the cascode transistors are brought out to a pin and thus may be voltage biased using either a resistive voltage divider or a variable voltage power supply.

5.3.3 Performance

Despite the complexity of this circuit, the relations previously developed for the two stage (pole splitter) CMOS operational amplifier can be applied here. This is because the cascode stage can be considered to fold over the current from the differential pair, while still obeying the relationships for a simple two stage opamp. This of course assumes that the parasitic pole at the source of M3/M4 does not intrude unduly, as well as any other parasitic poles not yet considered.

The performance of this opamp has been simulated, including the effect of post layout parasitic capacitances, using a proprietary set of process parameters. The results of these simulations will be presented graphically and compared with theoretical performance predicted using the relations developed in Chapter 4. Where needed, values of transistor parameters such as transconductance and output resistance will be taken from the simulator output rather than calculated using the simplified relations presented previously.

Bandwidth.

The unity gain bandwidth of a CMOS two stage operational amplifier was presented in equation 3.4 as

$$\omega_o = \frac{g_{mi}}{C_C} \quad (5.1)$$

assuming that the phase shift is near 90 degrees, ie; the dominant pole is truly dominant. In the case of a folded cascode topology such as in fig. 5.4, some of the transconductance of the differential pair will be shunted by the conductance of the cascode transistor (M3 in fig. 5.4), thus equation 5.1 will be somewhat optimistic. In a circuit such as this where the transistors are physically large and operate at high current levels, it is entirely possible that the frequencies of the non-dominant poles will be low and thus violate the assumption of equation 5.1. Due to both the possible existence of relatively low frequency non-dominant poles and the shunting action of the conductance of the cascode transistors, it is thus not unreasonable to expect a discrepancy between equation 5.1 and simulation results. In order to see the magnitude of this discrepancy, equation 5.1 will be evaluated using the actual load capacitance and the transconductance of the differential pair reported by the simulator. The transconductance of the transistors of the differential pair was reported by the simulator to be

$$g_{mi} = 4.96 \times 10^{-3} \Omega^{-1}. \quad (5.2)$$

The unity gain bandwidth of the opamp is then predicted using equation 5.1 to be

$$\omega_o = \frac{g_{mi}}{C_C} \quad (5.3)$$

$$= \frac{4.96 \times 10^{-3}}{5 \times 10^{-12}} \quad (5.4)$$

$$= 992 \times 10^6 \text{ Rad/s} \quad (5.5)$$

$$f_o = \frac{\omega_o}{2\pi} \quad (5.6)$$

$$= 158 \text{ MHz.} \quad (5.7)$$

The simulation results for frequency domain analysis are presented in fig. 5.7. The

Figure 5.7: NMOS Input Opamp Frequency Response

scale of fig. 5.7 is not appropriate to read bandwidth, phase and gain margins and DC gain from the graph with any accuracy; however, these quantities can easily be read from the simulator output file and are

$$f_o = 129 \text{ MHz} \quad (5.8)$$

$$\phi M = 61.0 \text{ degrees} \quad (5.9)$$

$$\text{Gain Margin} = 9.7 \text{ dB} \quad (5.10)$$

$$A_{\text{DC}} = 101.8 \text{ dB} \quad (5.11)$$

Other recent examples in the literature include a CMOS opamp with a unity gain bandwidth of 120 MHz into a load of 16 pF [3] and a bipolar opamp with a unity gain bandwidth of 100 MHz into a 100 pF load [4]. The bandwidth achieved by this opamp cannot be considered a major improvement over the results of others, however, the information learned during its development should allow significant improvements in bandwidth to be made in future designs.

The unity gain bandwidth reported by the simulator is thus somewhat less than that predicted by equation 5.1. As was discussed above, this likely indicates that one or more non-dominant poles exist at a sufficiently low frequency to have sufficient effect on the opamp at frequencies near the unity gain bandwidth, thus violating some of the assumptions on which equation 5.1 is based. This assumption of non-dominant poles at relatively low frequencies is supported by the small phase and gain margins which are only sufficient to ensure stability of the opamp.

The most likely non dominant pole has already been mentioned; it is that formed by the transconductance of M3 and the parasitic capacitances at the source of M3. One can approximate the capacitance seen by the source of M3 by summing the drain-substrate (C_{db7}) and drain-gate (C_{dg7}) capacitances of M7, the source-gate (C_{sg3}) and source-substrate (C_{sb3}) capacitances of M3 and the drain-substrate (C_{db1}) capacitance of the input transistor M1. The values for these capacitances and the transconductance of M3 were reported by the simulator to be

$$g_{m3} = 1.42 \times 10^{-3} \Omega^{-1} \quad (5.12)$$

$$C_{db7} = 144.2 \times 10^{-15} F \quad (5.13)$$

$$C_{dg7} = 56.9 \times 10^{-15} F \quad (5.14)$$

$$C_{sg3} = 247.8 \times 10^{-15} F \quad (5.15)$$

$$C_{sb3} = 89.0 \times 10^{-15} F \quad (5.16)$$

$$C_{db1} = 524.5 \times 10^{-15} F \quad (5.17)$$

The approximate location of this pole is then

$$\omega_{p2} = \frac{g_{m3}}{C_{db7} + C_{dg7} + C_{sg3} + C_{sb3} + C_{db1}} \quad (5.18)$$

$$= \frac{1.42 \times 10^{-3}}{(144.2 + 56.9 + 247.8 + 89.0 + 524.5) \times 10^{-15}} \quad (5.19)$$

$$= 1.34 \times 10^9 \text{ rad/s} \quad (5.20)$$

$$f_{p2} = \frac{\omega_{p2}}{2\pi} \quad (5.21)$$

$$= 213 \text{ MHz.} \quad (5.22)$$

Assuming that all other non-dominant poles are at somewhat higher frequencies, the phase shift due to the pole at the source of M3 will be

$$\phi_{p2} = -\arctan\left(\frac{\omega_o}{\omega_{p2}}\right) = -\arctan\left(\frac{f_o}{f_{p2}}\right) = -\arctan\left(\frac{129}{213}\right) = -31.2 \text{ degrees.} \quad (5.23)$$

Thus one would expect that the phase margin would be reduced from the single pole value of 90 degrees to about 60 degrees, which agrees with the actual value of 61 degrees reported above. Thus the non-dominant pole limiting the bandwidth of this opamp is that due to the cascode transistor M3 as described above.

DC Gain.

As this opamp can be considered to be a two stage opamp, the DC gain is simply the product of the gains of the two stages, and is easily determined once the significant active devices and loads are identified. The DC gain of a MOS differential pair driving a resistive load has been presented in equation 4.10 and is

$$A_{dm} = -\frac{g_{m1} R_L r_{o1}}{r_{o1} + R_L} \quad (5.24)$$

Equation 5.24 is not applicable in this case as it was derived with the load resistors connected directly to the drain of the differential pair transistors. In fig. 5.4, the output current of the differential pair is "folded over" by current source transistors M7 and M8, is impedance transformed by the cascode transistors M3 and M4, and is then applied to a load. The Wilson current mirrors are diode connected, and make the common mode load resistance much less than the difference mode resistance.

The gain of a folded cascode operational amplifier was presented as equation 3.19 which, with the transistor numbering adjusted to correspond to that of fig. 5.4 and the terms relating to the resistance presented by the Wilson current mirror (formed by Q4, Q5, Q6 and Q7) replaced by R_{out} , is

$$A_{dm} = -\frac{g_{m1}}{\frac{1}{R_{out}} + \frac{g_{o1} + g_{o8}}{g_{m4} r_{ol}}} \quad (5.25)$$

In the absence of an output stage, R_{out} would be the output resistance of the Wilson current mirror formed by Q4, Q5, Q6 and Q7. Since the emitter follower - common emitter output stage has a low input impedance, the load resistance R_{out} in equation 5.25 will be the parallel combination of the output resistance of the Wilson current mirror (R_{mir}) and the input resistance of the output stage (R_{ef}). The total load resistance seen by the transconductance of the differential pair is the parallel combination of R_{out} and the conductance seen looking into the drain of the PMOS cascodes which is the $\frac{g_{o1} + g_{o8}}{g_{m4} r_{ol}}$ term in equation 5.25. The output resistance of the cascode current mirror was presented previously as equation 4.77 and is

$$R_{mir} \approx \frac{1}{2} \beta r_{ol} \quad (5.26)$$

where r_{ol} is the output resistance of the transistors forming the current mirror. The output resistance of the NPN transistors that form the current mirror was reported by the simulator to be

$$r_{ol} = 173 K\Omega \quad (5.27)$$

Equation 5.26 evaluates to

$$R_{mir} = \frac{1}{2} \beta r_{ol} \quad (5.28)$$

$$= (0.5)(100)(173 \times 10^3) \quad (5.29)$$

$$= 8.65 \times 10^6 \Omega \quad (5.30)$$

The input resistance of the emitter follower - common emitter output stage, presented previously as equation 4.81 and expressed using the transistor numbering of fig. 5.4, is

$$R_{ef} = r_{\pi 8} + \frac{\beta + 1}{g_{o8} + g_{o9} + g_{\pi 10}} \quad (5.31)$$

In general, g_o is about an order of magnitude less than g_π . In this particular case, it is known that the collector current of Q8 and Q9 is an order of magnitude less than that of Q10, thus g_{o8} and g_{o9} will be about two orders of magnitude less than $g_{\pi10}$ and can be ignored. The values of the parameters in $g_{\pi8}$ and $g_{\pi10}$ reported by the simulator were

$$g_{\pi8} = 48.28 \times 10^{-6} \Omega^{-1} \quad (5.32)$$

$$g_{\pi10} = 692.7 \times 10^{-6} \Omega^{-1} \quad (5.33)$$

Equation 5.31 then evaluates to

$$R_{ef} = \frac{1}{g_{\pi8}} + \frac{\beta + 1}{g_{\pi10}} \quad (5.34)$$

$$= \frac{1}{48.3 \times 10^{-6}} + \frac{101}{692.7 \times 10^{-6}} \quad (5.35)$$

$$= 166.5 K\Omega \quad (5.36)$$

It can be seen from the above calculations that R_{ef} of the output stage is much smaller than R_{mir} of the Wilson current mirror. That is, the difference mode resistance of the Wilson current mirror is so high as to not be a factor in the difference mode gain. Thus R_L in equation 5.25 can be replaced by the value of R_{ef} . The values of the parameters in equation 5.25 reported by the simulator are

$$g_{oi} = 183 \times 10^{-6} \Omega^{-1} \quad (5.37)$$

$$g_{mi} = 4.96 \times 10^{-3} \Omega^{-1} \quad (5.38)$$

$$g_{o4} = 35.8 \times 10^{-6} \Omega^{-1} \quad (5.39)$$

$$g_{m4} = 1.42 \times 10^{-3} \Omega^{-1} \quad (5.40)$$

$$g_{o8} = 54.8 \times 10^{-6} \Omega^{-1} \quad (5.41)$$

By substitution of the above values into equation 5.25, the difference mod · DC gain of the first stage becomes

$$A_1 = - \frac{g_{mi}}{\frac{1}{R_{ef}} + \frac{(g_{oi} + g_{o8})g_{o4}}{g_{m4}}} \quad (5.42)$$

$$= - \frac{4.96 \times 10^{-3}}{\frac{1}{166.5 \times 10^3} + \frac{((183 + 54.8) \times 10^{-6})(35.8 \times 10^{-6})}{1.42 \times 10^{-3}}} \quad (5.43)$$

$$= -413 \quad (5.44)$$

$$|A_1| \approx 52.3 \text{ dB} \quad (5.45)$$

The DC gain of the emitter follower - common emitter output stage was presented in equation 4.80 as

$$A_2 = -\frac{g_{m8}g_{m10}}{(g_{o10} + g_{o5})(g_{m8} + g_{o9} + \frac{g_{m10}}{\beta})} \quad (5.46)$$

where the transistor identifications correspond to device names in fig. 5.4. The values of the parameters in equation 5.46 reported by the simulator are

$$g_{m8} = 4.387 \times 10^{-3} \Omega^{-1} \quad (5.47)$$

$$g_{m10} = 59.97 \times 10^{-3} \Omega^{-1} \quad (5.48)$$

$$g_{o5} = 120.6 \times 10^{-6} \Omega^{-1} \quad (5.49)$$

$$g_{o9} = 1.930 \times 10^{-6} \Omega^{-1} \quad (5.50)$$

$$g_{o10} = 30.20 \times 10^{-6} \Omega^{-1} \quad (5.51)$$

Substitution of these values into equation 5.46 gives

$$A_2 = -\frac{g_{m8}g_{m10}}{(g_{o10} + g_{o5})(g_{m8} + g_{o9} + \frac{g_{m10}}{\beta})} \quad (5.52)$$

$$= -\frac{(4.4 \times 10^{-3})(58 \times 10^{-3})}{((30 + 120) \times 10^{-6})((4.4 \times 10^{-3}) + (1.93 \times 10^{-6}) + \frac{58 \times 10^{-3}}{100})} \quad (5.53)$$

$$= -342 \quad (5.54)$$

$$|A_2| = 50.7 \text{ dB} \quad (5.55)$$

The DC gain of the complete opamp is then

$$|A_{DC}| = |A_1| \times |A_2| = 52.3 + 50.7 = 103.0 \text{ dB}. \quad (5.56)$$

As was stated previously, the DC gain reported by the simulator was 101.8 dB which compares quite well with the calculated DC gain of 103 dB above. The DC gains reported for recent examples of opamps in the literature include 90 dB for a CMOS opamp with a unity gain bandwidth of 120 MHz into a load of 16 pF [3] and 100 dB for a bipolar opamp with a unity gain bandwidth of 100 MHz into a 100 pF load [4]. The DC gain of the opamp presented here is of the same order, and slightly better than either of the two examples from the literature.

Common Mode Range

Figure 5.8: NMOS Input Opamp DC Common Mode Response (Closed Loop)

In order to determine the common mode range of the opamp, a simulation of the DC transfer function of the opamp connected as a voltage follower was performed. Due to the voltage follower connection, the signal applied to the non-inverting output will also appear at the inverting input and is thus common mode.⁴ The results of this simulation are shown in fig. 5.8. The limit of common mode range has been reached when the curve of fig. 5.8 departs from a straight line. This occurs approximately -1.9 V and +1.7 V which are the lower and upper limits of common mode range.

The common mode range for a MOS differential pair was calculated for an arbitrary load in Chapter 4. This calculation will be repeated for the actual opamp circuit using voltages obtain from the simulation. The relevant transistor voltages

reported by the simulator were:

$$V_{GS,1} = 0.96V \quad (5.57)$$

$$V_{DS,1} = 1.73V \quad (5.58)$$

$$|V_{DS,7}| = 1.73V \quad (5.59)$$

$$V_T = 1.06V \quad (5.60)$$

The lower limit of common mode range can then be calculated, assuming that the saturation voltage of Q3 is $0.2V$ as

$$V_{cm,min} = V_{SS} + V_{CE,sat} + V_{GS,1} \quad (5.61)$$

$$= -2.5 + 0.2 + 0.96 \quad (5.62)$$

$$= -1.34V \quad (5.63)$$

The above relation is based on simple square law relationships that are only approximate in the real world. The lower limit of common mode range determined by simulation can be seen in fig. 5.8 and was stated above to be $-1.9V$, suggesting that the square law relations are pessimistic. The upper limit of common mode range can be calculated by assuming that the drain-source voltage of the current source transistors M7 and M8 is approximately constant. Simulation results (not shown) indicate that this is indeed the case. Using a similar approach to the calculation of the lower limit of common mode range, the upper limit is

$$V_{cm,max} = V_{DD} - |V_{DS,7}| - V_{DS,1,min} + V_{GS,1} \quad (5.64)$$

$$= V_{DD} - |V_{DS,7}| - (V_{GS,1} - V_T) + V_{GS,1} \quad (5.65)$$

$$= V_{DD} - |V_{DS,7}| + V_T \quad (5.66)$$

$$= 2.5 - 1.73 + 1.06 \quad (5.67)$$

$$= 1.83V \quad (5.68)$$

As this relation is also based on the simple square law equations, it agrees reasonably well the result, from fig. 5.8, of $1.7V$.

Common Mode Rejection

The common mode rejection ratio of an opamp is dependent upon the output resistance of the current source supplying the tail current to the input differential pair and upon differences in the small signal resistance presented by the load of the differential pair to difference and common mode signals. If the differential pair is loaded by linear resistors as in fig. 4.5 (repeated as fig. 5.9), the load resistance seen by the differential pair is the same for both difference and common mode signals.

The difference and common mode gains and the common mode rejection ratio

Figure 5.9: MOS Differential Pair

were presented as equations 4.10, 4.11, and 4.13 respectively. These equations are repeated below with transistor output resistances expressed as conductances rather than resistances as was done in Chapter 4. The difference mode gain of a MOS

Figure 5.10: MOS Differential Pair with Active Load

differential pair loaded by resistors is

$$A_{dm} = -\frac{g_{mi}}{g_{ol} + g_{oi}} \quad (5.69)$$

where g_{mi} and g_{oi} are the transconductance and output conductance of the input transistors and g_{ol} is the conductance of the load resistor. The common mode gain of a MOS differential pair loaded by resistors is

$$A_{cm} = -\frac{G_S g_{mi}}{2g_{mi}g_{ol} + G_S g_{oi}} \quad (5.70)$$

where G_S is the output conductance of the current source biasing the differential pair and g_{mi} , g_{oi} and g_{ol} are as above. The common mode rejection ratio is then

$$CMRR = \frac{A_{dm}}{A_{cm}} = \frac{G_S g_{oi} + 2g_{mi}g_{ol}}{G_S (g_{ol} + g_{oi})} \quad (5.71)$$

It should be noted that the above equations apply for both resistor loads and independent active loads. In the latter case g_{ol} in the preceding equations would be the output conductance of externally biased load transistors or would be replaced by the

transconductance of the load transistors if they are diode connected. If it can be assumed that g_{ol} and g_{oi} are similar in value, and noting that most likely $g_{mi} \gg G_S$, equation 5.71 reduces to

$$CMRR = \frac{g_{mi}}{G_S} \quad (5.72)$$

If g_{oi} is large, equation 5.71 reduces to

$$CMRR = \frac{2g_{mi}}{G_S} \quad (5.73)$$

Using the transconductance of the differential pair of $4.96 \times 10^{-3} \Omega^{-1}$ and the value of G_S reported by the simulator of $11.6 \times 10^{-6} \Omega^{-1}$, equations 5.72 and 5.73 evaluate to 53 dB and 59 dB respectively.

It is possible (and desirable) for the load resistance seen by the differential pair to be different for common mode and difference mode signals as was mentioned already in Section 5.3.3. As a simple example, a MOS differential pair with an active load in the form of a current mirror is shown in fig. 5.10. The difference and common mode gains can be shown to be [8]

$$A_{dm} = -\frac{g_{mi}}{g_{ol} + g_{oi}} \quad (5.74)$$

and

$$A_{cm} = -\frac{G_S g_{mi}}{2g_{mi}(g_{oi} + g_{ol})} \quad (5.75)$$

where g_{mi} and g_{oi} are the transconductance and output conductance of the input transistors M1 and M2, g_{ol} is the output conductance of the load transistors M3 and M4, and G_S is the output conductance of the current source transistor M5. The common mode rejection ratio is then

$$CMRR = \frac{A_{dm}}{A_{cm}} = \frac{2g_{mi}g_{mi}}{G_S g_{oi}} \quad (5.76)$$

By comparing equations 5.76 and 5.73, the replacement of single load resistors by a current mirror increases the common mode rejection ratio by a factor of $\frac{g_{mi}}{g_{oi}}$. Depending on the types, sizes and currents of the various devices, $\frac{g_{mi}}{g_{oi}}$ can range from about 20 dB to 60 dB.

It can be seen from equations 5.70 and 5.75 that the replacement of individual load resistors with a current mirror causes a significant difference in the common mode gain expression with no change in the difference mode gain expression. Thus although the equations developed in Chapter 4 are useful for comparing different differential input circuits, they cannot be readily used to predict the common mode rejection ratio of a specific opamp such as that shown in fig. 5.4. The Wilson current mirror load chosen will have a beneficial effect in decreasing common mode output resistance in a manner similar to that of the simple mirror of fig. 5.10 as was verified by the frequency domain simulation of the common mode rejection ratio. The result of this simulation is shown graphically in fig. 5.11. The DC common mode rejection ratio reported by the simulator was 114.3 dB, whereas the calculation assuming there was no mirroring effect, ie; the same load for difference and common mode, gave a much lower value of 59 dB. The difference between the common mode rejection ratio of 114 dB for the complete opamp and 59 dB predicted due to the differential pair alone is 55 dB which is within the range of improvement in common mode rejection that can be expected due to the use of a current mirror load over simple resistors. The cascode transistors will serve to decrease the effective output conductance of the differential pair and are thus also responsible for some of the improvement in common mode rejection.

Power Supply Rejection

It is stated in the literature [16, 6] that for opamps employing a pole splitting output stage, the high frequency power supply rejection ratio to the power supply to which the output transistor is connected will be poor. This is due to the fact that, as the frequency increases, the impedance of the compensation capacitor becomes small and effectively diode connects the output transistor, thus coupling power supply noise to the output. In the case of the opamp here, the output transistor is connected to the negative supply and thus the high frequency power supply rejection ratio to V_{SS} will be poor. As the frequency increases, the V_{SS} power supply gain will approach unity and the V_{SS} power supply rejection ratio at a given frequency will be approximately

Figure 5.11: NMOS Input Opamp Common Mode and Power Supply Rejection Ratio

the same as the difference mode gain of the opamp at the frequency in question.

It is expected that the high frequency power supply rejection to the V_{DD} supply will be somewhat better than to the V_{SS} supply. Any V_{DD} noise entering through the cascode current source transistors (M7 and M8 in fig. 5.4) should appear as a common mode signal and be rejected by the circuit. V_{DD} noise will also enter the opamp through the output stage current source transistor (M5 in fig. 5.4) and the emitter follower driving the output transistor (Q8 in fig. 5.4). Due to the high output resistance of bipolar transistors, the effect of the latter should be small. Also any effect of power supply noise entering via Q8 would tend to cancel noise entering through M5 due to the inverting nature of the common emitter output stage (Q10). V_{DD} noise entering through M5 will appear at the output, although the amplitude

will be small and will decrease with frequency at about the same rate as the opamp gain due to the load capacitance, C_L . When the effects of V_{DD} noise entering through the cascode current source transistors M7 and M8 and entering through the output stage current source transistor M5 are combined, the V_{DD} power supply rejection ratio plotted against frequency should look like the common mode rejection ratio shifted down to a slightly lower value.

Simulation results depicting common mode rejection ratio and the power supply rejection ratios are shown graphically in fig. 5.11. As predicted, the V_{SS} power supply rejection ratio decreases to unity at approximately the unity gain bandwidth of the opamp and is less than the V_{DD} power supply rejection ratio. Also as predicted, the V_{DD} power supply rejection ratio shown in fig. 5.11 is approximately a downward shifted copy of the common mode rejection ratio. The simulator reported the following low frequency power supply and common mode rejection ratios

$$V_{DD} \text{ PSRR} = 105.8 \text{ dB}$$

$$V_{SS} \text{ PSRR} = 93.9 \text{ dB}$$

$$\text{CMRR} = 114.3 \text{ dB}$$

Although the V_{SS} power supply rejection ratio is still reasonably good, the fact that it is reduced by the use of the common emitter output stage is of some concern. One method of increasing the V_{SS} power supply rejection ratio while keeping the topology of a single ended opamp is to use a MOS transistor biased in saturation as the compensation resistor [8]. This is accomplished by biasing the MOS transistor with a current source connected to both its source and drain. This transistor and bias network can be designed so that the small signal resistance is the same as the fixed compensation resistor. However the use of a transistor serves to decouple the input of the output stage from the compensation capacitor and the power supply gain will be increased. An alternative approach to improving the V_{SS} power supply rejection ratio is to convert to a fully differential op amp. In that case, there will be two output stages, each of which will be affected similarly by an V_{SS} power supply noise which will be manifested as a common mode signal on the output. This common mode signal will be cancelled by the opamp in the next stage of the filter.

DC Transfer Function

Figure 5.12: NMOS Input Opamp DC transfer Function

The simulated open loop DC transfer function of this operational amplifier is presented in fig. 5.12. In order to more easily examine the behavior of the opamp about its normal operating point, the voltage range over which the inverting input was swept in fig. 5.12 has been limited to $\pm 50\mu V$. By measuring the slope of a line tangent to the curve where V_{out} is near ground, it can be seen that the DC gain of the opamp is approximately 101 dB, agreeing well with the results obtained from frequency domain simulation.

Transient Response

Figure 5.13: NMOS Input Opamp Open Loop Transient Response

The simulated transient response to a square wave applied to the input of the opamp is presented for the opamp running open loop in fig. 5.13 and for the opamp connected as a voltage follower in fig. 5.14. The open loop simulation was performed to investigate the output slew rate of the opamp and to observe its behavior under large output voltage excursions. The voltage follower simulation was performed to investigate the settling behavior of the opamp, and was run with an input square wave of $\pm 0.5V$ in order to avoid any effects due to large output excursions.

An opamp will reach its maximum slew rate when a large differential voltage is applied to the inputs, and almost all of the tail current (collector current of Q3 in fig. 5.4) will flow in one of the two transistors of the differential pair. As the bias

Figure 5.14: NMOS Input Opamp Closed Loop Transient Response

current levels of the PMOS current source transistors M7 and M8 are each the same as the tail current, one leg of the cascode stage will have almost no current flowing down it and the other leg will conduct a current, approximately equal to the tail current, which will charge or discharge the compensation capacitor C_C . Assuming that the output current of all three current sources remains equal and constant, the slew rate of the opamp may be calculated to be

$$S_r = \frac{I_o}{C_C} \quad (5.77)$$

$$= \frac{600\mu A}{5pF} \quad (5.78)$$

$$= 120V/\mu s \quad (5.79)$$

where I_o is the tail current of $600\mu A$.

It was found that the rising and falling slew rates are unequal. By measuring the slope of fig. 5.13 or by calculation directly from the simulator output, it can be determined that the rising output slew rate is $100\text{V}/\mu\text{s}$ and the falling slew rate is $117\text{V}/\mu\text{s}$. In the case of the falling output, the entire drain current of M7 will be taken by M1 and thus the current mirror formed by Q4, Q5, Q6 and Q7 will be turned off. The entire drain current of M8 will flow through M4 and through the compensation capacitor, with a small amount flowing into the base of the emitter follower transistor Q8. Thus equation 5.77 applies and the simulated falling slew rate of $117\text{V}/\mu\text{s}$ compares well with $120\text{V}/\mu\text{s}$ predicted by equation 5.77. In the case of the rising output, one would expect the opposite to occur with the current mirror active and a current equal to the bias current of $600\mu\text{A}$ flowing through the compensation capacitor and into the collector of Q5. However in order to allow the output voltage to rise, the voltage applied to the output stage (at the base of Q8,) must be reduced, thus partially turning off the output transistor Q10. The reduction in the base voltage of Q8 will reduce the collector - emitter voltage of Q5, possibly forcing it into saturation. Once this occurs, the collector current of Q5 will decrease, thus resulting in a lower slew rate for rising output. The simulated rising slew rate of $100\text{V}/\mu\text{s}$ is somewhat less than either the falling slew rate or the slew rate predicted by equation 5.77, which is consistent with the preceding discussion.

The simulated transient response of the opamp in closed loop (voltage follower) mode is shown in fig. 5.14. The settling time and degree of overshoot of the opamp can be determined from the results of this simulation. The settling time of an opamp is usually defined as the time between when a voltage step being applied to the input of the opamp and when the output ringing is confined to within $\pm 5\%$ or $\pm 1\%$ of transition about the final value. Knowledge of the settling time of an opamp is useful in the design of a switched capacitor circuit as it is directly related to the maximum frequency at which the filter may be clocked. The amount of overshoot at the end of a transition is usually specified as a percentage of the step size, ie.

$$\text{Percent Overshoot} = \frac{\text{Maximum Value} - \text{Step Size}}{\text{Step Size}} \quad (5.80)$$

The values of settling time and percent overshoot calculated from the simulator output

are

	Settling Time		Percent
	5%	1%	Overshoot
Fall	11.3 ns	12.4 ns	0.18 %
Rise	12.3 ns	13.3 ns	7.8 %

Noise

Figure 5.15: NMOS Input Opamp Input Referred Noise

In any analog system, noise is of concern and thus it was decided to simulate the noise behavior of this opamp. The simulated input referred noise versus frequency of the NMOS input opamp is presented in fig. 5.15. The input referred noise is presented against a linear $V/\sqrt{\text{Hz}}$ scale in the top half of fig. 5.15 and against a logarithmic

$V/\sqrt{\text{Hz}}$ scale in the bottom half of fig. 5.15. It can be seen from the figure that the significant noise at low frequencies is $1/f$ noise, most likely primarily from the NMOS input devices. This large $1/f$ noise is one of the negative consequences of choosing MOS devices for the input. At higher frequencies, the logarithmic plot shows the effect of the opamp gain on an approximately constant level of noise generated within the circuit.

5.4 Test and Measurement

In order to fully verify the opamp designed above, it is necessary to test actual silicon samples. A number of test structures using this opamp have been designed and submitted for fabrication. Each of the structures will be described, along with proposed tests, in the following paragraphs.

5.4.1 Test Circuits for BiCMOS Operational Amplifier with NMOS Differential Pair

Due to silicon area constraints, it was only possible to provide a small number of test structures for the opamp. Three structures were designed and submitted for fabrication. These were open loop as shown in fig 5.16, closed loop as shown in fig 5.18, and an emitter follower structure as shown in fig 5.19. Plots of the actual layouts may be found in Appendix B. The intent and features of each of these test structures is described below.

Common Features

All input and output pins of the opamp test structures include simple ESD protection circuitry in order to minimize the risk of chip damage during handling. The protection structures add approximately 50 ohms of series resistance to the opamp inverting and non-inverting inputs and approximately 1000 ohms of series resistance to the bias inputs of the opamp. The protection network used on output pins does

Figure 5.16: Open Loop Test Structure for BiCMOS Op Amp with NMOS Differential Pair

not add any series resistance, due to the desire to avoid distorting any measurements and to the fact that large output devices are generally less susceptible to ESD damage than input devices, particularly MOS gates.

Both opamp structures feature an on chip load capacitor and an emitter follower buffering the output. This was done both to ensure that a predictable load was presented to the opamp under test and to isolate the opamp from the relatively large loads presented by most test equipment. The load capacitor is identical to the compensation capacitor in the opamp both in terms of value and geometry, and both are arranged such that their value may be reduced or increased by 50% with microsurgery, should the original value prove to be unsuitable during testing. The physical arrangement of the compensation capacitor can be seen in fig. 5.5, the plot of the opamp above. The emitter follower is intended to be biased using an off chip resistor, thus simplifying layout and allowing the bias conditions to be easily varied during testing. In addition to the open and closed loop opamp test structures, a test structure containing two emitter followers has also been submitted. This should allow the emitter followers to be characterized and allow the open loop opamp to be operated closed loop, using off chip components and one of the test emitter followers to compensate for the voltage shift introduced by the emitter follower on the output of the opamp.

As submitted for fabrication, only the input, output and bias points labeled in figs 5.16 and 5.18 and the power supplies are connected to bonding pads. The layout has been arranged such that the node at the output of the opamp can be brought to an unused bonding pad reserved for this purpose with microsurgery. It was decided to not connect this pad due to the relatively large capacitance it would add to the opamp output. Similarly the node common to the output and the inverting input of the opamp in the closed loop test structure may also be brought to a bonding pad with microsurgery. Each of the three test structures and tests planned for them will now be described.

5.4.2 Open Loop Test Structure

Figure 5.17: Low Frequency Test Structure

A schematic of the open loop test structure is shown in fig. 5.16. A computer plot of the actual layout is shown in Appendix B, fig. B.3. This structure can be used to measure the low frequency parameters of the opamp, specifically DC gain and offset.

It should also be possible to realize a filter or other closed loop circuit using suitable external components.

Due to the high gain of this opamp, it is not likely to be possible to operate it truly open loop as even a minute differential voltage between the two inputs would be sufficient to cause the output to swing to one of the power supply rails. It is then necessary to connect an external feedback network to the opamp such as is shown in fig. 5.17. Those circuit elements enclosed by the dotted lines in fig. 5.17 are on the chip. Those outside the box are external and must be provided on the test board. The second emitter follower is obtained from the emitter follower test structure described later, and is used to cancel the voltage shift due to the emitter follower on the output of the opamp. The external opamp can either be operated open loop, or with its own feedback network to give a specific DC gain, A_2 and offset, V_{off2} . The operation of the low frequency test circuit shown in fig. 5.17 can be described as follows. The voltage at the output of the test opamp is

$$V_{O1} = A_1(V_{+1} - V_{-1} + V_{off1}) \quad (5.81)$$

where V_{off1} is the input referred offset voltage of the test opamp and A_1 is the DC gain of the test opamp. The voltage $V_{+1} = 0$ as the non-inverting input of the test opamp is grounded and V_{-1} is determined by the voltage divider formed by R_1 and R_2 , eg;

$$V_{-1} = \frac{R_1}{R_1 + R_2} \cdot V_O \quad (5.82)$$

giving

$$V_{O1} = A_1(V_{off1} - \frac{R_1 \cdot V_O}{R_1 + R_2}) \quad (5.83)$$

Now considering the external opamp

$$V_O = A_2(V_{+2} - V_{-2} + V_{off2}) \quad (5.84)$$

where V_{off2} is the input referred offset voltage of the external opamp and A_2 is the DC gain of the external opamp. From fig. 5.17, it can be seen that

$$V_{-2} = V_{IN} - V_{EF} \quad (5.85)$$

and

$$V_{+2} = V_{O1} - V_{EF} \quad (5.86)$$

where V_{EF} is the voltage drop across each emitter follower. Equation 5.84 may now be stated as

$$V_O = A_2(V_{O1} - V_{EF} - (V_{IN} - V_{EF}) + V_{off2}) \quad (5.87)$$

which reduces to

$$V_O = A_2(V_{O1} - V_{IN} + V_{off2}) \quad (5.88)$$

thus showing that any effects of the emitter followers should to a first order be cancelled by the external opamp, and can be ignored. Since the second opamp is external, its characteristics are known, or can be measured and thus A_2 is known and if the opamp has offset nulling terminals, V_{off2} can be set to zero. Similarly the actual values for R_1 and R_2 can be measured and are then known. The quantities that are easily measured from the test circuit of fig. 5.17 are the DC output voltage with the input grounded (V_O), and the gain of the circuit when a small amplitude low frequency signal is applied to the input (V_O/V_{IN}). Equations 5.83 and 5.88 can then be manipulated to give the DC gain of the test opamp, which is

$$A_1 = \frac{[(V_O/V_{IN}) + A_2](R_1 + R_2)}{(V_O/V_{IN})A_2R_1} \quad (5.89)$$

and the input referred offset of the test opamp which is

$$V_{off1} = \frac{A_2R_1V_O}{[(V_O/V_{IN}) + A_2](R_1 + R_2)} \quad (5.90)$$

All of the terms in equations 5.89 and 5.90 are known or easily measured, thus the DC gain A_1 and the input referred offset voltage V_{off1} can be determined.

5.4.3 Closed Loop Test Structure

A schematic of the closed loop test structure is shown in fig. 5.18. A computer plot of the actual layout is shown in Appendix B, fig. B.2. This structure will be

Figure 5.18: Closed Loop Test Structure for BiCMOS Op Amp with NMOS Differential Pair

used to measure the unity gain bandwidth, settling time and common mode range. Isolating the on chip load capacitance from off chip loads using an emitter follower will make bandwidth measurements more quantifiable, as previously explained. The bandwidth of the opamp can be determined by connecting the input and output to a network analyzer and observing the frequency at which the closed loop gain is reduced by 3 dB. This assumes that the emitter follower is not too heavily capacitively loaded, ie; has a relatively high frequency pole. The settling time of the opamp may be determined by applying a square wave to the input and observing the settling behavior on an oscilloscope. The common mode range can be measured by applying a small amplitude signal to the input of the test structure and varying the DC bias of the input signal until distortion can be observed at the output on an oscilloscope.

5.4.4 Emitter Follower Test Structure

The emitter follower test structure, the schematic of which is shown in fig. 5.19, is intended to allow the emitter followers to be characterized and to be used to cancel the DC voltage shift due to the output emitter follower in the other test structures. A computer plot of the actual layout is shown in Appendix B, fig. B.4. The test structure consists of an emitter follower with the base protected by a 50 ohm protection network similar to those used for the opamp inputs and an emitter follower with a non-resistive

Figure 5.19: Emitter Follower Test Structure

input protection network. The measurement most of interest would be the frequency response, as interpretation of results from the other test structures could be affected if the frequency response of the emitter followers is not flat to well beyond the unity gain bandwidth of the opamp. The frequency response of the emitter followers will most likely need to be measured with a network analyzer due to their very high bandwidth.

5.4.5 Measurements

At the time of writing, the test circuits described above are still in the process of fabrication. Thus it is not possible to include any actual measurements in this thesis. It is planned to perform these measurements later. After initial validation, it is expected that the circuits will be used as the basis for an undergraduate student project co-supervised by the author.

5.5 Additional Topologies

In the course of the research for this thesis, a number of variations on the proposed BiCMOS operational amplifier topology shown in fig. 5.2 were considered, before it was concluded that a simple MOS differential pair was the appropriate choice of input stage for the opamp to be used in sampled data applications. The circuits to be presented in this section will all employ some sort of bipolar differential pair

and will otherwise have the same topology as fig. 5.2. Most of these opamps were designed and submitted for fabrication but were unstable due to the unavailability of fabricated parts or because the process had shifted significantly between when the circuit was designed and when the samples were fabricated. This is because the thesis work began relatively early in the development of the process. However the layouts performed were valuable in establishing expected parasitic capacitances and in giving specific parameter estimates that allowed the final design choice to be made. There were a number of ideas that still may be worthy of further investigation later, and thus these circuits will be presented and discussed below as possibly of benefit to later investigators.

5.5.1 A BiCMOS Operational Amplifier with NPN Differential Pair and Input Current Cancellation

Figure 5.20: BiCMOS Operational Amplifier with Bipolar Inputs and Input Current Cancellation Circuitry

One of the drawbacks of the bipolar transistor when used as an input transistor is the requirement for a DC base current. This is particularly a concern in switched capacitor applications where the DC input current would partially discharge the sampling capacitor during a clock period. One method of minimizing this problem is to use a suitable network to supply a current to the input terminals that is as close as possible to the expected DC input current. The network employed in this operational amplifier will be described below. The circuit is essentially that of the proposed topology shown in fig. 5.4 with the NMOS differential pair replaced by an NPN differential pair and with the addition of the DC input current cancellation circuitry and a simple biasing network. This op amp is shown in fig. 5.20.

Circuit Operation

The operation of this opamp is similar to that of the NMOS input opamp described in Section 5.3, with the exception of effects due to the input characteristics of the bipolar transistors of the differential pair. The approximate relations for bandwidth and DC gain presented in Section 5.3 are equally applicable here. When numerical values were substituted into these equations, it was found that both the gain and bandwidth can be greatly increased over that of the NMOS input opamp due to the higher transconductance of the NPN transistors. However the issues of bandwidth and stability limitation due to non-dominant poles then become very critical.

The base current cancellation circuitry should have minimal impact on the performance of the opamp, other than adding to the input referred noise, and will now be described.

Operation of the Input Current Cancellation Circuitry

The function of the input cancellation circuit is to supply a current to the base terminals of the input transistors that is as close as possible to that which they will require in operation, so as to prevent DC current draw from switched capacitors. This is done by generating a reference current proportional to the tail current used to bias the differential pair, then mirroring and scaling this current to the required

level. This assumes that the β of the NPN devices is relatively constant.

The input current cancellation circuitry can be considered to be composed of three sections. The differential pair tail current is replicated by Q12 which is connected to the same current mirror as Q3, the transistor biasing the differential pair. Thus the collector current of Q12 will be approximately the collector current of Q3 multiplied by the emitter area ratio between Q12 and Q3. Transistor Q12 was made minimum size in order to minimize the current drawn by and hence the power dissipated in the cancellation circuitry. The sample of the tail current must be divided by the β of the NPN transistors and then scaled to account for (i) the size ratio between Q12 and Q3, and (ii) the fact that the collector current of Q3 is split between the two input transistors. This division by β is done by transistor Q13, while the necessary scaling is accomplished by selecting appropriate (W/L) ratios for transistors M7, M8 and M9, which form a current mirror that attempts to supply the base current to the input transistors Q1 and Q2. Transistor M10 is used to disable the input current cancellation circuitry to allow the actual input currents to be determined during testing. This is done by pulling the gates of M8 and M9 to Vdd, thus shutting them off.

5.5.2 A BiCMOS Operational Amplifier with Darlington Differential Pair and Emitter Degeneration

The employment of a technique to reduce the DC base current drawn by its inputs was presented in section 5.5.1 above; however this approach does not improve the low input impedance of the bipolar differential pair. One approach that both increases the input impedance and reduces the DC base current requirement is the use of a Darlington connected input pair. An example of this approach is the operational amplifier shown in fig 5.21, which was designed following the approach of this thesis for use in a high speed oversampled analog to digital converter experiment being performed by another student [2]. In order to allow a rapid slew rate capability, it is necessary to operate the differential pair at a high current and thus the transconductance of the differential pair will be high. In order to avoid excessive bandwidth (and the stability

Figure 5.21: BiCMOS Operational Amplifier with Darlington Inputs and Emitter Degeneration

difficulties that would thus arise) due to the high transconductance of the bipolar transistors, it was chosen to use emitter degeneration to limit the transconductance of the differential pair while maintaining a high current level suitable for the desired slew rate.

Circuit Description

In bipolar design, the use of the Darlington configuration is a traditional method of increasing the effective β of an input transistor, increasing the input impedance and reducing the magnitude of the DC base current required by that input. A discussion of the Darlington differential pair and a comparison of it to the simple bipolar differential pair have been presented previously in Chapter 4. The input impedance of a Darlington differential pair as is used here was presented as equation 4.47 and is

$$R_{id} = 2\beta_o \left[\frac{1}{g_{mo}} + \beta_i \left(\frac{1}{g_{mi}} + R_E \right) \right] \quad (5.91)$$

3

OF/DE

3

PM-1 3½"x4" PHOTOGRAPHIC MICRO TARGET
NBS 1010a ANSI/ISO #2 EQUIVALENT

PRECISION[™] RESOLUTION TARGETS

including the effect of emitter degeneration. It was discussed previously that the transconductance of a bipolar transistor cannot be adjusted without changing the current through the device. In the application for which this particular opamp has been designed, it was desirable to have a fast slewing behavior while maintaining adequate stability. The rapid slew rate requirement necessitates that the differential pair operate at a high current. However increasing the current in the differential pair will increase the transconductance which will increase the gain and hence, possibly reduce the stability of the opamp. A solution is to use emitter degeneration, as has been done here.

The approximate DC gain and bandwidth of the opamp shown in fig. 5.21 can be calculated using the relations presented in Section 5.3 if the transconductance of the Darlington differential pair is used instead of the transconductance of the NMOS transistors forming the differential pair in fig. 5.4. From analysis presented in Chapter 4, it can be determined that the transconductance of a Darlington differential pair with emitter degeneration is

$$g_{m,dp} \approx \frac{g_{mi}}{1 + g_{mi}R_E + \frac{g_{mi}}{\beta g_{mo}}} \quad (5.92)$$

where g_{mi} and g_{mo} are the transconductances of the inner and outer transistors respectively of the differential pair. Depending on the exact bias levels of the inner and outer transistors and on the value of the emitter degeneration resistor R_E , the transconductance of the differential pair $g_{m,dp}$ will be reduced to between $\frac{1}{3}$ and $\frac{1}{4}$ of g_{mi} . Thus the gain and bandwidth of an opamp based on the topology shown in fig. 5.21 are roughly one quarter to one third that available from the topology shown in fig. 5.20.

5.5.3 BiCMOS Operational Amplifiers with NPN Differential Pair and PMOS Source Follower Inputs

As was stated previously, one of the objects of these experiments is to investigate the suitability of using the high available gain of bipolar transistors in monolithic analog systems. Many such systems involve switched capacitor circuits, which are

Figure 5.22: BiCMOS Operational Amplifier with Bipolar Differential Pair and PMOS Source Follower Inputs

particularly sensitive to the input impedance of the operational amplifiers used. The DC base current drawn by the opamp inputs can be reduced but not eliminated with a suitable mirror network as was done in fig. 5.20. The input impedance of a bipolar input opamp can be increased by the use of emitter degeneration and a Darlington input as was done in fig. 5.21. Unfortunately these approaches while improving the situation significantly, fall short of the effectively infinite impedance that a MOS input opamp presents to the external circuit. Thus an opamp using MOS input devices is still a very attractive choice for switched capacitor applications. The drawback of this approach is the relatively low transconductance available from a reasonably sized MOS device operating at a reasonable current. Typically, the current at which the input stage operates is restricted by slew rate and power consumption requirements. If a large transconductance is obtained by increasing the area of a MOS transistor, the parasitic capacitances become large, possibly causing the frequency of the non dominant pole at the source of the cascode transistors (M3 and M4 in

fig. 5.22 to decrease until they begin to affect stability. One alternative that should offer both the infinite input impedance of the MOS differential pair and most of the transconductance of the bipolar differential pair is to use a bipolar differential pair with MOS source followers driving the NPN base terminals. The MOS source followers can operate at lower currents than the input stage and thus can be smaller than a MOS differential pair would need to be, and the parasitic capacitances are thus kept reasonably small.

An operational amplifier using such an input stage with PMOS transistors is shown in fig. 5.22. It is possible to utilize either NMOS or PMOS source followers. The main

Figure 5.23: BiCMOS Operational Amplifier with Source Follower Inputs and Emitter Degeneration

advantage of using NMOS source followers is that they could be smaller for a given source follower gain due to their larger transconductance parameter. Unfortunately, the gate-source voltage of an NMOS transistor will be larger than the base-emitter voltage of an NPN transistor and thus the common mode range of a differential pair driven by NMOS source followers would be less than that of any other differential

pair, including the Darlington differential pair. If a PMOS source follower is used, the gate-source voltage will be no smaller, however it will be in the opposite direction. Provided that the voltage at the base of the NPN differential pair transistors is not raised high enough to cause them to enter saturation, the gate-source voltage of the PMOS source follower transistors will act opposite to the base-emitter voltage of the NPN transistors of the differential pair and the common mode range should be acceptable. Another reason to choose PMOS source followers is the lower levels of flicker noise exhibited by PMOS transistors as compared to NMOS devices. The primary drawback of using source followers on the input is that they have gains of less than unity. However with reasonably sized devices, the reduction in overall input stage gain should not be substantially worse than would occur with a Darlington input stage. In the case of a switched capacitor circuit, the reduction in the gain and bandwidth of the operational amplifier by the use of the source follower inputs may well be offset by the improvement in the system performance due to the much higher input impedance.

A minor variation on the opamp shown in fig. 5.22 is shown in fig. 5.23. This amplifier was designed for use in an oversampled analog to digital converter experiment performed by another student [2] and is the circuit of fig. 5.22 with the addition of emitter degeneration. As was discussed in Section 5.5.2, the use of emitter degeneration gives additional control over the transconductance of the input stage while keeping the tail current high to ensure an adequate slew rate.

5.6 Summary

In this chapter, two possible topologies for a BiCMOS operational amplifier were proposed and discussed. Following this, the topology shown in fig. 5.2 was chosen and a detailed design was presented. Simulation results were presented and compared to performance predicted by simple square law equations developed in Chapter 4. Test structures using the opamp presented in fig. 5.4 were presented and have been submitted for fabrication. Planned measurements to be performed when fabricated parts

become available were then presented. A number of alternative BiCMOS operational amplifier topologies that were investigated early in the research for this thesis were presented and briefly discussed. Conclusions to this research and recommendations for future investigation will now be presented in Chapter 6.

Chapter 6

Conclusions and Future Work

6.1 Summary and Conclusions

The objective of this thesis was to explore the use of BiCMOS technology in the realization of improved operational amplifiers for use in monolithic analog signal processing. The main focus was the maximization of bandwidth while still achieving high DC gain and compatibility with switched capacitor circuits.

In Chapter 1, the motivation for this thesis was presented. Traditionally, bipolar and CMOS integrated circuit technologies have evolved and been employed separately. Bipolar analog integrated circuits have typically been subsystems, for example operational amplifiers, that are connected together on a printed circuit board to form complete analog systems. CMOS analog integrated circuits have typically been complete signal processing systems on a chip, often in combination with a large amount of digital circuitry to form a mixed analog/digital chip such as a codec (COder - DECoder). The availability of BiCMOS integrated circuit process combining both bipolar and MOS transistors on the same chip offers opportunities to exploit the high speed of bipolar transistors while retaining the ability to combine analog systems and low power CMOS digital circuitry on a single integrated circuit.

The primary building block for monolithic analog system design is the operational amplifier. It is thus reasonable to investigate the opportunities to increase operational amplifier performance over available CMOS opamps by the use of bipolar transistors

where advantageous.

In Chapter 2, the characteristics, design equations and small signal models for MOS field effect transistors and bipolar junction transistors were presented and compared. The comparison included transconductance, current handling ability, output resistance, noise, input impedance, silicon area required and the availability and relative performance of complementary devices. The process flow of typical MOS and bipolar integrated circuits were discussed. Similarities and differences between the two types of process were identified and possible approaches to a BiCMOS process were discussed. A detailed description of a hypothetical 1 micron analog/digital BiCMOS process was presented.

In the comparison between bipolar and MOS transistors, several advantages of bipolar transistors were noted for analog applications. It was noted that the transconductance and output resistance of bipolar transistors is greater than that available from MOS transistors of reasonable size operating under similar conditions. The capacitances associated with bipolar transistors are considerably smaller than those associated with MOS transistors performing a similar function on account of the smaller size of the bipolar transistor. Thus bipolar transistors offer greater gain and potentially greater bandwidth than MOS transistors. It was shown that bipolar transistors have superior noise performance due to their relative lack of flicker ($1/f$) noise and their lower levels of thermal noise.

Several drawbacks of bipolar transistors when compared to MOS transistors were noted, particularly in the context of switched capacitor circuits. The input impedance of bipolar transistors is low, typically on the order of a few thousand Ohms. A MOS gate is only capacitively coupled to the rest of the device, thus the input impedance of a MOS transistor is effectively infinite at low and moderate frequencies. It was noted that the transconductance of a bipolar transistor cannot be easily varied independently of the current level, except by the use of emitter degeneration resistors. The transconductance of a MOS transistor can be varied at a given current level by varying the W/L ratio. It is necessary to operate an opamp input stage at a high current level in order to achieve rapid slew rates. This may lead to the transconductance

of a non-degenerated NPN differential pair being so high as to cause the bandwidth of the opamp to be wide enough that parasitic poles intrude and reduce the stability of the opamp.

In many BiCMOS technologies, PNP transistors are not available due to cost constraints or process control difficulties. The lack of a complementary bipolar device means that circuits must be designed without the use of complementary devices, or the lower performance PMOS transistor must be used, possibly negating some of the advantages of the bipolar transistor because of the much greater stray capacitances introduced.

In Chapter 3, general information about operational amplifiers was presented, including a definition of an operational amplifier. Non-idealities of real operational amplifiers were discussed, including the effects on each of bipolar and MOS opamps. The differences in end application between typical bipolar and MOS operational amplifiers and the resulting design differences between the two types of opamps were discussed.

As bipolar operational amplifiers have typically been used in non-monolithic systems, they have output stages that are capable of driving large off chip capacitive and resistive loads, and are typically not integrated with other components and subsystems on chip. CMOS operational amplifiers have typically been employed in creating complete analog systems on an integrated circuit, and thus need an output stage that is designed to drive only the small capacitive loads that are found on chip, and not resistive loads. An important advantage of CMOS opamps is their effectively infinite input impedance, which allows the use of switched capacitor techniques in the implementation of analog systems.

The topologies and operation of the CMOS two stage (also known as the pole-splitter) and folded cascode operational amplifiers were discussed with emphasis on the design details of the CMOS pole-splitter opamp. The topology and operation of a practical bipolar operational amplifier was discussed, along with techniques to reduce the DC bias current that must be supplied to the input terminals of a bipolar differential pair.

In Chapter 4, MOS, bipolar and where appropriate, BiCMOS implementations of several analog "building block" circuits were presented and analyzed using simple square law relations. Current mirror circuits were analyzed and compared with emphasis on their effect on the common mode performance and DC gain of an opamp when used as the current source for the differential pair and when used as an active load for a gain stage. It was concluded that bipolar transistors are preferred over MOS transistors in both cases due to their higher output resistance, except when matching between the reference and output currents is particularly important. In order to avoid limiting the common mode range of an opamp, it was decided that the current source for a differential pair should be a single transistor as opposed to a cascode. For active loads followed by a gain stage, it is preferable to use cascode loads as the small signal resistance is considerably larger and the voltage swing will be small due to the subsequent gain stage.

Several differential input circuits were presented and compared. It was shown that both the DC gain and bandwidth of a complete opamp are proportional to the transconductance of the input differential pair. It was shown that the common mode small signal gain of a differential pair is dependent upon the output impedance of the current source used to bias the differential pair and the small signal resistance that loads the differential pair. Hence if the resistive load and bias current source are the same for all differential pairs under consideration, the common mode rejection ratio will also be proportional to the transconductance of the differential pair. Thus differential pairs having high transconductance will also have high bandwidth, DC gain and common mode rejection. Also compared between the differential pairs were the input impedance calculated from small signal equivalent circuits and the common mode range, which was considered to be reached when MOS devices began to leave saturation or bipolar devices began to enter deep saturation. It was determined that the NPN differential pair offered the highest transconductance and hence bandwidth. However the need to supply a DC base current and the low input impedance of bipolar transistors renders them unsuitable as an input stage for an opamp intended for use in switched capacitor applications. More complicated differential pairs such

as the Darlington and the NPN differential pair with PMOS source follower inputs were considered but rejected due to their limited common mode range. It was then concluded that a MOS differential pair would be used.

Output stage options consisting of the common source amplifier, the common emitter amplifier, and a folded cascode topology having no output stage were considered. It was determined that an output gain stage is desirable due to the large increase in DC gain provided and that an NPN common emitter configuration should be used due to the higher transconductance and smaller parasitic capacitances of bipolar transistors. As the base current and low input impedance of a common emitter stage would load down the previous stage, it was decided that an emitter follower should be used to drive the common emitter stage.

In Chapter 5, the findings of Chapter 4 were further discussed and two possible topologies for a BiCMOS operational amplifier were proposed and discussed. These topologies were a folded cascode topology with a NMOS differential pair and a pole-splitter output stage, and a flipped two stage topology with a PMOS differential pair. The NMOS input topology was determined to hold the most promise for high bandwidth and a detailed design was pursued. The design process of an opamp based on this proposed topology was then described. Simulation results were presented and compared to performance predicted by simple square law equations. The simulation results, for the most part, agree with performance predicted by theory. Test structures and the measurements to be performed when parts are available were presented. Alternative BiCMOS operational amplifier topologies that were investigated early in the research for this thesis were presented and briefly discussed.

6.2 Recommendations

The test chip intended to study the operational amplifier presented in fig. 5.4 is still in fabrication. It should be studied once available, both as a final verification of the design and to verify the accuracy of the simulation approach. Following this, it is appropriate to consider possibilities of extending BiCMOS operational amplifier

bandwidth further.

Further study of high bandwidth operational amplifier topologies should concentrate on the combination of alternative topologies that are less sensitive to parasitics and on the reduction of load capacitances. Simply increasing the transconductance of the input pair is not likely to be fruitful as device sizes would need to be increased, either to provide the increased transconductance or to handle the higher currents that would be required. Larger devices have larger associated capacitances which will push non-dominant poles to even lower frequencies, and thus adequate stability would be difficult to obtain. Similarly, reducing the load capacitance with no other changes will increase the bandwidth of the opamp without changing the location of non-dominant poles, resulting in questionable stability.

The non-dominant poles of concern in the opamp of fig. 5.4 are those due to capacitive loading at the low impedance nodes, particularly those common to the PMOS current source and the PMOS cascode transistors. Unless new devices such as complementary PNP transistors become available, it may be difficult to substantially reduce the capacitances at these nodes. Thus it is worthwhile to investigate opportunities to further reduce the impedance at the low impedance nodes, thus causing the non-dominant poles to be at higher frequencies. It has been suggested that flipping the folded cascode structure and using NPN transistors as the cascode devices may offer some improvement due to the very low impedance seen looking into the emitter of a bipolar transistor [15].

Another approach worthy of investigation is to simplify the opamp topology, thus minimizing the number of possible parasitic poles. Use of an NPN differential pair loaded with linear resistors and an emitter follower output may hold promise as the elimination of MOS devices decreases the capacitances considerably. Thus much higher bandwidth capabilities are available, at the expense of lower DC gain and of low input impedance. Such circuits are already well known in the context of classic RF analog signal processing. It may be, for higher bandwidth, that it is better to look for contexts where low DC gain and low input impedance are acceptable, eg, continuous time transconductance filtering, rather than to try to extend opamps drawing on

CMOS topologies to much higher frequencies.

It is suggested that any further high bandwidth opamp designs be fully differential, except perhaps those at the output of a monolithic system that is intended to drive off chip. The use of fully differential opamp topologies will allow emphasis to be placed solely on bandwidth, as any common mode input not fully rejected by one opamp should be rejected by the next opamp in the system. This approach coupled with NPN differential pairs and linear resistors, in a continuous time configuration, would be the way to approach much higher bandwidth on-chip filtering. It is assumed that in circuit tuning would then be useful to deal with the non-idealities implied in such a simple amplifier context.

Appendix A

SPICE Models for Hypothetical BiCMOS Process

The following are the actual spice parameters from the hypothetical BiCMOS process that was used for most of the early design in this work [9].

* 1 u BiCMOS Device Models: for course projects 97.589W 31 Jan 1989

** Bipolar Models:

* Junction capacitance calculations @ 0 bias (for bipolars and resistors):

* $C_{je} = 1A + 0.7P$ * $C_{jc} = 0.3A + 0.3P$ $A = \text{area (u **2)}$ * $C_{js} = .15A$

* $P = \text{non-walled periphery (u)}$

* $C_{sch} = 0.4A$

** Min NPN 1x3 u emitter, single base contact; Tub: 5x7u

.MODEL N1X3 NPN IS=3E-18 RB=300 RC=300 RE=10 EG=1.11 XTI=3 XTB=1

+ BF=100 VAF=50 IKF=1E-3 TF=15P NF=1 ISE=4E-17 NE=1.5

+ BR=4 VAR=10 IKR=1E-4 TR=1N NR=1

+ CJE=8.6FF VJE=.9 MJE=.45 FC=.5

+ CJC=5.4FF VJC=.6 MJC=.4 XCJC=.2

+ CJS=5.25FF VJS=.5 MJS=.33

+ XTF=10 VTF=1 ITF=4E-3

** Min Lat PNP: 1.5u dia emitter, WV=1.5u Tub=7.5x10u **

.MODEL P1 PNP IS=2E-17 RB=1000 RC=200 RE=100 EG=1.11 XTI=3 XTB=1

+ BF=9 VAF=25 IKF=30E-6 NF=1 ISE=9E-16 NE=1.5

+ BR=1 VAR=10 IKR=40E-6 NR=1

+ CJE=4FF VJE=.6 MJE=.4 FC=.5

+ CJC=20FF VJC=.6 MJC=.4 XCJC=.1

+ TF=3N TR=10N

** Min Lat PNP tub: (substrate diode: connect anode to -ve supply,
* cathode to PNP base) Tub size=7.5x10u

.MODEL PTUB D CJO=15FF VJ=.5 M=.33

** 4x5u Schottky, Current density 2uA/u**2 max for NPN clamp

.MODEL SD20 D IS=5E-13 N=1 RS=150 TT=1P EG=.8

+ CJO=9FF VJ=.6 M=.5 XTI=2

* 4x5u Schottky tub capacitance: (substrate diode: connect anode to
* -ve supply, cathode to Schottky cathode) Tub size=7x8.5u

.MODEL SDTUB D CJO=9FF VJ=.5 M=.33

** MOS models:

** Est NMOS model for 1u process * RAH version 1, Jan 12 89 **

.MODEL NM1 NMOS LEVEL=3 CGS0=0.17E-9 CGD0=0.17E-9 CGB0=0.07E-9

+ TOX=20E-9 NSUB=2E16 NFS=0.7E12 U0=575 VMAX=1E5 CBD=1.8E-15

+ CBS=1.8E-15 CJ=0.3E-3 CJSW=4E-10 XJ=2E-7 LD=1E-7

+ THETA=0.1 ETA=0.07 KAPPA=0.1 VTO=0.7

** Est PMOS model for 1u process * RAH version 1, Jan 12 89 **

.MODEL PM1 PMOS LEVEL=3 CGS0=0.17E-9 CGD0=0.17E-9 CGB0=0.07E-9

+ TOX=20E-9 NSUB=2.5E16 NFS=1.5E12 U0=190 VMAX=0.7E5 CBD=3.0E-15

+ CBS=3.0E-15 CJ=0.5E-3 CJSW=4.5E-10 XJ=2.0E-7 LD=0.0E-7

+ THETA=0.15 ETA=0.07 KAPPA=1.0 VTO=-0.7

Appendix B

Layout of Opamp and Test Structures

The following pages contain plots of the opamp and test structures generated by the Cadence layout tool. Photomicrographs of the actual circuits are not available as the circuits are still in fabrication. The circuits in the following figures are described in Section 5.4.

Figure B.1: NMOS Input Operational Amplifier

Figure B.2: Closed Loop Test Structure

Figure B.3: Open Loop Test Structure

Figure B.4: Emitter Follower Test Structure

References

- [1] Asad A. Abidi. *On the Operation of Cascode Gain Stages*. IEEE J. Solid State Circuits, vol 23 pp. 1434-1437.
- [2] Bill Bereza. *Dynamic Range and Bandwidth Limitations in Sampled-data and Clocked Continuous-time BiCMOS Sigma-Delta Modulators*. Ph.D. proposal, Carleton University, 1991
- [3] Klaas Bult, Govert J. G. M. Geelen *A Fast-Settling CMOS Op Amp for SC Circuits with 90-dB DC Gain* IEEE J. Solid State Circuits, vol 25 pp. 1379-1384.
- [4] Ruud G. G. Eschauzier et. al. *A 100 MHz 100 dB Operational Amplifier with Multipath Nested Miller Compensation Structure* 1992 IEEE International Solid State Circuits Conference, pp. 196-197
- [5] Paul R. Gray, Robert G. Meyer. *Analysis and Design of Analog Integrated Circuits, Second Edition*. John Wiley & Sons, New York, 1984 pp. 625-647.
- [6] Paul R. Gray, Robert G. Meyer. *MOS Operational Amplifier Design - A Tutorial Overview*. IEEE J. Solid State Circuits, vol. SC-17 pp. 969-982.
- [7] Alan B. Grebene. *Bipolar and MOS Analog Integrated Circuit Design*. John Wiley & Sons, New York, 1984 pp. 66-69, 113-115, 180-182, 217-223.
- [8] R. Gregorian, G. Temes. *Analog MOS Integrated Circuits for Signal Processing*. John Wiley & Sons, New York, 1986 pp. 96-99, 153-155, 218-219, 222-227, 251-252.

- [9] R. Hadaway, R. Brown, M. A. Copeland. *Course Notes from BiCMOS Technology and Circuits Course*. Department of Electronics, Carleton University, 1989.
- [10] R. Hadaway et al. *A Sub-Micron BiCMOS Technology for Telecommunications*. European Solid State Device Research Conference 1991. pp. 513-516.
- [11] H. Klose. *BiCMOS - The Technology for Integrating Systems onto One Silicon IC*. European Solid State Device Research Conference 1991. pp. 501-512.
- [12] Joel Lachance and Francois Duval. *Subthreshold Models for CMOS Analog Devices*. Analogue in the VLSI world: 1990 CMC Workshop.
- [13] Hae-Sung Lee. *1987 IEDM Short Course on BiCMOS Analog Design Techniques*. Department of Electrical Engineering and Computer Science Massachusetts Institute of Technology.
- [14] Seiki Ogura, et al. *Design and Characteristics of the Lightly Doped Drain-Source (LDD) Insulated Gate Field-Effect Transistor*. IEEE Transactions on Electron Devices, Aug. 1980, pp. 1359-1367.
- [15] Frank Singer. Conversation with author. January 1993.
- [16] David B. Ribner and Miles A. Copeland. *Design Techniques for Cascoded CMOS Op Amps with Improved PSRR and Common-Mode Input Range*. IEEE J. Solid State Circuits, vol. SC-19 pp.919-925.
- [17] Adel S. Sedra, Kenneth C. Smith. *Microelectronic Circuits*. Holt Rinehart and Winston, New York, 1982 p. 448.
- [18] James E. Solomon. *The Monolithic Op Amp: A Tutorial Study*. IEEE J. Solid State Circuits, vol. SC-9 pp. 314-332.

END

10-11-94

FIN