
**AN INTRODUCTION
TO
LOW-VOLTAGE, LOW-POWER
ANALOG CMOS DESIGN**

Summary

These course notes provide an introduction to topics in the design of Low-Voltage Low-Power (LV-LP) Analog CMOS design. The course is suitable for professional designers of analog circuits and graduate engineers wishing to acquire knowledge in this area.

Chapter 1

System related design issues for low-voltage low-power CMOS analog circuit design

1.1	Introduction	4
1.2	Limits to Low Power for Analog Circuit Design	4
1.2.1	Fundamental Limits	4
1.2.2	Practical Limits	7
1.2.3	Other obstacles to low-power	8
1.2.4	Implications of supply voltage reduction.....	9

Chapter 2

General notes on the design of low-voltage, low-power operational amplifier cells

2.1	Introduction to low-voltage low-power CMOS circuits.....	11
2.2	Design Issues	11
2.3	Low Voltage Analog Design Considerations	12
2.3.1	Introduction.....	12
2.3.2	Classification of CMOS low-voltage circuits.....	12
2.4	Electrical Properties of MOS transistors	13
2.4.1	Strong Inversion.....	13
2.4.2	Weak Inversion	15
2.4.3	Moderate Inversion	16
2.5	Rail-to-Rail Signals	17
2.5.1	The Inverting Amplifier.....	17
2.5.2	Non-Inverting Amplifier Configuration	18
2.5.3	Summary of Amplifier Configurations.....	18
2.6	Rail-to-rail stages	19

Chapter 3

Input stage design of low-voltage, low-power operational amplifier design

3.1	Introduction	20
3.2	Single Differential Input Stage.....	20
3.3	Rail-to-Rail Input stage	23
3.4	Constant gm rail-to-rail input stages	30
3.4.1	Rail-to-rail input stages with current based gm control.....	30

Chapter 4

Output stage design of low-voltage, low power operational amplifiers

4.1	Introduction	38
4.2	Common-Source Output Voltage	38
4.3	Class AB output stages.....	41
4.4	Feedforward class AB output stages	45
4.5	Feedback class AB output stage.....	51

List of References	55
--------------------------	----

TABLE 1. List of Symbols used in these course notes

Symbol	Quantity	Unit
β	transconductance factor	A/V^2
γ	bulk-threshold parameter	$V^{-1/2}$
Θ	gate electric field parameter	V^{-1}
λ_R	bandwidth reduction factor	-
μ	charge carrier mobility	cm^2/Vs
ξ	source-drain electric field parameter	$\mu m/V$
ϕ_f	flat-band voltage	V
σ	real part of the complex frequency	rad/s
ω	pole frequency	rad/s
ω_μ	unity-gain frequency	rad/s
A_0	DC open-loop gain	dB
A_v	closed loop voltage gain	-
B	Bandwith	Hz
c_{gs}	gate-source capacitance	F
CMRR	common-mode rejection ratio	dB
C	capacitor	F
C_M	Miller Capacitor	F
C_L	load capacitor	F
C_{ox}	normalized oxide capacitance	F/m^2
E_{tr}	energy dissipation in a binary gate transition	J
f	frequency	Hz
g_m	transconductance	Ω^{-1}
g_0	output conductance	Ω^{-1}
I_d	drain current	A
I_{pull}	pull current	A
I_{push}	push current	A
I_q	quiescent current	A
I_s	specific current	A
$j\omega$	imaginary part of the complex frequency	rad/s
k	Boltzmann's constant, $1.3805 \cdot 10^{-23}$	J/K
K_f	flicker noise component	V^2F
L	channel length	m
m	number of binary gate transition cycles	-
n	weak inversion slope factor	-
N	number of bits	-
p	pole	rad/s
PSRR	Power Supply Rejection Ratio	dB

TABLE 1. List of Symbols used in these course notes

Symbol	Quantity	Unit
r_{ds}	drain-source resistance	Ω
R	resistor	Ω
R_L	load resistor	Ω
SNR	Signal-to-Noise Ratio	dB
T	absolute temperature	K
U_T	Thermodynamic Voltage	V
V_B	supply voltage	V
$\overline{v_{eq}^2}$	squared equivalent noise voltage	V ² /HZ
V_{common}	common-mode input voltage	V
V_{DD}	positive supply	V
V_{ds}	drain-source voltage	V
V_{dsat}	saturation voltage	V
V_{gs}	gate-source voltage	V
$V_{gs,eff}$	effective gate-source voltage	V
V_{os}	offset voltage	V
V_P	Pinch-off voltage	V
V_{PP}	Peak-to-Peak Voltage	V
V_{sb}	source-bulk voltage	V
V_{SS}	negative supply voltage	V
V_{sup}	supply voltage	V
$V_{sup,min}$	minimum supply voltage	V
V_{th}	thermal voltage	V
V_T	threshold voltage	V
V_{T0}	threshold voltage at $V_{sb}=0V$	V
$V(t)$	time varying voltage	V
W	channel width	m
z	zero	rad/s
Z	impedance	Ω

Chapter 1

System related design issues for low-voltage low-power CMOS analog circuit design

1.1 Introduction

The current trend towards low-power design is mainly driven by two forces: the growing demand for long-life autonomous portable equipment, and the technological limitations of high-performance VLSI systems. For the first category of products, low-power is the major goal for which speed and/or dynamic range might have to be sacrificed. High speed and high integration density are the objectives for the second application category, which has experienced a dramatic increase of heat dissipation that is now reaching a fundamental limit. These two forces are now merging as portable equipment grows to encompass high-throughput computationally intensive products such as portable computers and cellular phones.

The most efficient way to reduce the power consumption of digital circuits is to reduce the supply voltage, since the average power consumption of CMOS digital circuits is proportional to the square of the supply voltage. The resulting performance loss can be overcome for standard CMOS technologies by introducing more parallelism and/or to modify the process and optimize it for low supply voltage operation.

The rules for analog circuits are quite different to those applied to digital circuits. In order to clarify these differences, the fundamental limits to the reduction of the power consumption are discussed in Section 1.1. It is shown that decreasing the supply voltage unfortunately does not reduce the power consumption of analog circuits. This is mainly due to the fact that the power consumption of analog circuits at a given temperature is basically set by the required signal-to-noise ratio (SNR) and the frequency of operation (or the required bandwidth). A first-order analysis also shows that the absolute minimum power consumption required to process analog signals is almost independent of the supply voltage reduction. In addition to these fundamental limits, some practical limits and additional obstacles to the power reduction are also discussed.

1.2 Limits to Low Power for Analog Circuit Design

1.2.1 Fundamental Limits

Power is consumed in analog signal processing circuits to maintain the signal energy above the fundamental thermal noise in order to achieve the required signal to noise ratio. A representative figure of merit of different signal processing systems is the power consumed to realize a single pole. This can be realised by considering the basic integrator shown in figure 1-1, assuming that all the current pulled from the power sup-

ply is used to charge the integrating capacitor (i.e an ideal 100% current efficient transconductor).

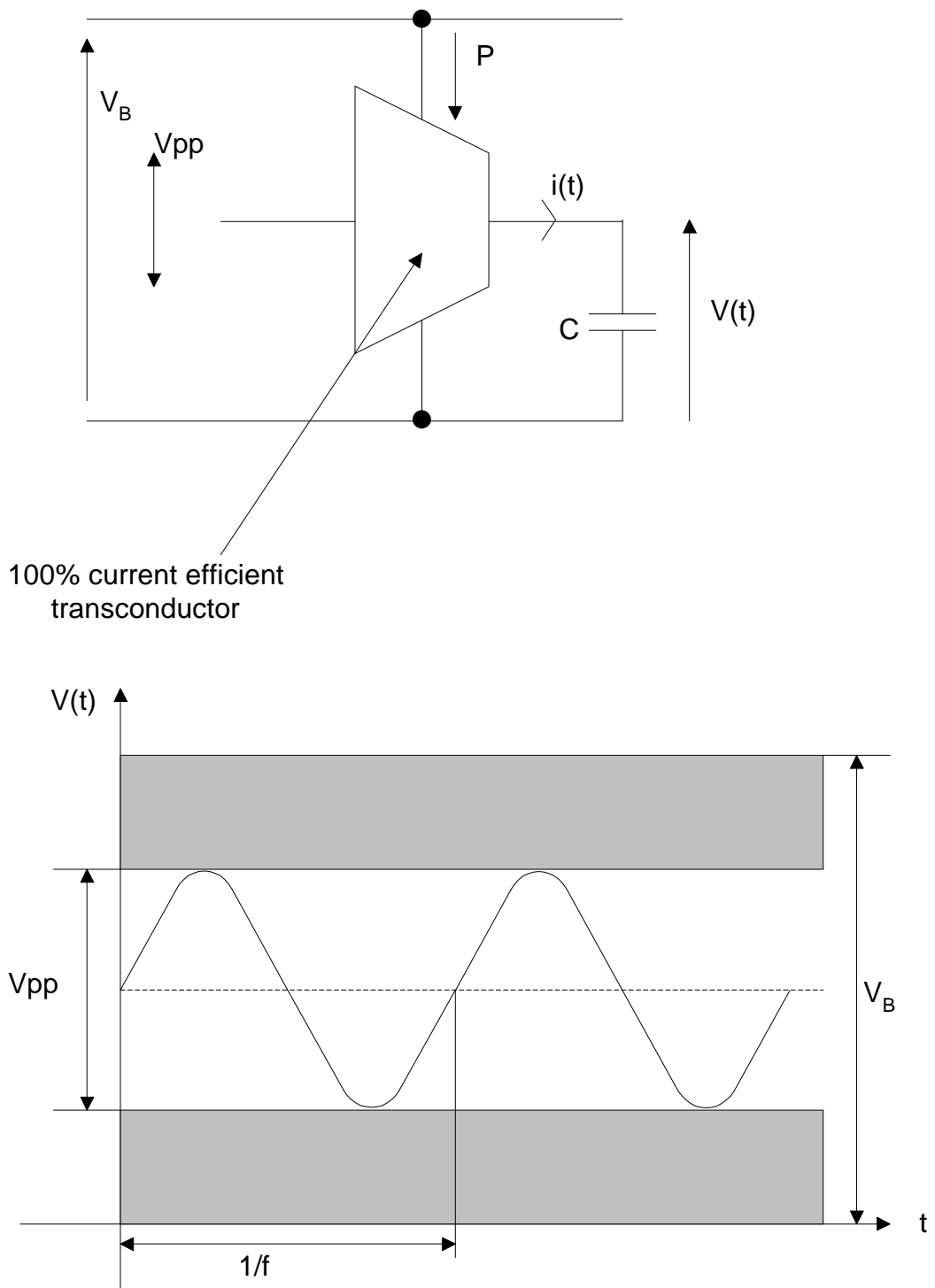


Figure 1-1 Basic Integrator used to evaluate the power necessary to realize a single pole

The power consumed from the supply voltage source, V_B , which is necessary to create a sinusoidal voltage $V(t)$ across capacitor C having a peak-to-peak amplitude V_{pp} and a frequency f can be expressed as:

$$P = 8kT \cdot f \cdot SNR \cdot \frac{V_B}{V_{pp}} \dots \dots \dots (1)$$

According to (1), the minimum power consumption of analog circuits at a given temperature is basically set by the required SNR and the frequency of operation (or the required bandwidth). Since this minimum power consumption is also proportional to the ratio between the supply voltage and the signal peak-to-peak amplitude, power efficient analog circuits should be designed to maximize the voltage swing. The minimum power for circuits that can handle rail-to-rail signal voltages, $V_{pp}=V_B$, reduces to:

$$P = 8kT \cdot f \cdot SNR \dots \dots \dots (2)$$

As shown in figure 1-2, the absolute limit is very steep, since it requires a factor of 10 of power increase for every 10dB of signal-to-noise ratio improvement. It applies to each breakpoint of any linear analog filter (continuous or sampled data as in switched capacitors). The limit is reached in the case of a simple RC filter, but the best existing active filters are still two orders of magnitude above this limit.

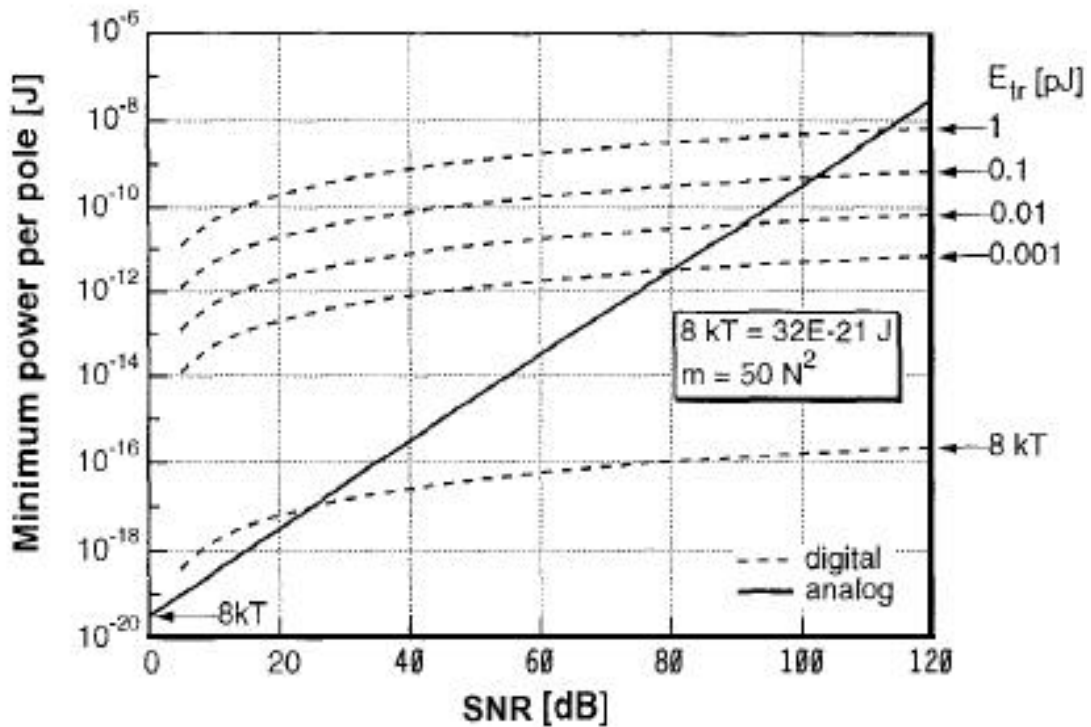


Figure 1-2 Minimum power for analog and digital circuits

The minimum power for an analog system can be compared to that of a digital system, in which each elementary operation requires a certain number m of binary gate transi-

tion cycles, each of which dissipates an amount of energy E_{tr} . The minimum power is then given by:

$$P_{min-digital} = m \cdot f \cdot E_{tr} \dots \dots \dots (3)$$

where f is the signal bandwidth.

The number m of transitions is only proportional to some power a of the number of bits N , and therefore power consumption is only weakly dependent on SNR:

$$m \cong N^a \sim [\log(SNR)] \dots \dots \dots (4)$$

It is possible to compare analog and digital technology by estimating the number of gate transitions that are required to compute each period of the signal, which for a single pole digital filter can be estimated to be approximately:

$$m \cong 50 \cdot N^2 \dots \dots \dots (5)$$

However, the need to recharge each gate capacitance to the supply voltage raises the minimum energy above the absolute $8kT$ limit. As shown in figure 1-2, the minimum power for digital is therefore much higher than the absolute limit.

Comparison of these fundamental limits are plotted in figure 1-2. They show clearly that analog systems may consume much less power than their digital counterpart, provided a small signal-to-noise ratio is acceptable. But for systems requiring large signal-to-noise ratios, analog becomes very power inefficient.

1.2.2 Practical Limits

The limits discussed so far are fundamental since they do not depend on the technology nor the choice of power supply voltage. However a number of obstacles or technological limitations are in the way to approaching these limits in practical circuits, and ways to reduce the effect of these various limitations can be found at all levels of analog design ranging from device to system:

1. Capacitors increase the power necessary to achieve a given bandwidth. They are only acceptable if their presence reduces the noise power by some amount (by reducing

the noise bandwidth). Therefore, ill placed parasitic capacitors very often increase power consumption.

2. The power spent in bias circuitry is wasted and should in principle be minimized. However, inadequate bias schemes may increase the noise and therefore require a proportional increase in power. For example, a bias current is more noisy if it is obtained by multiplying a smaller current.

3. According to equation (1), power is increased if the signal at any node corresponding to a functional pole (pole within the bandwidth, or state variable) has a peak-to-peak voltage amplitude smaller than the supply voltage V_B . Thus, care must be taken to amplify the signal as early as possible to its maximum possible voltage value, and to maintain this level all along the processing path. Using current-mode circuits with limited voltage swings is therefore not a good approach to reduce power, as long as the energy is supplied by a voltage source.

4. The presence of additional sources of noise implies an increase in power consumption. These include $1/f$ noise in the devices, and noise coming from the power supply or generated on chip by other blocks of the circuit.

5. When capacitive loads are imposed (for example by parasitic capacitors), the power supply current I necessary to obtain a given bandwidth is inversely proportional to the transconductance-to-current ratio gm/I of the active device. The small value of gm/I inherent to MOS transistors operated in strong inversion may therefore cause an increase in power consumption.

6. The need for precision usually leads to the use of larger dimensions for active and passive components, with a resulting increase in parasitic capacitors and power.

7. All switched capacitors must be clocked at a frequency higher than twice the signal-frequency. The power consumed by the clock itself may be dominant in some applications.

1.2.3 Other obstacles to low-power

In addition to the fundamental and practical limitations discussed previously, there are also historical or even psychological barriers to the efficient design of low power analog circuits. The most important can be listed as:

1. Analog blocks must often be taken from existing libraries with bias currents at the milliamper level and with architectures that are not compatible with low-voltage or low-current.

2. The use of very low bias currents is often discarded due to a lack of adequate transistor models and correct characterization of transistors parameters as well as worst case leakage currents. Another obstacle is the fear of breaking the psychological microampere barrier.

3. The requirements on PSRR are often exaggerated and mistaken for insensitivity to noise generated on chip.

1.2.4 Implications of supply voltage reduction

According to equation (1), reducing the supply voltage of analog circuits while preserving the same bandwidth and SNR, has no fundamental effect on their minimum power consumption. However, this absolute limit was obtained by neglecting the possible limitation of bandwidth B due to the limited transconductance g_m of the active device. The maximum value of B is proportional to g_m/C . It can be shown that by replacing the capacitor value C by g_m/B yields:

$$SNR \cdot B = V_{PP}^2 \cdot \frac{g_m}{8kT} \dots\dots\dots(6)$$

In most cases, scaling the supply voltage V_B , by a factor K requires a proportional reduction of the signal swing V_{pp} . Maintaining the bandwidth and the SNR is therefore only possible if the transconductance g_m is increased by a factor K^2 . If the active device is a bipolar transistor (or a MOS transistor biased in weak inversion), its transconductance can only be increased by increasing the bias current I by the same factor K^2 ; power ($V_B \times I$) is therefore increased by K . The situation is different if the active device is a MOS transistor biased in strong inversion. Its transconductance can be shown to be proportional to I/V_p , where V_p is the pinch-off or saturation voltage of the device. Since this saturation voltage has to be reduced proportionally with V_B , increasing g_m by K^2 only requires an increase of current I by a factor K and hence the power remains unchanged.

However, the maximum frequency of operation may be affected by the value of the supply voltage. For a MOS transistor in strong inversion, the frequency f_{max} for which the current gain falls to unity is approximately given by:

$$f_{max} \cong \frac{\mu \cdot V_P}{L^2} \dots\dots\dots(7)$$

Therefore, if the process is fixed (channel length L constant) a reduction of V_B , and V_p by a factor K causes a proportional reduction of f_{max} . However, there is no fundamental reason to reduce the supply voltage of an analog circuit in a given process. On the other hand, a reduction of V_B is unavoidable to maintain the electric fields constant when scaling down a process. Both V_p and L are then scaled by the same factor K and the maximum frequency f_{max} is increased by K . For a bipolar transistor, V_p in equation (7) is replaced by $U_T = kT/q$ and f_{max} does not, in first approximation, depend on the supply voltage V_B .

Low-voltage limitations are not restricted to power or frequency problems. Reducing V_p increases the transconductance-to-current ratio of MOS transistors which in turn increases the noise content of current sources and drastically degrades their precision. In analog switches, when the supply voltage falls below approximately the sum of the p-channel and n-channel transistor threshold voltages, it is difficult to ensure the conductance. The absolute value of charge injection in a switch for a given value of time constant does not depend on V_B , but it increases in relative value if V_B and V_{pp} are decreased. The same is true for any constant voltage overhead such as the base-emitter voltage in bipolar transistors or the threshold voltage in MOS transistors.

As already illustrated in the previous discussion on the fundamental limits to LP and LV, many of the problems and solutions encountered in the design of LP-LV analog circuits are directly related to the properties and limitations of the MOS transistor itself, which must therefore be properly understood and correctly modelled down to very low currents.

Chapter 2

General notes on the design of low-voltage, low-power operational amplifier cells

2.1 Introduction to low-voltage low-power CMOS circuits

Low voltage circuits are needed because:

1. As the device channel length is scaled down into submicrons and the gate oxide thickness becomes only several nanometers thick, the supply voltage has to be reduced in order to ensure device reliability. With deep submicron processes now available, the maximum allowable supply voltage is decreasing from 5V to 3V and even to 2V.
2. The increasing density of the components on chip dictates low power. A silicon chip can only dissipate a limited amount of power per unit area. Since the increasing density of components allows more electronic functions per unit area, the power per electronic function has to be lowered in order to prevent overheating.
3. Portable, battery-powered equipment needs low power to ensure an acceptable operation period from a battery, and the supply voltage must be as low as possible to reduce the number of batteries used.

2.2 Design Issues

Unlike digital circuits, the power dissipated by analog circuitry does not necessarily decrease when the supply voltage lowers, as the traditional stacking of transistors has to be replaced by folding techniques, which inevitably increases the current drawn from the supply. Hence to decrease the power dissipated in low-voltage analog circuits, the design has to be kept as simple as possible, while maintaining good circuit specifications.

The low-voltage low-power demand has an enormous impact on the dynamic range of an amplifier:

1. The dynamic range is lowered because of the lower allowable signal swing and it is reduced because of the larger noise voltages that arise as a result of the smaller supply currents.
2. In order to maximise the dynamic range, a low voltage amplifier must be able to deal with signal voltages that extend from rail-to-rail. This requires classical circuit solutions to be replaced by new configurations.
3. The unity-gain frequency of operational amplifiers is greatly effected by the low power condition. The lower supply currents will drastically reduce the bandwidth for cases where the load capacitance cannot be lowered.
4. To obtain sufficient low frequency gain, low voltage amplifiers often require cascaded gain stages, which results in more complex frequency compensation

schemes. In a low voltage low power environment, these frequency compensation schemes have to be as power efficient as possible.

2.3 Low Voltage Analog Design Considerations

2.3.1 Introduction

Low voltage design methods are used for circuits which are able to run on supply voltages somewhere between 1 and 5 Volts. These low supply voltages put an upper limit on the number of gate-source voltages and saturation voltages which can be stacked. The gate source and the saturation voltage of an MOS transistor (and therefore the minimum supply voltage of a circuit) depend on the specific design parameters, such as the threshold voltage and biasing levels. The lowest supply voltage can be obtained by biasing MOS transistors in weak inversion, since this gives the smallest gate-source voltage for a given transistor.

However, relatively high frequency or high slew rate applications require transistors biased in strong inversion rather than in weak inversion. This raises the gate-source voltage of a device, and therefore the minimum supply voltage. Section 2.4 focuses on the gate-source voltage of an MOS transistor. Expressions will be given for an transistor operating in weak inversion and in strong inversion. Associated with the gate-source voltage is the transconductance of an MOS transistor which also will be discussed in this section.

Section 2.5 zooms in on a specific problem that is encountered in low voltage operational amplifier design. The lower supply voltage drastically reduces the dynamic range of operational amplifiers. In order to maximize the dynamic range, the signal voltages have to be as large as possible, preferably from rail-to-rail. This will pose specific demands on the common mode input range and output voltage range of low-voltage amplifiers. Section 2.5 discusses these requirements from the point of view of two widely used applications, the non-inverting and inverting feedback applications. Since the signal voltage can extend from rail-to-rail, the input and output stage of an amplifier must be able to process these signals. This requires traditional circuit solutions to be abandoned.

2.3.2 Classification of CMOS low-voltage circuits

In order to enable a designer to predict the feasibility of a circuit application, this work gives a relationship between low-voltage and the number of stacked gate-source voltages and saturation voltages. Here the term low voltage is used for circuits that are able to operate on a supply voltage of two stacked gate-source voltages and two saturation voltages. In a formula

$$V_{sup, min} = 2(V_{gs} + V_{dsat}) \dots \dots \dots (8)$$

where V_{gs} and V_{dsat} are the gate-source voltage and the saturation voltage of an MOS device, respectively.

Circuits that only need a minimum supply voltage of one gate-source voltage and a saturation voltage will be referred to as extremely low-voltage circuits. This yields

$$V_{sup, min} = V_{gs} + V_{dsat} \dots \dots \dots (9)$$

It should be noted that extremely low-voltage circuits require a minimum supply voltage that is about half the supply voltage which is necessary for low-voltage circuits.

2.4 Electrical Properties of MOS transistors

One of the most important electrical properties of an MOS transistor, when designing low -voltage amplifiers, is the gate source voltage, as it determines the minimum supply voltage at which the amplifier is able to operate. Associated with this gate source voltage is the transconductance. Since the MOS transistor is a voltage driven device, the required transconductance determines the gate source voltage of a transistor. In this section the gate-source voltage and the transconductance of an MOS transistor will be dealt with.

2.4.1 Strong Inversion

An MOS transistor is said to be operating in the strong inversion region when it's gate-source voltage is larger than its threshold voltage. In this region the transistor saturates when

$$V_{ds} > V_{gs} - V_T \dots \dots \dots (10)$$

where V_{ds} and V_T are the drain-source voltage and the threshold voltage respectively. The drain-source voltage at which a transistor begins to saturate is called the saturation voltage, V_{dsat} . In the operational amplifier design practice almost all transistors are biased in the saturation region, because this provides the largest voltage gain for a given drain current and device geometry.

In the saturation region, the relationship between the drain current, I_d , and the gate-source voltage, V_{gs} , can be expressed by:

$$I_d = \frac{1}{2} \cdot \frac{\mu \cdot C_{ox}}{1 + (V_{gs} - V_T) \cdot \left(\Theta + \frac{\xi}{L} \right)} \cdot \frac{W}{L} \cdot (V_{gs} - V_T)^2 \dots \dots \dots (11)$$

where μ is the mobility of the charge carriers, C_{ox} is the normalised oxide capacitance, V_{gs} is the gate-source voltage, V_T is the threshold voltage of the device, W is the gate width, L is the gate length, Θ models the effect of the gate electrical field (typical value is 0.1 V^{-1}), ξ expresses the effect of the source-drain electrical field (typical value is $0.3 \mu\text{m/V}$).

In order to determine the total gate-source voltage of an MOS transistor, it can be divided into two parts, the threshold voltage and the effective gate-source voltage which actually drives the transistor. This yields

$$V_{gs} = V_T + V_{gs, eff} \dots \dots \dots (12)$$

The threshold voltage, V_T , can be expressed by::

$$V_T = V_{T0} + \gamma \cdot (\sqrt{2\phi_f + V_{sb}} - \sqrt{2\phi_f}) \dots \dots \dots (13)$$

where: V_{sb} is the source-bulk voltage, V_{T0} is the threshold voltage at zero bulk-source voltage, γ is the bulk-threshold parameter (typical value of $0.7 \text{ V}^{-1/2}$), and ϕ_f is the surface potential (typical value of 0.6 V).

Using these values, figure 2-1 shows the threshold voltage verses the source-bulk voltage. From this it is clear that the threshold voltage increases when the source-bulk voltage grows. Therefore in low voltage design the source-bulk voltage should be kept as low as possible. Also, it can be concluded that transistors which have to match must be biased at the same source-bulk potential

For low voltage design practice, where most of the transistors in the amplifier operate on the verge of saturation, it can be shown that in devices that are required to supply a high driving current, for instance in the output transistors of an amplifier, that:

$$I_d = \frac{1}{2} \cdot \mu \cdot C_{ox} \cdot \frac{W}{L} \cdot V_{gs, eff}^2 \dots \dots \dots (14)$$

This can be re-arranged to show that:

$$V_{gs, eff} = \sqrt{\frac{2}{\mu \cdot C_{ox}} \cdot \frac{L}{W} \cdot I_d} \dots \dots \dots (15)$$

The key small signal parameter of an MOS transistor is the transconductance:

$$g_m = \frac{\partial I_d}{\partial V_{gs, eff}} \dots \dots \dots (16)$$

Which yields:

$$g_m = \mu \cdot C_{ox} \cdot \frac{W}{L} \cdot V_{gs, eff} \dots \dots \dots (17)$$

And by substitution, it can be shown that:

$$g_m = \sqrt{2 \cdot \mu \cdot C_{ox} \cdot \frac{W}{L} \cdot I_d} \dots \dots \dots (18)$$

The above equations imply that g_m is determined by $V_{gs,eff}$. However, if we wish to increase g_m it may not be possible to increase $V_{gs,eff}$ since this would require a higher supply voltage. It is therefore better to increase g_m , say by a factor n , by increasing both W/L and I_d by the same factor n .

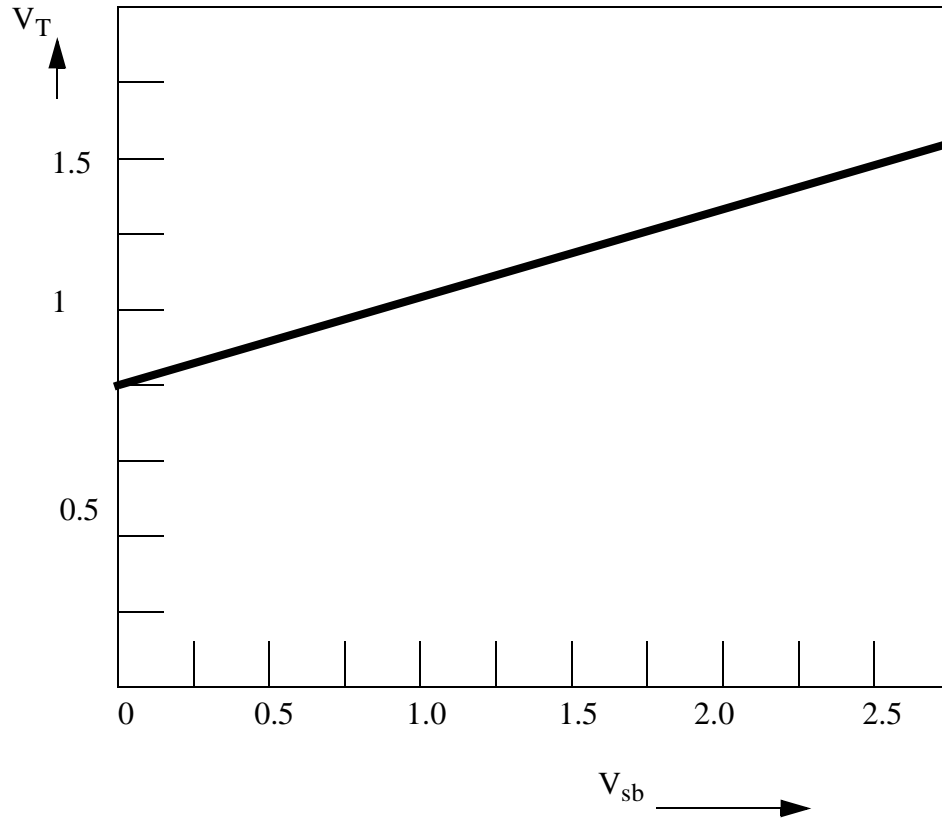


Fig. 2-1 The threshold voltage versus the source-bulk voltage. V_{T0} , ϕ_f and γ are chosen to be 0.8V, 0.6V and $0.7 \text{ V}^{-1/2}$, respectively.

2.4.2 Weak Inversion

An MOS transistor operates in Weak Inversion when $V_{gs} < V_T$. In this mode, the transistor saturates when $V_{ds} > 3$ to $4 V_{th}$. V_{th} is the thermal voltage kT/q , which is about 25mV at room temperature. In general, the saturation voltage of an MOS transistor operating in this mode is lower than that of a device operating in strong inversion. The current I_d in saturation is described as follows:

$$I_d = I_s \cdot e^{\frac{V_{gs} - V_T}{n \cdot V_{th}}} \dots\dots\dots(19)$$

Where n is the weak inversion slope factor, and I_s is the specific current given by:

$$I_s = 2 \cdot n \cdot \mu \cdot C_{ox} \cdot V_{th}^2 \cdot \frac{W}{L} \dots \dots \dots (20)$$

I_s typically ranges from 2nA to 200nA. These equations can be re-arranged to as:

$$V_{gs,eff} = n \cdot V_{th} \cdot \ln\left(\frac{I_d}{I_s}\right) \dots \dots \dots (21)$$

Because I_d is less than I_s , $V_{gs,eff}$ has a negative value. Therefore, V_{gs} ($V_{gs}=V_T+V_{gs,eff}$) for a MOS transistor operating in weak inversion, is smaller than the V_{gs} of a device operating in strong inversion. Thus a device biased under the weak inversion regime is more suitable for low voltage operation.

The transconductance in weak inversion is defined as:

$$g_m = \frac{I_d}{n \cdot V_{th}} \dots \dots \dots (22)$$

In order to achieve a large g_m , we must increase I_d . However, increasing this current may push the device into the strong inversion mode, and this must be avoided to preserve low voltage operation. We can maintain the device in the low voltage mode by adjusting the W/L but the increased device geometries lead to increased device parasitic capacitances, thus affecting the high frequency performance.

2.4.3 Moderate Inversion

There is a smooth transition region between the weak inversion and the strong inversion, which is called the moderate inversion. The moderate inversion region extends the drain currents approximately as follows:

$$\frac{1}{8} \cdot I_s < I_d < 8 \cdot I_s \dots \dots \dots (23)$$

In this region, simple analytic models are not available, therefore it is advisable to use computer simulations to analyse the transistor operation.

2.5 Rail-to-Rail Signals

The lowering of the power supply and voltage has an enormous impact on the SNR of analog circuits. SNR decreases because of the lower allowable signal voltages and also because of the higher noise voltages due to lower supply currents.

To maximise SNR, we must make the signals as large as possible, ideally from rail-to-rail. This imposes special demands on the common mode input range and the output voltage range of an amplifier. We will discuss these demands by employing two widely used op-amp applications, the inverting and non-inverting amplifier.

2.5.1 The Inverting Amplifier

The inverting amplifier is described in figure 2-2.

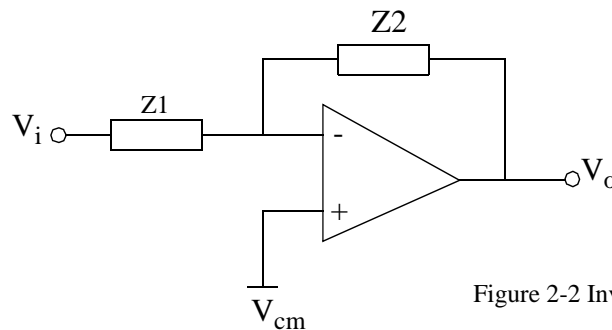


Figure 2-2 Inverting Amplifier

The gain of this amplifier is as follows:

$$A_V = \frac{V_O}{V_i} = -\frac{Z_2}{Z_1} \dots \dots \dots (24a)$$

To maximise the SNR, the output voltage swing should be as large as possible, ideally from rail-to-rail.

The demand on the common mode input voltage range is less severe. This range can be small because the positive input of the amplifier is biased at a fixed common mode voltage, V_{cm} . This input can be biased at any voltage, but it is more efficient to connect it to half the supply voltage, because then the output voltage can have its maximum positive and negative swing. If the common mode voltage is biased at another value, a level shift is required to obtain the maximum signal swing at the output and the additional level shift will induce noise and thus degrade the SNR further. Very careful design of the level shift is required to minimise this noise.

2.5.2 Non-Inverting Amplifier Configuration

The non-inverting amplifier configuration is described in figure 2-3.

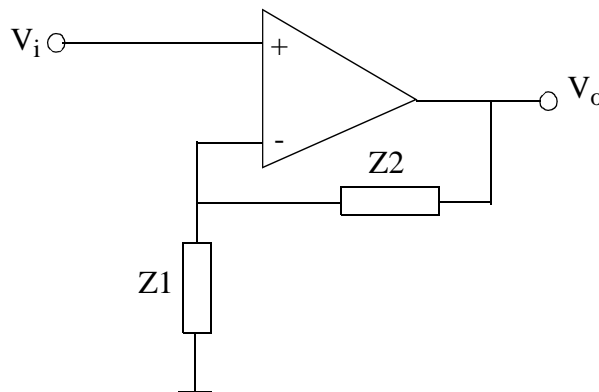


Figure 2-3 Non-Inverting Amplifier Configuration

The gain for the non-inverting amplifier is described as follows:

$$A_v = \frac{V_o}{V_i} = 1 + \frac{Z_2}{Z_1} \dots\dots\dots(24b)$$

To maximise the SNR the output voltage swing has to be able to swing from rail-to-rail. The demands on the common-mode input voltage range are more severe than for the inverting feedback application. This is because the feedback network is connected in series with the output and therefore the common-mode input voltage swing is as large as the input signal itself. Assuming that V_o can swing from rail-to-rail, it can be shown that the common-mode input voltage range of the amplifier has to be at least:

$$V_{common} = V_{sup}/A_v \dots\dots\dots(25)$$

where V_{common} = Common Mode Input Voltage

V_{sup} = Supply Voltage

In the limiting case, that is when Z_1 approaches infinity and Z_2 reaches zero, the non-inverting amplifier operates as the well-known voltage buffer with a gain of one. For this type of application, the common-mode input voltage range of the amplifier has to be from rail-to-rail.

2.5.3 Summary of Amplifier Configurations

To obtain a maximum SNR in the inverting feedback applications:

- The output voltage has to be able to swing from rail-to-rail
- The common mode input voltage has only to be biased at a fixed voltage

To obtain a maximum SNR in the non-inverting feedback applications:

- The output voltage has to be able to swing from rail-to-rail
- The common mode input voltage range has to extend from rail-to-rail

2.6 Rail-to-rail stages

To maximise SNR in LV-LP requires rail-to-rail input and output ranges, although input range is required only in voltage follower applications.

This rail to rail demand on the amplifier stages requires classical solutions to be abandoned.

Chapter 3

Input stage design of low-voltage, low-power operational amplifier design

3.1 Introduction

The main purpose of the input stage is to amplify differential signals and reject common-mode input voltages. An important specification of an input stage is the common-mode input range. If the common mode voltage is kept within this range, the input stage will properly respond to small differential signals. Hence an application has to be designed such that the common mode input voltage stays within the common-mode input range. Other important specifications of the input stage are the input referred noise, offset, and the common-mode rejection ratio.

3.2 Single Differential Input Stage

The most commonly used input stage is a single differential pair. This can typically be composed of a p-channel pair, M1-M2, or an n-channel pair M3-M4, as shown in figure 3-1.

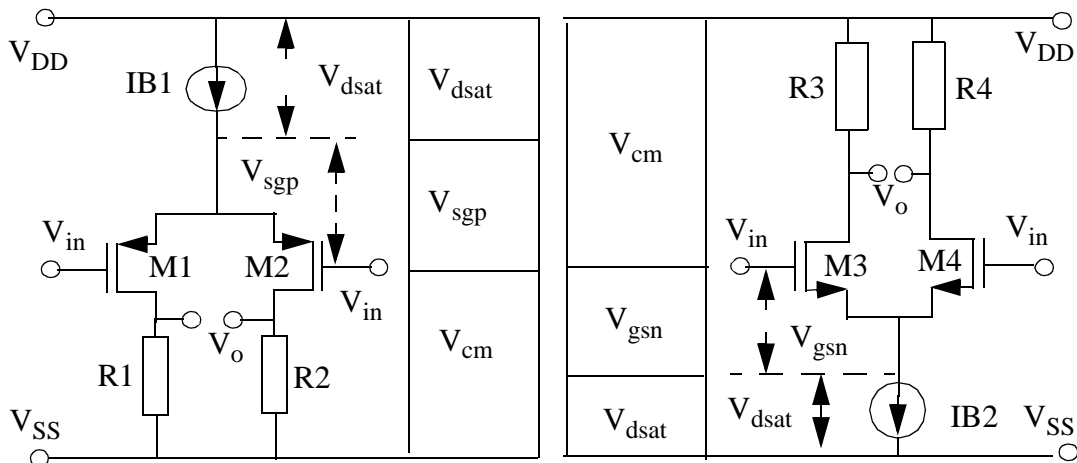


Figure 3-1 The common-mode input range of a p-channel and an n-channel differential pair

For a p-channel pair, the common mode input voltage range is given by:

$$V_{SS} < V_{common} < V_{DD} - V_{dsat} - V_{sgp} \dots \dots \dots (26a)$$

where:

V_{common} is the common mode input voltage, V_{sgp} is the source-gate voltage of an input transistor, V_{dsat} is the voltage across a current source, V_{DD} is the positive supply voltage, and V_{SS} is the negative supply voltage.

For an n-channel input pair, the common mode input voltage is given by:

$$V_{SS} + V_{gsn} + V_{dsat} < V_{common} < V_{DD} \dots \dots \dots (26b)$$

where V_{gsn} is the gate source voltage of an n-channel input transistor.

In practical amplifiers, a differential pair is often loaded with a current mirror instead of resistors. This is shown in figure 3-2.

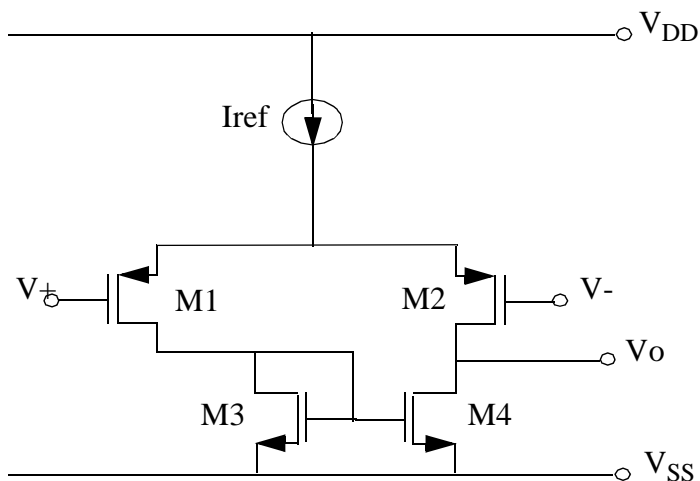


Figure 3-2 Single differential input pair with a current mirror as a load

This input stage consists of a p-channel differential pair M1-M2 and a current mirror M3-M4, providing a differential to single ended conversion. This current mirror load drastically reduces the common-mode input range, because the drain voltage of M3 can only reach the negative supply rail within one gate-source voltage.

To understand this, suppose the common mode input voltage decreases. As a result of this, the current mirror will finally push M1 out of saturation. This yields a common mode input voltage range which is limited to:

$$V_{SS} + V_{gsn} + V_{TP} < V_{common} < V_{DD} - V_{sgp} - V_{dsat} \dots \dots \dots (27)$$

Other factors raise this gate source voltage, V_{gsn} , i.e.:

- V_{gsn} of the mirror is often increased to minimise noise and offset
- 300mV increase typical due to threshold voltage mismatch

As a consequence, the current mirror can easily raise the lower limit of the common mode input range by 600mV above the negative supply rail. This also raises the minimum supply voltage of the input stage by 600mV.

The folded cascode input stage, as shown in figure 3-3, overcomes these problems.

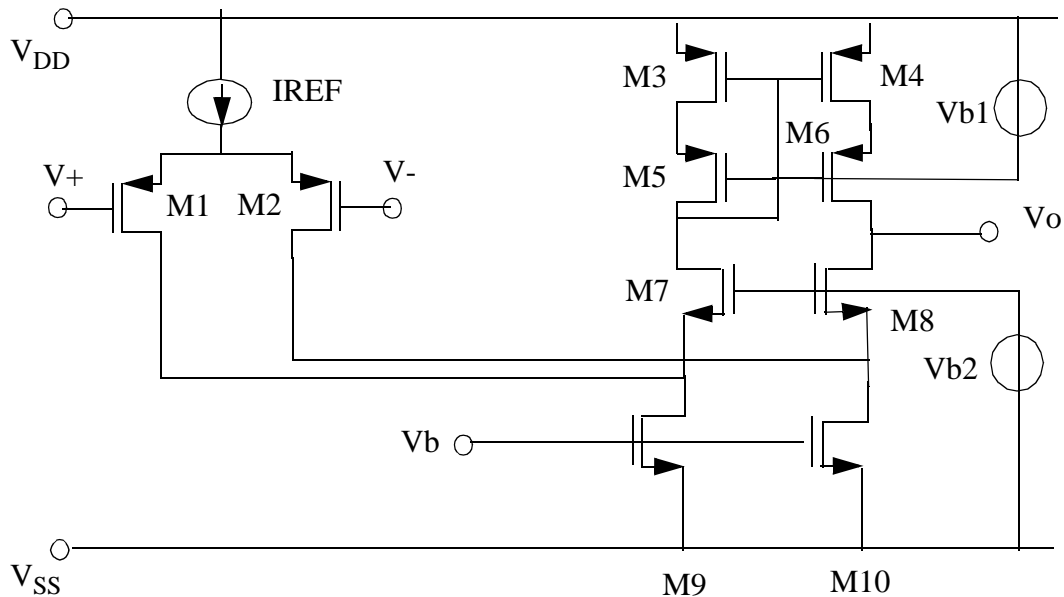


Figure 3-3 Folded Cascode Input Stage

The operation of the folded cascode stage can be described briefly as follows; the input stage consists of a p-channel differential pair M1-M2; the folded cascodes M7-M8 provide a level shift function; current mirror M3-M6, providing differential to single ended conversion; M9-M10 function as bias current sources (in order to maximise the output current of the input stage these current sources are biased at the same values as the tail current *IREF*).

In this configuration, the drain voltage of both input transistors can reach the negative supply voltage within one saturation voltage of the current sources M9 and M10. This saturation voltage is generally much smaller than the gate-source voltage. Thus the folded cascode input stage has a common mode input voltage which includes the negative supply rail.

An important parameter of the folded cascode input stage is the input referred offset voltage. The most significant mismatches that give rise to offset are those of threshold voltages, V_T , and those of transconductance factor $\beta = \frac{1}{2} \cdot \mu \cdot C_{ox} \cdot \frac{W}{L}$(28)

The offset of the folded cascode input stage can be minimised by making:

- The area of the transistors as large as possible
- the effective gate source voltage of the input transistors as small as possible
- the W/L ratio of the current mirror and the current sources as small as possible

The thermal noise can be minimised by making:

- The g_m of the input transistors as large as possible
- The W/L ratio of the current mirror and the current sources as large as possible

The flicker noise of the input stage can be minimised by making:

- The area of the input transistors as large as possible
- The length of the current mirror and the current source as long as possible
- Use of input transistors that display the smallest flicker noise possible, often p-channel devices

The single differential input pair has a common mode input range which includes only one of the supply rails. Therefore it is only suitable for applications in which the common mode input range of the amplifier can be small, as in the inverting amplifier configuration. In voltage follower applications, the amplifier requires a common mode input voltage range that extends from rail-to-rail.

3.3 Rail-to-Rail Input stage

The input stage of an amplifier intended for use in a voltage follower configuration has to have a common mode input range which extends from rail to rail. In order to achieve this, an n-channel and a p-channel input pair can be placed in parallel, as shown in figure 3-4.

The n-channel input pair M3-M4 is able to reach the positive supply rail while the p-channel one, M1-M2, can sense common-mode voltages around the negative supply rail.

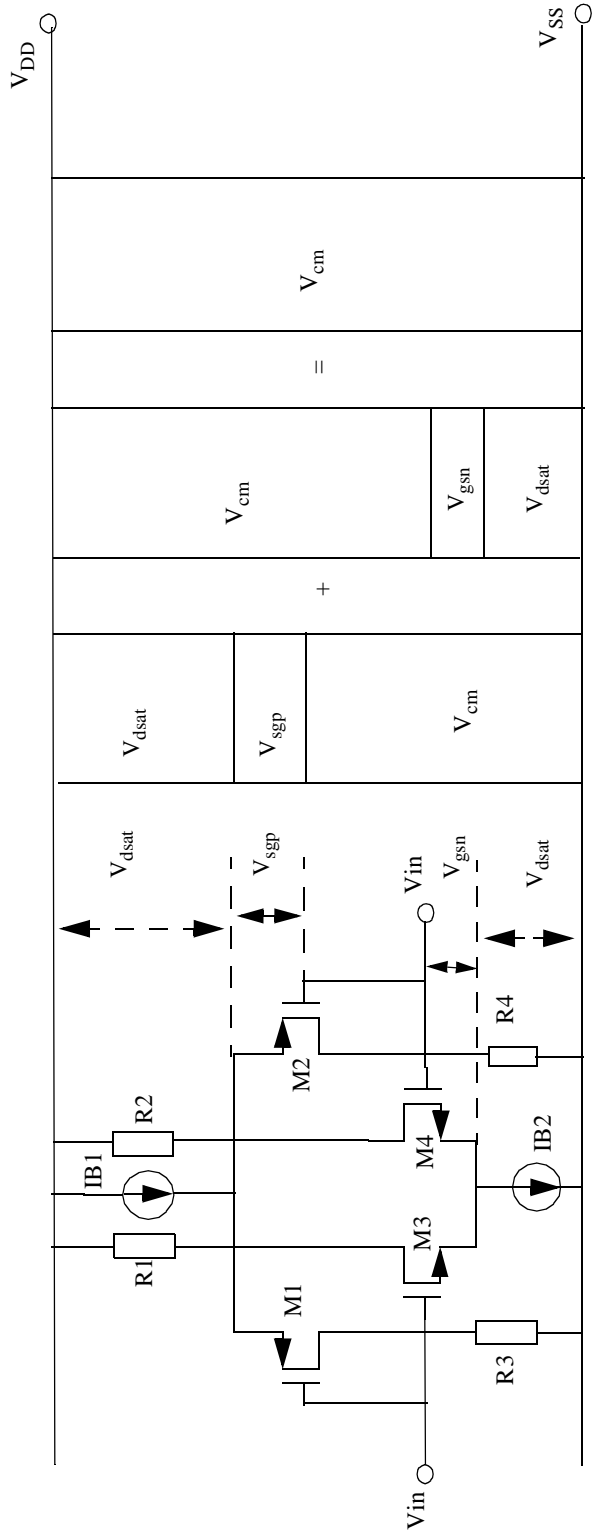


Figure 3-4 Common-Mode Input Range of a rail-to-rail input stage. The supply voltage is larger than $V_{sgp} + V_{gsn} + 2V_{dsat}$

In order to ensure a full rail-to-rail common mode input voltage range, the supply voltage of the rail-to-rail input stage has to be at least:

$$V_{sup, min} = V_{sgp} + V_{gsn} + 2V_{dsat} \dots \dots \dots (29)$$

If the supply voltage is above the minimum supply voltage, the common mode input voltage range can be divided into the following three parts:

- Low common mode input voltages; only p-channel input pair operates
- Intermediate common mode input voltages; the p-channel as well as the n-channel input pair operate.
- High common mode input voltage; only the n-channel input pair operate

If the rail to rail input stage is biased at a supply voltage below $V_{sup, min}$, a gap appears in the intermediate part of the common mode input voltage range. This is clearly depicted in figure 3-5.

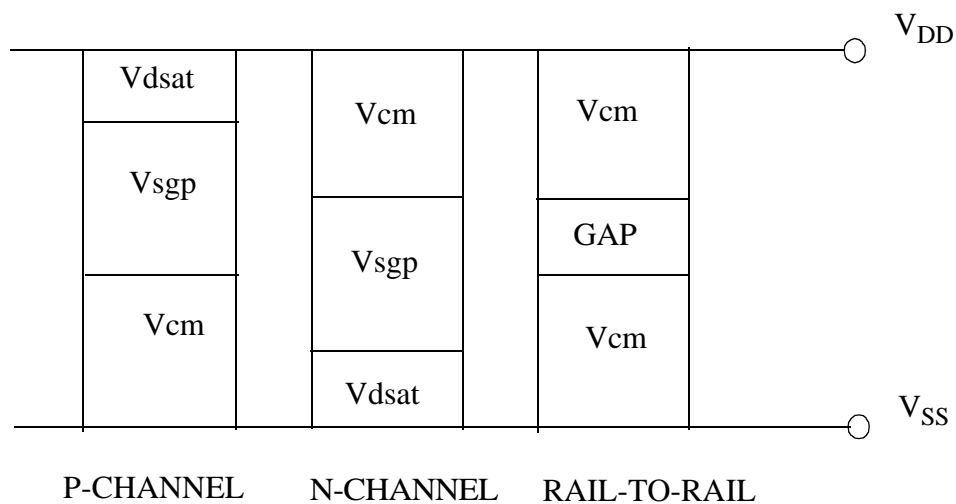


Figure 3-5 Common-Mode input range of a rail-to-rail input stage. The supply voltage is smaller than $V_{sgp} + V_{gsn} + 2V_{dsat}$

Some important points about the rail-to-rail input stage:

- The minimum supply voltage of this stage can be as low as 1.6V, assuming that it operates in weak inversion and that the threshold voltages of the p-channel and n-channel devices have a value of 0.8V.
- The minimum required supply voltage increases when the input stage operates in strong inversion. It will be about 2.5V for a gate source voltage of 1V and a V_{dsat} of 0.25V.

- In order to maintain the rail-to-rail capabilities of the input stage, the complementary input pairs have to be loaded with folded cascodes instead of current mirrors, which is shown in figure 3-6.

Rail-to-Rail M1 through M4 are loaded with a summing circuit. The summing circuit comprises folded cascode M9-M10, which together with the low-voltage current mirror, M5-M8, add the signals from the input stage. M11-M12 function as bias current sources.

The key parameter in the rail-to-rail input stage is the input referred offset voltage. The most important mismatches that give rise to this offset are mismatches in the threshold voltages, V_T , and mismatches in the transconductance factors, b . The strategies adopted for reducing these are similar to those identified for the single input pair differential amplifier, section 3.2.

A major drawback with the rail-to-rail input stage is that its transconductance varies by a factor of two over the common mode input range. If this input stage is part of an operational amplifier in feedback, the loop gain of this configuration will also vary by a factor of two. This in turn causes an undesired additional distortion. To understand this, consider the non-inverting feedback application, as shown in figure 3-7. The gain of this configuration is given by:

$$\frac{V_o}{V_i} = \frac{1}{B} \cdot \left(\frac{A \cdot B}{1 + A \cdot B} \right) \dots \dots \dots (30)$$

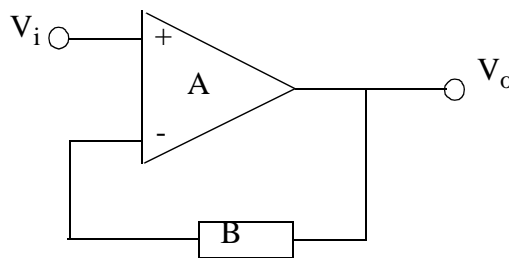


Figure 3-7 Amplifier connected in a non-inverting feedback configuration

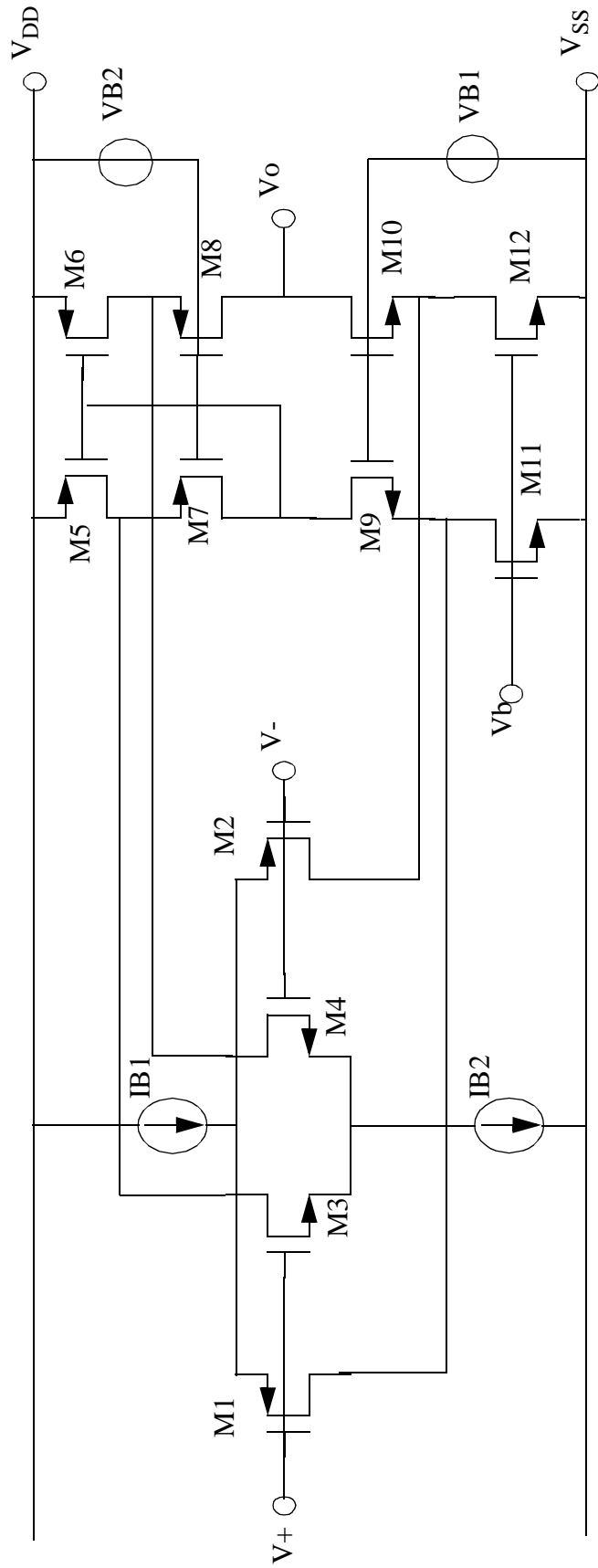


Figure 3-6 Rail-to-Rail folded cascoded input stage

As can be concluded from this a varying loop gain, AB , results in a change of the gain of the non-inverting feedback configuration. At heavy resistive loads, the nominal value of AB can be as low as 100 in the upper and lower part of the common mode input range. At intermediate common mode input voltages the gain increases to a value of 200. As a result the gain of the non-inverting feedback application varies about 0.5% over the common-mode input range, which in turn gives rise to distortion.

Another drawback of the varying load transconductance is that it impedes an optimal frequency compensation. This can be explained by examining the two stage op-amp shown in figure 3-8. This circuit consists of a rail-to-rail input stage M1-M4, a summing circuit M5-M10, a simple class A output stage. The capacitor C_M provides the frequency compensation of the amplifier.

A two stage amplifier has to be dimensioned such that ω_u (it's unity gain frequency) always obeys:

$$\omega_u = \frac{g_{mi}}{C_M} = \frac{1}{2} \cdot \frac{g_{mo}}{C_L} \dots\dots\dots(31)$$

where: g_{mi} is the transconductance of the input stage, g_{mo} is the transconductance of the output transistor M13, and C_L is the load capacitor

Suppose that the amplifier should have a unity gain frequency of at least ω_{mmin} over the total common-mode input range. Since the g_m of the input stage is two times larger in the intermediate than in the outer parts of the common mode input voltage range, this indicates that for intermediate common-mode input voltages the unity gain frequency is $2\omega_{mmin}$. From the above equation it can be understood that the transconductance of the output transistor has to be two times larger than required in order to guarantee stability for each common mode input level. Assuming that the output transistor is biased in strong inversion, this results in a bias current of the output transistor which has to be four times larger than necessary. Often the bias current of the output transistor largely determines the the total current of an amplifier, and thus the total power drawn from supplies. Thus the power dissipation of the amplifier is four times larger than necessary.

In order to overcome the aforementioned drawbacks of rail-to-rail amplifiers, the transconductance has to be regulated at a constant value.

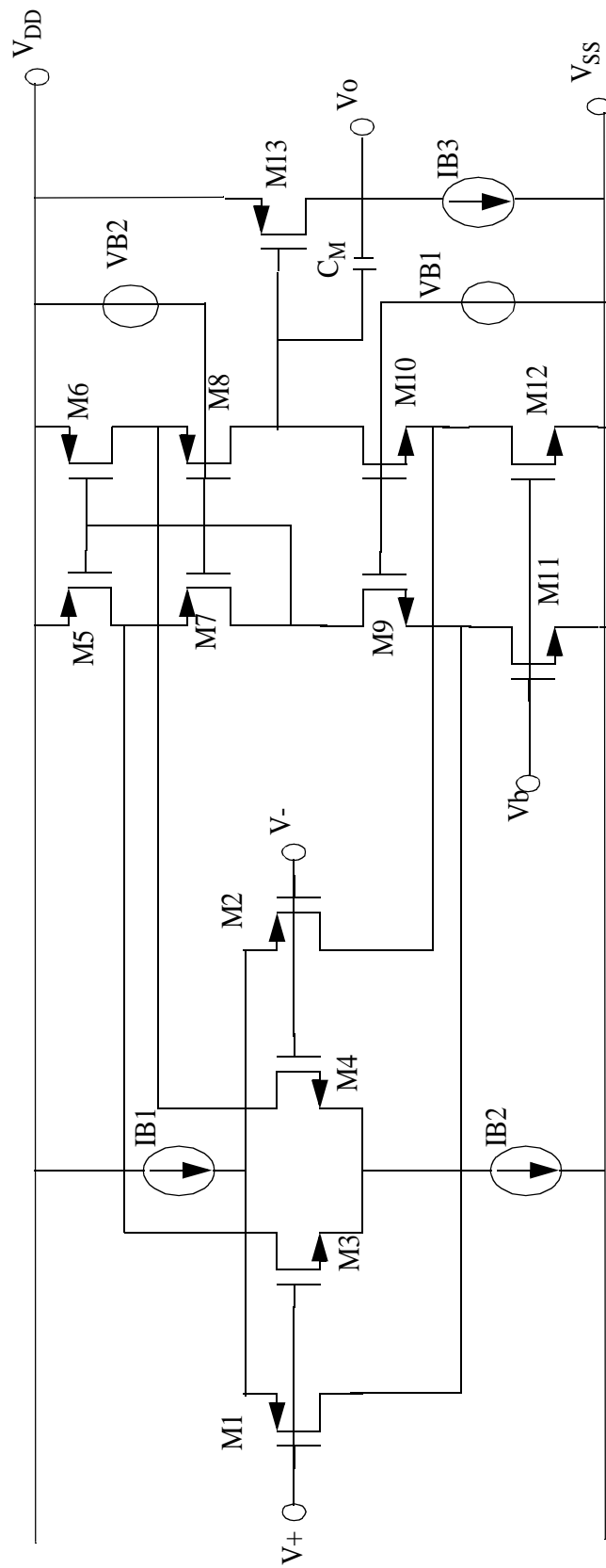


Figure 3-8 Two stage op-amp with class-A output stage

3.4 Constant g_m rail-to-rail input stages

The rail-to-rail folded cascode input stage, as shown in figure 3-6, can be biased in weak or in strong inversion. If it operates in weak inversion, the total transconductance is given by:

$$g_{mi,weak} = \frac{I_p}{2 \cdot n_p \cdot V_{th}} + \frac{I_n}{2 \cdot n_n \cdot V_{th}} \dots\dots\dots(32)$$

where:

- I_p is the tail current of the p-channel input pair
- I_n is the tail current of the n-channel input pair

If the input stage is biased in strong inversion, then the total g_m is given by:

$$g_{mi,strong} = \sqrt{\mu_p \cdot C_{ox} \cdot \left(\frac{W}{L}\right)_p \cdot I_p} + \sqrt{\mu_n \cdot C_{ox} \cdot \left(\frac{W}{L}\right)_n \cdot I_n} \dots\dots\dots(33)$$

or from a voltage point of view:

$$g_{mi,strong} = \mu_p \cdot C_{ox} \cdot \left(\frac{W}{L}\right)_p \cdot V_{sgp,eff} + \mu_n \cdot C_{ox} \cdot \left(\frac{W}{L}\right)_n \cdot V_{gsn,eff} \dots\dots\dots(34)$$

where $V_{gs,eff}$ is the effective gate-source voltage of an input transistor.

It can be seen that $g_{mi,weak}$ can be controlled by changing the tail currents of the input transistors. $g_{mi,strong}$ can be controlled by either changing the tail currents, the gate-source voltages or the W/L ratios.

3.4.1 Rail-to-rail input stages with current based g_m control

g_m control by one-times current mirror

In weak inversion the g_m of an MOS transistor is proportional to its drain current. This indicates that the g_m of a rail-to-rail input stage operating in weak inversion can be made constant by keeping the sum of the tail currents of the complementary input pairs

constant. Thus for a constant g_m the sum of the tail currents has to obey the following expression:

$$I_p + I_n = I_{ref} \dots \dots \dots (35)$$

where it is assumed that the weak inversion slope factors of both transistor types are equal.

The above mentioned principle is realized in the rail-to-rail input stage as shown in figure 3-9. It consists of complementary input pairs M1-M4, and a summing circuit M5-M10. The g_m control of the input stage is implemented by means of the current switch M13 and the current mirror M14-M15.

If low common mode input voltages are applied to this input stage, the current source I_{ref} , biases the p-channel input pair M1-M2. As a consequence, the p-channel input pair can process the input signal. If the common-mode voltage is now raised by about $(V_{DD}-VB3)$, the current switch M13 takes away a part of the current I_{ref} and feeds it through the current mirror M14-M15, into the n-channel input stage. In this way, the sum of the tail currents of the input pairs is kept equal to I_{ref} . If the common mode input range is further increased the current switch directs the complete current I_{ref} via the current mirror, into the n-channel input pair. The result is a largely constant g_m over the entire common mode input range, as shown in figure 3-10.

As can be observed from equation (32), the transconductance of an input pair depends on the weak inversion slope factor, n . If this factor is different for the n-channel and the p-channel input pair, it will result in a variation of the g_m . This difference in n can be largely compensated by modifying the gain factor of the current mirror.

When the current switch gradually steers the current I_{ref} from the p-channel to the n-channel pair, the offset of the rail-to-rail input stage will change, because the offset of the input pairs tends to be different. In order to maximise the CMRR of the rail-to-rail input stage, this offset change should be spread out over a large part of the common-mode input range. To achieve this, either the W/L ratio of the current switch has to be made small when compared to that of the input transistors, or a resistor can be placed in series with the source of the current switch.

Apart from a constant g_m , the main advantages of the g_m control by a current switch are its small die area and its low power consumption. The g_m control hardly increases the size of the input stage, because the current switch and the current mirrors are small in comparison.

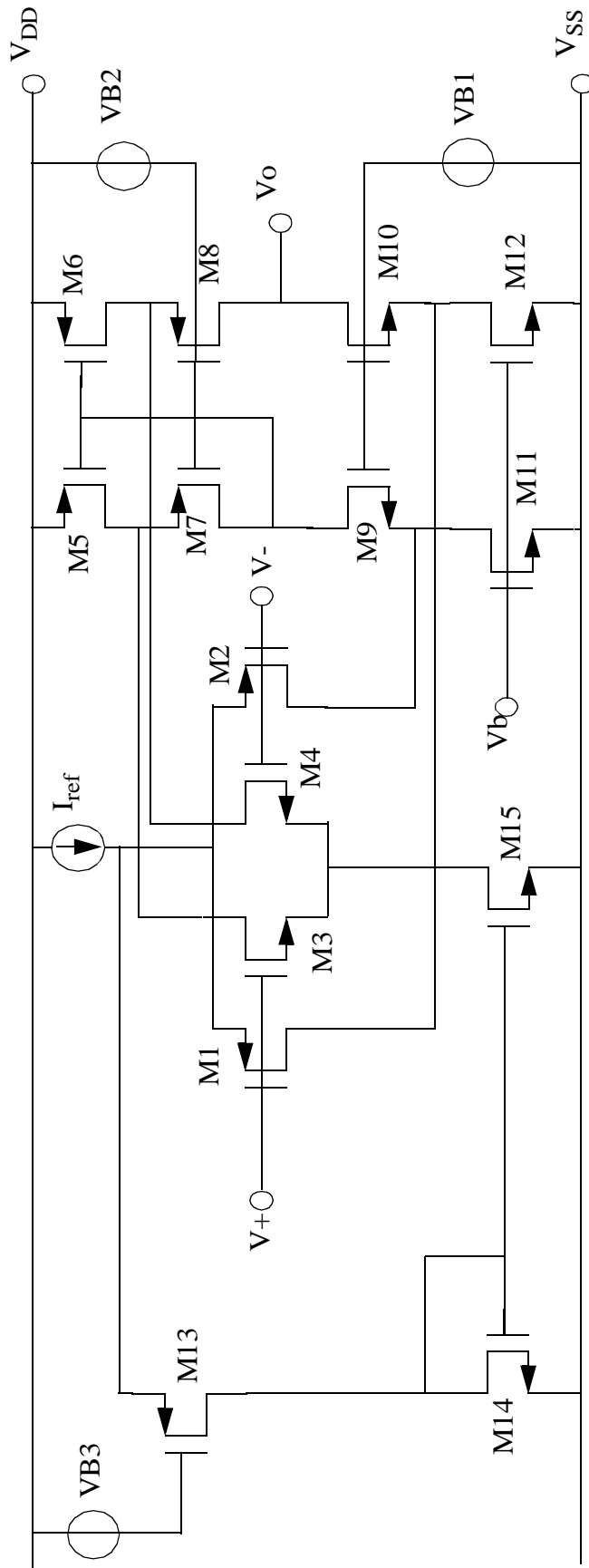


Figure 3-9 Rail-to-Rail Input Stage with g_m control by a current switch and a current mirror

Another advantage is that g_m control does not increase the noise of the input stage. This is because the noise generated in the g_m control circuit is inserted into the tails of the complementary input pairs, and thus can be considered as a common-mode signal. As a consequence, the noise contribution of the g_m control to the rail-to-rail input stage can be neglected, assuming that the input transistors are matched.

If the input stage shown in figure 3-9 is biased in strong inversion, the transconductance of an input pair is proportional to the square root of its tail current. As a result, the g_m displays a variation of about 41% over the common-mode input range, as shown in figure 3-11. Although this variation is less than that of a rail-to-rail input stage without g_m control, it is still too large for a power optimal frequency compensation of the amplifier. Thus for input stages operating in strong inversion, a different g_m control has to be designed.

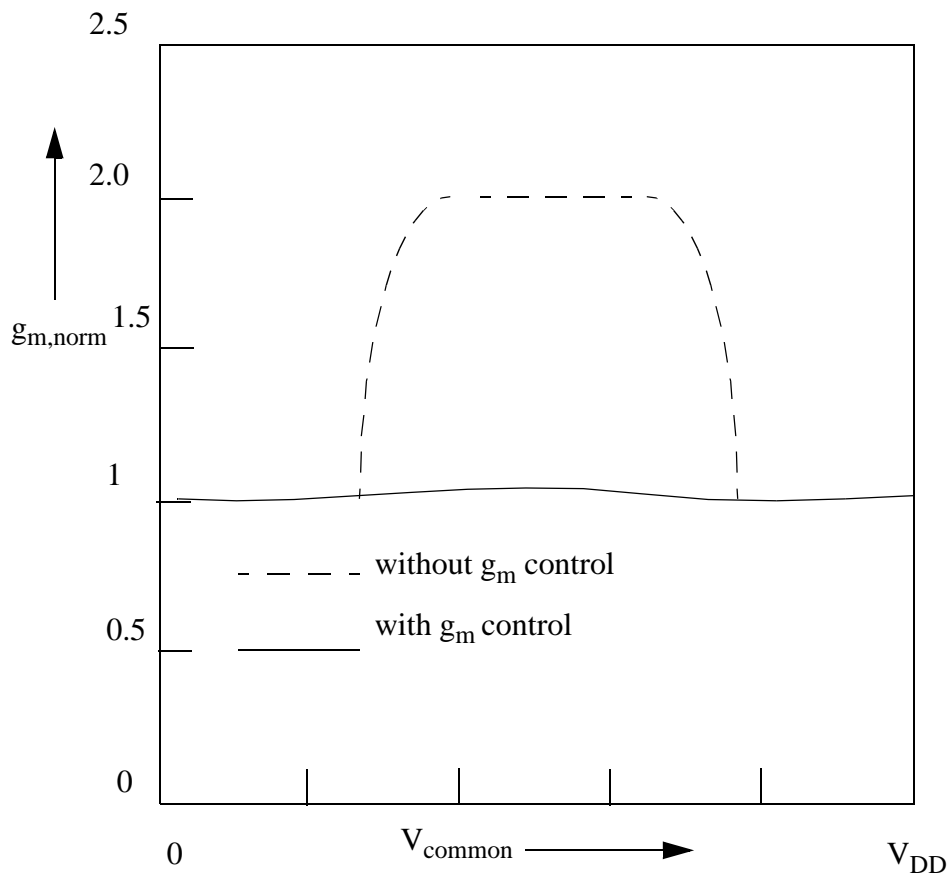


Figure 3-10 Normalized g_m versus the common mode input voltage for the rail-to-rail input stage as shown in figure 3-9. The input pairs are biased in weak inversion.

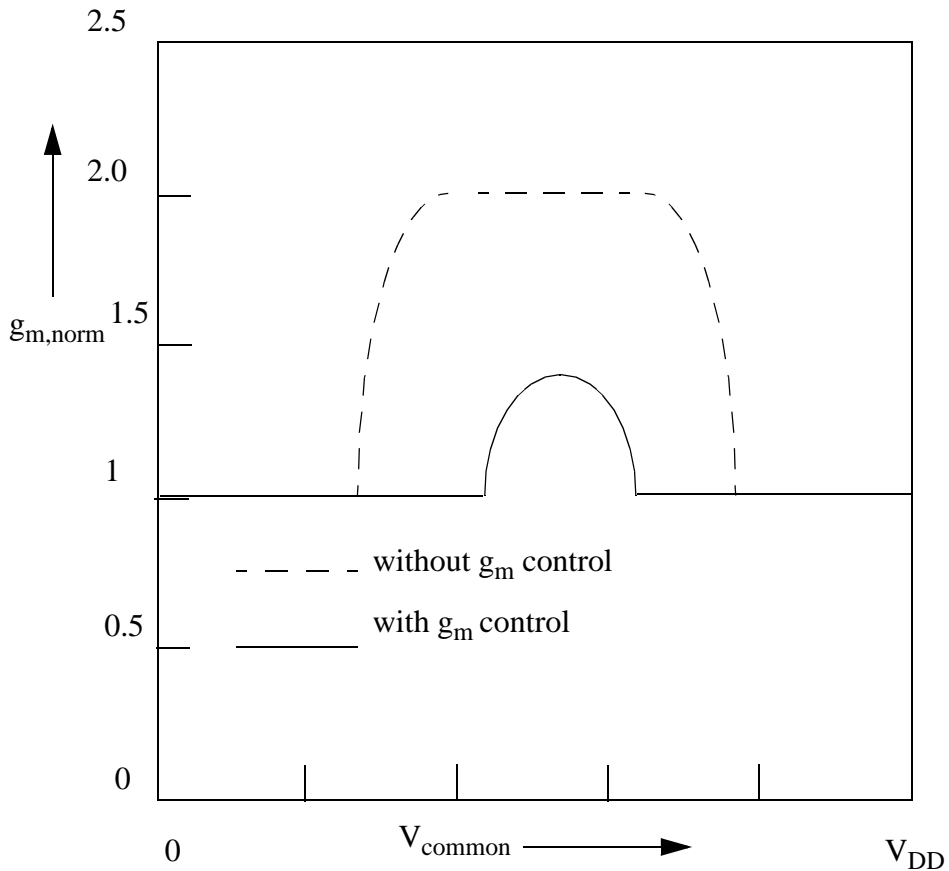


Figure 3-11 Normalized g_m versus the common mode input voltage for the rail-to-rail input stage as shown in figure 3-9. The input pairs are biased in strong inversion.

g_m control by three-times current mirrors

The transconductance of a rail-to-rail input stage operating in strong inversion can be made constant by keeping the sum of the square roots of the tail currents of the complementary input pairs constant, as readily follows from equation (33). This yields:

$$\sqrt{I_p} + \sqrt{I_n} = 2 \cdot \sqrt{I_{ref}} \dots \dots \dots (36)$$

where it is assumed that the W/L ratios of the input transistors obey the condition:

$$\frac{\left(\frac{W}{L}\right)_p}{\left(\frac{W}{L}\right)_n} = \frac{\mu_n}{\mu_p} \dots \dots \dots (37)$$

A brute force implementation of equation (37) is applied to the rail-to-rail input stage as shown in figure 3-12. The input stage consists of a rail-to-rail input stage M1-M4

and a folded cascoded summing circuit, M5-M10. The g_m of this input stage is regulated by means of two current switches, M13 and M16, and two current mirrors, M14-M15 and M17-M18, each with a gain of three. For the sake of simplicity, these current mirrors will be called three times current mirrors, in the remaining part of this section. The sizes of the input transistors are chosen such that relation (37) is obeyed.

In the intermediate part of the common mode input range both current switches are off. The result is that the complementary input pairs are biased with a current of I_{ref} and thus the tail currents obey expression (36) in this part of the common-mode input range.

If the common-mode input voltage decreases below $VB4$, the current switch M16 takes away the tail current of the n-channel input pair and feeds it into the current mirror M17-M18. Here it is multiplied by a factor of three and added to the tail current of the p-channel input pair. The result is that the tail current of the p-channel input pair is equal to $4I_{ref}$. Since the tail current of the n-channel input pair is zero in this part of the common-mode input range, equation (36) is fulfilled. Similarly, it can be explained that, for large common mode input voltages, the g_m control regulates the tail current of the n-channel input pair at a value of $4I_{ref}$. As a consequence, the tail currents again comply with expression (36).

From the above, it follows that the transconductance of the rail-to-rail input stage with g_m control by three times current mirrors has the same value for each part of the common-mode input range. This transconductance is given by:

$$g_{m,r-r} = \sqrt{2 \cdot \mu \cdot C_{ox} \cdot \frac{W}{L} \cdot I_{ref}} \dots \dots \dots (38)$$

Figure 3-13 shows the normalised transconductance of the rail-to-rail input stage as shown in figure 3-12 versus the common-mode input voltage. In this figure the solid line depicts the transconductance of the input stage with g_m control, while the dashed line shows the transconductance of the rail-to-rail input stage as if it was biased by two tail current sources, each with a value of I_{ref} . From figure 3-13 it can be concluded that the transconductance of the input stage with g_m control by two three-times current mirrors is nearly constant over the entire common-mode input range, except for two take-over regions where it varies only 15%. In the take-over regions, one of the current switches gradually steers the tail current out of one of the input pairs to the other.

The main disadvantage of the rail-to-rail input stage with g_m control by three times current mirrors is that both current switches can conduct at very low supply voltage. As a consequence, the current switches together with the three times current mirrors form a positive feedback loop with a gain larger than one. Obviously this cannot be tolerated. The positive feedback loop can be avoided by, for instance, ensuring that the current switch M13 is always turned off at these very low supply voltages. This can be noted that when M13 is always off at a very low supply voltage, the g_m control does not obey equation (36). As a consequence, the g_m is not constant at these supply voltages.

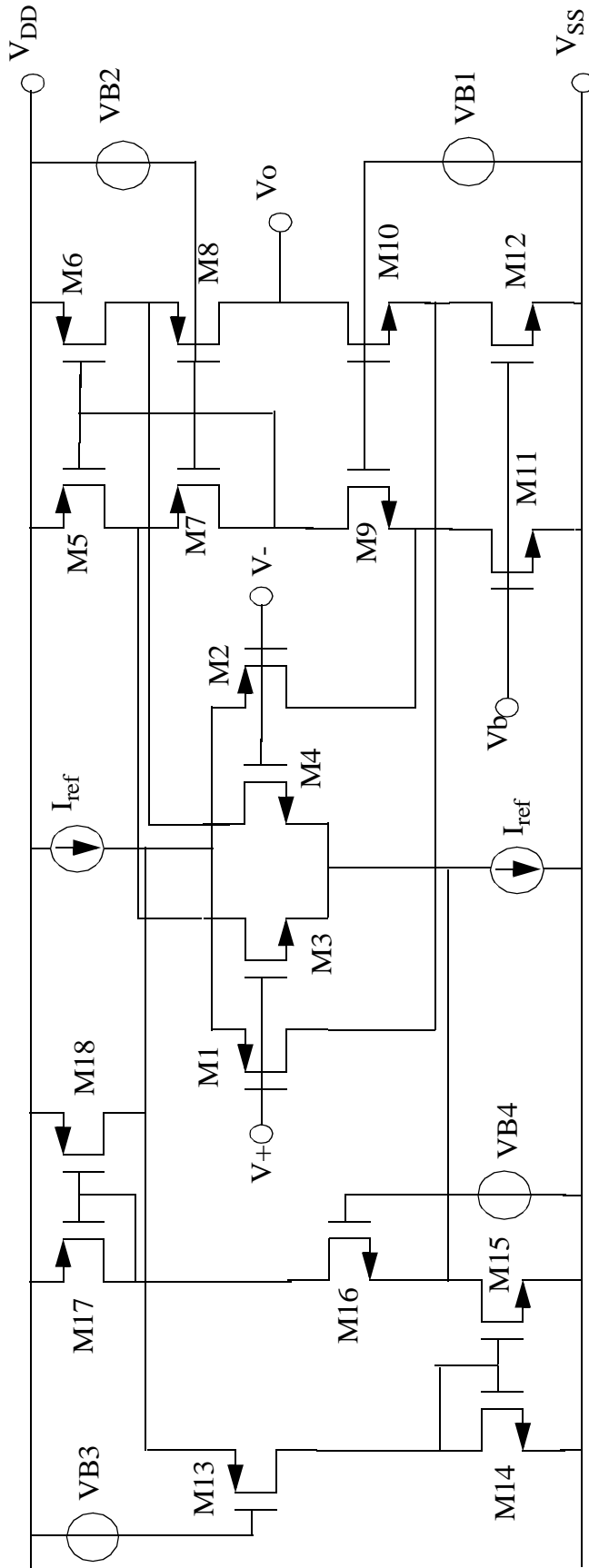


Figure 3-12 Rail-to-Rail input stage with g_m control by two three-times current mirrors

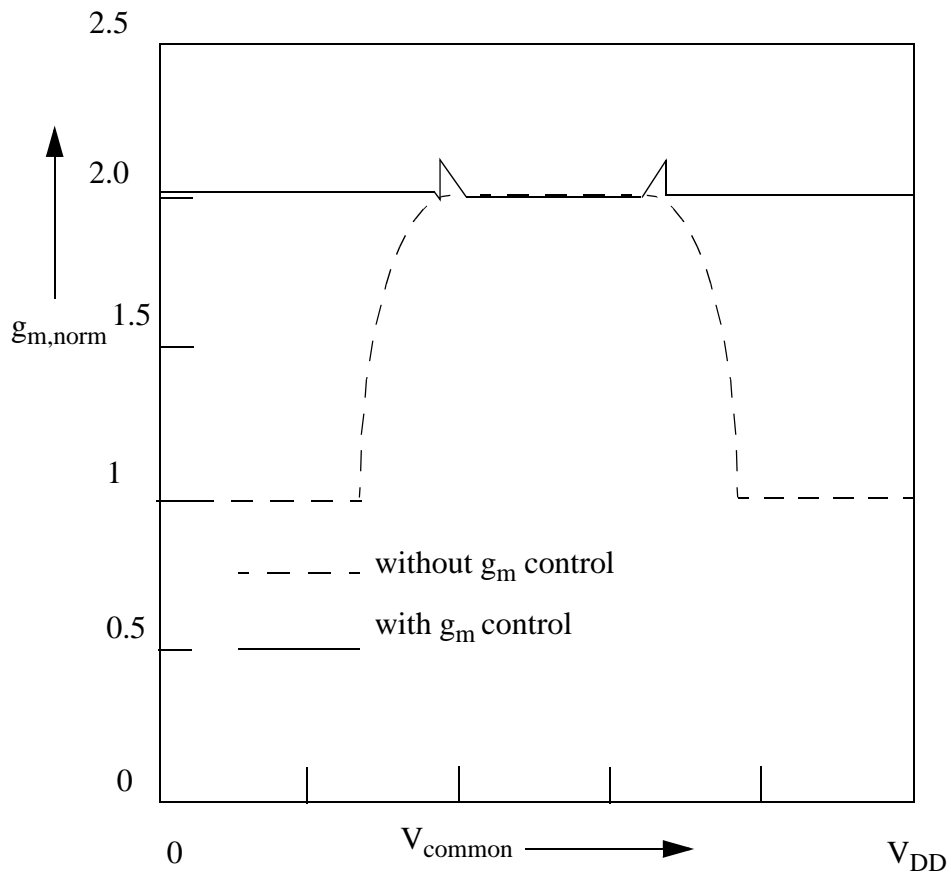


Figure 3-13 Normalised g_m versus the common mode input voltage for the rail-to-rail input stage as shown in figure 3-12. The input pairs are biased in strong inversion.

Other g_m control strategies

- g_m control by square root current control
- voltage based g_m control
- W/L based g_m control

Chapter 4

Output stage design of low-voltage, low power operational amplifiers

4.1 Introduction

The main purpose of the output stage of an operational amplifier is to deliver a certain amount of signal power into a load with acceptably low levels of signal distortion.

In a low-voltage, low-power environment, this has to be achieved by efficiently using the supply voltage as well as the supply current.

To implement this the output voltage range must be as large as possible, preferably from rail-to-rail. To achieve this the output transistors have to be connected in common source configuration.

An efficient use of the supply current requires a high ratio between the maximum signal current that can be delivered to a load, and the quiescent current of the output stage. To accomplish this the transistors have to be class AB biased.

When designing a low voltage output stage, it is of prime importance to know:

- What is the minimum supply voltage at which an output stage is able to operate. This will be determined by the gate-source voltage of the output transistors, which can become large, particularly when the output stage has to drive large output signal currents.

Two classes of class AB control circuits will be discussed:

- Feedforward class AB control is suitable for operation in a low-voltage environment.
- Feedback class AB control is suitable in op-amps which are part of a system that operates under extremely low voltage conditions

4.2 Common-Source Output Voltage

Common Source Output Stage

The most rudimentary output stage that can be used in a low voltage op-amp is a common source connected transistor as shown in figure 4-1.

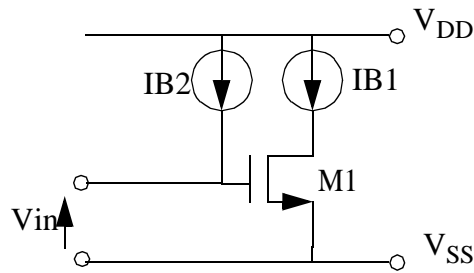


Figure 4-1 Common-Source Output Stage

The minimum required supply voltage for this output stage is given by:

$$V_{sup, min} = V_{gso} + V_{dsat}$$

where:

V_{gso} is the gate source voltage of the output transistor M1 and V_{dsat} is the voltage across the current source $IB2$, which is necessary to have a current flowing out of it.

The minimum supply voltage can be kept low by making the gate-source voltage of the output transistor M1 and the saturation voltage of the current source $IB2$ small.

1. The gate source voltage of the output transistor can be reduced by minimising both the threshold and the effective gate-source voltage. The latter can be kept small by maximising the W/L ratio of the output transistor and by minimising the maximum required output current (in most cases this maximum output current is predetermined by the application for which the amplifier is intended, and this cannot be lowered by the designer).
2. In practical amplifiers the saturation voltage of the current source $IB2$ varies between 100mV for a simple current source biased in weak inversion, and 500mV for a cascoded current source operating in strong inversion.
3. The maximum gate-source voltage of the output transistor varies between 1V, for output stages designed for mediocre output currents and 2V for output stages that have to deliver very large output currents, assuming that the output transistors have a threshold voltage of about 0.8V.
4. As a result the output stage can run on supply voltages between 1.1V and 2.5V, depending on the implementation of the current source and the driving capability of the output stage.

5. In the operational amplifier design practice a push-pull stage, as shown in figure 4-2, is often used as an output stage.

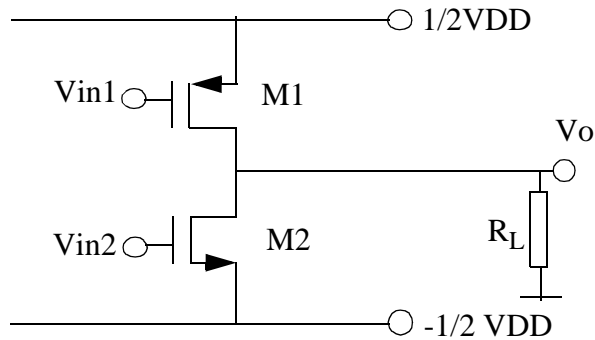


Figure 4-2 Rail-to-Rail Push-Pull output stage

It consists of two complementary common-source connected transistors M1-M2, allowing a rail-to-rail output voltage range. The output transistors are driven by two in-phase signal voltages. If these input signal voltages are above their DC value, the drain current of the n-channel output transistor will be larger than that of the p-channel output transistor, and thus the output stage pulls a current from the load. Similarly, if the input signal voltages are below their DC value, the output stage pushes a current into the load.

In order to determine the output voltage range of the rail-to-rail push-pull output stage, first suppose that the signal voltage is increasing. As a result, the output stage pulls a progressively increasing current from the load, and thus the output voltage decreases. The output voltage continues to decrease until the n-channel output transistor is no longer in the saturation region (the region where I_d changes with V_{gs} only) and ends up in its triode region (the region where I_d changes with V_{ds} and V_{gs}), and the output voltage is limited. Similarly, it can be explained that the p-channel output transistor ends up in the triode region, when the input signal voltage decreases.

The drain current of a transistor operating in its triode region is given by:

$$I_d = \frac{1}{2} \cdot \beta \cdot ((2 \cdot V_{gs, eff} \cdot V_{ds}) - V_{ds}^2) \dots \dots \dots (39)$$

Using this equation and neglecting the term V_{ds} squared the output voltage swing can be estimated by equation (40) as follows::

$$-\frac{1}{2} \cdot V_{DD} \cdot \left(1 - \frac{1}{\beta_n \cdot V_{gsn, eff} \cdot R_L + 1}\right) < V_o < \frac{1}{2} \cdot V_{DD} \cdot \left(1 - \frac{1}{\beta_p \cdot V_{gsp, eff} \cdot R_L + 1}\right)$$

From this equation it can be concluded that the output voltage swing can be optimised by maximising the gate-voltage swing as well as the transconductance factor of the output transistors. The latter can be achieved by choosing the largest possible W/L ratio of the output transistors. As an example, suppose that R_L is $10\text{K}\Omega$, b_p is $7.5\text{mA}/\text{V}^2$, b_n is $7.5\text{mA}/\text{V}^2$, and both output devices have a $V_{gs,eff}$ of 200mV . The latter value corresponds to the maximum effective gate-source voltage of a 1V output stage, designed with devices having a threshold voltage of 0.8V . Using this data, it can be calculated that the output voltage of the common source push-pull output stage can reach both supply rails within 31mV , which almost corresponds to a rail-to-rail output swing. The output voltage swing reduces when the output voltage stage is loaded with a small resistive load. For instance, if the same output stage has to drive a resistor of $1\text{K}\Omega$, its output voltage can only reach the supply rails within 200mV .

4.3 Class AB output stages

In order to efficiently use the power supply, an output stage should combine high maximum output current with a low quiescent current. To fulfil this requirement class-B biasing can be used, because an output stage equipped with this type of biasing unites a large output current with a quiescent current which is approximately zero.

In order to determine the power efficiency of a class B output stage, the following definition can be used. The power efficiency of an output stage is equal to the average signal power divided by the power drawn from the supplies. Using this definition, it can be calculated that the power efficiency of a rail-to-rail class B output stage is about 75% for a rail-to-rail output sine wave.

A drawback of class B biasing is that it introduces a large cross-over distortion. To minimize this distortion, class A biasing can be used. However, the maximum output current of a class A biased output stage is equal to its quiescent current, which leads to power efficiency of only 25% for a rail-to-rail output sine wave. Thus from a power point of view class A biasing is highly undesirable.

To achieve a good compromise between distortion and quiescent dissipation, the output stage has to be biased between class A and class B. The solution to this is called class AB biasing. Figure 4-3 shows the desired class AB transfer function.

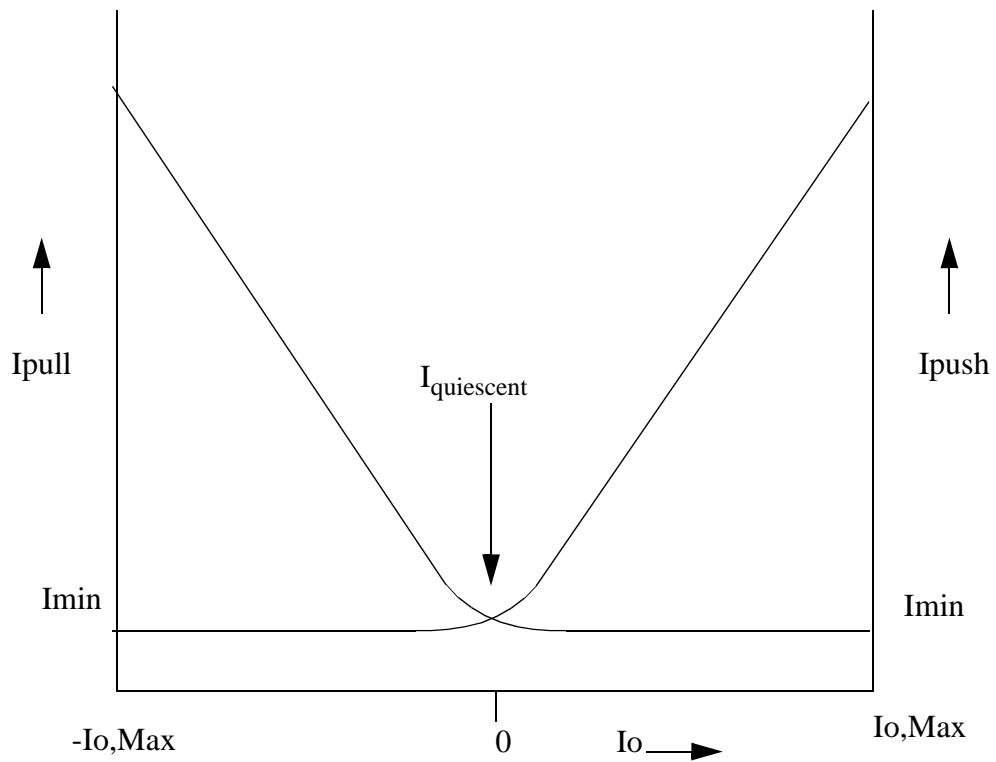


Figure 4-3 Desired class-AB transfer function

As can be readily seen, the output transistors are biased at a small quiescent current, which decreases the crossover distortion compared to that of class B biased output transistors. The maximum output current of the output stage is much larger than its quiescent current, which increases its efficiency compared to that of a class A biased output stage. Figure 4-3 also shows that the output transistor which is not delivering the output current is biased with a small current I_{min} . This minimum current prevents a turn-on delay of the non-active output transistor, which in turn reduces the crossover distortion.

In a rail-to-rail output stage, the class AB transfer function can be realized by keeping the voltage between the gates of the output transistors constant. This principle is shown in figure 4-4a. In order to make the relation between the push and the pull currents of the output transistors insensitive to supply voltage and process variations, the voltage source V_{ab} has to track these parameters. To achieve this it can be modelled by the circuit shown in figure 4-4b. In this schematic, two diodes M3-M4 are biased by a constant current I_{ref} and two constant voltage sources each with a value of $V_{sup}/2$, replace the function of the source V_{ab} . As a result, the relation between the push current, I_{d1} , and the pull current I_{d2} of the output transistors is given by:

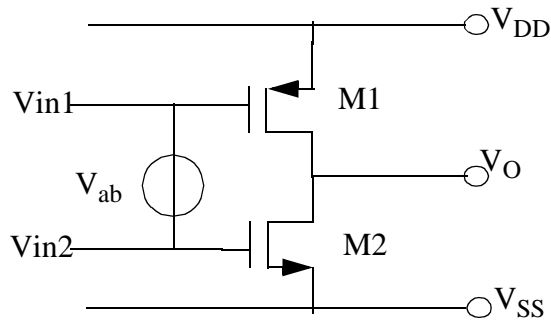


Figure 4-4a Basic Principle of a class-AB control circuit

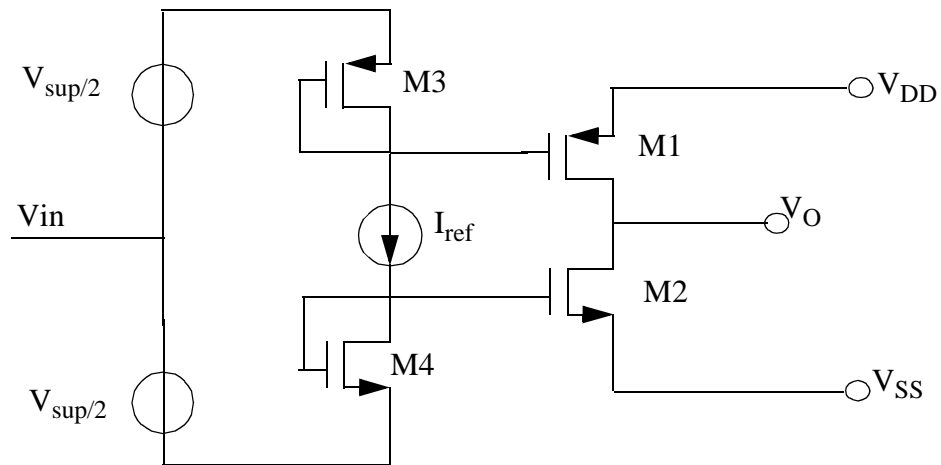


Figure 4-4b Class-AB control circuit with insensitivity to supply voltage and process variations

$$\sqrt{I_{push}} + \sqrt{I_{pull}} = 2 \cdot \sqrt{I_q} \dots \dots \dots (41)$$

where it is assumed that the output transistors operate in strong inversion, and their transconductance factors obey:

$$K \cdot \frac{W}{L} = \frac{1}{2} \cdot \mu_n \cdot C_{ox} \cdot \left(\frac{W}{L}\right)_n = \frac{1}{2} \cdot \mu_p \cdot C_{ox} \cdot \left(\frac{W}{L}\right)_p \dots \dots \dots (42)$$

The quiescent current, I_q , of the output stages is given by:

$$I_q = \frac{\left(\frac{W}{L}\right)_1}{\left(\frac{W}{L}\right)_3} \cdot I_{ref} \dots \dots \dots (43)$$

which is insensitive to process and supply voltage variations.

From equation (41) it can be concluded that the current through the output transistor which conducts the lowest current, slowly reduces to zero when the other output transistor approaches a value of four times the quiescent current. Note that according to figure 4-3, the non-active output transistor should preferably conduct a minimum current larger than zero, in order to minimise turn-on delay of the non-active output transistor.

The maximum output current of the class AB output stage is determined by the allowable gate voltage drive of the output transistors. The gate voltages of the output transistors, as shown in figure 4-4a and 4-4b are capable of extending beyond the supply rails. In practice this is of course not feasible, because the gate voltage will be limited by the driving circuit of the output stage and the supply voltage. A well designed class AB circuit should not substantially further limit the maximum allowable gate-source voltage of an output transistor.

In weak inversion, the circuit as shown in figure 4-4b realizes a relation between the push and pull current which resembles the well known bipolar class AB relation, i.e the product of the currents which flow through the output transistors is constant. This yields:

$$I_{push} \cdot I_{pull} = I_q^2 \dots \dots \dots (44)$$

where it is assumed that both transistor types have the same weak inversion slope factor. The quiescent current of an output stage biased in weak inversion is also given by (43).

The next step is to replace the function of the voltage source by an actual circuit implementation. When designing these real life class AB circuits for low voltage, low power environment, the most important design parameters are inevitably the minimum required supply voltage and the quiescent current. In a low voltage environment the output stage should still be able to run on a supply voltage of two stacked gate-source voltages and two saturation voltages, while for extremely low voltage applications only one saturation voltage on top of a gate-source voltage can be allowed. In order to meet the low power condition the class AB control circuit should not raise the quiescent current of the output stage too much.

Other requirements of a class AB output stage are:

- A sufficiently large maximum output current
- A good high frequency performance
- A small die area

- A quiescent current which has to have a low sensitivity for process and supply voltage variations.

In the next section, several class AB control circuits will be discussed with respect to these specifications. Two main types of control circuits will be distinguished, the feedforward class AB control for use in low voltage op amps, and the feedback class AB control for use in extremely low voltage amplifiers.

4.4 Feedforward class AB output stages

In the previous section it was shown that class AB biasing of an output stage can be achieved by setting the voltage between the gates of the output transistors. A straight-forward implementation of this principle is shown in figure 4-5.

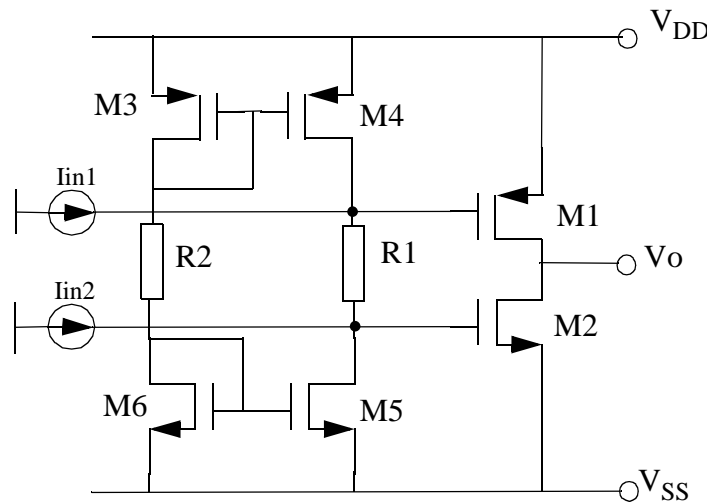


Figure 4-5 Rail-to-rail output stage with resistive feedforward class AB control

This circuit contains a rail-to-rail output stage M1-M2, and a resistance coupled class AB control, R1-R2 and M3-M6. The output stage is driven by two in-phase signal currents I_{in1} and I_{in2} . The diode connected transistors M3 and M6 together with R2 build up a reference chain which generates a bias current I_{ref} . This current is copied by current mirrors M3-M4 and M5-M6 and fed into resistor R1. This resistor in turn sets the voltage between the gates of the output transistors.

Figure 4-6 depicts the push and pull current as a function of the transistors. In order to make this function independent of the supply voltage the resistors R1 and R2 must have the same value. Using this assumption, the relation between the push I_{d1} and the pull I_{d2} current can be described by:

$$\sqrt{I_{push}} + \sqrt{I_{pull}} = 2 \cdot \sqrt{I_q} \dots \dots \dots (45)$$

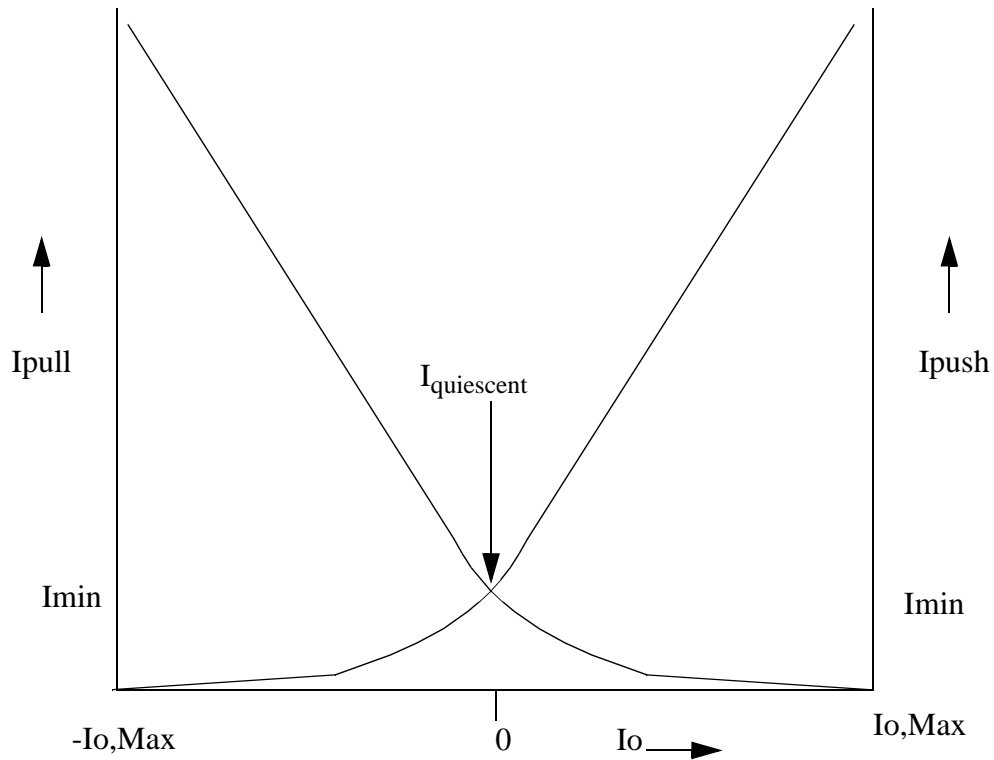


Figure 4-6 Push and pull current versus the output current for the rail-to-rail output stage with resistive class-AB control

in which the quiescent current I_q is given by:

$$I_q = \frac{\left(\frac{W}{L}\right)_1}{\left(\frac{W}{L}\right)_2} \cdot I_{ref} \dots \dots \dots (46)$$

Please note that the class AB relation of the resistive class AB controlled output stage is identical to that of the basic class AB output stage, as discussed previously.

The maximum current of this output stage is limited to mediocre values, because the gates of the output transistors can only reach the supply rails within one saturation voltage and the DC voltage across the resistor R1. For example with the following process parameters: $\mu_n = 440 \text{cm}^2/\text{V}$, $\mu_p = 147 \text{cm}^2/\text{V}$, $C_{ox} = 1.77 \times 10^{-3} \text{F/m}^2$, $V_{TP} = -0.8 \text{V}$, $V_{TN} = 0.8 \text{V}$, $\Theta = 0.1$, $\xi = 0.3 \text{mm/V}$, $V_{sup} = 3 \text{V}$, $V_{dsat4} = 200 \text{mV}$, $V_{dsat5} = 200 \text{mV}$, $V_{RI} = 1.1 \text{V}$, $W1 = 200 \mu\text{m}$, $L1 = 2 \mu\text{m}$, $W2 = 600 \mu\text{m}$, and $L2 = 2 \mu\text{m}$.

We previously discussed that:

$$I_D = \frac{1}{2} \cdot \frac{\mu C_{ox}}{1 + (V_{gs} - V_T) \cdot \left(Q - \frac{x}{L}\right)} \cdot \frac{W}{L} \cdot (V_{gs} - V_{gt})^2 \dots\dots\dots(47)$$

Using this equation (47), it can be calculated that the maximum current for the design example has a value of about 2.5mA, which is about 30 times the quiescent current.

The minimum supply of this class AB output stage can be as low as two stacked gate-source voltages and one saturation voltage, which makes class AB control suitable for operation under low voltage conditions.

The main advantage of this circuit is that the class AB hardly increases the quiescent current of the output stage because the current through the reference chain can be relatively low. To achieve this the resistors have to be large, which deteriorates the high frequency behaviour of the class AB control. This problem can be overcome by inserting a capacitor parallel to the class AB resistor R1.

The main disadvantage of this circuit is that it is sensitive to supply voltage variations. If the supply voltage increases, the current in the reference chain increases, and so will the quiescent current in the output stage.

Another disadvantage is that this circuit can occupy considerable die area. This is mainly due to the resistors which have, in practical cases, values between 10KΩ and 100KΩ.

To overcome these problems, the transistor coupled feedforward class AB control as shown in figure 4-7 can be used.

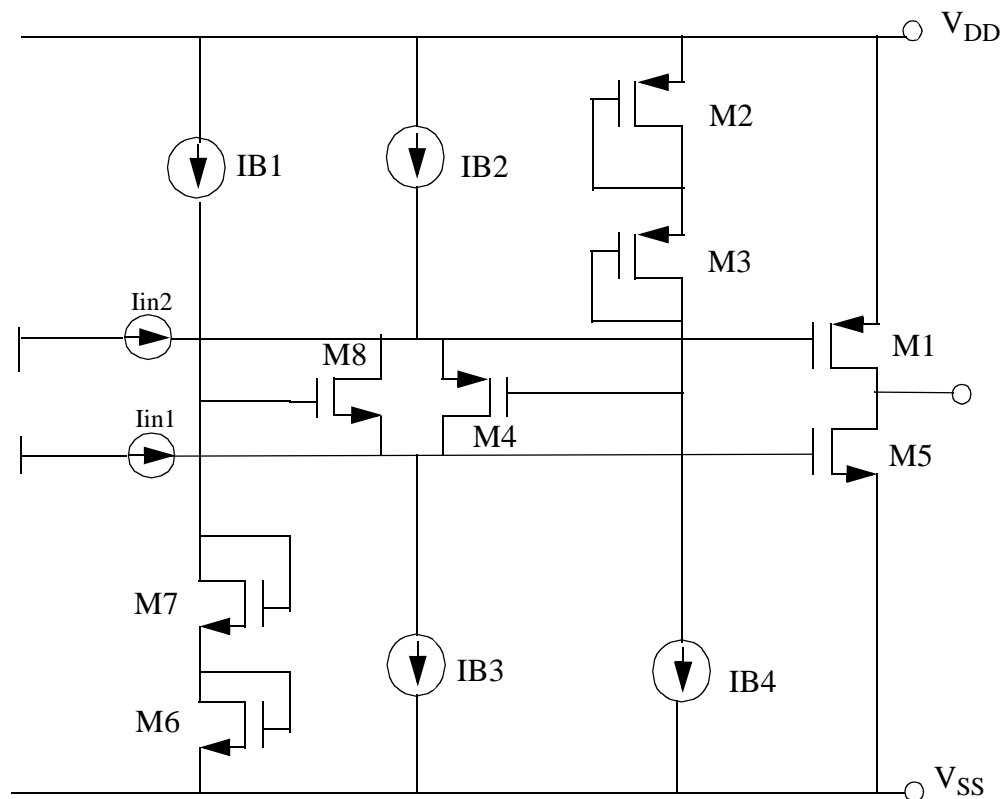


Figure 4-7 Rail-to-rail output stage with transistor coupled feedforward class AB control

The circuit consists of a rail-to-rail output stage M1 and M5, and a class AB control circuit M4 and M8. Since the class AB control consists only of transistors, it occupies potentially less die area than the resistive class AB control.

The class AB control sets up two translinear loops M1 through M4 and M5 through M8, which fix the voltage between the gates of the output transistors, in accordance to the principle shown in figure 4-4a.

The resulting behaviour of the push, I_{d1} and the pull I_{d5} is shown in figure 4-8. In quiescence the current $IB2$ is equally divided into M4 and M8. To compensate for the body effect, M7-M8 and M3-M4 are biased at the same gate-source voltage and thus M5-M6 as well as M1-M2 also have equal gate-source voltage. Now it is calculated that the quiescent current, I_q , in the output transistors is given by:

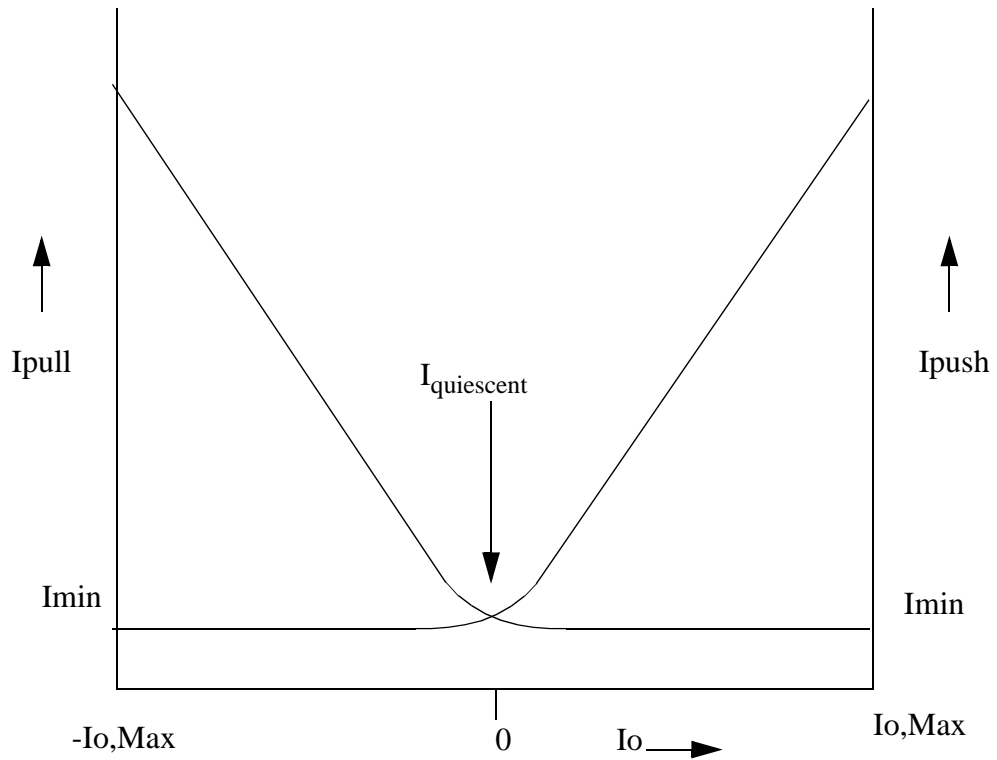


Figure 4-8 Push and Pull current versus the output current for the rail-to-rail output stage with transistor coupled feedforward class-AB control

$$I_q = \frac{\left(\frac{W}{L}\right)_5}{\left(\frac{W}{L}\right)_6} \cdot IB1 \dots \dots \dots (48)$$

where it is assumed that $IB1$ and $IB4$ have the same value and that transistor sizes obey:

$$\frac{\left(\frac{W}{L}\right)_5}{\left(\frac{W}{L}\right)_1} = \frac{\left(\frac{W}{L}\right)_6}{\left(\frac{W}{L}\right)_2} = \frac{\left(\frac{W}{L}\right)_7}{\left(\frac{W}{L}\right)_3} = \frac{\left(\frac{W}{L}\right)_8}{\left(\frac{W}{L}\right)_4} \dots \dots \dots (49)$$

If the output stage operates in strong inversion, then the relation between the push and the pull current can be described by:

$$\left(\sqrt{I_{push}} - (\alpha \cdot \sqrt{I_q})\right)^2 + \left(\sqrt{I_{pull}} - (\alpha \cdot \sqrt{I_q})\right)^2 = 2 \cdot \left(\frac{L}{W}\right)_7 \cdot \left(\frac{W}{L}\right)_6 \cdot I_q \dots \dots \dots (50)$$

where:

$$\alpha = 1 + \sqrt{\left(\frac{W}{L}\right)_6 \cdot \left(\frac{L}{W}\right)_7} \dots\dots\dots(51)$$

The push and pull current obey relation (50) until either push or pull current exceeds a value of:

$$I_{max} = \alpha^2 \cdot I_q \dots\dots\dots(52)$$

where α is given by (51).

If for example the push current exceeds this value, the complete bias current I_{B2} flows through M8, while M4 is cut off. As a result the current through the output transistor M5 is kept at a minimum value. This minimum value immediately follows from equation (50), and is given by:

$$I_{min} = (\alpha - \sqrt{2} \cdot (\alpha - 1))^2 \cdot I_q \dots\dots\dots(53)$$

In the case that M6 and M7 are the same size, the minimum current will be about $0.34I_q$. The body effect of M3-M4, M7-M8 will slightly increase this value. Similarly, it can be explained that if the push current exceeds the current I_{max} , the current through M1 is kept at a minimum value of I_{min} .

In weak inversion the relation between the push and the pull currents is given by:

$$\frac{I_{push} \cdot I_{pull}}{I_{pull} + I_{push}} = \left(\frac{1}{2} \cdot I_q\right) \dots\dots\dots(54)$$

If either the push or the pull current becomes large, the current through the transistor is kept at a minimum value of $0.5I_q$

In both weak and strong inversion, the class AB control is able to operate until one of the output transistors pushes either M4 or M8 out of saturation. As a result, the gate voltage of the output transistors can reach the supply rail within one saturation voltage and one minimum gate-source voltage, which restricts the output current to mediocre values. Using equation previously used to calculate I_d , a typical value for the maximum current would be about 5mA, which is about two times larger than for the resistive class AB output stage. This is because the gate swing of the output transistor in the transistor coupled class AB control is about 300mV larger.

The minimum supply voltage of the output stage equals two stacked gate-source voltages and one saturation voltage, which makes it suitable for low-voltage operation. Note that the output stage with resistive class AB control requires the same minimum supply voltage.

An advantage of the transistor coupled class AB control is that it hardly increases the dissipation of the output stage, although it consumes slightly more power than the resistance coupled class AB control. In addition a good high frequency behaviour is achieved, because the coupling between the gates is realised by a single transistor. This is particularly advantageous when only one gate of the output transistor is driven, as in this case the signal has only to pass one single class AB transistor in order to drive the other gates.

4.5 Feedback class AB output stage

In the previous section, several implementations of feedforward class AB control circuits have been discussed. The minimum supply voltage of these feedforward class AB control circuits is limited to two stacked gate-source voltages and one saturation voltage, which impedes these control circuits to operate under extremely low-voltage conditions.

The aforementioned limitation of feedforward class AB control can be overcome by using feedback class AB control. In contrast to feedforward control, this type of biasing does not directly control the current of the output stage. Instead, the push and the pull currents are measured first and then regulated in a class AB way. This allows the output to run on extremely low supply voltages.

Figure 4-9 shows a straightforward implementation of a feedback class AB controlled output stage. In this output stage, the current through the output transistors M1-M2 are measured by M7 and M3 respectively.

The measured currents are fed into the resistors R1 and R2. As a result, the voltage across R1 represents the drain current of the n-channel output transistor while the voltage across R2 mimics the drain current through the p-channel output transistor.

If the output stage is at rest, the currents through the output transistors and therefore the voltages across R1 and R2 are equal. As a consequence, the tail current of the decision pair M8-M9 is equally divided over M8 and M9, and thus the common source voltage of the drain pair represents the quiescent current through the output transistors. This common source voltage is compared to a reference voltage, which is set by M12,R3 and $IB1$. If a difference occurs between the reference voltage and the common source voltage of M8 and M9, the feedback amplifier feeds a connection signal to the gates of the output transistors. In this way, the quiescent current in the output stage is set.

In order to obtain a quiescent current which is insensitive to process and temperature variations, the resistor R3 has to match R2 and R1, the current source $IB1$ should have half the value of $IB2$, and the W/L ratio of M12 should be half the W/L of M8 or M9. Using these values, it can be calculated that the quiescent current is given by:

$$\frac{\left(\frac{W}{L}\right)_1}{\left(\frac{W}{L}\right)_7} \cdot \frac{R_3}{R_2} \cdot IB1 \dots \dots \dots (55)$$

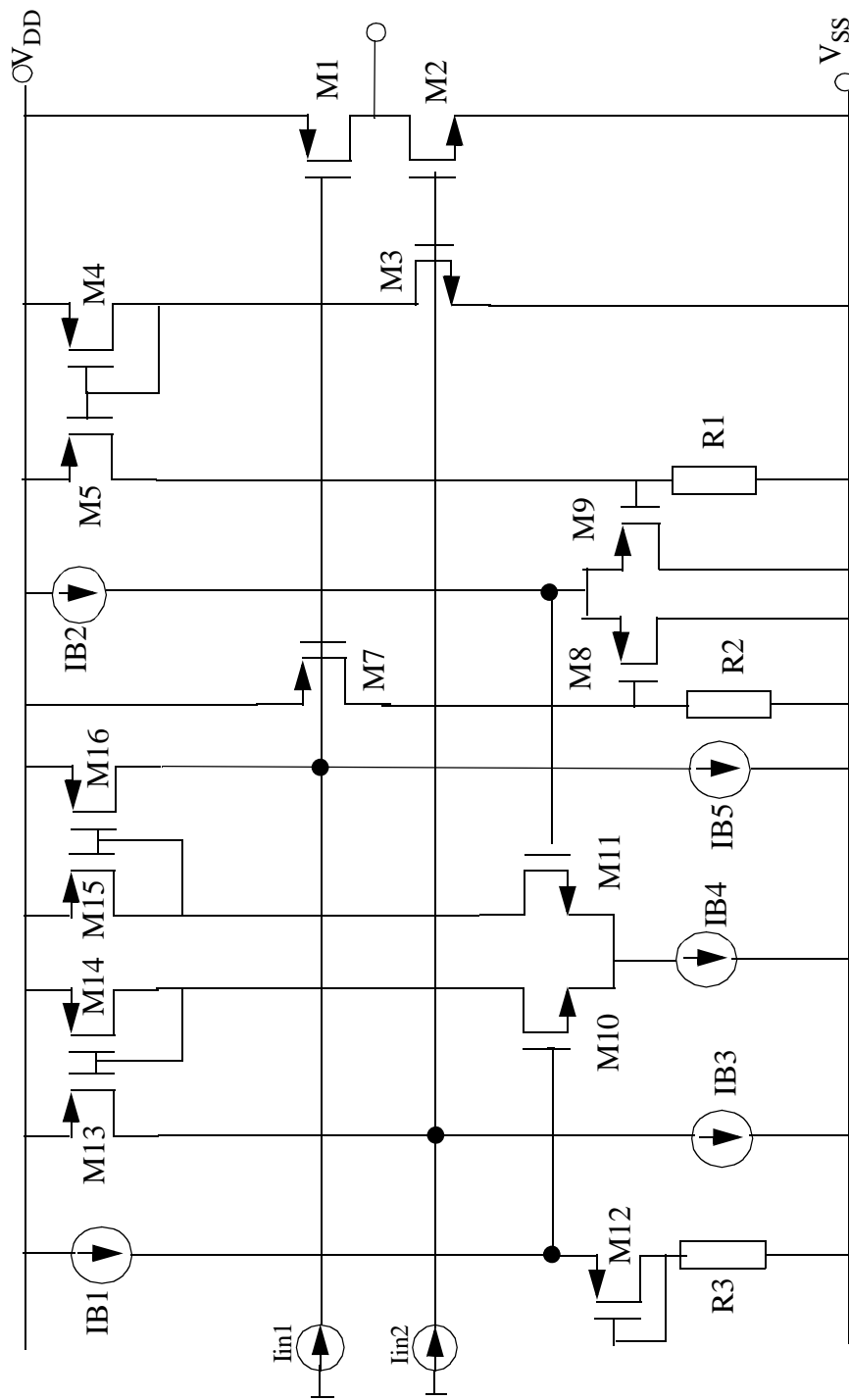


Figure 4-9 Feedback-biased class AB rail-to-rail output stage. The currents through the output transistors are measured by resistors.

It should be noted that the quiescent current slightly depends on the supply voltage because the outputs of the feedback amplifier M10-M15 refer to different supply rails. This problem can be overcome by using cascoded mirrors. However this will increase the minimum supply voltage of the output stage with one saturation voltage.

The class AB control also sets the maximum current of the output transistors, that is the current through the output transistor which is not delivering the output current. Consider for instance that M1 is pushing a large current into the output node, then the voltage across R2 is much larger than the voltage across R1. As a result, the tail current of the decision pair flows completely through M9. Thus the common source voltage of the decision pair represents the minimum current flowing through the output transistor M2. Again if a difference occurs between this voltage and the reference voltage, the feedback amplifier feeds a correction signal to the output stage. In this way, the minimum current of the n-channel output transistor can be controlled. Similarly it can be explained that the class AB control regulates the minimum current of M1, when the output transistor M2 is pulling a large current from the output node. The minimum current of both output transistors is given by:

$$I_{min} = I_q - (\sqrt{2} - 1) \cdot \frac{\left(\frac{W}{L}\right)_1}{\left(\frac{W}{L}\right)_7} \cdot \frac{V_{gs12,eff}}{R_2} \dots\dots\dots(56)$$

As follows from the formula., the minimum current can be controlled accurately because it depends on process parameters and the absolute value of R2. Using typical data ($I_q=75\mu A$, $V_{gs12,eff}$ of 150mV, R2 of 20K Ω and W/L for M1 approximately ten times larger than that of M7) gives a minimum current equal to $0.4I_q$.

Figure 4-10 shows the push and pull current as a function of the output current. In this figure the quiescent and the minimum currents are determined by the above equations.

The maximum output current of the output stage can be large because the gate of the output transistors are able to reach one of the supply rails within one saturation voltage. For example using typical output stage data, it can be calculated that the output stage is able to drive a minimum current of 10mA.

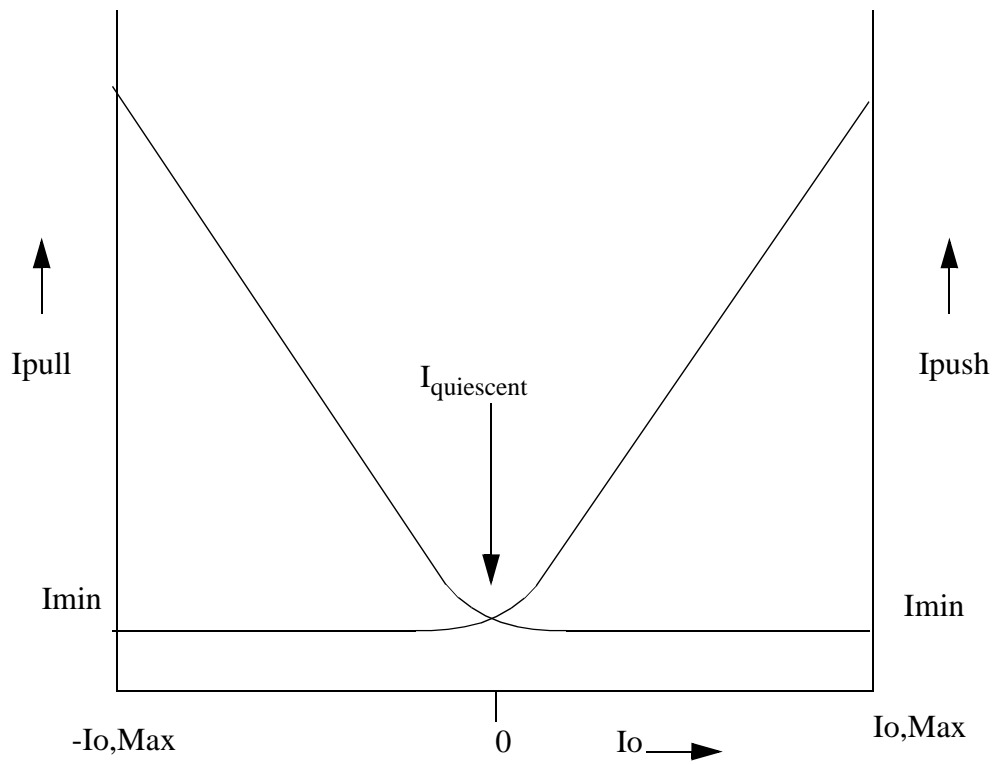


Figure 4-10 Push and Pull currents as a function of the output current for the class AB output stage as shown in figure 4-9.

List of References

1. R. Hogervorst, J.H. Huijsing, Design of Low-Voltage, Low-Power Operational Amplifier Cells, Kluwer Academic Publishers, 1996.
2. R. Cavin, W. Liu, Emerging Technologies: Designing Low Power Digital Systems, IEEE Press, 1996.
3. A. Chandrakasan, R. Brodersen, Low Power CMOS Design, IEEE Press, 1998.
4. E. Sanchez-Sinencio, A. G. Andreou, Low-Voltage/Low-Power Integrated Circuits and Systems, IEEE Press, 1999.