Chapter 1

Introduction to Analog Design

1.1 Why Analog?

It was in the early 1980s that many experts predicted the demise of analog circuits. Digital signal processing algorithms were becoming increasingly more powerful while advances in integrated-circuit (IC) technology provided compact, efficient implementation of these algorithms in silicon. Many functions that had traditionally been realized in analog form were now easily performed in the digital domain, suggesting that, with enough capability in IC fabrication, all processing of signals would eventually occur digitally. The future looked quite bleak to analog designers and they were seeking other jobs.

But, why are analog designers in such great demand today? After all, digital signal processing and IC technologies have advanced tremendously since the early 1980s, making it possible to realize processors containing millions of transistors and performing billions of operations per second. Why did this progress not confirm the earlier predictions?

While many types of signal processing have indeed moved to the digital domain, analog circuits have proved *fundamentally* necessary in many of today's complex, high-performance systems. Let us consider a few applications where it is very difficult or even impossible to replace analog functions with their digital counterparts regardless of advances in technology.

Processing of Natural Signals Naturally occurring signals are analog—at least at a macroscopic level. A high-quality microphone picking up the sound of an orchestra generates a voltage whose amplitude may vary from a few microvolts to hundreds of millivolts. The photocells in a video camera produce a current that is as low as a few electrons per microsecond. A seismographic sensor has an output voltage ranging from a few microvolts for very small vibrations of the earth to hundreds of millivolts for heavy earthquakes. Since all of these signals must eventually undergo extensive processing in the digital domain, we observe that each of these systems consists of an analog-to-digital converter (ADC) and a digital signal processor (DSP) [Fig. 1.1(a)]. The design of ADCs for high speed, high precision, and low power dissipation is one of many difficult challenges in analog design.

In practice, the electrical version of natural signals may be prohibitively small for direct digitization by the ADC. The signals are also often accompanied by unwanted, out-of-band

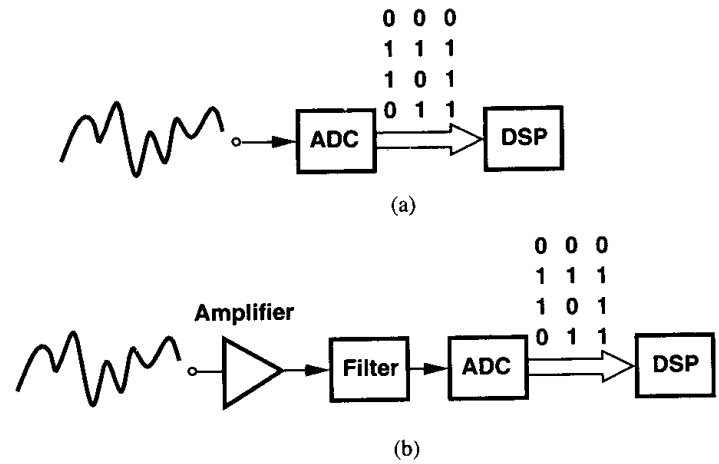


Figure 1.1 (a) Digitization of a natural signal, (b) addition of amplification and filtering for higher sensitivity.

interferers. The front end of Fig. 1.1(a) may therefore be modified as shown in Fig. 1.1(b), where an amplifier boosts the signal level and an analog filter suppresses the out-of-band components. The design of high-performance amplifiers and filters is also a topic of active research today.

Digital Communications Binary data generated by various systems must often be transmitted over long distances. For example, computer networks in large office buildings may transmit the data over cables that are hundreds of meters long.

What happens if a high-speed stream of binary data travels through a long cable? As illustrated in Fig. 1.2, the signal experiences both attenuation and "distortion," no longer resembling a digital waveform. Thus, a receiver similar to that of Fig. 1.1(b) may be necessary here.

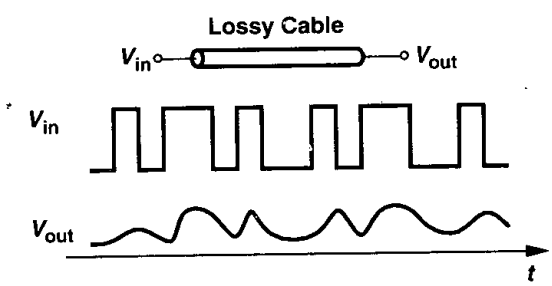


Figure 1.2 Attentuation and distortion of data through a lossy cable.

In order to improve the quality of communication, the above system may incorporate "multi-level"—rather than binary—signals. For example, if, as shown in Fig. 1.3, every two consecutive bits in the sequence are grouped and converted to one of four levels, then

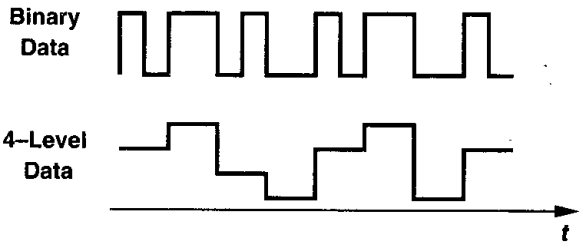


Figure 1.3 Use of multi-level signalling to reduce the required bandwidth.

each level is twice as long as a bit period, demanding only *half* the bandwidth required for transmission of the binary stream. Utilized extensively in today's communication systems, multi-level signals necessitate a digital-to-analog converter (DAC) in the transmitter to produce multiple levels from the grouped binary data and an ADC in the receiver to determine which level has been transmitted. The key point here is that increasing the number of levels relaxes the bandwidth requirements while demanding a higher precision in the DAC and the ADC.

Disk Drive Electronics The data stored magnetically on a computer hard disk is in binary form. However, when the data is read by a magnetic head and converted to an electrical signal, the result appears as shown in Fig. 1.4. The amplitude is only a few millivolts, the noise content is quite high, and the bits experience substantial distortion.

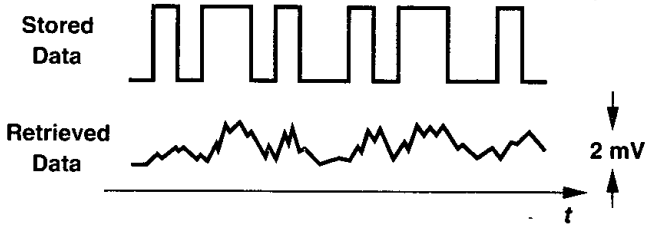


Figure 1.4 Data stored in and retrieved from a hard disk.

Thus, as illustrated in Fig. 1.1, the signal is amplified, filtered, and digitized for further processing. Depending on the overall system architecture, the analog filter in this case may in fact serve to remove a significant portion of the noise and the distortion of the signal. The design of each of these building blocks poses great challenges as the speed of computers and their storage media continues to increase every year. For example, today's disk drives require a speed of 500 Mb/s.

Wireless Receivers The signal picked up by the antenna of a radio-frequency (RF) receiver, e.g., a pager or a cellular telephone, exhibits an amplitude of only a few microvolts and a center frequency of 1 GHz or higher. Furthermore, the signal is accompanied by large

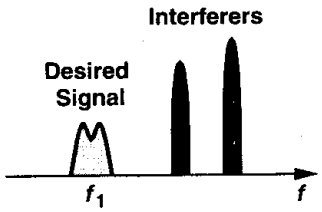


Figure 1.5 Signal and interferers received by the antenna of a wireless receiver.

interferers (Fig. 1.5). The receiver must therefore amplify the low-level signal with minimal noise, operate at a high frequency, and withstand large unwanted components. Note that these requirements are necessary even if the desired signal is not in "analog" form. The trade-offs between noise, frequency of operation, tolerance of interferers, power dissipation, and *cost* constitute the principal challenge in today's wireless industry.

Optical Receivers For transmission of high-speed data over very long distances, cables generally prove inadequate because of their limited bandwidth and considerable attenuation. Thus, as illustrated in Fig. 1.6, the data is converted to light by means of a laser diode and transmitted over an optical fiber, which exhibits an extremely wide band and a very low

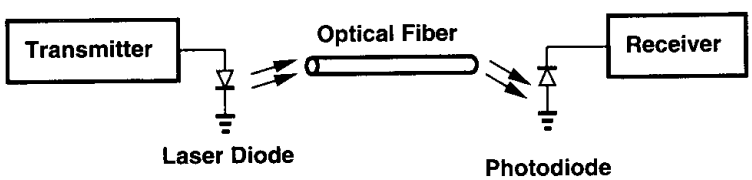


Figure 1.6 Optical fiber system.

loss. At the receive end, the light is converted to a small electrical current by a photodiode. The receiver must then process a low-level signal at a very high speed, requiring low-noise, broadband circuit design. For example, state-of-the-art optical receivers operate in the range of 10 to 40 Gb/s.

Sensors Mechanical, electrical, and optical sensors play a critical role in our lives. For example, video cameras incorporate an array of photodiodes to convert an image to current and ultrasound systems use an acoustic sensor to generate a voltage proportional to the amplitude of the ultrasound waveform. Amplification, filtering, and A/D conversion are essential functions in these applications.

An interesting example of sensors is the accelerometers employed in automobiles to activate air bags. When the vehicle hits an obstacle, the drop in the speed is measured as acceleration and, if exceeding a certain threshold, it triggers the air bag release mechanism. Modern accelerometers are based on a variable capacitor consisting of a fixed plate and a deflectable plate [Fig. 1.7(a)]. The deflection and hence the value of the capacitor are proportional to the acceleration, requiring a circuit that accurately measures the change in capacitance. The design of such interface circuits is quite difficult because for typical

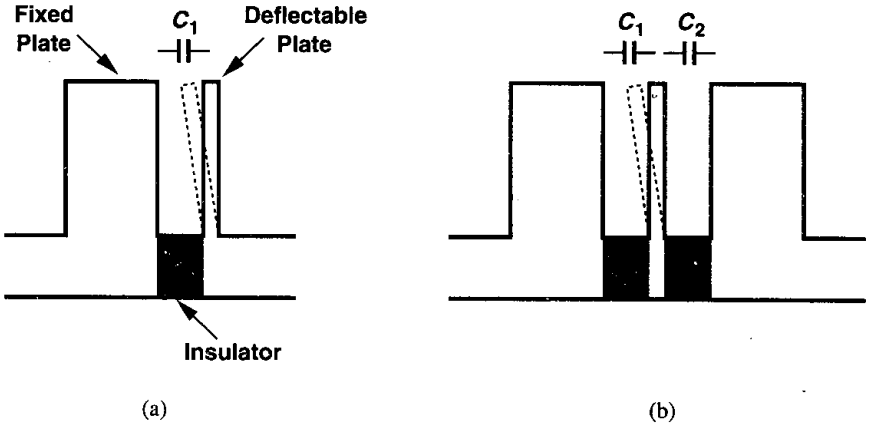


Figure 1.7 (a) Simple accelerometer, (b) differential accelerometer.

accelerations, the interplate capacitance may change by less than 1%, demanding a high precision in the measurement. In practice, the structure of Fig. 1.7(b) is used to provide two capacitors that change in opposite directions, reducing the task to the measurement of the *difference* between two capacitances rather than the absolute value of one.

Microprocessors and Memories Today's microprocessors and memories draw upon a great deal of analog design expertise. Many issues related to the distribution and timing of data and clocks across a large chip or among chips mandate that high-speed signals be viewed as analog waveforms. Furthermore, nonidealities in signal and power interconnects on the chip as well as package parasitics require a solid understanding of analog design. In addition, semiconductor memories employ high-speed "sense amplifiers" extensively, necessitating many analog techniques. For these reasons, it is often said "high-speed digital design is in fact analog design."

The foregoing applications demonstrate the wide and inevitable spread of analog circuits in modern industry. But, why is analog design difficult? We make the following observations. (1) Whereas digital circuits entail primarily one trade-off between speed and power dissipation, analog design must deal with a multi-dimensional trade-off consisting of speed, power dissipation, gain, precision, supply voltage, etc. (2) With the speed and precision required in processing analog signals, analog circuits are much more sensitive to noise, crosstalk, and other interferers than are digital circuits. (3) Second-order effects in devices influence the performance of analog circuits much more heavily than that of digital circuits. (4) The design of high-performance analog circuits can rarely be automated, usually requiring that every device be "hand-crafted." By contrast, many digital circuits are automatically synthesized and laid out. (5) Despite tremendous progress, modeling and simulation of many effects in analog circuits continue to pose difficulties, forcing the designers to draw upon experience and intuition when analyzing the results of a simulation. (6) An important thrust in today's semiconductor industry is to design analog circuits in mainstream IC technologies used to fabricate digital products. Developed and characterized for digital applications,

such technologies do not easily lend themselves to analog design, requiring novel circuits and architectures to achieve a high performance.

1.2 Why Integrated?

The idea of placing multiple electronic devices on the same substrate was conceived in the late 1950s. In 40 years, the technology has evolved from producing simple chips containing a handful of components to fabricating memories accommodating more than one billion transistors as well as microprocessors comprising more than 10 million devices. As Gordon Moore (one of the founders of Intel) predicted in the early 1970s, the number of transistors per chip has continued to double approximately every one and a half years. At the same time, the minimum dimension of transistors has dropped from about 25 μm in 1960 to about 0.18 μm in the year 2000, resulting in a tremendous improvement in the speed of integrated circuits.

Driven by primarily the memory and microprocessor market, integrated-circuit technologies have also embraced analog design extensively, affording a complexity, speed, and precision that would be impossible to achieve using discrete implementations. Analog and mixed analog/digital integrated circuits containing tens of thousands of devices now routinely appear in consumer products. We can no longer build a discrete prototype to predict the behavior and performance of modern analog circuits.

1.3 Why CMOS?

The idea of metal-oxide-silicon field-effect transistors (MOSFETs) was patented by J. E. Lilienfeld in the early 1930s—well before the invention of the bipolar transistor. Owing to fabrication limitations, however, MOS technologies became practical much later, in the early 1960s, with the first several generations producing only n-type transistors. It was in the mid-1960s that complementary MOS (CMOS) devices (i.e., both n-type and p-type transistors) were introduced, initiating a revolution in the semiconductor industry.

CMOS technologies rapidly captured the digital market: CMOS gates dissipated power only during switching and required very few devices, two attributes in sharp contrast to their bipolar or GaAs counterparts. It was also soon discovered that the dimensions of MOS devices could be scaled down more easily than those of other types of transistors. Furthermore, CMOS circuits proved to have a lower fabrication cost.

The next obvious step was to apply CMOS technology to analog design. The low cost of fabrication and the possibility of placing both analog and digital circuits on the same chip so as to improve the overall performance and/or reduce the cost of packaging made CMOS technology attractive. However, MOSFETs were quite slower and noisier than bipolar transistors, finding limited application.

How did CMOS technology come to dominate the analog market as well? The principal force was device scaling because it continued to improve the speed of MOSFETs. The intrinsic speed of MOS transistors has increased by more than three orders of magnitude in the past 30 years, becoming comparable with that of bipolar devices even though the latter

have also been scaled (but not as fast). Multi-gigahertz analog CMOS circuits are now in production.

1.4 Why This Book?

The design of analog circuits itself has evolved together with the technology and the performance requirements. As the device dimensions shrink, the supply voltage of integrated circuits drops, and analog and digital circuits are fabricated on one chip, many design issues arise that were unimportant only a decade ago. Such trends demand that the analysis and design of circuits be accompanied by an in-depth understanding of their advantages and disadvantages with respect to new technology-imposed limitations.

Good analog design requires intuition, rigor, and creativity. As analog designers, we must wear our engineer's hat for a quick and intuitive understanding of a large circuit, our mathematician's hat for quantifying subtle, yet important effects in a circuit, and our artist's hat for inventing new circuit topologies.

This book describes modern analog design from both intuitive and rigorous angles. It also fosters the reader's creativity by carefully guiding him/her through the evolution of each circuit and presenting the thought process that occurs during the development of new circuit techniques.

1.5 General Concepts

1.5.1 Levels of Abstraction

Analysis and design of integrated circuits often require thinking at various levels of abstraction. Depending on the effect or quantity of interest, we may study a complex circuit at device physics level, transistor level, architecture level, or system level. In other words, we may consider the behavior of individual devices in terms of their internal electric fields and charge transport [Fig. 1.8(a)], the interaction of a group of devices according to their electrical characteristics [Fig. 1.8(b)], the function of several building blocks operating as a unit [Fig. 1.8(c)], or the performance of the system in terms of that of its constituent subsystems [Fig. 1.8(d)]. Switching between levels of abstraction becomes necessary in both understanding the details of the operation and optimizing the overall performance. In fact, in today's IC industry, the interaction between all groups, from device physicists to system designers, is essential to achieving a high performance and a low cost. In this book, we begin with device physics and develop increasingly more complex circuit topologies.

1.5.2 Robust Analog Design

Many device and circuit parameters vary with the fabrication process, supply voltage, and ambient temperature. We denote these effects by PVT and design circuits such that their performance remains in an acceptable range for a specified range of PVT variations. For example, the supply voltage may vary from 2.7 V to 3.3 V and the temperature from 0° to 70°. Robust analog design in CMOS technology is a challenging task because device parameters vary significantly from wafer to wafer.

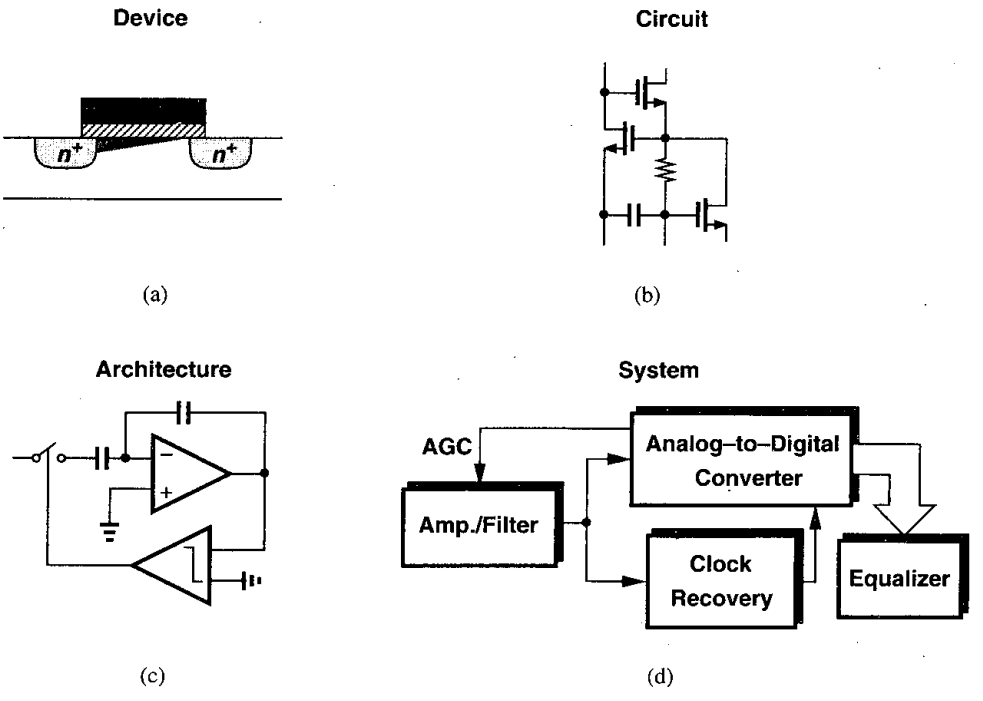


Figure 1.8 Abstraction levels in circuit design: (a) device level, (b) circuit level, (c) architecture level, (d) system level.

1.5.3 Notations

The voltages and currents in integrated circuits typically contain a bias component and a signal component. While it is desirable to employ a notation that distinguishes between these quantities, in practice other difficulties arise. For example, if the drain bias current of a transistor is denoted by I_D and the drain signal current by i_D, then the Laplace transform of i_D, $I_D(s)$, may be confused with I_D unless it is always accompanied by s. Furthermore, it is confusing to write the low-frequency gain of a circuit as $v_{out}/v_{in} = -g_m R_D$ and the high-frequency gain as $V_{out}/V_{in} = -g_m R_D/(1 + R_D C_L s)$.

In this book, we denote most voltages and currents by uppercase letters, making it clear from the context which component they represent. For example, I_D, V_{GS}, and V_X denote bias, signal, or bias+signal quantities. For input and output voltages, we use V_{in} and V_{out}, respectively.

Chapter 2

Basic MOS Device Physics

In studying the design of integrated circuits, one of two extreme approaches can be taken: (1) begin with quantum mechanics and understand solid-state physics, semiconductor device physics, device modeling, and finally the design of circuits; (2) treat each semiconductor device as a black box whose behavior is described in terms of its terminal voltages and currents and design circuits with little attention to the internal operation of the device. Experience shows that neither approach is optimum. In the first case, the reader cannot see the relevance of all of the physics to designing circuits, and in the second, he/she is constantly mystified by the contents of the black box.

In today's IC industry, a solid understanding of semiconductor devices is essential, more so in analog design than in digital design because in the former, transistors are not considered as simple switches and many of their second-order effects directly impact the performance. Furthermore, as each new generation of IC technologies scales the devices, these effects become more significant. Since the designer must often decide which effects can be neglected in a given circuit, insight into device operation proves invaluable.

In this chapter, we study the physics of MOSFETs at an elementary level, covering the bare minimum that is necessary for basic analog design. The ultimate goal is still to develop a circuit model for each device by formulating its operation, but this is accomplished with a good understanding of the underlying principles. After studying many analog circuits in Chapters 3 through 13 and gaining motivation for a deeper understanding of devices, we return to the subject in Chapter 16 and deal with other aspects of MOS operation.

We begin our study with the structure of MOS transistors and derive their I/V characteristics. Next, we describe second-order effects such as body effect, channel-length modulation, and subthreshold conduction. We then identify the parasitic capacitances of MOSFETs, derive a small-signal model, and present a simple SPICE model. We assume that the reader is familiar with such basic concepts as doping, mobility, and *pn* junctions.

2.1 General Considerations

2.1.1 MOSFET as a Switch

Before delving into the actual operation of the MOSFET, we consider a simplistic model of the device so as to gain a feeling for what the transistor is expected to be and which aspects of its behavior are important.

Shown in Fig. 2.1 is the symbol for an n-type MOSFET, revealing three terminals: gate (G), source (S), and drain (D). The latter two are interchangeable because the device is

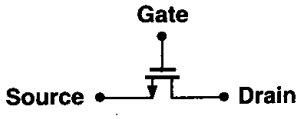


Figure 2.1 Simple view of a MOS device.

symmetric. When operating as a switch, the transistor "connects" the source and the drain together if the gate voltage, V_G, is "high" and isolates the source and the drain if V_G is "low."

Even with this simplified view, we must answer several questions. For what value of V_G does the device turn on? In other words, what is the "threshold" voltage? What is the resistance between S and D when the device is on (or off)? How does this resistance depend on the terminal voltages? Can we always model the path between S and D by a simple linear resistor? What limits the speed of the device?

While all of these questions arise at the circuit level, they can be answered only by analyzing the structure and physics of the transistor.

2.1.2 MOSFET Structure

Fig. 2.2 shows a simplified structure of an n-type MOS (NMOS) device. Fabricated on a p-type substrate (also called the "bulk" or the "body"), the device consists of two heavily-doped n regions forming the source and drain terminals, a heavily-doped (conductive) piece

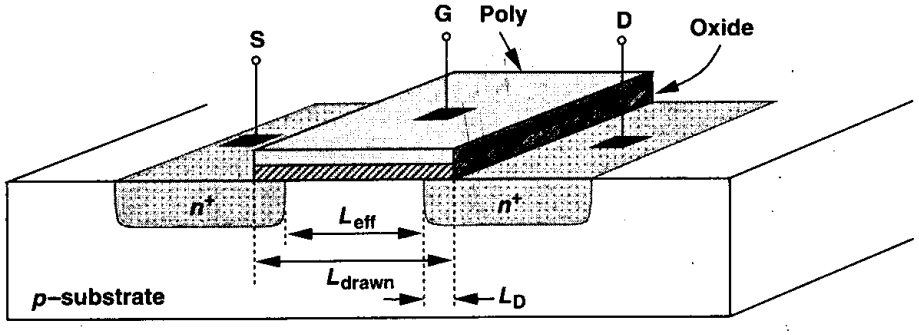


Figure 2.2 Structure of a MOS device.

of polysilicon[1] (often simply called "poly") operating as the gate, and a thin layer of silicon dioxide (SiO_2) insulating the gate from the substrate. The useful action of the device occurs in the substrate region under the gate oxide. Note that the structure is symmetric with respect to S and D.

The dimension of the gate along the source-drain path is called the length, L, and that perpendicular to the length is called the width, W. Since during fabrication the S/D junctions "side-diffuse," the actual distance between the source and the drain is slightly less than L. To avoid confusion, we write, $L_{eff} = L_{drawn} - 2L_D$, where L_{eff} is the "effective" length, L_{drawn} is the total length,[2] and L_D is the amount of side diffusion. As we will see later, L_{eff} and the gate oxide thickness, t_{ox}, play an important role in the performance of MOS circuits. Consequently, the principal thrust in MOS technology development is to reduce both of these dimensions from one generation to the next without degrading other parameters of the device. Typical values at the time of this writing are $L_{eff} \approx 0.15\ \mu$m and $t_{ox} \approx 50$ Å. In the remainder of this book, we denote the effective length by L.

If the MOS structure is symmetric, why do we call one n region the source and the other the drain? This becomes clear if the source is defined as the terminal that provides the charge carriers (electrons in the case of NMOS devices) and the drain as the terminal that collects them. Thus, as the voltages at the three terminals of the device vary, the source and the drain may exchange roles. These concepts are practiced in the problems at the end of the chapter.

We have thus far ignored the substrate on which the device is fabricated. In reality, the substrate potential greatly influences the device characteristics. That is, the MOSFET is a *four*-terminal device. Since in typical MOS operation the S/D junction diodes must be reverse-biased, we assume the substrate of NMOS transistors is connected to the most negative supply in the system. For example, if a circuit operates between zero and 3 volts, $V_{sub,NMOS} = 0$. The actual connection is usually provided through an ohmic p^+ region, as depicted in the side view of the device in Fig. 2.3.

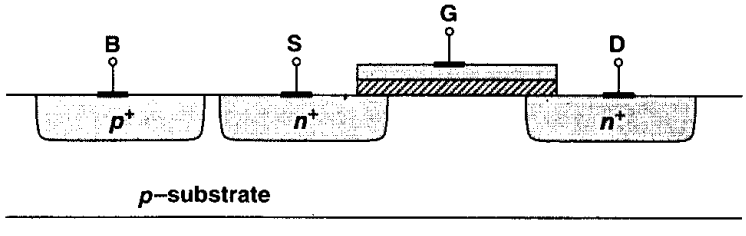


Figure 2.3 Substrate connection.

In complementary MOS (CMOS) technologies, both NMOS and PMOS transistors are available. From a simplistic view point, the PMOS device is obtained by negating all of

[1]Polysilicon is silicon in amorphous (non-crystal) form. As explained in Chapter 17, when the gate silicon is grown on top of the oxide, it cannot form a crystal.

[2]The subscript "drawn" is used because this is the dimension that we draw in the layout of the transistor (Section 2.4.1).

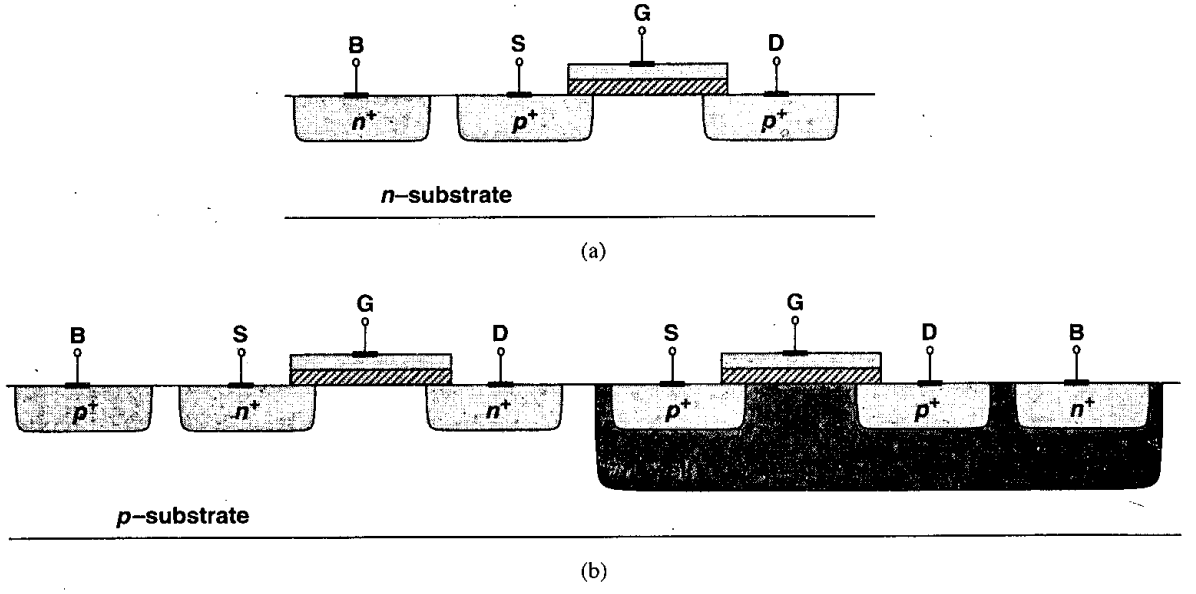


Figure 2.4 (a) Simple PMOS device, (b) PMOS inside an n-well.

the doping types (including the substrate) [Fig. 2.4(a)], but in practice, NMOS and PMOS devices must be fabricated on the same wafer, i.e., the same substrate. For this reason, one device type can be placed in a "local substrate," usually called a "well." In most of today's CMOS processes, the PMOS device is fabricated in an n-well [Fig. 2.4(b)]. Note that the n-well must be connected to a potential such that the S/D junction diodes of the PMOS transistor remain reverse-biased under all conditions. In most circuits, the n-well is tied to the most positive supply voltage. For the sake of brevity, we sometimes call NMOS and PMOS devices "NFETs" and "PFETs," respectively.

Fig. 2.4(b) indicates an interesting difference between NMOS and PMOS transistors: while all NFETs share the same substrate, each PFET can have an independent n-well. This flexibility of PFETs is exploited in some analog circuits.

2.1.3 MOS Symbols

The circuit symbols used to represent NMOS and PMOS transistors are shown in Fig. 2.5. The symbols in Fig. 2.5(a) contain all four terminals, with the substrate denoted by "B" (bulk) rather than "S" to avoid confusion with the source. The source of the PMOS device is positioned on top as a visual aid because it has a higher potential than its gate. Since in most circuits the bulk terminals of NMOS and PMOS devices are tied to ground and V_{DD}, respectively, we usually omit these connections in drawing [Fig. 2.5(b)]. In digital circuits, it is customary to use the "switch" symbols depicted in Fig. 2.5(c) for the two types, but we prefer those in Fig. 2.5(b) because the visual distinction between S and D proves helpful in understanding the operation of circuits.

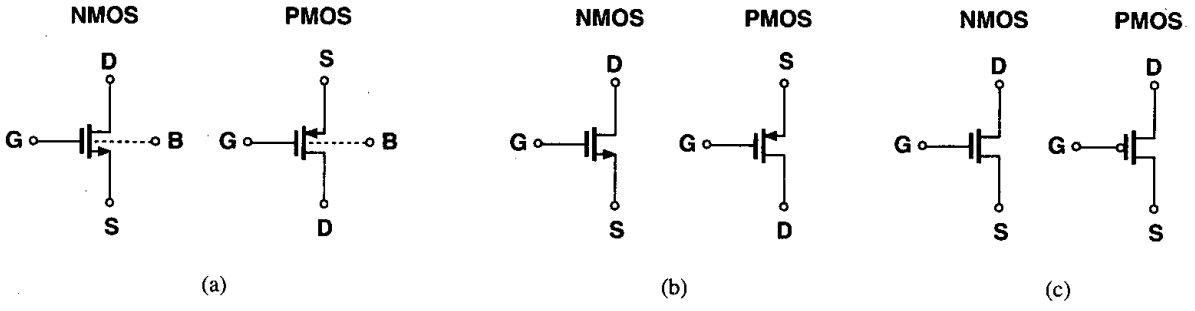


Figure 2.5 MOS symbols.

2.2 MOS I/V Characteristics

In this section, we analyze the generation and transport of charge in MOSFETs as a function of the terminal voltages. Our objective is to derive equations for the I/V characteristics such that we can elevate our abstraction from device physics level to circuit level.

2.2.1 Threshold Voltage

Consider an NFET connected to external voltages as shown in Fig. 2.6(a). What happens as the gate voltage, V_G, increases from zero? Since the gate and the substrate form a capacitor,

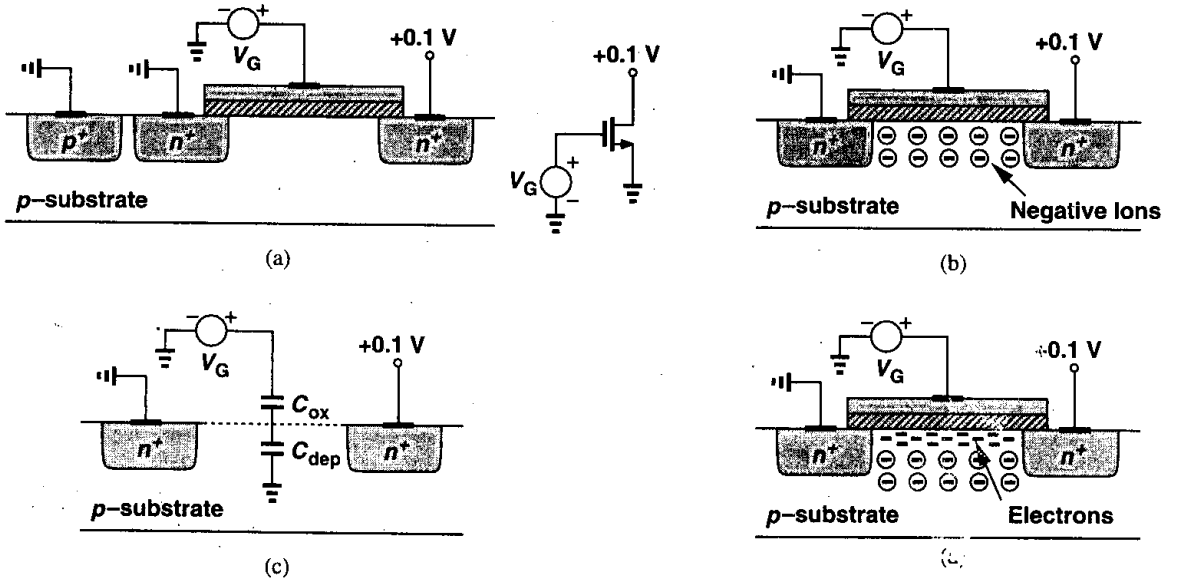


Figure 2.6 (a) A MOSFET driven by a gate voltage, (b) formation of depletion region, (c) onset of inversion, (d) formation of inversion layer.

as V_G becomes more positive, the holes in the p-substrate are repelled from the gate area, leaving negative ions behind so as to mirror the charge on the gate. In other words, a depletion region is created [Fig. 2.6(b)]. Under this condition, no current flows because no charge carriers are available.

As V_G increases, so do the width of the depletion region and the potential at the oxide-silicon interface. In a sense, the structure resembles two capacitors in series: the gate oxide capacitor and the depletion region capacitor [Fig. 2.6(c)]. When the interface potential reaches a sufficiently positive value, electrons flow from the source to the interface and eventually to the drain. Thus, a "channel" of charge carriers is formed under the gate oxide between S and D, and the transistor is "turned on." We also say the interface is "inverted." The value of V_G for which this occurs is called the "threshold voltage," V_{TH}. If V_G rises further, the charge in the depletion region remains relatively constant while the channel charge density continues to increase, providing a greater current from S to D.

In reality, the turn-on phenomenon is a gradual function of the gate voltage, making it difficult to define V_{TH} unambiguously. In semiconductor physics, the V_{TH} of an NFET is usually defined as the gate voltage for which the interface is "as much n-type as the substrate is p-type." It can be proved [1] that[3]

$$V_{TH} = \Phi_{MS} + 2\Phi_F + \frac{Q_{dep}}{C_{ox}}, \tag{2.1}$$

where Φ_{MS} is the difference between the work functions of the polysilicon gate and the silicon substrate, $\Phi_F = (kT/q)\ln(N_{sub}/n_i)$, q is electron charge, N_{sub} is the doping concentration of the substrate, Q_{dep} is the charge in the depletion region, and C_{ox} is the gate oxide capacitance per unit area. From pn junction theory, $Q_{dep} = \sqrt{4q\epsilon_{si}|\Phi_F|N_{sub}}$, where ϵ_{si} denotes the dielectric constant of silicon. Since C_{ox} appears very frequently in device and circuit calculations, it is helpful to remember that for $t_{ox} \approx 50$ Å, $C_{ox} \approx 6.9$ fF/μm^2. The value of C_{ox} can then be scaled proportionally for other oxide thicknesses.

In practice, the "native" threshold value obtained from the above equation may not be suited to circuit design, e.g., $V_{TH} = 0$ and the device does not turn off for $V_G \geq 0$. For this reason, the threshold voltage is typically adjusted by implantation of dopants into the channel area during device fabrication, in essence altering the doping level of the substrate near the oxide interface. For example, as shown in Fig. 2.7, if a thin sheet of p^+ is created, the gate voltage required to deplete this region increases.

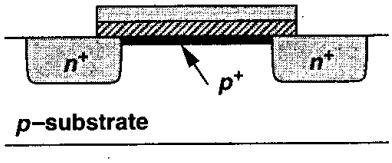


Figure 2.7 Implantation of p^+ dopants to alter the threshold.

The above definition is not directly applicable to the *measurement* of V_{TH}. In Fig. 2.6(a), only the drain current can indicate whether the device is "on" or "off," thus failing to reveal at what V_{GS} the interface is as much n-type as the bulk is p-type. As a result, the calculation

[3]Charge trapping in the oxide is neglected here.

of V_{TH} from I/V measurements is somewhat ambiguous. We return to this point later but assume in our preliminary analysis that the device turns on *abruptly* for $V_{GS} \geq V_{TH}$.

The turn-on phenomenon in a PMOS device is similar to that of NFETs but with all of the polarities reversed. As shown in Fig. 2.8, if the gate-source voltage becomes sufficiently

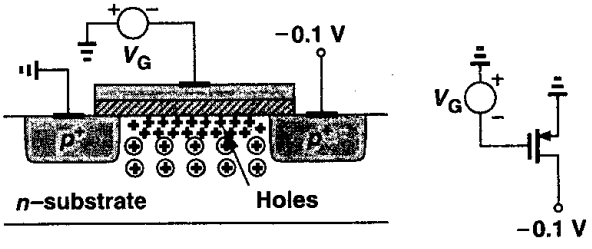


Figure 2.8 Formation of inversion layer in a PFET.

negative, an inversion layer consisting of holes is formed at the oxide-silicon interface, providing a conduction path between the source and the drain.

2.2.2 Derivation of I/V Characteristics

In order to obtain the relationship between the drain current of a MOSFET and its terminal voltages, we make two observations.

First, consider a semiconductor bar carrying a current I [Fig. 2.9(a)]. If the charge density along the direction of current is Q_d coulombs per meter and the velocity of the charge is v meters per second, then

$$I = Q_d \cdot v. \tag{2.2}$$

To understand why, we measure the total charge that passes through a cross section of the bar in unit time. With a velocity v, all of the charge enclosed in v meters of the bar must flow through the cross section in one second [Fig. 2.9(b)]. Since the charge density is Q_d, the total charge in v meters equals $Q_d \cdot v$. This lemma proves useful in analyzing semiconductor devices.

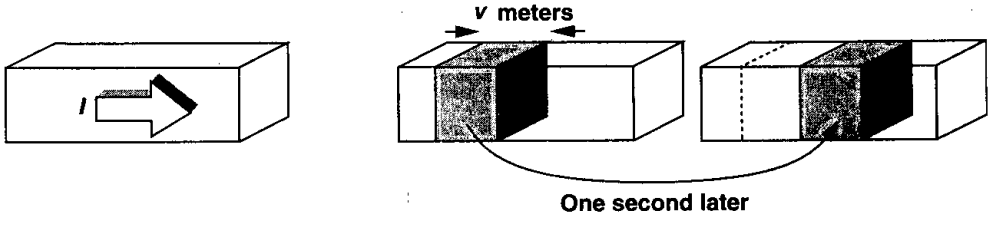


Figure 2.9 (a) A semiconductor bar carrying a current I, (b) snapshots of the carriers one second apart.

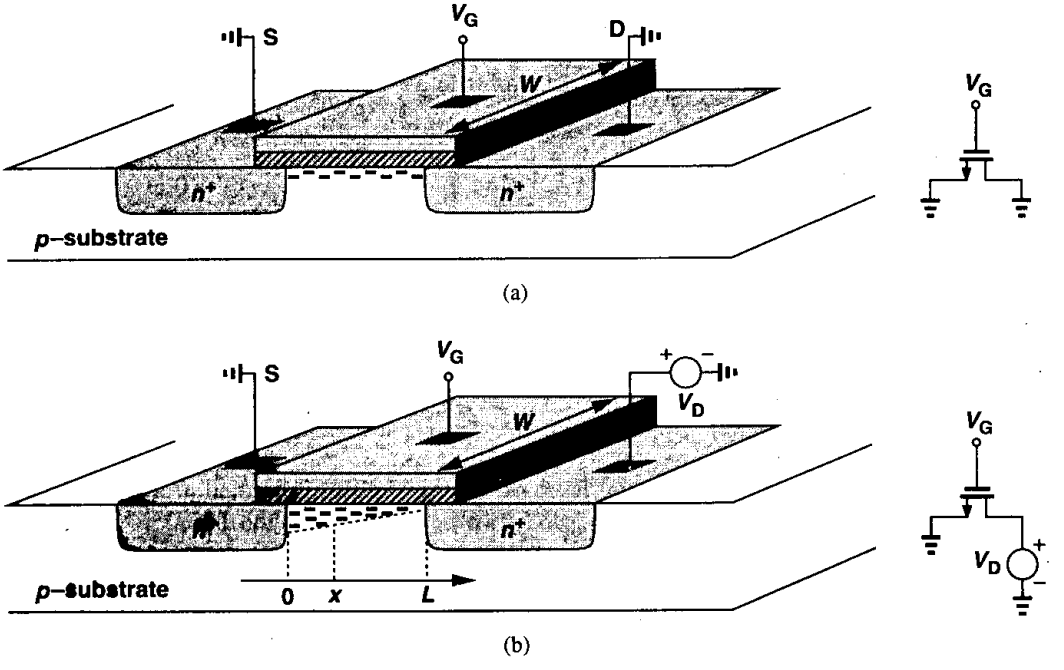


Figure 2.10 Channel charge with (a) equal source and drain voltages, (b) unequal source and drain voltages.

Second, consider an NFET whose source and drain are connected to ground [Fig. 2.10(a)]. What is the charge density in the inversion layer? Since we assume the onset of inversion occurs at $V_{GS} = V_{TH}$, the inversion charge density produced by the gate oxide capacitance is proportional to $V_{GS} - V_{TH}$. For $V_{GS} \geq V_{TH}$, any charge placed on the gate must be mirrored by the charge in the channel, yielding a uniform channel charge density (charge per unit length) equal to

$$Q_d = WC_{ox}(V_{GS} - V_{TH}), \tag{2.3}$$

where C_{ox} is multiplied by W to represent the total capacitance per unit length.

Now suppose, as depicted in Fig. 2.10(b), the drain voltage is greater than zero. Since the channel potential varies from zero at the source to V_D at the drain, the local voltage *difference* between the gate and the channel varies from V_G to $V_G - V_D$. Thus, the charge density at a point x along the channel can be written as

$$Q_d(x) = WC_{ox}[V_{GS} - V(x) - V_{TH}], \tag{2.4}$$

where $V(x)$ is the channel potential at x.

From (2.2), the current is given by

$$I_D = -WC_{ox}[V_{GS} - V(x) - V_{TH}]v, \tag{2.5}$$

where the negative sign is inserted because the charge carriers are negative and v denotes the velocity of the electrons in the channel. For semiconductors, $v = \mu E$, where μ is the mobility of charge carriers and E is the electric field. Noting that $E(x) = -dV/dx$ and representing the mobility of electrons by μ_n, we have

$$I_D = WC_{ox}[V_{GS} - V(x) - V_{TH}]\mu_n \frac{dV(x)}{dx}, \tag{2.6}$$

subject to boundary conditions $V(0) = 0$ and $V(L) = V_{DS}$. While $V(x)$ can be easily found from this equation, the quantity of interest is in fact I_D. Multiplying both sides by dV and performing integration, we obtain

$$\int_{x=0}^{L} I_D dx = \int_{V=0}^{V_{DS}} WC_{ox}\mu_n[V_{GS} - V(x) - V_{TH}]dV. \tag{2.7}$$

Since I_D is constant along the channel:

$$I_D = \mu_n C_{ox}\frac{W}{L}\left[(V_{GS} - V_{TH})V_{DS} - \frac{1}{2}V_{DS}^2\right]. \tag{2.8}$$

Note that L is the effective channel length.

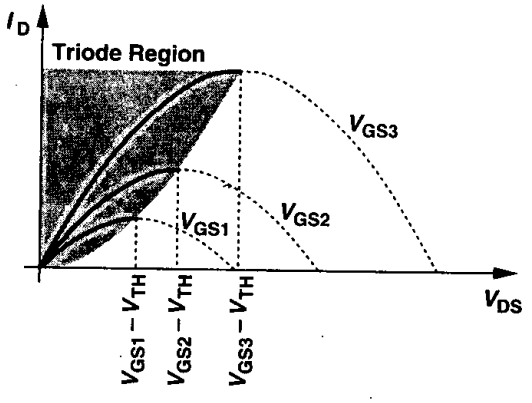


Figure 2.11 Drain current versus drain-source voltage in the triode region.

Fig. 2.11 plots the parabolas given by (2.8) for different values of V_{GS}, indicating that the "current capability" of the device increases with V_{GS}. Calculating $\partial I_D/\partial V_{DS}$, the reader can show that the peak of each parabola occurs at $V_{DS} = V_{GS} - V_{TH}$ and the peak current is

$$I_{D,max} = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{GS} - V_{TH})^2. \tag{2.9}$$

We call $V_{GS} - V_{TH}$ the "overdrive voltage"[4] and W/L the "aspect ratio." If $V_{DS} \leq V_{GS} - V_{TH}$, we say the device operates in the "triode region."[5]

[4]Sometimes called the "effective voltage."

[5]This is also called the "linear region."

Equations (2.8) and (2.9) serve as the foundation for analog CMOS design, describing the dependence of I_D upon the constant of the technology, $\mu_n C_{ox}$, the device dimensions, W and L, and the gate and drain potentials with respect to the source. Note that the integration in (2.7) assumes μ_n and V_{TH} are independent of x and the gate and drain voltages, an approximation that we will revisit in Chapter 16.

If in (2.8), $V_{DS} \ll 2(V_{GS} - V_{TH})$, we have

$$I_D \approx \mu_n C_{ox} \frac{W}{L}(V_{GS} - V_{TH})V_{DS}, \tag{2.10}$$

that is, the drain current is a *linear* function of V_{DS}. This is also evident from the characteristics of Fig. 2.11 for small V_{DS}: as shown in Fig. 2.12, each parabola can be approximated by a straight line. The linear relationship implies that the path from the source to the drain can be represented by a linear resistor equal to

$$R_{on} = \frac{1}{\mu_n C_{ox} \dfrac{W}{L}(V_{GS} - V_{TH})}. \tag{2.11}$$

A MOSFET can therefore operate as a resistor whose value is controlled by the overdrive voltage [so long as $V_{DS} \ll 2(V_{GS} - V_{TH})$]. This is conceptually illustrated in Fig. 2.13. Note that in contrast to bipolar transistors, a MOS device may be on even if it carries no

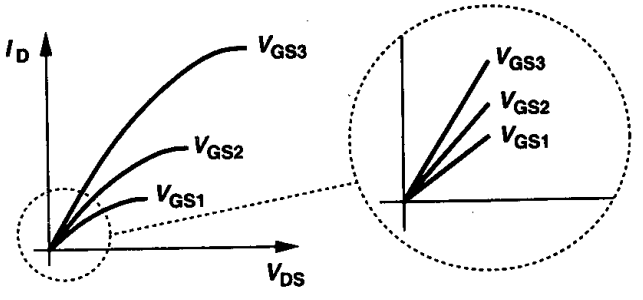


Figure 2.12 Linear operation in deep triode region.

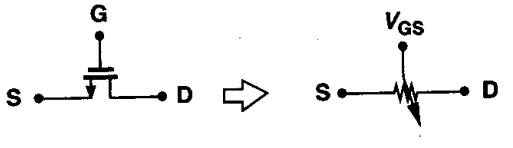


Figure 2.13 MOSFET as a controlled linear resistor.

current. With the condition $V_{DS} \ll 2(V_{GS} - V_{TH})$, we say the device operates in deep triode region.

Example 2.1

For the arrangement in Fig. 2.14(a), plot the on-resistance of M_1 as a function of V_G. Assume $\mu_n C_{ox} = 50\ \mu\text{A/V}^2$, $W/L = 10$, and $V_{TH} = 0.7$ V. Note that the drain terminal is open.

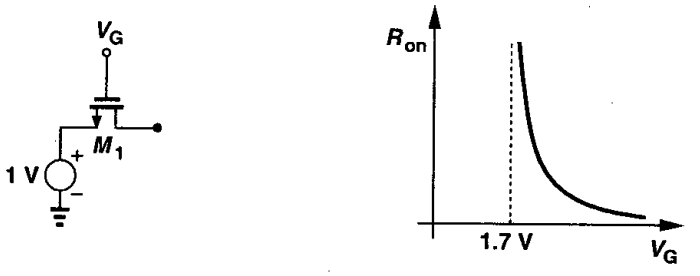


Figure 2.14

Solution

Since the drain terminal is open, $I_D = 0$ and $V_{DS} = 0$. Thus, if the device is on, it operates in the deep triode region. For $V_G < 1\text{ V} + V_{TH}$, M_1 is off and $R_{on} = \infty$. For $V_G > 1\text{ V} + V_{TH}$, we have

$$R_{on} = \frac{1}{50\,\mu\text{A/V}^2 \times 10(V_G - 1\text{ V} - 0.7\text{ V})}. \tag{2.12}$$

The result is plotted in Fig. 2.14(b).

The utility of MOSFETs as controllable resistors and hence switches plays a crucial role in many analog circuits. This is studied in Chapter 12.

What happens if in Fig. 2.11 the drain-source voltage exceeds $V_{GS} - V_{TH}$? In reality, the drain current does *not* follow the parabolic behavior for $V_{DS} > V_{GS} - V_{TH}$. In fact, as shown in Fig. 2.15, I_D becomes relatively constant and we say the device operates in the "saturation" region.[6] To understand this phenomenon, recall from (2.4) that the local

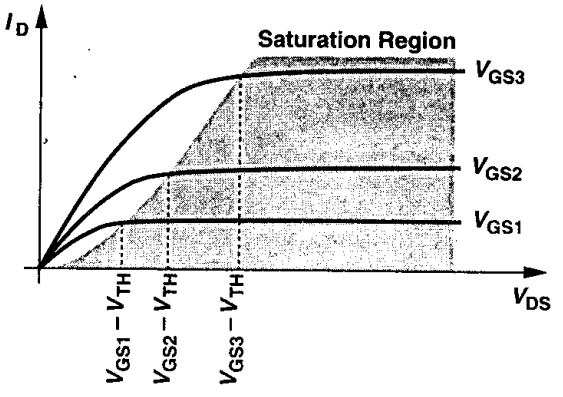


Figure 2.15 Saturation of drain current.

[6]Note the difference between saturation in bipolar and MOS devices.

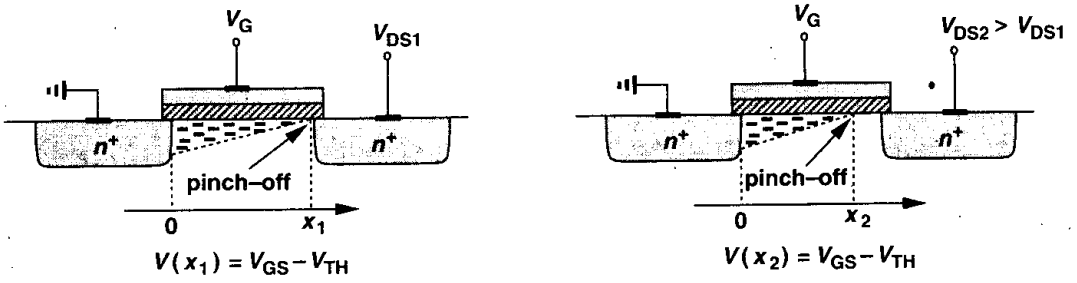


Figure 2.16 Pinch-off behavior.

density of inversion layer charge is proportional to $V_{GS} - V(x) - V_{TH}$. Thus, if $V(x)$ approaches $V_{GS} - V_{TH}$, then $Q_d(x)$ drops to zero. In other words, as depicted in Fig. 2.16, if V_{DS} is slightly greater than $V_{GS} - V_{TH}$, then the inversion layer stops at $x \leq L$, and we say the channel is "pinched off." As V_{DS} increases further, the point at which Q_d equals zero gradually moves toward the source. Thus, at some point along the channel, the local potential difference between the gate and the oxide-silicon interface is not sufficient to support an inversion layer.

With the above observations, we re-examine (2.7) for a saturated device. Since Q_d is the density of *mobile* charge, the integral on the left-hand side of (2.7) must be taken from $x = 0$ to $x = L'$, where L' is the point at which Q_d drops to zero, and that on the right from $V(x) = 0$ to $V(x) = V_{GS} - V_{TH}$. As a result:

$$I_D = \frac{1}{2}\mu_n C_{ox}\frac{W}{L'}(V_{GS} - V_{TH})^2, \tag{2.13}$$

indicating that I_D is relatively independent of V_{DS} if L' remains close to L.

For PMOS devices, Eqs. (2.8) and (2.13) are respectively written as

$$I_D = -\mu_p C_{ox}\frac{W}{L}\left[(V_{GS} - V_{TH})V_{DS} - \frac{1}{2}V_{DS}^2\right] \tag{2.14}$$

and

$$I_D = -\frac{1}{2}\mu_p C_{ox}\frac{W}{L'}(V_{GS} - V_{TH})^2. \tag{2.15}$$

The negative sign appears here because we assume I_D flows from the drain to the source, whereas holes flow in the reverse direction. Since the mobility of holes is about one-half to one-fourth of the mobility of electrons, PMOS devices suffer from lower "current drive" capability.

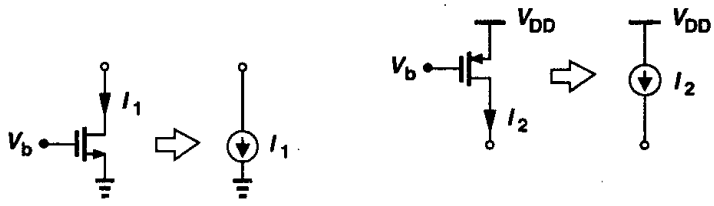


Figure 2.17 Saturated MOSFETs operating as current sources.

With the approximation $L \approx L'$, a saturated MOSFET can be used as a current source connected between the drain and the source (Fig. 2.17), an important component in analog design. Note that the current sources inject current into ground or draw current from V_{DD}. In other words, only one terminal of each current source is "floating."

Since a MOSFET operating in saturation produces a current in response to its gate-source overdrive voltage, we may define a figure of merit that indicates how well a device converts a voltage to a current. More specifically, since in processing signals we deal with the *changes* in voltages and currents, we define the figure of merit as the change in the drain current divided by the change in the gate-source voltage. Called the "transconductance" and denoted by g_m, this quantity is expressed as:

$$g_m = \frac{\partial I_D}{\partial V_{GS}}\bigg|_{V_{DS},\text{const.}} \tag{2.16}$$

$$= \mu_n C_{ox} \frac{W}{L}(V_{GS} - V_{TH}). \tag{2.17}$$

In a sense, g_m represents the sensitivity of the device: for a high g_m, a small change in V_{GS} results in a large change in I_D. Interestingly, g_m in the saturation region is equal to the inverse of R_{on} in deep triode region.

The reader can prove that g_m can also be expressed as

$$g_m = \sqrt{2\mu_n C_{ox} \frac{W}{L} I_D} \tag{2.18}$$

$$= \frac{2I_D}{V_{GS} - V_{TH}}. \tag{2.19}$$

Plotted in Fig. 2.18, each of the above expressions proves useful in studying the behavior of g_m as a function of one parameter while other parameters remain constant. For example, (2.17) suggests that g_m increases with the overdrive if W/L is constant whereas (2.19) implies that g_m decreases with the overdrive if I_D is constant. The concept of transconductance

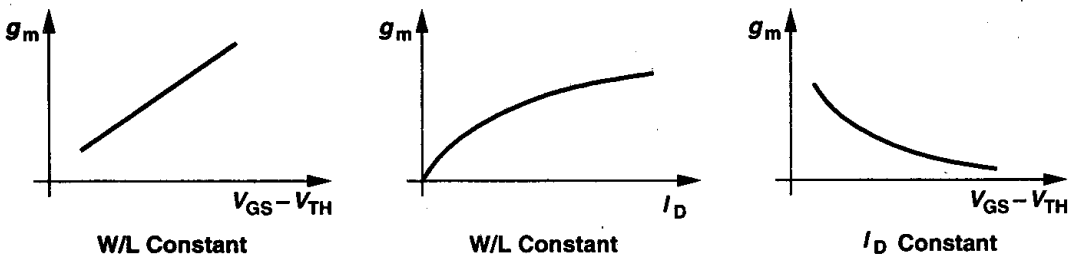


Figure 2.18 MOS transconductance as a function of overdrive and drain current.

can also be applied to a device operating in the triode region, as illustrated in the following example.

Example 2.2

For the arrangement shown in Fig. 2.19, plot the transconductance as a function of V_{DS}.

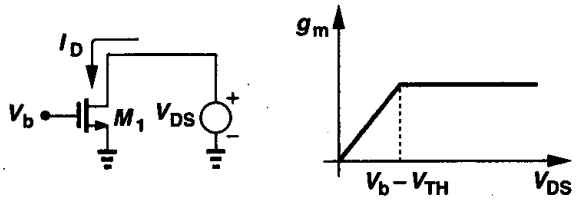


Figure 2.19

Solution

It is simpler to study g_m as V_{DS} decreases from infinity. So long as $V_{DS} \geq V_b - V_{TH}$, M_1 is in saturation, I_D is relatively constant, and, from (2.18), so is g_m. For $V_{DS} < V_b - V_{TH}$, M_1 is in the triode region and:

$$g_m = \frac{\partial}{\partial V_{GS}} \left\{ \frac{1}{2} \mu_n C_{ox} \frac{W}{L} \left[2(V_{GS} - V_{TH}) V_{DS} - V_{DS}^2 \right] \right\} \tag{2.20}$$

$$= \mu_n C_{ox} \frac{W}{L} V_{DS}. \tag{2.21}$$

Thus, as plotted in Fig. 2.19, the transconductance drops if the device enters the triode region. For amplification, therefore, we usually employ MOSFETs in saturation.

The distinction between saturation and triode regions can be confusing, especially for PMOS devices. Intuitively, we note that the channel is pinched off if the difference between the gate and drain voltages is not sufficient to create an inversion layer. As depicted conceptually in Fig. 2.20, as $V_G - V_D$ of an NFET drops below V_{TH}, pinch-off occurs. Similarly,

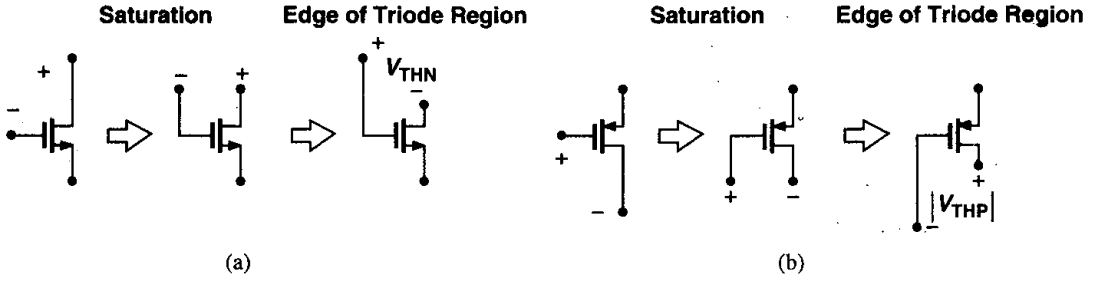


Figure 2.20 Conceptual visualization of saturation and triode regions.

if $V_D - V_G$ of a PFET is not large enough ($< |V_{THP}|$), the device is saturated. Note that this view does not require knowledge of the source voltage. This means we must know a priori which terminal operates as the drain.

2.3 Second-Order Effects

Our analysis of the MOS structure has thus far entailed various simplifying assumptions, some of which are not valid in many analog circuits. In this section, we describe three second-order effects that are essential in our subsequent circuit analyses. Other phenomena that appear in submicron devices are studied in Chapter 16.

Body Effect In the analysis of Fig. 2.10, we tacitly assumed that the bulk and the source of the transistor were tied to ground. What happens if the bulk voltage of an NFET drops below the source voltage (Fig. 2.21)? Since the S and D junctions remain reverse-biased, we surmise that the device continues to operate properly but certain characteristics may

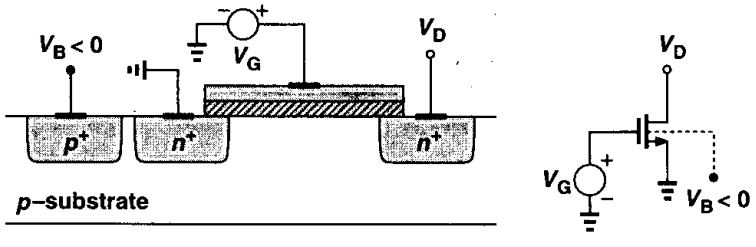


Figure 2.21 NMOS device with negative bulk voltage.

change. To understand the effect, suppose $V_S = V_D = 0$, and V_G is somewhat less than V_{TH} so that a depletion region is formed under the gate but no inversion layer exists. As V_B becomes more negative, more holes are attracted to the substrate connection, leaving a larger negative charge behind, i.e., as depicted in Fig. 2.22, the depletion region becomes wider. Now recall from Eq. (2.1) that the threshold voltage is a function of the total charge in the depletion region because the gate charge must mirror Q_d before an inversion layer is

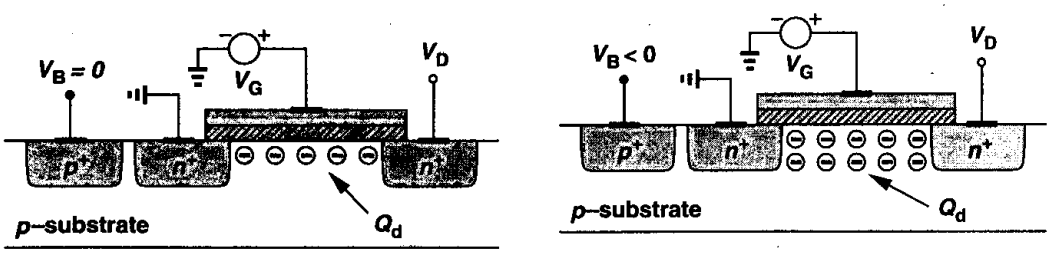


Figure 2.22 Variation of depletion region charge with bulk voltage.

formed. Thus, as V_B drops and Q_d increases, V_{TH} also increases. This is called the "body effect" or the "backgate effect."

It can be proved that with body effect:

$$V_{TH} = V_{TH0} + \gamma \left(\sqrt{|2\Phi_F + V_{SB}|} - \sqrt{|2\Phi_F|} \right), \tag{2.22}$$

where V_{TH0} is given by (2.1), $\gamma = \sqrt{2q\epsilon_{si}N_{sub}}/C_{ox}$ denotes the body effect coefficient, and V_{SB} is the source-bulk potential difference [1]. The value of γ typically lies in the range of 0.3 to 0.4 $V^{1/2}$.

Example 2.3

In Fig. 2.23(a), plot the drain current if V_X varies from $-\infty$ to 0. Assume $V_{TH0} = 0.6$ V, $\gamma = 0.4$ $V^{1/2}$, and $2\Phi_F = 0.7$ V.

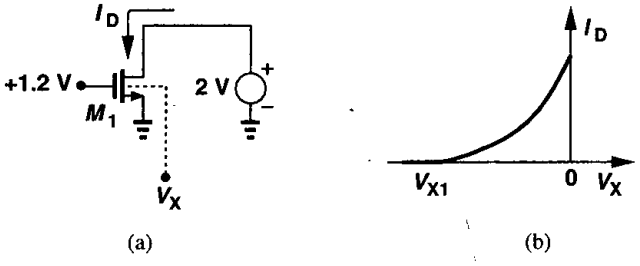


Figure 2.23

Solution

If V_X is sufficiently negative, the threshold voltage of M_1 exceeds 1.2 V and the device is off. That is,

$$1.2 \text{ V} = 0.6 + 0.4 \left(\sqrt{0.7 - V_{X1}} - \sqrt{0.7} \right), \tag{2.23}$$

and hence $V_{X1} = -4.76$ V. For $V_{X1} < V_X < 0$, I_D increases according to

$$I_D = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}\left[V_{GS} - V_{TH0} - \gamma\left(\sqrt{2\Phi_F - V_X} - \sqrt{2\Phi_F}\right)\right]^2. \quad (2.24)$$

Fig. 2.23(b) shows the resulting behavior.

For body effect to manifest itself, the bulk potential, V_{sub}, need not change: if the source voltage varies with respect to V_{sub}, the same phenomenon occurs. For example, consider the circuit in Fig. 2.24(a), first ignoring body effect. We note that as V_{in} varies, V_{out} closely follows the input because the drain current remains equal to I_1. In fact, we can write

$$I_1 = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{in} - V_{out} - V_{TH})^2, \quad (2.25)$$

concluding that $V_{in} - V_{out}$ is constant if I_1 is constant [Fig. 2.24(b)].

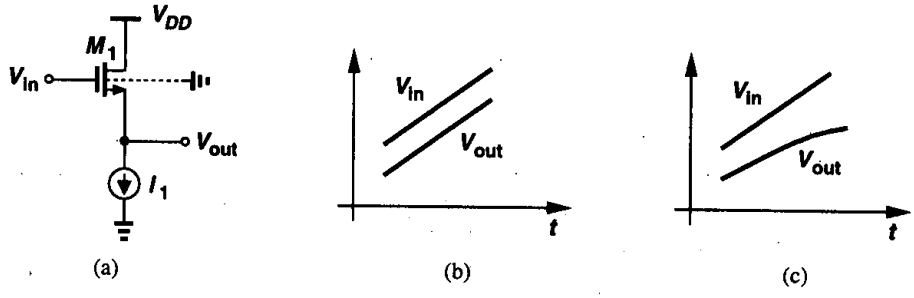


Figure 2.24 (a) A circuit in which the source-bulk voltage varies with input level, (b) input and output voltages with no body effect, (c) input and output voltages with body effect.

Now suppose the substrate is tied to ground and body effect is significant. Then, as V_{in} and hence V_{out} become more positive, the potential difference between the source and the bulk increases, raising the value of V_{TH}. Eq. (2.25) therefore implies that $V_{in} - V_{out}$ must increase so as to maintain I_D constant [Fig. 2.24(c)].

Body effect is usually undesirable. The change in the threshold voltage, e.g., as in Fig. 2.24(a), often complicates the design of analog (and even digital) circuits. Device technologists balance N_{sub} and C_{ox} to obtain a reasonable value for γ.

Channel-Length Modulation In the analysis of channel pinch-off in Section 2.2, we noted that the actual length of the inverted channel gradually decreases as the potential difference between the gate and the drain increases. In other words, in (2.13), L' is in fact a function of V_{DS}. This effect is called "channel-length modulation." Writing $L' = L - \Delta L$, i.e., $1/L' \approx (1 + \Delta L/L)/L$, and assuming a first-order relationship between $\Delta L/L$ and V_{DS} such as $\Delta L/L = \lambda V_{DS}$, we have, in saturation,

$$I_D \approx \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{GS} - V_{TH})^2(1 + \lambda V_{DS}), \quad (2.26)$$

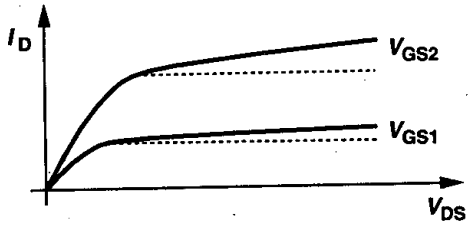


Figure 2.25 Finite saturation region slope resulting from channel-length modulation.

where λ is the channel-length modulation coefficient. Illustrated in Fig. 2.25, this phenomenon results in a nonzero slope in the I_D/V_{DS} characteristic and hence a nonideal current source between D and S in saturation. The parameter λ represents the *relative* variation in length for a given increment in V_{DS}. Thus, for longer channels, λ is smaller.

With channel-length modulation, some of the expressions derived for g_m must be modified. Equations (2.17) and (2.18) are respectively rewritten as

$$g_m = \mu_n C_{ox} \frac{W}{L}(V_{GS} - V_{TH})(1 + \lambda V_{DS}). \tag{2.27}$$

$$= \sqrt{\frac{2\mu_n C_{ox}(W/L) I_D}{1 + \lambda V_{DS}}}, \tag{2.28}$$

while Eq. (2.19) remains unchanged.

Example 2.4

Keeping all other parameters constant, plot I_D/V_{DS} characteristic of a MOSFET for $L = L_1$ and $L = 2L_1$.

Solution

Writing

$$I_D = \frac{1}{2}\mu_n C_{ox} \frac{W}{L}(V_{GS} - V_{TH})^2(1 + \lambda V_{DS}) \tag{2.29}$$

and $\lambda \propto 1/L$, we note that if the length is doubled, the slope of I_D vs. V_{DS} is divided by *four* because $\partial I_D/\partial V_{DS} \propto \lambda/L \propto 1/L^2$ (Fig. 2.26). For a given gate-source overdrive, a larger L gives a more

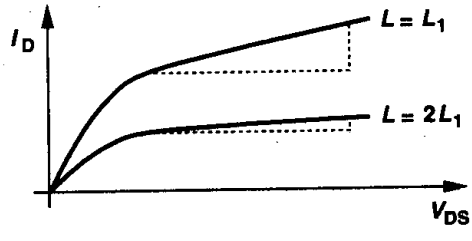


Figure 2.26 Effect of doubling channel length.

ideal current source while degrading the current capability of the device. Thus, W may need to be increased proportionally.

The linear approximation $\Delta L/L \propto V_{DS}$ becomes less accurate in short-channel transistors, resulting in a *variable* slope in the saturated I_D/V_{DS} characteristics. We return to this issue in Chapter 16.

The dependence of I_D upon V_{DS} in saturation may suggest that the bias current of a MOSFET can be defined by the proper choice of the drain-source voltage, allowing freedom in the choice of $V_{GS} - V_{TH}$. However, since the dependence on V_{DS} is much weaker, the drain-source voltage is not used to set the current. The effect of V_{DS} on I_D is usually considered an *error* and it is studied in Chapter 5.

Subthreshold Conduction In our analysis of the MOSFET, we have assumed that the device turns off abruptly as V_{GS} drops below V_{TH}. In reality, for $V_{GS} \approx V_{TH}$, a "weak" inversion layer still exists and some current flows from D to S. Even for $V_{GS} < V_{TH}$, I_D is finite, but it exhibits an *exponential* dependence on V_{GS} [2, 3]. Called "subthreshold conduction," this effect can be formulated for V_{DS} greater than roughly 200 mV as

$$I_D = I_0 \exp \frac{V_{GS}}{\zeta V_T}, \tag{2.30}$$

where $\zeta > 1$ is a nonideality factor and $V_T = kT/q$. We also say the device operates in "week inversion." Except for ζ, (2.30) is similar to the exponential I_C/V_{BE} relationship in a bipolar transistor. The key point here is that as V_{GS} falls below V_{TH}, the drain current drops at a finite rate. With typical values of ζ, at room temperature V_{GS} must decrease by approximately 80 mV for I_D to decrease by one decade (Fig. 2.27). For example, if a

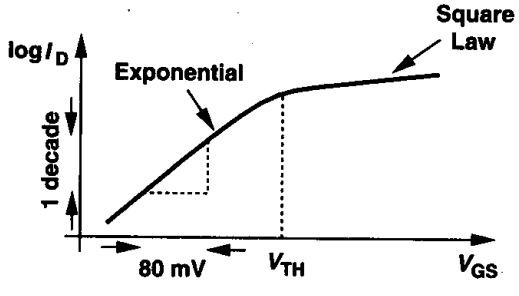


Figure 2.27 MOS subthreshold characteristics.

threshold of 0.3 V is chosen in a process to allow low-voltage operation, then when V_{GS} is reduced to zero, the drain current decreases by only a factor of $10^{3.75}$. Especially problematic in large circuits such as memories, subthreshold conduction can result in significant power dissipation (or loss of analog information).

It is appropriate at this point to return to the definition of the threshold voltage. One definition is to plot the inverse on-resistance of the device $R_{on}^{-1} = \mu C_{ox}(W/L)(V_{GS} - V_{TH})$ as a function of V_{GS} and extrapolate the result to zero, for which $V_{GS} = V_{TH}$. In rough calculations, we often view V_{TH} as the gate-source voltage yielding $I_D/W = 1\mu A/\mu m$ in saturation. For example, if a device with $W = 100\ \mu m$ operates with $I_D = 100\ \mu A$, it is in the vicinity of the subthreshold region. This view is nonetheless vague, especially as device length scales down in every technology generation.

We now re-examine Eq. (2.18) for the transconductance of a MOS device operating in the subthreshold region. Is it possible to achieve an arbitrarily high transconductance by increasing W while maintaining I_D constant? Is it possible to obtain a *higher* transconductance than that of a bipolar transistor (I_C/V_T) biased at the same current? Equation (2.18) was derived from the square-law characteristics $I_D = (1/2)\mu_n C_{ox}(W/L)(V_{GS} - V_{TH})^2$. However, if W increases while I_D remains constant, then $V_{GS} \rightarrow V_{TH}$ and the device enters the subthreshold region. As a result, the transconductance is calculated from (2.30) to be $g_m = I_D/(\zeta V_T)$, revealing that MOSFETs are inferior to bipolar transistors in this respect.

The exponential dependence of I_D upon V_{GS} in subthreshold operation may suggest the use of MOS devices in this regime so as to achieve a higher gain. However, since such conditions are met by only a large device width or low drain current, the speed of subthreshold circuits is severely limited.

Voltage Limitations MOSFETs experience various breakdown effects if their terminal voltage differences exceed certain limits. At high gate-source voltages, the gate oxide breaks down irreversibly, damaging the transistor. In short-channel devices, an excessively large drain-source voltage widens the depletion region around the drain so much that it touches that around the source, creating a very large drain current. (This effect is called "punchthrough.") Other limitations relate to "hot electron effects" and are described in Chapter 16.

2.4 MOS Device Models

2.4.1 MOS Device Layout

For the developments in subsequent sections, it is beneficial to have some understanding of the layout of a MOSFET. We describe only a simple view here, deferring the fabrication details and structural subtleties to Chapters 17 and 18.

The layout of a MOSFET is determined by both the electrical properties required of the device in the circuit and the "design rules" imposed by the technology. For example, W/L is chosen to set the transconductance or other circuit parameters, while the minimum L is dictated by the process. In addition to the gate, the source and drain areas must be defined properly as well.

Shown in Fig. 2.28 are the "bird eye's view" and the top view of a MOSFET. The gate polysilicon and the source and drain terminals are typically tied to metal (aluminum) wires that serve as interconnects with low resistance and capacitance. To accomplish this, one or more "contact windows" must be opened in each region, filled with metal, and connected to the upper metal wires. Note that the gate poly extends beyond the channel area by some amount to ensure reliable definition of the "edge" of the transistor.

The source and drain junctions play an important role in the performance. To minimize the capacitance of S and D, the total area of each junction must be minimized. We see from Fig. 2.28 that one dimension of the junctions is equal to W. The other dimension must be large enough to accommodate the contact windows and is specified by the technology design rules.[7]

[7]This dimension is typically three to four times the minimum allowable channel length.

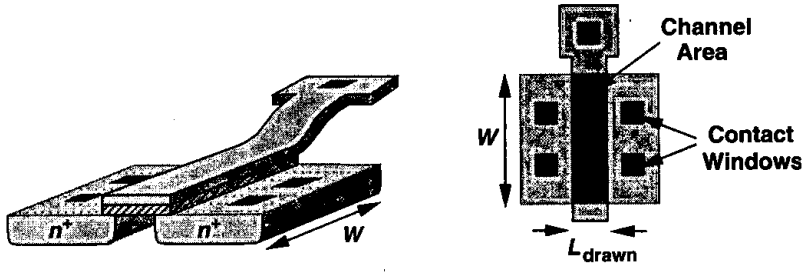


Figure 2.28 Bird's eye and vertical views of a MOS device.

Example 2.5

Draw the layout of the circuit shown in Fig. 2.29(a).

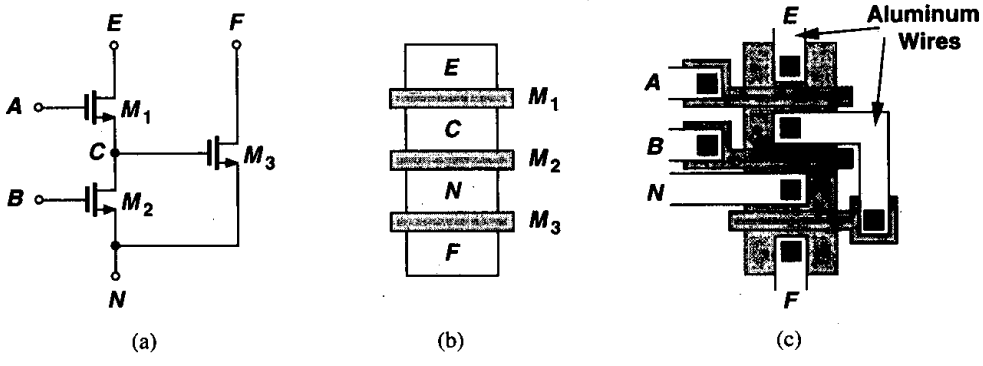


Figure 2.29

Solution

Noting that M_1 and M_2 share the same S/D junctions at node C and M_2 and M_3 also do so at node N, we surmise that the three transistors can be laid out as shown in Fig. 2.29(b). Connecting the remaining terminals, we obtain the layout in Fig. 2.29(c). Note that the gate polysilicon of M_3 cannot be directly tied to the source material of M_1, thus requiring a metal interconnect.

2.4.2 MOS Device Capacitances

The basic quadratic I/V relationships derived in the previous section along with corrections for body effect and channel-length modulation provide a reasonable model for understanding the "dc" behavior of CMOS circuits. In many analog circuits, however, the capacitances associated with the devices must also be taken into account so as to predict the "ac" behavior as well.

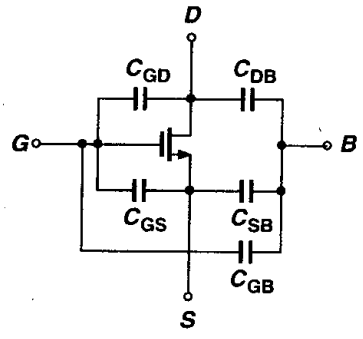
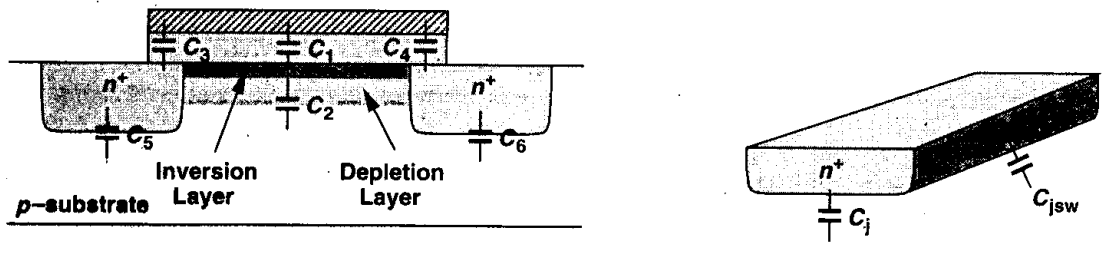


Figure 2.30 MOS capacitances.

We expect that a capacitance exists between every two of the four terminals of a MOSFET (Fig. 2.30).[8] Moreover, the value of each of these capacitances may depend on the bias conditions of the transistor. Considering the physical structure in Fig. 2.31(a), we identify the following. (1) Oxide capacitance between the gate and the channel, $C_1 = WLC_{ox}$; (2) Depletion capacitance between the channel and the substrate, $C_2 = WL\sqrt{q\epsilon_{si}N_{sub}/(4\Phi_F)}$; (3) Capacitance due to the overlap of the gate poly with the source and drain areas, C_3 and C_4. Owing to fringing electric field lines, C_3 and C_4 cannot be simply written as WL_DC_{ox}, and are usually obtained by more elaborate calculations. The overlap capacitance per unit width is denoted by C_{ov}; (4) Junction capacitance between the source/drain areas and the substrate. As shown in Fig. 2.31(b), this capacitance is usually decomposed into two components: bottom-plate capacitance associated with the bottom of the junction, C_j, and sidewall capacitance due to the perimeter of the junction, C_{jsw}. The distinction is necessary because different transistor geometries yield different area and perimeter values for the S/D junctions. We typically specify C_j and C_{jsw} as capacitance per unit area and unit length, respectively. Note that each junction capacitance can be expressed as $C_j = C_{j0}/[1 + V_R/\Phi_B]^m$, where V_R is the reverse voltage across the junction, Φ_B is the junction built-in potential, and m is a power typically in the range of 0.3 and 0.4.

[8]The capacitance between S and D is negligible.



(a) (b)

Figure 2.31 (a) MOS device capacitances, (b) decomposition of S/D junction capacitance into bottom-plate and sidewall components.

Example 2.6

Calculate the source and drain junction capacitances of the two structures shown in Fig. 2.32.

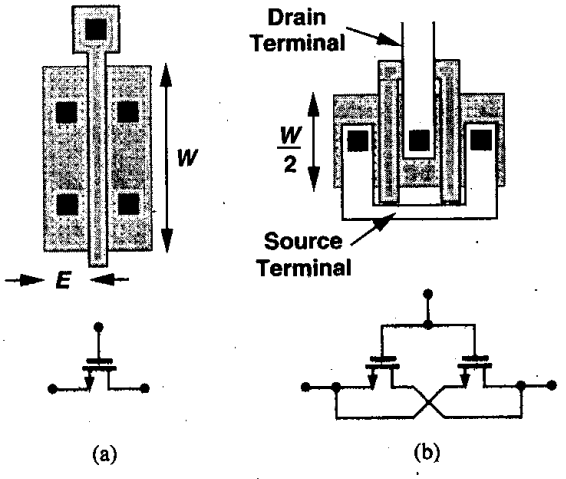


Figure 2.32

Solution

For the transistor in Fig. 2.32(a), we have

$$C_{DB} = C_{SB} = WEC_j + 2(W + E)C_{jsw}, \tag{2.31}$$

whereas for that in Fig. 2.32(b),

$$C_{DB} = \frac{W}{2}EC_j + 2\left(\frac{W}{2} + E\right)C_{jsw}, \tag{2.32}$$

$$C_{SB} = 2\left[\frac{W}{2}EC_j + 2\left(\frac{W}{2} + E\right)C_{jsw}\right] \tag{2.33}$$

$$= WEC_j + 2(W + 2E)C_{jsw}. \tag{2.34}$$

Called a "folded" structure, the geometry in Fig. 2.32(b) exhibits substantially less drain junction capacitance than that in Fig. 2.32(a) while providing the same W/L.

In the above calaculations, we have assumed that the total source or drain perimeter, $2(W + E)$, is multiplied by C_{jsw}. In reality, the capacitance of the sidewall facing the channel may be less than that of the other three sidewalls because of the channel-stop implant (Chapter 17). Nonetheless, we typically assume all four sides have the same unit capacitance. The error resulting from this assumption is negligible because each node in a circuit is connected to a number of other device capacitances as well.

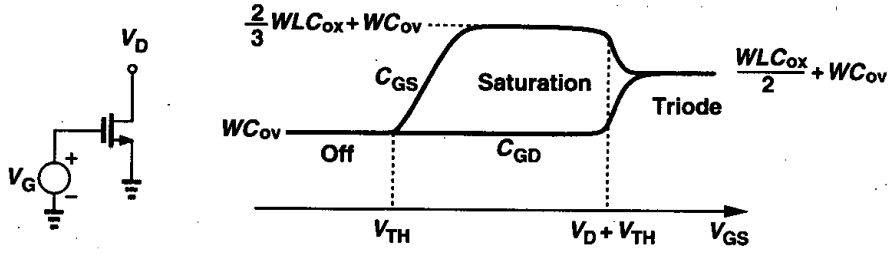


Figure 2.33 Variation of gate-source and gate-drain capacitances versus V_{GS}.

We now derive the capacitances between terminals of a MOSFET in different regions of operation. If the device is off, $C_{GD} = C_{GS} = C_{ov}W$, and the gate-bulk capacitance consists of the series combination of the gate oxide capacitance and the depletion region capacitance, i.e., $C_{GB} = (WLC_{ox})C_d/(WLC_{ox} + C_d)$, where L is the effective length and $C_d = WL\sqrt{q\epsilon_{si}N_{sub}/(4\Phi_F)}$. The value of C_{SB} and C_{DB} is a function of the source and drain voltages with respect to the substrate.

If the device is in deep triode region, i.e., if S and D have approximately equal voltages, then the gate-channel capacitance, WLC_{ox}, is divided equally between the gate and source terminals and the gate and drain terminals. This is because a change ΔV in the gate voltage draws equal amounts of charge from S and D. Thus, $C_{GD} = C_{GS} = WLC_{ox}/2 + WC_{ov}$.

If in saturation, a MOSFET exhibits a gate-drain capacitance of roughly WC_{ov}. The potential difference between the gate and the channel varies from V_{GS} at the source to $V_{GS} - V_{TH}$ at the pinch-off point, resulting in a nonuniform vertical electric field in the gate oxide along the channel. It can be proved that the equivalent capacitance of this structure excluding the gate-source overlap capacitance equals $2WLC_{ox}/3$ [1]. Thus, $C_{GS} = 2WL_{eff}C_{ox}/3 + WC_{ov}$. The behavior of C_{GD} and C_{GS} in different regions of operation is plotted in Fig. 2.33. Note that the above equations do not provide a smooth transition from one region of operation to another, creating convergence difficulties in simulation programs. This issue is revisited in Chapter 16.

The gate-bulk capacitance is usually neglected in the triode and saturation regions because the inversion layer acts as a "shield" between the gate and the bulk. In other words, if the gate voltage varies, the charge is supplied by the source and the drain rather than the bulk.

Example 2.7

Sketch the capacitances of M_1 in Fig. 2.34 as V_X varies from zero to 3 V. Assume $V_{TH} = 0.6$ V and $\lambda = \gamma = 0$.

Solution

To avoid confusion, we label the three terminals as shown in Fig. 2.34. For $V_X \approx 0$, M_1 is in the triode region, $C_{EN} \approx C_{EF} = (1/2)WLC_{ox} + WC_{ov}$, and C_{FB} is maximum. The value of C_{NB} is independent of V_X. As V_X exceeds 1 V, the role of the source and drain is exchanged [Fig. 2.35(a)],

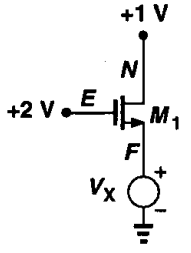


Figure 2.34

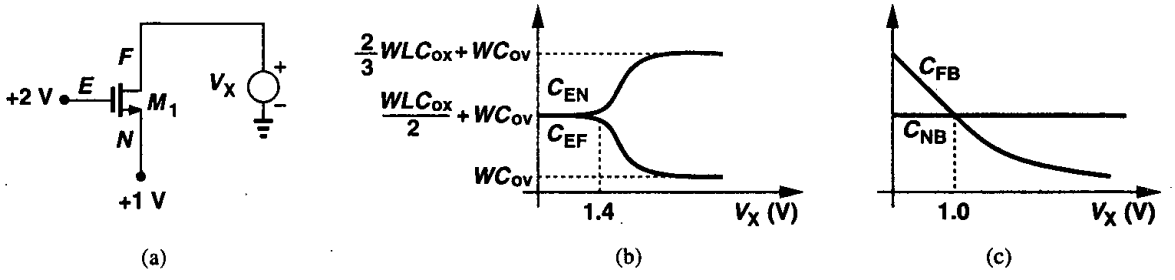


Figure 2.35

eventually bringing M_1 out of the triode region for $V_X \geq 2\,\text{V} - 0.6\,\text{V}$. The variation of the capacitances is plotted in Figs. 2.35(b) and (c).

2.4.3 MOS Small-Signal Model

The quadratic characteristics described by (2.8) and (2.9) along with the voltage-dependent capacitances derived above form the large-signal model of MOSFETs. Such a model proves essential in analyzing circuits in which the signal significantly disturbs the bias points, particularly if nonlinear effects are of concern. By contrast, if the perturbation in bias conditions is small, a small-signal model, i.e., an approximation of the large-signal model around the operating point, can be employed to simplify the calculations. Since in many analog circuits, MOSFETs are biased in the saturation region, we derive the corresponding small-signal model here. For transistors operating as switches, a linear resistor given by (2.11) together with device capacitances serves as a rough small-signal equivalent.

We derive the small-signal model by producing a small increment in a bias point and calculating the resulting increment in other bias parameters. Since the drain current is a function of the gate-source voltage, we incorporate a voltage-dependent current source equal to $g_m V_{GS}$ [Fig. 2.36(a)]. Note that the low-frequency impedance between G and S is very high. This is the small-signal model of an ideal MOSFET.

Owing to channel-length modulation, the drain current also varies with the drain-source voltage. This effect can also be modeled by a voltage-dependent current source [Fig. 2.36(b)], but a current source whose value linearly depends on the voltage across it is equivalent to

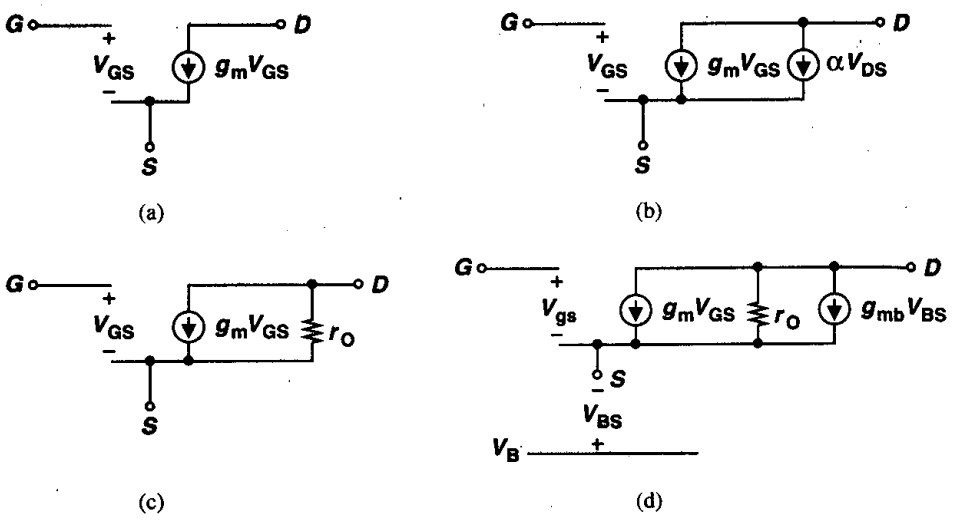


Figure 2.36 (a) Basic MOS small-signal model, (b) channel-length modulation represented by a dependent current source, (c) channel-length modulation represented by a resistor, (d) body effect represented by a dependent current source.

a linear resistor [Fig. 2.36(c)]. Tied between D and S, the resistor is given by

$$r_O = \frac{\partial V_{DS}}{\partial I_D} \tag{2.35}$$

$$= \frac{1}{\partial I_D / \partial V_{DS}}. \tag{2.36}$$

$$= \frac{1}{\frac{1}{2}\mu_n C_{ox} \frac{W}{L}(V_{GS} - V_{TH})^2 \cdot \lambda} \tag{2.37}$$

$$\approx \frac{1}{\lambda I_D}. \tag{2.38}$$

As seen throughout this book, the output resistance, r_O, impacts the performance of many analog circuits. For example, r_O limits the maximum voltage gain of most amplifiers.

Now recall that the bulk potential influences the threshold voltage and hence the gate-source overdrive. As demonstrated in Example 2.3, with all other terminals held at a constant voltage, the drain current is a function of the bulk voltage. That is, the bulk behaves as a second gate. Modeling this dependence by a current source connected between D and S [Fig. 2.36(d)], we write the value as $g_{mb} V_{bs}$, where $g_{mb} = \partial I_D / \partial V_{BS}$. In the saturation region, g_{mb} can be expressed as:

$$g_{mb} = \frac{\partial I_D}{\partial V_{BS}} \tag{2.39}$$

$$= \mu_n C_{ox} \frac{W}{L}(V_{GS} - V_{TH})\left(-\frac{\partial V_{TH}}{\partial V_{BS}}\right). \tag{2.40}$$

We also have

$$\frac{\partial V_{TH}}{\partial V_{BS}} = -\frac{\partial V_{TH}}{\partial V_{SB}} \tag{2.41}$$

$$= -\frac{\gamma}{2}(2\Phi_F + V_{SB})^{-1/2}. \tag{2.42}$$

Thus,

$$g_{mb} = g_m \frac{\gamma}{2\sqrt{2\Phi_F + V_{SB}}} \tag{2.43}$$

$$= \eta g_m, \tag{2.44}$$

where $\eta = g_{mb}/g_m$. As expected, g_{mb} is proportional to γ. Equation (2.43) also suggests that incremental body effect becomes less pronounced as V_{SB} increases. Note that $g_m V_{GS}$ and $g_{mb} V_{BS}$ have the same polarity, i.e., raising the gate voltage has the same effect as raising the bulk potential.

The model in Fig. 2.36(d) is adequate for most low-frequency small-signal analyses. In reality, each terminal of a MOSFET exhibits a finite ohmic resistance resulting from the resistivity of the material (and the contacts), but proper layout can minimize such resistances. For example, consider the two structures of Fig. 2.32, repeated in Fig. 2.37 along with the gate distributed resistance. We note that folding reduces the gate resistance by a factor of four.

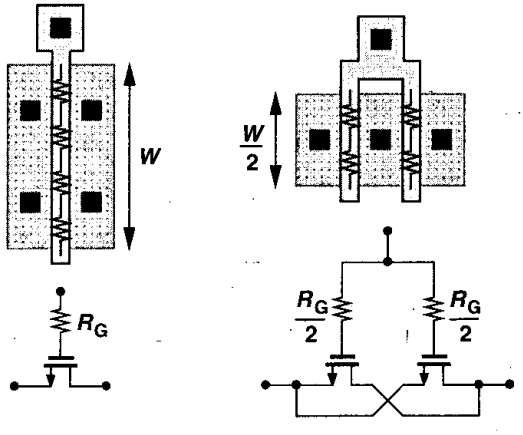


Figure 2.37 Reduction of gate resistance by folding.

Shown in Fig. 2.38, the complete small-signal model includes the device capacitances as well. The value of each capacitance is calculated according to the equations derived in Section 2.4.2. The reader may wonder how a complex circuit is analyzed intuitively if each transistor must be replaced by the model of Fig. 2.38. The first step is to determine

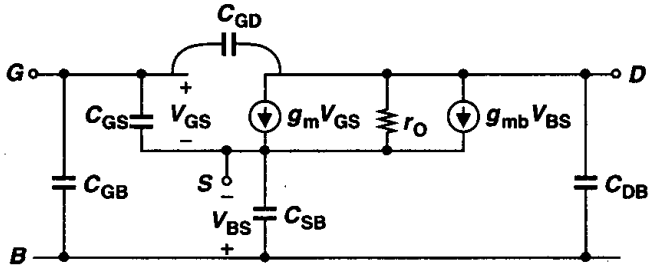
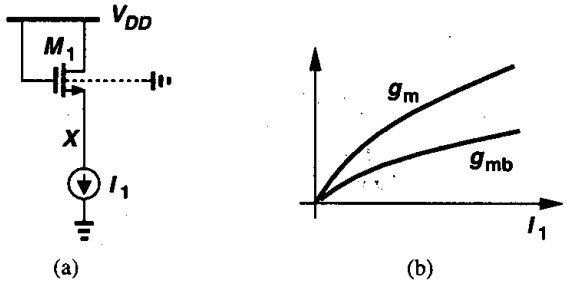


Figure 2.38 Complete MOS small-signal model.

the *simplest* device model that can represent the role of each transistor with reasonable accuracy. We provide some guidelines for this task at the end of Chapter 3.

Example 2.8

Sketch g_m and g_{mb} of M_1 in Fig. 2.39 as a function of the bias current I_1.



(a) (b)

Figure 2.39

Solution

Since $g_m = \sqrt{2\mu_n C_{ox}(W/L)I_D}$, we have $g_m \propto \sqrt{I_1}$. The dependence of g_{mb} upon I_1 is less straightforward. As I_1 increases, V_X decreases and so does V_{SB}.

Unless otherwise stated, in this book we assume the bulk of all NFETs is tied to the most negative supply (usually the ground) and that of PFETs to the most positive supply (usually V_{DD}).

2.4.4 MOS SPICE models

In order to represent the behavior of transistors in circuit simulations, SPICE requires an accurate model for each device. Over the last two decades, MOS modeling has made tremendous progress, reaching quite sophisticated levels so as to represent high-order effects in short-channel devices.

Table 2.1 Level 1 SPICE Models for NMOS and PMOS Devices.

NMOS Model			
LEVEL = 1	VTO = 0.7	GAMMA = 0.45	PHI = 0.9
NSUB = 9e+14	LD = 0.08e−6	UO = 350	LAMBDA = 0.1
TOX = 9e−9	PB = 0.9	CJ = 0.56e−3	CJSW = 0.35e−11
MJ = 0.45	MJSW = 0.2	CGDO = 0.4e−9	JS = 1.0e−8
PMOS Model			
LEVEL = 1	VTO = −0.8	GAMMA = 0.4	PHI = 0.8
NSUB = 5e+14	LD = 0.09e−6	UO = 100	LAMBDA = 0.2
TOX = 9e−9	PB = 0.9	CJ = 0.94e−3	CJSW = 0.32e−11
MJ = 0.5	MJSW = 0.3	CGDO = 0.3e−9	JS = 0.5e−8

In this section, we describe the simplest MOS SPICE model, known as "Level 1," and provide typical values for each parameter in the model corresponding to a 0.5-μm technology. Chapter 16 describes more accurate SPICE models. Table 2.1 shows the model parameters for NMOS and PMOS devices. The parameters are defined as below:

VTO: threshold voltage with zero V_{SB} (unit: V)
GAMMA: body effect coefficient (unit: $V^{1/2}$)
PHI: $2\Phi_F$ (unit: V)
TOX: gate oxide thickness (unit: m)
NSUB: substrate doping (unit: cm^{-3})
LD: source/drain side diffusion (unit: m)
UO: channel mobility (unit: $cm^2/V/s$)
LAMBDA: channel-length modulation coefficient (unit: V^{-1})
CJ: source/drain bottom-plate junction capacitance per unit area (unit: F/m^2)
CJSW: source/drain sidewall junction capacitance per unit length (unit: F/m)
PB: source/drain junction built-in potential (unit: V)
MJ: exponent in CJ equation (unitless)
MJSW: exponent in CJSW equation (unitless)
CGDO: gate-drain overlap capacitance per unit width (unit: F/m)
CGSO: gate-source overlap capacitance per unit width (unit: F/m)
JS: source/drain leakage current per unit area (unit: A/m^2)

2.4.5 NMOS versus PMOS Devices

In most CMOS technologies, PMOS devices are quite inferior to NMOS transistors. For example, due to the lower mobility of holes, $\mu_p C_{ox} \approx 0.25\mu_n C_{ox}$ in modern processes, yielding low current drive and transconductance. Moreover, for given dimensions and bias currents, NMOS transistors exhibit a higher output resistance, providing more ideal current sources and higher gain in amplifiers. For these reasons, it is preferred to incorporate NFETs rather than PFETs wherever possible.

2.4.6 Long-Channel versus Short-Channel Devices

In this chapter, we have employed a very simple view of MOSFETs so as to understand the basic principles of their operation. Most of our treatment is valid for "long-channel" devices, e.g., transistors having a minimum length of about 4 μm. Many of the relationships derived here must be reexamined and revised for short-channel MOSFETs. Furthermore, the SPICE models necessary for simulation of today's devices need to be much more sophisticated than the Level 1 model. For example, the intrinsic gain, $g_m r_O$, calculated from the device parameters in Table 2.1 is quite higher than actual values. These issues are studied in Chapter 16.

The reader may wonder why we begin with a simplistic view of devices if such a view does not lead to a high accuracy in predicting the performance of circuits. The key point is that the simple model provides a great deal of intuition that is necessary in analog design. As we will see throughout this book, we often encounter a trade-off between intuition and rigor, and our approach is to establish the intuition first and gradually complete our understanding so as to achieve rigor as well.

Appendix A: Behavior of MOS Device as a Capacitor

In this chapter, we have limited our treatment of MOS devices to a basic level. However, the behavior of a MOSFET as a capacitor merits some attention. Recall that if the source, drain, and bulk of an NFET are grounded and the gate voltage rises, an inversion layer begins to form for $V_{GS} \approx V_{TH}$. We also noted that for $0 < V_{GS} < V_{TH}$, the device operates in the subthreshold region.

Now consider the NFET of Fig. 2.40. The transistor can be considered a two-terminal

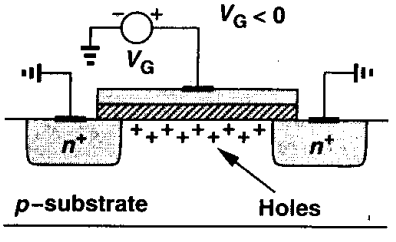


Figure 2.40 NMOS operating in accumulation mode.

device and hence its capacitance can be examined for different gate voltages. Let us begin with a very *negative* gate-source voltage. The negative potential on the gate attracts the holes in the substrate to the oxide interface. We say the MOSFET operates in the "accumulation" region. The two-terminal device can be viewed as a capacitor having a unit-area capacitance of C_{ox} because the two "plates" of the capacitor are separated by t_{ox}.

As V_{GS} rises, the density of holes at the interface falls, a depletion region begins to form under the oxide, and the device enters weak inversion. In this mode, the capacitance consists of the series combination of C_{ox} and C_{dep}. Finally, as V_{GS} exceeds V_{TH}, the oxide-silicon interface sustains a channel and the unit-area capacitance returns to C_{ox}. Figure 2.41 plots the behavior.

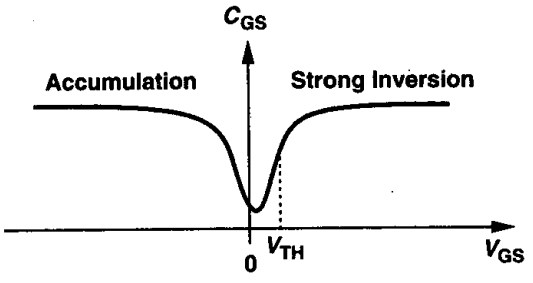


Figure 2.41 Capacitance-voltage characteristic of an NMOS device.

Problems

Unless otherwise stated, in the following problems, use the device data shown in Table 2.1 and assume $V_{DD} = 3$ V where necessary.

2.1. For $W/L = 50/0.5$, plot the drain current of an NFET and a PFET as a function of $|V_{GS}|$ as $|V_{GS}|$ varies from 0 to 3 V. Assume $|V_{DS}| = 3$ V.

2.2. For $W/L = 50/0.5$, and $|I_D| = 0.5$ mA, calculate the transconductance and output impedance of both NMOS and PMOS devices. Also, find the "intrinsic gain," defined as $g_m r_O$.

2.3. Derive expressions for $g_m r_O$ in terms of I_D and W/L. Plot $g_m r_O$ as a function of I_D with L as a parameter. Note that $\lambda \propto 1/L$.

2.4. Plot I_D versus V_{GS} for an MOS transistor (a) with V_{DS} as a parameter, (b) with V_{BS} as a parameter. Identify the break points in the characteristics.

2.5. Sketch I_X and the transconductance of the transistor as a function of V_X for each circuit in Fig. 2.42 as V_X varies from 0 to V_{DD}. For part (a), assume V_X varies from 0 to 1.5 V.

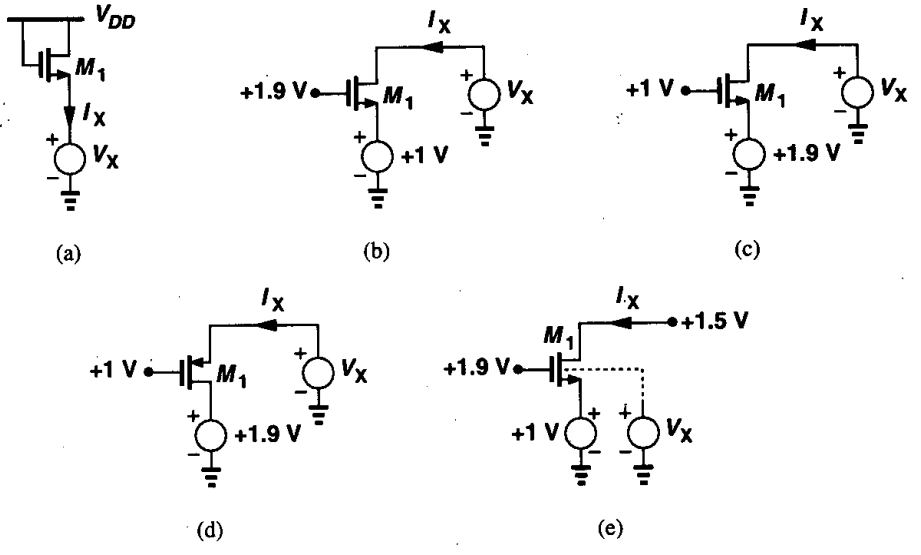


Figure 2.42

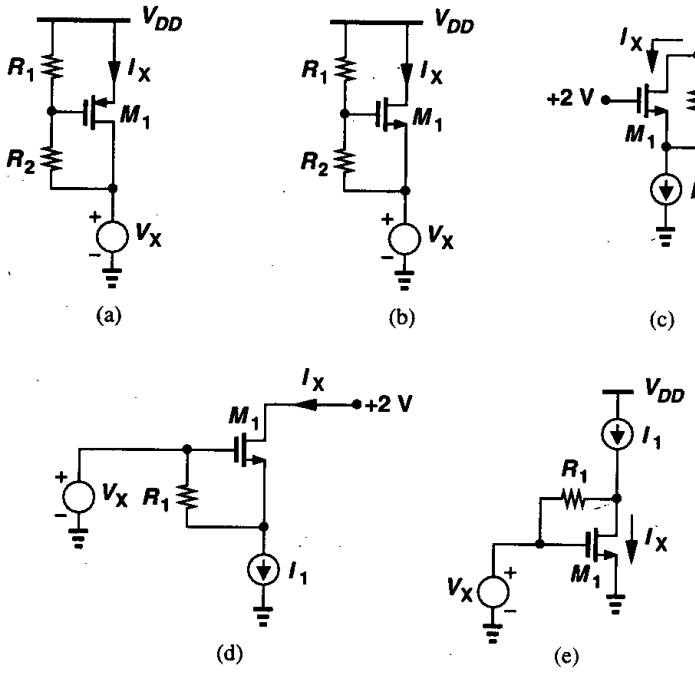


Figure 2.43

2.6. Sketch I_X and the transconductance of the transistor as a function of V_X for each circuit in Fig. 2.43 as V_X varies from 0 to V_{DD}.

2.7. Sketch V_{out} as a function of V_{in} for each circuit in Fig. 2.44 as V_{in} varies from 0 to V_{DD}.

2.8. Sketch V_{out} as a function of V_{in} for each circuit in Fig. 2.45 as V_{in} varies from 0 to V_{DD}.

2.9. Sketch V_X and I_X as a function of time for each circuit in Fig. 2.46. The initial voltage of C_1 is equal to 3 V.

2.10. Sketch V_X and I_X as a function of time for each circuit in Fig. 2.47. The initial voltages of C_1 and C_2 are equal to 1 V and 3 V, respectively.

2.11. Sketch V_X as a function of time for each circuit in Fig. 2.48. The initial voltage of each capacitor is shown.

2.12. Sketch V_X as a function of time for each circuit in Fig. 2.49. The initial voltage of each capacitor is shown.

2.13. The transit frequency, f_T, of a MOSFET is defined as the frequency at which the small-signal current gain of the device drops to unity while the source and drain terminals are held at ac ground.

(a) Prove that

$$f_T = \frac{g_m}{2\pi(C_{GD} + C_{GS})}. \tag{2.45}$$

Note that f_T does not include the effect of the S/D junction capacitance.

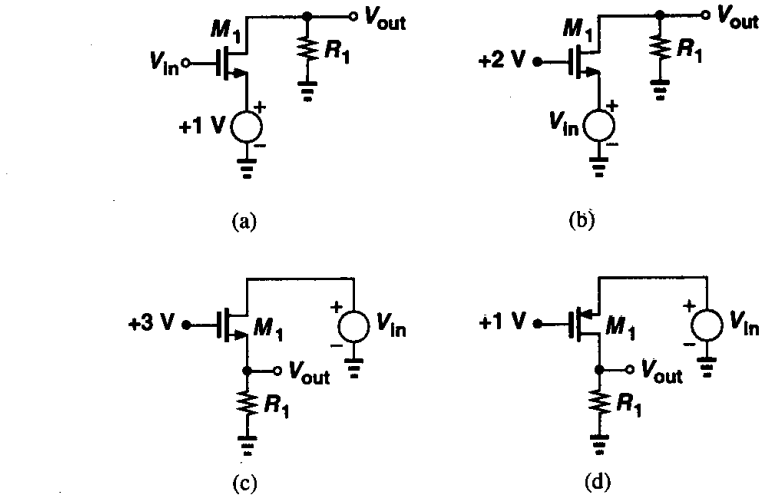


Figure 2.44

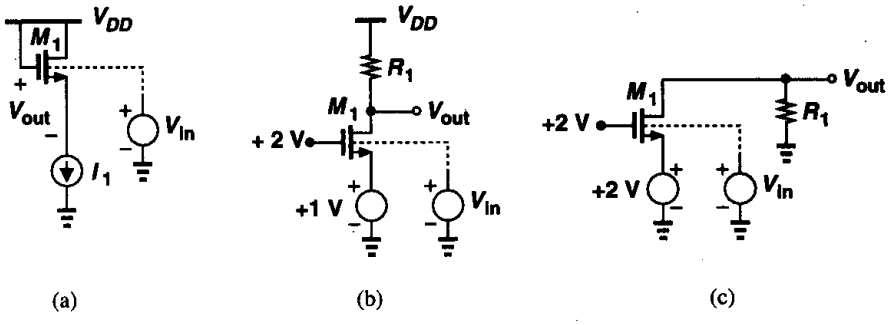


Figure 2.45

(b) Suppose the gate resistance, R_G, is significant and the device is modeled as a distributed set of n transistors each with a gate resistance equal to R_G/n. Prove that the f_T of the device is independent of R_G and still equal to the value given above.

(c) For a given bias current, the minimum allowable drain-source voltage for operation in saturation can be reduced only by increasing the width and hence the capacitances of the transistor. Using square-law characteristics, prove that

$$f_T = \frac{\mu_n}{2\pi} \frac{V_{GS} - V_{TH}}{L^2}. \tag{2.46}$$

This relation indicates how the speed is limited as a device is designed to operate with lower supply voltages.

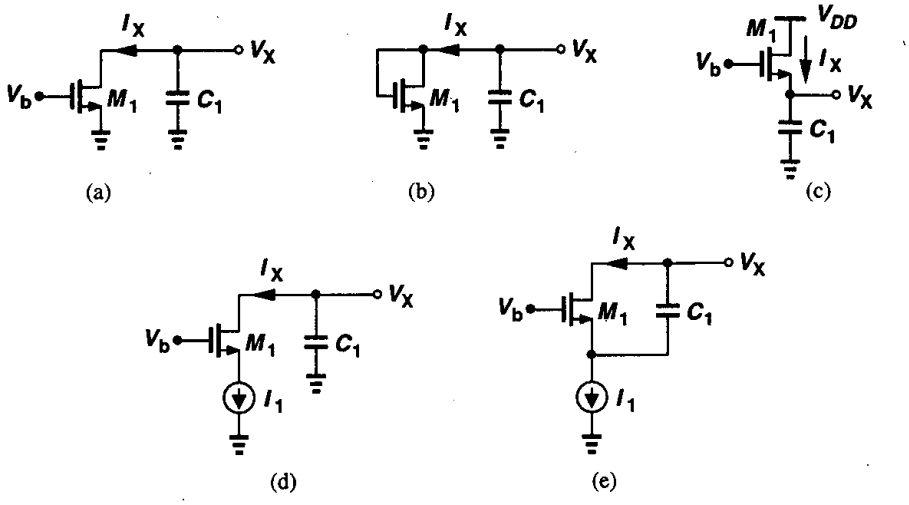


Figure 2.46

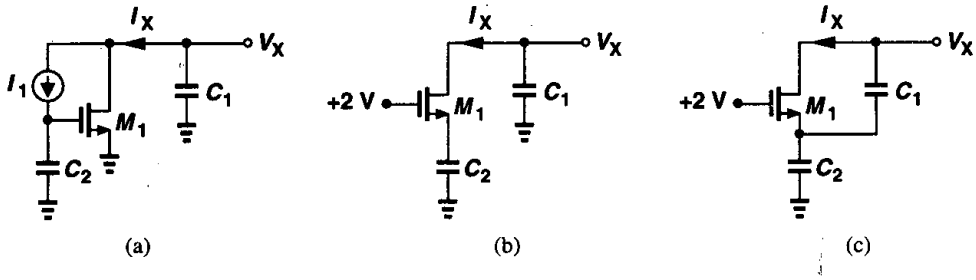


Figure 2.47

2.14. Calculate the f_T of a MOS device in the subthreshold region and compare the result with those obtained in Problem 2.13.

2.15. For a saturated NMOS device having $W = 50\ \mu$m and $L = 0.5\ \mu$m, calculate all of the capacitances. Assume the minimum (lateral) dimension of the S/D areas is 1.5 μm and the device is folded as shown in Fig. 2.32(b). What is the f_T if the drain current is 1 mA?

2.16. Consider the structure shown in Fig. 2.50. Determine I_D as a function of V_{GS} and V_{DS} and prove that the structure can be viewed as a single transistor having an aspect ratio $W/(2L)$. Assume $\lambda = \gamma = 0$.

2.17. For an NMOS device operating in saturation, plot W/L versus $V_{GS} - V_{TH}$ if (a) I_D is constant, (b) g_m is constant.

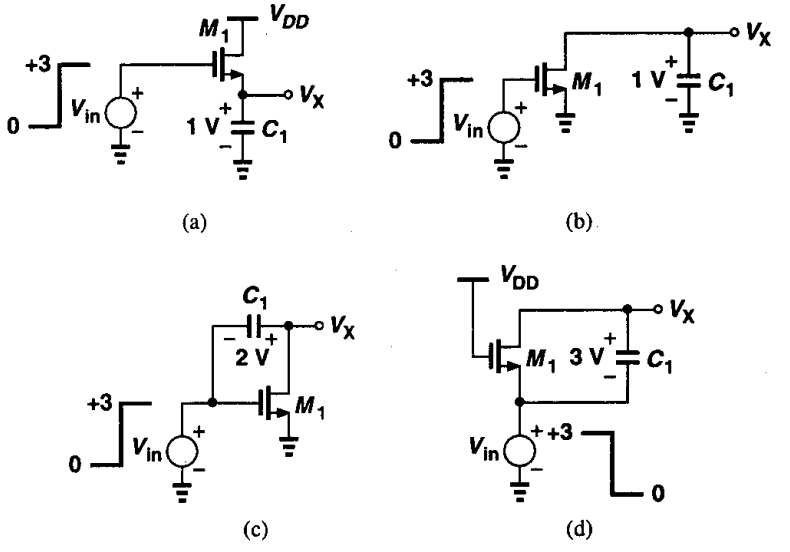


Figure 2.48

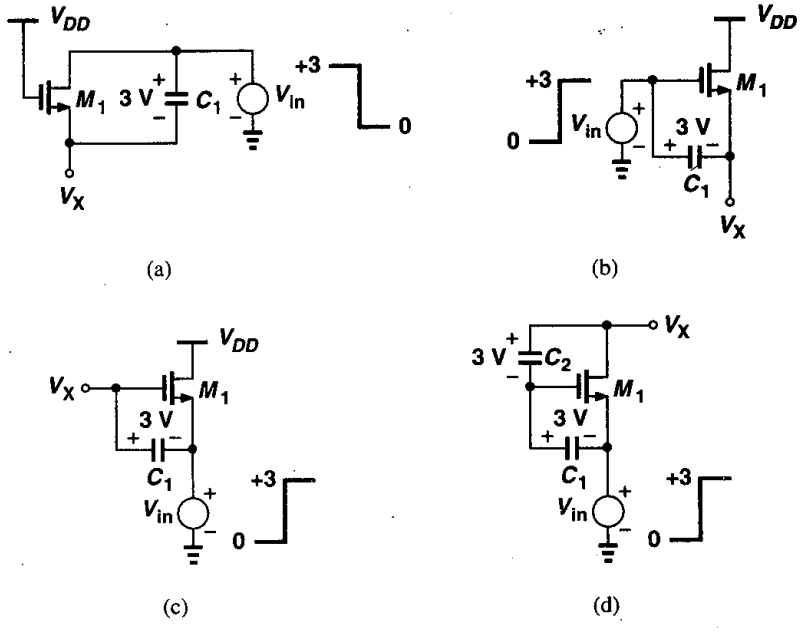


Figure 2.49

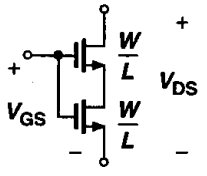
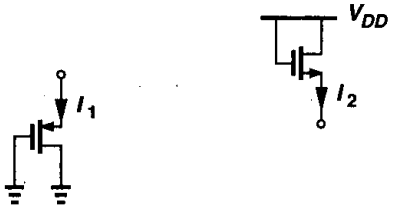


Figure 2.50

2.18. Explain why the structures shown in Fig. 2.51 cannot operate as current sources even though the transistors are in saturation.



(a) (b)

Figure 2.51

2.19. Considering the body effect as "backgate effect," explain intuitively why γ is directly proportional to $\sqrt{N_{sub}}$ and inversely proportional to C_{ox}.

2.20. A "ring" MOS structure is shown in Fig. 2.52. Explain how the device operates and estimate its equivalent aspect ratio. Compare the drain junction capacitance of this structure with that of the devices shown in Fig. 2.32.

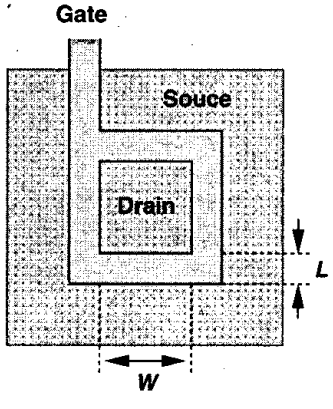


Figure 2.52

2.21. Suppose we have received an NMOS transistor in a package with four unmarked pins. Describe the minimum number of dc measurement steps using an ohmmeter necessary to determine the gate, source/drain, and bulk terminals of the device.

2.22. Repeat Problem 2.21 if the type of the device (NFET or PFET) is not known.

2.23. For an NMOS transistor, the threshold voltage is known but $\mu_n C_{ox}$ and W/L are not. Assume $\lambda = \gamma = 0$. If we cannot measure C_{ox} independently, is it possible to devise a sequence of dc measurement tests to determine $\mu_n C_{ox}$ and W/L? What if we have two transistors and we know one has twice the aspect ratio of the other?

2.24. Sketch I_X versus V_X for each of the composite structures shown in Fig. 2.53 with V_G as a parameter. Also, sketch the equivalent transconductance. Assume $\lambda = \gamma = 0$.

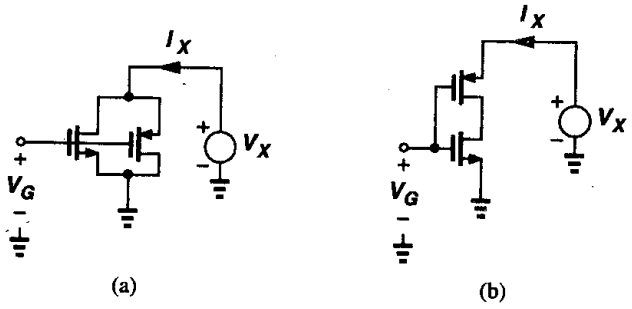


Figure 2.53

2.25. An NMOS current source with $I_D = 0.5$ mA must operate with drain-source voltages as low as 0.4 V. If the minimum required output impedance is 20 kΩ, determine the width and length of the device. Calculate the gate-source, gate-drain, and drain-substrate capacitance if the device is folded as in Fig. 2.32 and $E = 3\ \mu$m.

2.26. Consider the circuit shown in Fig. 2.54, where the initial voltage at node X is equal to V_{DD}. Assuming $\lambda = \gamma = 0$ and neglecting other capacitances, plot V_X and V_Y versus time if (a) V_{in} is a positive step with amplitude $V_0 > V_{TH}$, (b) V_{in} is a negative step with amplitude $V_0 = V_{TH}$.

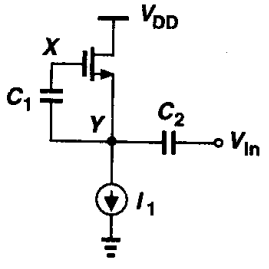


Figure 2.54

2.27. An NMOS device operating in the subthreshold region has a ζ of 1.5. What variation in V_{GS} results in a ten-fold change in I_D? If $I_D = 10\ \mu$A, what is g_m?

2.28. Consider an NMOS device with $V_G = 1.5$ V and $V_S = 0$. Explain what happens if we continually decrease V_D below zero or increase V_{sub} above zero.

References

1. R.S. Muller and T. I. Kamins, *Device Electronics for Integrated Circuits,* Second Ed., New York: Wiley, 1986.
2. Y. Tsividis, *Operation and Modeling of the MOS Transistor,* Second Ed., Boston: McGraw-Hill, 1999.
3. Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices,* New York: Cambridge University Press, 1998.

Chapter 3

Single-Stage Amplifiers

Amplification is an essential function in most analog (and many digital) circuits. We amplify an analog or digital signal because it may be too small to drive a load, overcome the noise of a subsequent stage, or provide logical levels to a digital circuit. Amplification also plays a critical role in feedback systems (Chapter 8).

In this chapter, we study the low-frequency behavior of single-stage CMOS amplifiers. Analyzing both the large-signal and the small-signal characteristics of each circuit, we develop intuitive techniques and models that prove useful in understanding more complex systems. An important part of a designer's job is to use proper approximations so as to create a simple mental picture of a complicated circuit. The intuition thus gained makes it possible to formulate the behavior of most circuits by inspection rather than by lengthy calculations.

Following a brief review of basic concepts, we describe in this chapter four types of amplifiers: common-source and common-gate topologies, source followers, and cascode configurations. In each case, we begin with a simple model and gradually add second-order phenomena such as channel-length modulation and body effect.

3.1 Basic Concepts

The input-output characteristic of an amplifier is generally a nonlinear function (Fig. 3.1) that can be approximated by a polynomial over some signal range:

$$y(t) \approx \alpha_0 + \alpha_1 x(t) + \alpha_2 x^2(t) + \cdots + \alpha_n x^n(t) \quad x_1 \leq x \leq x_2. \tag{3.1}$$

The input and output may be current or voltage quantities. For a sufficiently narrow range of x,

$$y(t) \approx \alpha_0 + \alpha_1 x(t), \tag{3.2}$$

where α_0 can be considered the operating (bias) point and α_1 the small-signal gain. So long as $\alpha_1 x(t) \ll \alpha_0$, the bias point is disturbed negligibly, (3.2) provides a reasonable

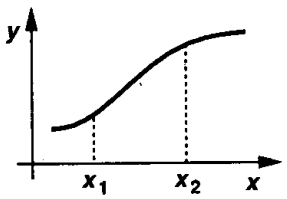


Figure 3.1 Input-output characteristic of a nonlinear system.

approximation, and higher order terms are insignificant. In other words, $\Delta y = \alpha_1 \Delta x$, indicating a linear relationship between the *increments* at the input and output. As $x(t)$ increases in magnitude, higher order terms manifest themselves, leading to nonlinearity and necessitating large-signal analysis. From another point of view, if the slope of the characteristic (the incremental gain) *varies* with the signal level, then the system is nonlinear. These concepts are described in detail in Chapter 13.

What aspects of the performance of an amplifier are important? In addition to gain and speed, such parameters as power dissipation, supply voltage, linearity, noise, or maximum voltage swings may be important. Furthermore, the input and output impedances determine how the circuit interacts with preceding and subsequent stages. In practice, most of these parameters trade with each other, making the design a multi-dimensional optimization problem. Illustrated in the "analog design octagon" of Fig. 3.2, such trade-offs present many challenges in the design of high-performance amplifiers, requiring intuition and experience to arrive at an acceptable compromise.

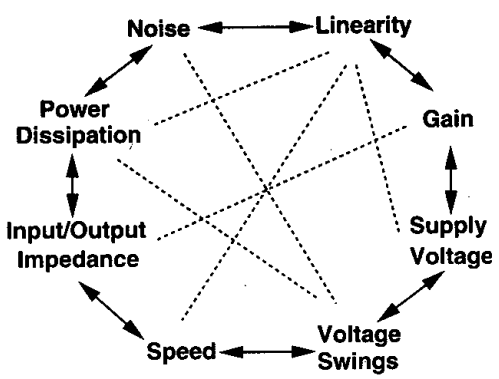


Figure 3.2 Analog design octagon.

3.2 Common-Source Stage

3.2.1 Common-Source Stage with Resistive Load

By virtue of its transconductance, a MOSFET converts variations in its gate-source voltage to a small-signal drain current, which can pass through a resistor to generate an output voltage. Shown in Fig. 3.3(a), the common-source (CS) stage performs such an operation.

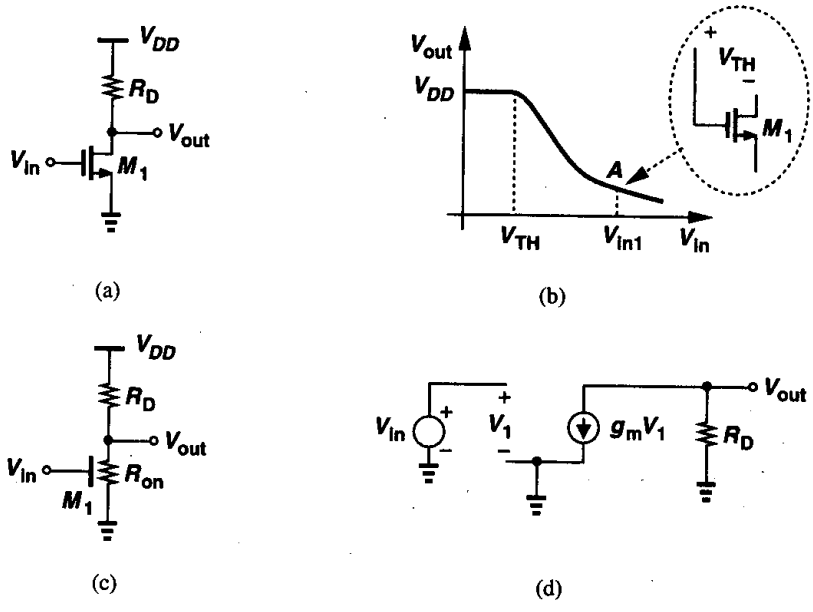


Figure 3.3 (a) Common-source stage, (b) input-output characteristic, (c) equivalent circuit in deep triode region, (d) small-signal model for the saturation region.

We study both the large-signal and the small-signal behavior of the circuit. Note that the input impedance of the circuit is very high at low frequencies.

If the input voltage increases from zero, M_1 is off and $V_{out} = V_{DD}$ [Fig. 3.3(b)]. As V_{in} approaches V_{TH}, M_1 begins to turn on, drawing current from R_D and lowering V_{out}. If V_{DD} is not excessively low, M_1 turns on in saturation, and we have

$$V_{out} = V_{DD} - R_D \frac{1}{2}\mu_n C_{ox} \frac{W}{L}(V_{in} - V_{TH})^2, \tag{3.3}$$

where channel-length modulation is neglected. With further increase in V_{in}, V_{out} drops more and the transistor continues to operate in saturation until V_{in} exceeds V_{out} by V_{TH} [point A in Fig. 3.3(b)]. At this point,

$$V_{in1} - V_{TH} = V_{DD} - R_D \frac{1}{2}\mu_n C_{ox} \frac{W}{L}(V_{in1} - V_{TH})^2, \tag{3.4}$$

from which $V_{in1} - V_{TH}$ and hence V_{out} can be calculated.

For $V_{in} > V_{in1}$, M_1 is in the triode region:

$$V_{out} = V_{DD} - R_D \frac{1}{2}\mu_n C_{ox} \frac{W}{L}\left[2(V_{in} - V_{TH})V_{out} - V_{out}^2\right]. \tag{3.5}$$

If V_{in} is high enough to drive M_1 into deep triode region, $V_{out} \ll 2(V_{in} - V_{TH})$, and, from the equivalent circuit of Fig. 3.3(c),

$$V_{out} = V_{DD}\frac{R_{on}}{R_{on} + R_D} \tag{3.6}$$

$$= \frac{V_{DD}}{1 + \mu_n C_{ox}\dfrac{W}{L}R_D(V_{in} - V_{TH})}. \tag{3.7}$$

Since the transconductance drops in the triode region, we usually ensure that $V_{out} > V_{in} - V_{TH}$, operating to the left of point A in Fig. 3.3(b). Using (3.3) as the input-output characteristic and viewing its slope as the small-signal gain, we have:

$$A_v = \frac{\partial V_{out}}{\partial V_{in}} \tag{3.8}$$

$$= -R_D\mu_n C_{ox}\frac{W}{L}(V_{in} - V_{TH}) \tag{3.9}$$

$$= -g_m R_D. \tag{3.10}$$

This result can be directly derived from the observation that M_1 converts an input voltage change ΔV_{in} to a drain current change $g_m \Delta V_{in}$, and hence an output voltage change $-g_m R_D \Delta V_{in}$. The small-signal model of Fig. 3.3(d) yields the same result.

Even though derived for small-signal operation, the equation $A_v = -g_m R_D$ predicts certain effects if the circuit senses a *large* signal swing. Since g_m itself varies with the input signal according to $g_m = \mu_n C_{ox}(W/L)(V_{GS} - V_{TH})$, the gain of the circuit changes substantially if the signal is large. In other words, if the gain of the circuit *varies* significantly with the signal swing, then the circuit operates in the large-signal mode. The dependence of the gain upon the signal level leads to nonlinearity (Chapter 13), usually an undesirable effect.

A key result here is that to minimize the nonlinearity, the gain equation must be a weak function of signal-dependent parameters such as g_m. We present several examples of this concept in this chapter and in Chapter 13.

Example 3.1

Sketch the drain current and transconductance of M_1 in Fig. 3.3(a) as a function of the input voltage.

Solution

The drain current becomes significant for $V_{in} > V_{TH}$, eventually approaching V_{DD}/R_D if $R_{on1} \ll R_D$ [Fig. 3.4(a)]. Since in saturation, $g_m = \mu_n C_{ox}(W/L)(V_{in} - V_{TH})$, the transconductance begins to rise for $V_{in} > V_{TH}$. In the triode region, $g_m = \mu_n C_{ox}(W/L)V_{DS}$, falling as V_{in} exceeds V_{in1} [Fig. 3.4(b)].

How do we maximize the voltage gain of a common-source stage? Writing (3.10) as

$$A_v = -\sqrt{2\mu_n C_{ox}\frac{W}{L}I_D}\frac{V_{RD}}{I_D}, \tag{3.11}$$

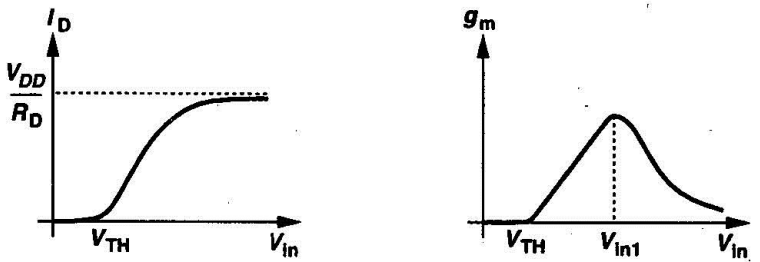


Figure 3.4

where V_{RD} denotes the voltage drop across R_D, we have

$$A_v = -\sqrt{2\mu_n C_{ox}\frac{W}{L}}\frac{V_{RD}}{\sqrt{I_D}}. \tag{3.12}$$

Thus, the magnitude of A_v can be increased by increasing W/L or V_{RD} or decreasing I_D if other parameters are constant. It is important to understand the trade-offs resulting from this equation. A larger device size leads to greater device capacitances, and a higher V_{RD} limits the maximum voltage swings. For example, if $V_{DD} - V_{RD} = V_{in} - V_{TH}$, then M_1 is at the edge of the triode region, allowing only very small swings at the output (and input). If V_{RD} remains constant and I_D is reduced, then R_D must increase, thereby leading to a greater time constant at the output node. In other words, as noted in the analog design octagon, the circuit exhibits trade-offs between gain, bandwidth, and voltage swings. Lower supply voltages further tighten these trade-offs.

For large values of R_D, the effect of channel length modulation in M_1 becomes significant. Modifying (3.4) to include this effect,

$$V_{out} = V_{DD} - R_D\frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{in} - V_{TH})^2(1 + \lambda V_{out}), \tag{3.13}$$

we have

$$\begin{aligned}\frac{\partial V_{out}}{\partial V_{in}} &= -R_D\mu_n C_{ox}\frac{W}{L}(V_{in} - V_{TH})(1 + \lambda V_{out}) \\ &\quad -R_D\frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{in} - V_{TH})^2\lambda\frac{\partial V_{out}}{\partial V_{in}}.\end{aligned} \tag{3.14}$$

Using the approximation $I_D \approx (1/2)\mu_n C_{ox}(W/L)(V_{in} - V_{TH})^2$, we obtain:

$$A_v = -R_D g_m - R_D I_D \lambda A_v \tag{3.15}$$

and hence

$$A_v = -\frac{g_m R_D}{1 + R_D \lambda I_D}. \tag{3.16}$$

Since $\lambda I_D = 1/r_O$,

$$A_v = -g_m \frac{r_O R_D}{r_O + R_D}. \tag{3.17}$$

The small-signal model of Fig. 3.5 gives the same result with much less effort. That is, since

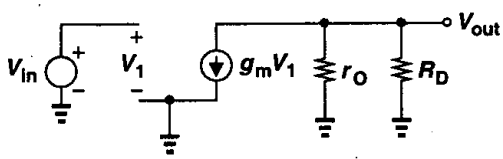


Figure 3.5 Small-signal model of CS stage including the transistor output resistance.

$g_m V_1(r_O \| R_D) = -V_{out}$ and $V_1 = V_{in}$, we have $V_{out}/V_{in} = -g_m(r_O \| R_D)$. Note that, as mentioned in Chapter 1, V_{in}, V_1, and V_{out} in this figure denote small-signal quantities.

Example 3.2

Assuming M_1 in Fig. 3.6 is biased in saturation, calculate the small-signal voltage gain of the circuit.

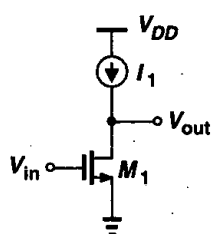


Figure 3.6

Solution

Since I_1 introduces an infinite impedance, the gain is limited by the output resistance of M_1:

$$A_v = -g_m r_O. \tag{3.18}$$

Called the "intrinsic gain" of a transistor, this quantity represents the maximum voltage gain that can be achieved using a single device. In today's CMOS technology, $g_m r_O$ of short-channel devices is between roughly 10 and 30. Thus, we usually assume $1/g_m \ll r_O$.

In Fig. 3.6, Kirchhoff's current law (KCL) requires that $I_{D1} = I_1$. Then, how can V_{in} change the current of M_1 if I_1 is constant? Writing the total drain current of M_1 as

$$I_{D1} = \frac{1}{2}\mu_n C_{ox}(V_{in} - V_{TH})^2(1 + \lambda V_{out}) \tag{3.19}$$

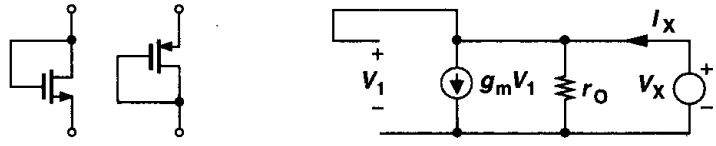
$$= I_1, \tag{3.20}$$

we note that V_{in} appears in the square term and V_{out} in the linear term. As V_{in} increases, V_{out} must decrease such that the product remains constant. We may nevertheless say "I_{D1} increases as V_{in} increases." This statement simply refers to the quadratic part of the equation.

3.2.2 CS Stage with Diode-Connected Load

In many CMOS technologies, it is difficult to fabricate resistors with tightly-controlled values or a reasonable physical size (Chapter 17). Consequently, it is desirable to replace R_D in Fig. 3.3(a) with a MOS transistor.

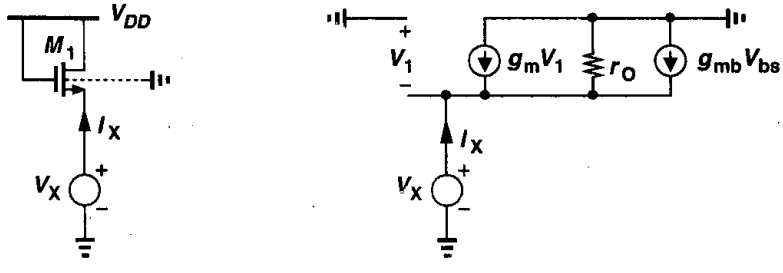
A MOSFET can operate as a small-signal resistor if its gate and drain are shorted [Fig. 3.7(a)]. Called a "diode-connected" device in analogy with its bipolar counterpart,



(a) (b)

Figure 3.7 (a) Diode-connected NMOS and PMOS devices, (b) small-signal equivalent circuit.

this configuration exhibits a small-signal behavior similar to a two-terminal resistor. Note that the transistor is always in saturation because the drain and the gate have the same potential. Using the small-signal equivalent shown in Fig. 3.7(b) to obtain the impedance of the device, we write $V_1 = V_X$ and $I_X = V_X/r_O + g_m V_X$. That is, the impedance of the diode is simply equal to $(1/g_m)\|r_O \approx 1/g_m$. If body effect exists, we can use the circuit in Fig. 3.8 to write $V_1 = -V_X$, $V_{bs} = -V_X$ and



(a) (b)

Figure 3.8 (a) Arrangement for measuring the equivalent resistance of a diode-connected MOSFET, (b) small-signal equivalent circuit.

$$(g_m + g_{mb})V_X + \frac{V_X}{r_O} = I_X. \tag{3.21}$$

It follows that

$$\frac{V_X}{I_X} = \frac{1}{g_m + g_{mb} + r_O^{-1}} \tag{3.22}$$

$$= \frac{1}{g_m + g_{mb}} \| r_O \tag{3.23}$$

$$\approx \frac{1}{g_m + g_{mb}}. \tag{3.24}$$

Interestingly, the impedance seen at the source of M_1 is *lower* when body effect is included. Intuitive explanation of this effect is left as an exercise for the reader.

We now study a common-source stage with a diode-connected load (Fig. 3.9). For negligible channel-length modulation, (3.24) can be substituted in (3.10) for the load impedance,

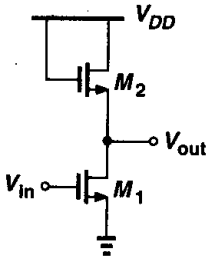


Figure 3.9 CS stage with diode-connected load.

yielding

$$A_v = -g_{m1} \frac{1}{g_{m2} + g_{mb2}} \tag{3.25}$$

$$= -\frac{g_{m1}}{g_{m2}} \frac{1}{1+\eta}, \tag{3.26}$$

where $\eta = g_{mb2}/g_{m2}$. Expressing g_{m1} and g_{m2} in terms of device dimensions and bias currents, we have

$$A_v = -\frac{\sqrt{2\mu_n C_{ox}(W/L)_1 I_{D1}}}{\sqrt{2\mu_n C_{ox}(W/L)_2 I_{D2}}} \frac{1}{1+\eta}, \tag{3.27}$$

and, since $I_{D1} = I_{D2}$,

$$A_v = -\sqrt{\frac{(W/L)_1}{(W/L)_2}} \frac{1}{1+\eta}. \tag{3.28}$$

This equation reveals an interesting property: if the variation of η with the output voltage is neglected, the gain is independent of the bias currents and voltages (so long as M_1 stays in saturation). In other words, as the input and output signal levels vary, the gain remains relatively constant, indicating that the input-output characteristic is relatively linear.

The linear behavior of the circuit can also be confirmed by large-signal analysis. Neglecting channel-length modulation for simplicity, we have in Fig. 3.9

$$\frac{1}{2}\mu_n C_{ox}\left(\frac{W}{L}\right)_1 (V_{in} - V_{TH1})^2 = \frac{1}{2}\mu_n C_{ox}\left(\frac{W}{L}\right)_2 (V_{DD} - V_{out} - V_{TH2})^2, \tag{3.29}$$

and hence

$$\sqrt{\left(\frac{W}{L}\right)_1}(V_{in} - V_{TH1}) = \sqrt{\left(\frac{W}{L}\right)_2}(V_{DD} - V_{out} - V_{TH2}). \tag{3.30}$$

Thus, if the variation of V_{TH2} with V_{out} is small, the circuit exhibits a linear input-output characteristic. The small-signal gain can also be computed by differentiating both sides with respect to V_{in}:

$$\sqrt{\left(\frac{W}{L}\right)_1} = \sqrt{\left(\frac{W}{L}\right)_2}\left(-\frac{\partial V_{out}}{\partial V_{in}} - \frac{\partial V_{TH2}}{\partial V_{in}}\right), \tag{3.31}$$

which, upon application of the chain rule $\partial V_{TH2}/\partial V_{in} = (\partial V_{TH2}/\partial V_{out})(\partial V_{out}/\partial V_{in}) = \eta(\partial V_{out}/\partial V_{in})$, reduces to

$$\frac{\partial V_{out}}{\partial V_{in}} = -\sqrt{\frac{(W/L)_1}{(W/L)_2}}\frac{1}{1+\eta}. \tag{3.32}$$

It is instructive to study the overall large-signal characteristic of the circuit as well. But let us first consider the circuit shown in Fig. 3.10(a). What is the final value of V_{out} if I_1 drops to zero? As I_1 decreases, so does the overdrive of M_2. Thus, for small I_1, $V_{GS2} \approx V_{TH2}$ and $V_{out} \approx V_{DD} - V_{TH2}$. In reality, the subthreshold conduction in M_2 eventually brings V_{out} to V_{DD} if I_D approaches zero, but at very low current levels, the finite capacitance at the output node slows down the change from $V_{DD} - V_{TH2}$ to V_{DD}. This is illustrated in the time-domain waveforms of Fig. 3.10(b). For this reason, in circuits that have frequent switching activity, we assume V_{out} remains around $V_{DD} - V_{TH2}$ when I_1 falls to small values.

Now we return to the circuit of Fig. 3.9. Plotted in Fig. 3.11 versus V_{in}, the output voltage equals $V_{DD} - V_{TH2}$ if $V_{in} < V_{TH1}$. For $V_{in} > V_{TH1}$, Eq. (3.30) holds and V_{out} follows an approximately straight line. As V_{in} exceeds $V_{out} + V_{TH1}$ (beyond point A), M_1 enters the triode region, and the characteristic becomes nonlinear.

The diode-connected load of Fig. 3.9 can be implemented with a PMOS device as well. Shown in Fig. 3.12, the circuit is free from body effect, providing a small-signal voltage gain equal to

$$A_v = -\sqrt{\frac{\mu_n(W/L)_1}{\mu_p(W/L)_2}}, \tag{3.33}$$

where channel-length modulation is neglected.

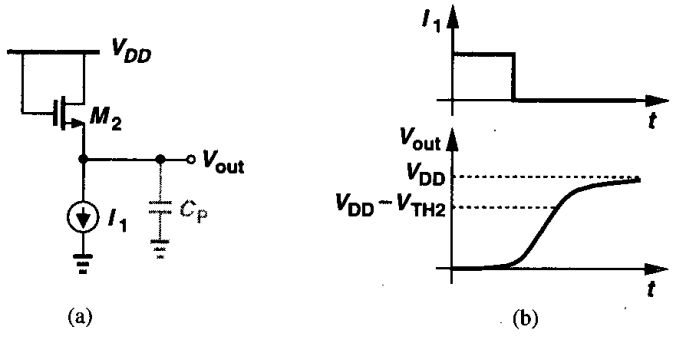


Figure 3.10 (a) Diode-connected device with stepped bias current, (b) variation of source voltage versus time.

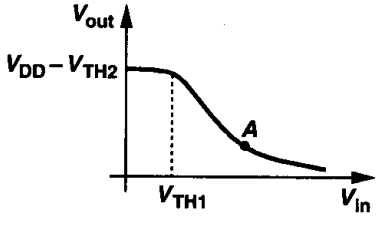


Figure 3.11 Input-output characteristic of a CS stage with diode-connected load.

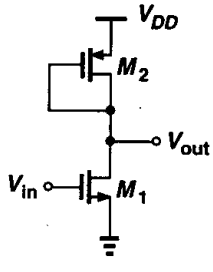


Figure 3.12 CS stage with diode-connected PMOS device.

Equations (3.28) and (3.33) indicate that the gain of a common-source stage with diode-connected load is a relatively weak function of the device dimensions. For example, to achieve a gain of 10, $\mu_n(W/L)_1/[\mu_p(W/L)_2] = 100$, implying that, with $\mu_n \approx 2\mu_p$, we must have $(W/L)_1 \approx 50(W/L)_2$. In a sense, a high gain requires a "strong" input device and a "weak" load device. In addition to disproportionately wide or long transistors (and hence a large input or load capacitance), a high gain translates to another important limitation: reduction in allowable voltage swings. Specifically, since in Fig. 3.12, $I_{D1} = |I_{D2}|$,

$$\mu_n \left(\frac{W}{L}\right)_1 (V_{GS1} - V_{TH1})^2 \approx \mu_p \left(\frac{W}{L}\right)_2 (V_{GS2} - V_{TH2})^2, \tag{3.34}$$

revealing that

$$\frac{|V_{GS2} - V_{TH2}|}{V_{GS1} - V_{TH1}} \approx A_v. \tag{3.35}$$

In the above example, the overdrive voltage of M_2 must be 10 times that of M_1. For example, with $V_{GS1} - V_{TH1} = 200$ mV, and $|V_{TH2}| = 0.7$ V, we have $|V_{GS2}| = 2.7$ V, severely limiting the output swing. This is another example of the trade-offs suggested by the analog design octagon. Note that, with diode-connected loads, the swing is constrained by both the required overdrive voltage and the threshold voltage. That is, even with a small overdrive, the output level cannot exceed $V_{DD} - |V_{TH}|$.

An interesting paradox arises here if we write $g_m = \mu C_{ox}(W/L)|V_{GS} - V_{TH}|$. The voltage gain of the circuit is then given by

$$A_v = \frac{g_{m1}}{g_{m2}} \tag{3.36}$$

$$= \frac{\mu_n C_{ox}(W/L)_1(V_{GS1} - V_{TH1})}{\mu_p C_{ox}(W/L)_2|V_{GS2} - V_{TH2}|}. \tag{3.37}$$

Equation (3.37) implies that A_v is *inversely* proportional to $|V_{GS2} - V_{TH2}|$. It is left for the reader to resolve the seemingly opposite trends suggested by (3.35) and (3.37).

Example 3.3

In the circuit of Fig. 3.13, M_1 is biased in saturation with a drain current equal to I_1. The current source $I_S = 0.75I_1$ is added to the circuit. How is (3.35) modified for this case?

Solution

Since $|I_{D2}| = I_1/4$, we have

$$A_v \approx -\frac{g_{m1}}{g_{m2}} \tag{3.38}$$

$$= -\sqrt{\frac{4\mu_n(W/L)_1}{\mu_p(W/L)_2}}. \tag{3.39}$$

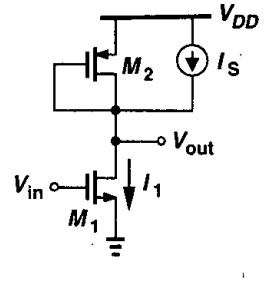


Figure 3.13

Moreover,

$$\mu_n \left(\frac{W}{L}\right)_1 (V_{GS1} - V_{TH1})^2 \approx 4\mu_p \left(\frac{W}{L}\right)_2 (V_{GS2} - V_{TH2})^2, \tag{3.40}$$

yielding

$$\frac{|V_{GS2} - V_{TH2}|}{V_{GS1} - V_{TH1}} \approx \frac{A_v}{4}. \tag{3.41}$$

Thus, for a gain of 10, the overdrive of M_2 need be only 2.5 times that of M_1. Alternatively, for a given overdrive voltage, this circuit achieves a gain four times that of the stage in Fig. 3.12. Intuitively, this is because for a given $|V_{GS2} - V_{TH2}|$, if the current decreases by a factor of 4, then $(W/L)_2$ must decrease proportionally, and $g_{m2} = \sqrt{2\mu_p C_{ox}(W/L)_2 I_{D2}}$ is lowered by the same factor.

We should also mention that in today's CMOS technology, channel-length modulation is quite significant and, more importantly, the behavior of transistors notably departs from the square law (Chapter 16). Thus, the gain of the stage in Fig. 3.9 must be expressed as

$$A_v = -g_{m1}\left(\frac{1}{g_{m2}} \| r_{O1} \| r_{O2}\right), \tag{3.42}$$

where g_{m1} and g_{m2} must be obtained as described in Chapter 16.

3.2.3 CS Stage with Current-Source Load

In applications requiring a large voltage gain in a single stage, the relationship $A_v = -g_m R_D$ suggests that we increase the load impedance of the CS stage. With a resistor or diode-connected load, however, increasing the load resistance limits the output voltage swing.

A more practical approach is to replace the load with a current source. Described briefly in Example 3.2, the resulting circuit is shown in Fig. 3.14, where both transistors operate in saturation. Since the total impedance seen at the output node is equal to $r_{O1} \| r_{O2}$, the gain is

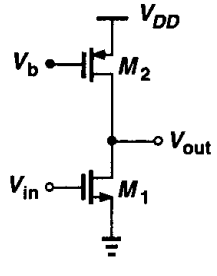


Figure 3.14 CS stage with current-source load.

$$A_v = -g_{m1}(r_{O1} \| r_{O2}). \tag{3.43}$$

The key point here is that the output impedance and the minimum required $|V_{DS}|$ of M_2 are less strongly coupled than the value and voltage drop of a resistor. The voltage

$|V_{DS2,min}| = |V_{GS2} - V_{TH2}|$ can be reduced to even a few hundred millivolts by simply increasing the width of M_2. If r_{O2} is not sufficiently high, the length and width of M_2 can be increased to achieve a smaller λ while maintaining the same overdrive voltage. The penalty is the large capacitance introduced by M_2 at the output node.

We should remark that the output bias voltage of the circuit in Fig. 3.14 is not well-defined. Thus, the stage is reliably biased only if a feedback loop forces V_{out} to a known value (Chapter 8). The large-signal analysis of the circuit is left as an exercise for the reader.

As explained in Chapter 2, the output impedance of MOSFETs at a given drain current can be scaled by changing the channel length, i.e., to the first order, $\lambda \propto 1/L$ and hence $r_O \propto L/I_D$. Since the gain of the stage shown in Fig. 3.14 is proportional to $r_{O1} \| r_{O2}$, we may surmise that longer transistors yield a higher voltage gain.

Let us consider M_1 and M_2 separately. If L_1 is scaled by a factor α (> 1), then W_1 may need to be scaled proportionally as well. This is because, for a given drain current, $V_{GS1} - V_{TH1} \propto 1/\sqrt{(W/L)_1}$, i.e., if W_1 is not scaled, the overdrive voltage increases, limiting the output voltage swing. Also, since $g_{m1} \propto \sqrt{(W/L)_1}$, scaling up only L_1 lowers g_{m1}.

In applications where these issues are unimportant, W_1 can remain constant while L_1 increases. Thus, the intrinsic gain of the transistor can be written as

$$g_{m1} r_{O1} = \sqrt{2 \left(\frac{W}{L}\right)_1 \mu_n C_{ox} I_D} \frac{1}{\lambda I_D}, \tag{3.44}$$

indicating that the gain *increases* with L because λ depends more strongly on L than g_m does. Also, note that $g_m r_O$ *decreases* as I_D increases.

Increasing L_2 while keeping W_2 constant increases r_{O2} and hence the voltage gain, but at the cost of higher $|V_{DS2}|$ required to maintain M_2 in saturation.

3.2.4 CS Stage with Triode Load

A MOS device operating in deep triode region behaves as a resistor and can therefore serve as the load in a CS stage. Illustrated in Fig. 3.15, such a circuit biases the gate of M_2 at a sufficiently low level, ensuring the load is in deep triode region for all output voltage swings.

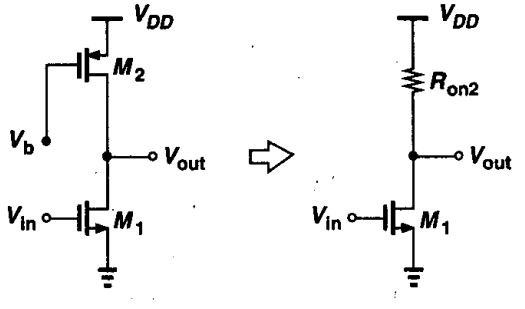


Figure 3.15 CS stage with triode load.

Since

$$R_{on2} = \frac{1}{\mu_p C_{ox}(W/L)_2(V_{DD} - V_b - |V_{THP}|)}, \tag{3.45}$$

the voltage gain can be readily calculated.

The principal drawback of this circuit stems from the dependence of R_{on2} upon $\mu_p C_{ox}$, V_b, and V_{THP}. Since $\mu_p C_{ox}$ and V_{THP} vary with process and temperature and since generating a precise value for V_b requires additional complexity, this circuit is difficult to use. Triode loads, however, consume less voltage headroom then do diode-connected devices because in Fig. 3.15 $V_{out,max} = V_{DD}$ whereas in Fig. 3.12, $V_{out,max} \approx V_{DD} - |V_{THP}|$.

3.2.5 CS Stage with Source Degeneration

In some applications, the square-law dependence of the drain current upon the overdrive voltage introduces excessive nonlinearity, making it desirable to "soften" the device characteristic. In Section 3.2.2, we noted the linear behavior of a CS stage using a diode-connected load. Alternatively, as depicted in Fig. 3.16, this can be accomplished by placing a "degeneration" resistor in series with the source terminal. Here, as V_{in} increases, so do I_D and the

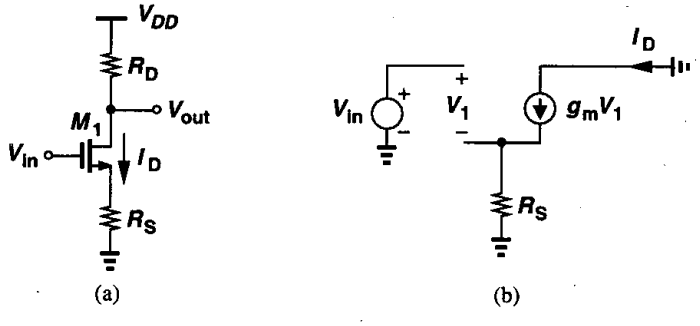


Figure 3.16 CS stage with source degeneration.

voltage drop across R_S. That is, a fraction of V_{in} appears across the resistor rather than as the gate-source overdrive, thus leading to a smoother variation of I_D. From another perspective, we intend to make the gain equation a weaker function of g_m. Since $V_{out} = -I_D R_D$, the nonlinearity of the circuit arises from the nonlinear dependence of I_D upon V_{in}. We note that $\partial V_{out}/\partial V_{in} = -(\partial I_D/\partial V_{in})R_D$, and define the equivalent transconductance of the circuit as $G_m = \partial I_D/\partial V_{in}$. Now, assuming $I_D = f(V_{GS})$, we write

$$G_m = \frac{\partial I_D}{\partial V_{in}} \tag{3.46}$$

$$= \frac{\partial f}{\partial V_{GS}} \frac{\partial V_{GS}}{\partial V_{in}}. \tag{3.47}$$

Since $V_{GS} = V_{in} - I_D R_S$, we have $\partial V_{GS}/\partial V_{in} = 1 - R_S \partial I_D/\partial V_{in}$, obtaining

$$G_m = \left(1 - R_S \frac{\partial I_D}{\partial V_{in}}\right) \frac{\partial f}{\partial V_{GS}}. \tag{3.48}$$

But, $\partial f/\partial V_{GS}$ is the transconductance of M_1, and

$$G_m = \frac{g_m}{1 + g_m R_S}. \tag{3.49}$$

The small-signal voltage gain is thus equal to

$$A_v = -G_m R_D \tag{3.50}$$

$$= \frac{-g_m R_D}{1 + g_m R_S}. \tag{3.51}$$

The same result can be derived using the small-signal model of Fig. 3.16(b). Equation (3.49) implies that as R_S increases, G_m becomes a weaker function of g_m and hence the drain current. In fact, for $R_S \gg 1/g_m$, we have $G_m \approx 1/R_S$, i.e., $\Delta I_D \approx \Delta V_{in}/R_S$, indicating that most of the change in V_{in} appears across R_S. We say the drain current is a "linearized" function of the input voltage. The linearization is obtained at the cost of lower gain [and higher noise (Chapter 7)].

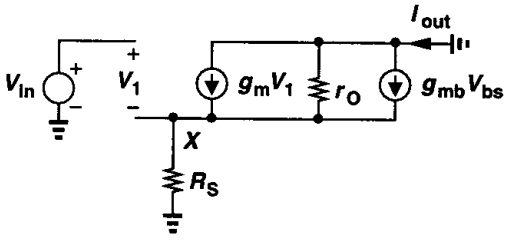


Figure 3.17 Small-signal equivalent circuit of a degenerated CS stage.

For our subsequent calculations, it is useful to determine G_m in the presence of body effect and channel-length modulation. With the aid of the equivalent circuit shown in Fig. 3.17, we recognize that the current through R_S equals I_{out} and, therefore, $V_{in} = V_1 + I_{out} R_S$. Summing the currents at node X, we have

$$I_{out} = g_m V_1 - g_{mb} V_X - \frac{I_{out} R_S}{r_O} \tag{3.52}$$

$$= g_m (V_{in} - I_{out} R_S) + g_{mb}(-I_{out} R_S) - \frac{I_{out} R_S}{r_O}. \tag{3.53}$$

It follows that

$$G_m = \frac{I_{out}}{V_{in}} \tag{3.54}$$

$$= \frac{g_m r_O}{R_S + [1 + (g_m + g_{mb}) R_S] r_O}. \tag{3.55}$$

Let us now examine the large-signal behavior of the CS stage with $R_S = 0$ and $R_S \neq 0$. For $R_S = 0$, our derivations in Chapter 2 indicate that I_D and g_m vary as shown in Fig. 3.18(a). For $R_S \neq 0$, the turn-on behavior is similar to that in Fig. 3.18(a) because,

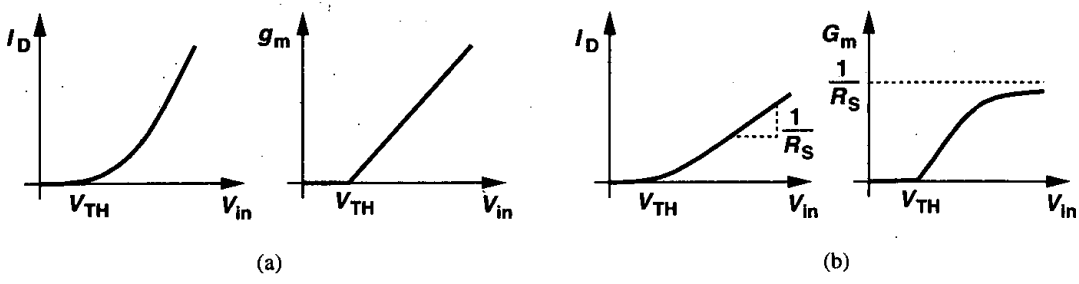


Figure 3.18 Drain current and transconductance of a CS device (a) without and (b) with source degeneration.

at low current levels, $1/g_m \gg R_S$ and hence $G_m \approx g_m$ [Fig. 3.18(b)]. As the overdrive and therefore g_m increase, the effect of degeneration, $1 + g_m R_S$ in (3.49), becomes more significant. For large values of V_{in} (if M_1 is still saturated), I_D is approximately linear and G_m approaches $1/R_S$.

Example 3.4

Plot the small-signal voltage gain of the circuit in Fig. 3.16 as a function of the input bias voltage.

Solution

Using the results derived above for the equivalent transconductance of M_1 and R_S, we arrive at the plot shown in Fig. 3.19. For V_{in} slightly greater than V_{TH}, $1/g_m \gg R_S$ and $A_v \approx -g_m R_D$.

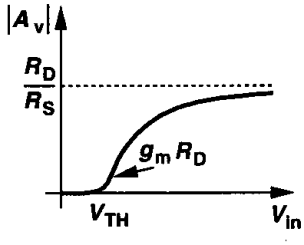


Figure 3.19

As V_{in} increases, degeneration becomes more significant and $A_v = -g_m R_D/(1 + g_m R_S)$. For large values of V_{in}, $G_m \approx 1/R_S$ and $A_v = -R_D/R_S$. However, if $V_{in} > V_{out} + V_{TH}$, that is, if $R_D I_D > V_{TH} + V_{DD} - V_{in}$, M_1 enters the triode region and A_v drops.

Equation (3.51) can be rewritten as

$$A_v = -\frac{R_D}{\frac{1}{g_m} + R_S}. \tag{3.56}$$

This result allows formulating the gain by inspection. First, let us examine the denominator of (3.56). The expression is equal to the *series* combination of the inverse transconductance of the device and the explicit resistance seen from the source to ground. We call the denominator "the resistance seen in the source path" because if, as shown in Fig. 3.20, we disconnect the bottom terminal of R_S from ground and calculate the resistance seen "looking up" (while setting the input to zero), we obtain $R_S + 1/g_m$.

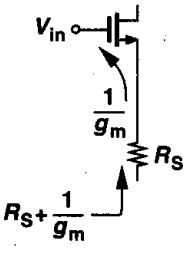


Figure 3.20 Resistance seen in the source path.

Noting that the numerator of (3.56) is the resistance seen at the drain, we view the magnitude of the gain as the resistance seen at the drain node divided by the total resistance in the source path. This method greatly simplifies the analysis of more complex circuits.

Example 3.5

Assuming $\lambda = \gamma = 0$, calculate the small-signal gain of the circuit shown in Fig. 3.21(a).

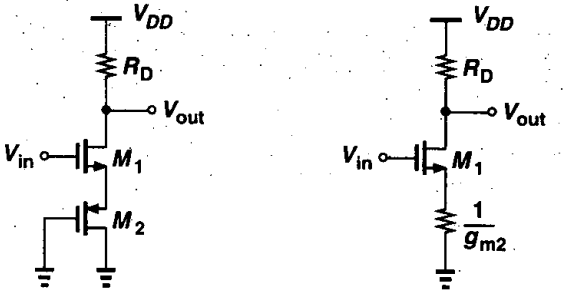


Figure 3.21

Solution

Noting that M_2 is a diode-connected device and simplifying the circuit to that shown in Fig. 3.21(b), we use the above rule to write

$$A_v = -\frac{R_D}{\frac{1}{g_{m1}} + \frac{1}{g_{m2}}}. \tag{3.57}$$

Another important consequence of source degeneration is the increase in the output resistance of the stage. We calculate the output resistance first with the aid of the equivalent circuit shown in Fig. 3.22. Note that body effect is also included to arrive at a general result.

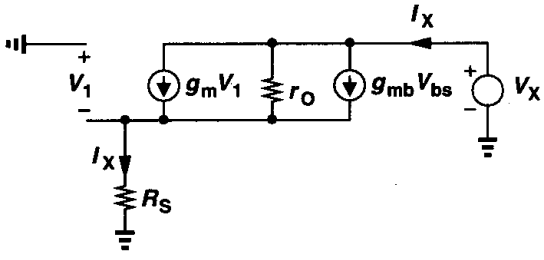


Figure 3.22 Equivalent circuit for calculating the output resistance of a degenerated CS stage.

Since the current through R_S is equal to I_X, $V_1 = -I_X R_S$ and the current flowing through r_O is given by $I_X - (g_m + g_{mb})V_1 = I_X + (g_m + g_{mb})R_S I_X$. Adding the voltage drops across r_O and R_S, we obtain

$$r_O[I_X + (g_m + g_{mb})R_S I_X] + I_X R_S = V_X. \tag{3.58}$$

It follows that

$$R_{out} = [1 + (g_m + g_{mb})R_S]r_O + R_S \tag{3.59}$$

$$= [1 + (g_m + g_{mb})r_O]R_S + r_O. \tag{3.60}$$

Since typically $(g_m + g_{mb})r_O \gg 1$, we have

$$R_{out} \approx (g_m + g_{mb})r_O R_S + r_O \tag{3.61}$$

$$= [1 + (g_m + g_{mb})R_S]r_O, \tag{3.62}$$

indicating that the output resistance has increased by a factor $1 + (g_m + g_{mb})R_S$. This is an important and useful result.

To gain more insight, let us consider the circuit of Fig. 3.22 with $R_S = 0$ and $R_S > 0$. If $R_S = 0$, then $g_m V_1 = g_{mb} V_{bs} = 0$ and $I_X = V_X / r_O$. On the other hand, if $R_S > 0$, we have $I_X R_S > 0$ and $V_1 < 0$, obtaining *negative* $g_m V_1$ and $g_{mb} V_{bs}$. Thus, the current supplied by V_X is less than V_X / r_O.

The relationships in (3.60) and (3.62) can also be derived by inspection. As shown in Fig. 3.23(a), we apply a voltage to the output node, change its value by ΔV, and measure the resulting change, ΔI, in the output current. Since the current through R_S must change by ΔI, we first compute the voltage change across R_S. To this end, we draw the circuit as shown in Fig. 3.23(b) and note that the resistance seen looking into the source of M_1 is equal to $1/(g_m + g_{mb})$ [Eq. (3.24)], thus arriving at the equivalent circuit in Fig. 3.23(c).

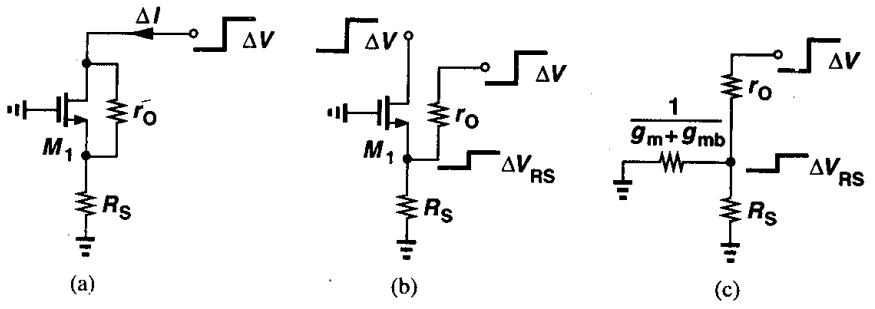


Figure 3.23 (a) Change in drain current in response to change in applied voltage to drain, (b) equivalent of (a), (c) small-signal model.

The voltage change across R_S is therefore equal to

$$\Delta V_{RS} = \Delta V \frac{\dfrac{1}{g_m + g_{mb}} \| R_S}{\dfrac{1}{g_m + g_{mb}} \| R_S + r_O}. \tag{3.63}$$

The change in the current is

$$\Delta I = \frac{\Delta V_{RS}}{R_S} \tag{3.64}$$

$$= \Delta V \frac{1}{[1 + (g_m + g_{mb})R_S]r_O + R_S}, \tag{3.65}$$

that is,

$$\frac{\Delta V}{\Delta I} = [1 + (g_m + g_{mb})R_S]r_O + R_S. \tag{3.66}$$

With the foregoing developments, we can now compute the gain of a degenerated CS stage in the general case, taking into account both body effect and channel-length modulation. In the equivalent circuit depicted in Fig. 3.24, the current through R_S must equal that through R_D, i.e., $-V_{out}/R_D$. Thus, the source voltage with respect to ground (and the bulk) is equal to $-V_{out}R_S/R_D$ and hence $V_1 = V_{in} + V_{out}R_S/R_D$. The current through r_O can therefore be written as

$$I_{ro} = -\frac{V_{out}}{R_D} - (g_m V_1 + g_{mb} V_{bs}) \tag{3.67}$$

$$= -\frac{V_{out}}{R_D} - \left[g_m \left(V_{in} + V_{out} \frac{R_S}{R_D} \right) + g_{mb} V_{out} \frac{R_S}{R_D} \right]. \tag{3.68}$$

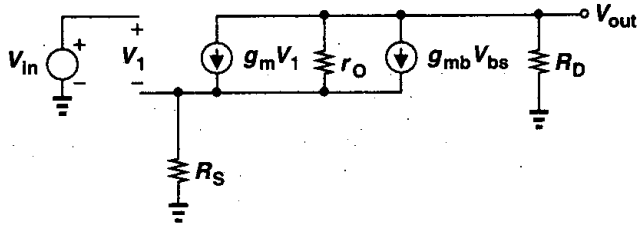


Figure 3.24 Small-signal model of degenerated CS stage with finite output resistance.

Since the voltage drop across r_O and R_S must add up to V_{out}, we have

$$V_{out} = I_{ro} r_O - \frac{V_{out}}{R_D} R_S \tag{3.69}$$

$$= -\frac{V_{out}}{R_D} r_O - \left[g_m \left(V_{in} + V_{out} \frac{R_S}{R_D} \right) + g_{mb} V_{out} \frac{R_S}{R_D} \right] r_O - V_{out} \frac{R_S}{R_D}. \tag{3.70}$$

It follows that

$$\frac{V_{out}}{V_{in}} = \frac{-g_m r_O R_D}{R_D + R_S + r_O + (g_m + g_{mb}) R_S r_O}. \tag{3.71}$$

To gain more insight into this result, we recognize that the last three terms in the denominator, namely, $R_S + r_O + (g_m + g_{mb}) R_S r_O$, represent the output resistance of a MOS device degenerated by a resistor R_S, as originally derived in (3.60). Let us now rewrite (3.71) as

$$A_v = \frac{-g_m r_O R_D [R_S + r_O + (g_m + g_{mb}) R_S r_O]}{R_D + R_S + r_O + (g_m + g_{mb}) R_S r_O} \cdot \frac{1}{R_S + r_O + (g_m + g_{mb}) R_S r_O} \tag{3.72}$$

$$= -\frac{g_m r_O}{R_S + r_O + (g_m + g_{mb}) R_S r_O} \cdot \frac{R_D [R_S + r_O + (g_m + g_{mb}) R_S r_O]}{R_D + R_S + r_O + (g_m + g_{mb}) R_S r_O}. \tag{3.73}$$

The two fractions in (3.73) represent two important parameters of the circuit: the first is identical to that in (3.55), i.e., the equivalent transconductance of a degenerated MOSFET; and the second denotes the parallel combination of R_D and $R_S + r_O + (g_m + g_{mb}) R_S r_O$, i.e., the overall output resistance of the circuit.

The above discussion suggests that in some circuits it may be easier to calculate the voltage gain by exploiting the following lemma.

Lemma. In a linear circuit, the voltage gain is equal to $-G_m R_{out}$, where G_m denotes the transconductance of the circuit when the output is shorted to ground and R_{out} represents the output resistance of the circuit when the input voltage is set to zero [Fig. 3.25(a)].

The lemma can be proved with the aid of Fig. 3.25 by noting that the output port of a linear circuit can be modeled by a Norton equivalent. That is, the output voltage is equal to $-I_{out} R_{out}$, and I_{out} can be obtained by measuring the short-circuit current at the output.

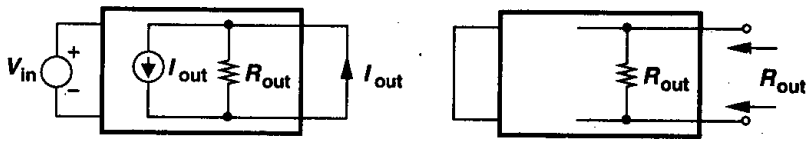


Figure 3.25 Modeling output port of an amplifier by a Norton equivalent.

Defining $G_m = I_{out}/V_{in}$, we have $V_{out} = -G_m V_{in} R_{out}$. This lemma proves useful if G_m and R_{out} can be determined by inspection.

Example 3.6

Calculate the voltage gain of the circuit shown in Fig. 3.26. Assume I_0 is ideal.

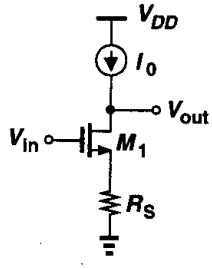


Figure 3.26

Solution

The transconductance and output resistance of the stage are given by Eqs. (3.55) and (3.60), respectively. Thus,

$$A_v = -\frac{g_m r_O}{R_S + [1 + (g_m + g_{mb})R_S]r_O}\{[1 + (g_m + g_{mb})r_O]R_S + r_O\} \tag{3.74}$$

$$= -g_m r_O. \tag{3.75}$$

Interestingly, the voltage gain is equal to the intrinsic gain of the transistor and independent of R_S. This is because, if I_0 is ideal, the current through R_S cannot change and hence the small-signal voltage drop across R_S is zero—as if R_S were zero itself.

3.3 Source Follower

Our analysis of the common-source stage indicates that, to achieve a high voltage gain with limited supply voltage, the load impedance must be as large as possible. If such a stage is to drive a low-impedance load, then a "buffer" must be placed after the amplifier so as to drive the load with negligible loss of the signal level. The source follower (also called the "common-drain" stage) can operate as a voltage buffer.

Illustrated in Fig. 3.27(a), the source follower senses the signal at the gate and drives

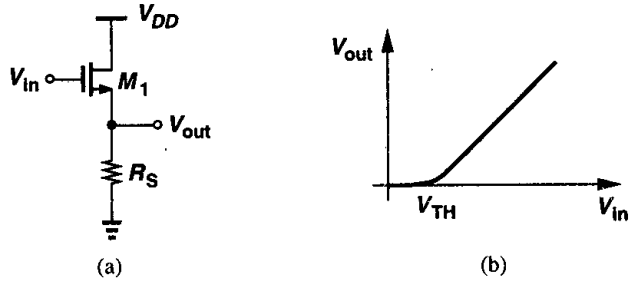


Figure 3.27 (a) Source follower, and (b) its input-output characteristic.

the load at the source, allowing the source potential to "follow" the gate voltage. Beginning with the large-signal behavior, we note that for $V_{in} < V_{TH}$, M_1 is off and $V_{out} = 0$. As V_{in} exceeds V_{TH}, M_1 turns on in saturation (for typical values of V_{DD}) and I_{D1} flows through R_S [Fig. 3.27(b)]. As V_{in} increases further, V_{out} follows the input with a difference (level shift) equal to V_{GS}. We can express the input-output characteristic as:

$$\frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{in} - V_{TH} - V_{out})^2 R_S = V_{out}. \tag{3.76}$$

Let us calculate the small-signal gain of the circuit by differentiating both sides of (3.76) with respect to V_{in}:

$$\frac{1}{2}\mu_n C_{ox}\frac{W}{L}2(V_{in} - V_{TH} - V_{out})\left(1 - \frac{\partial V_{TH}}{\partial V_{in}} - \frac{\partial V_{out}}{\partial V_{in}}\right)R_S = \frac{\partial V_{out}}{\partial V_{in}}. \tag{3.77}$$

Since $\partial V_{TH}/\partial V_{in} = \eta \partial V_{out}/\partial V_{in}$,

$$\frac{\partial V_{out}}{\partial V_{in}} = \frac{\mu_n C_{ox}\dfrac{W}{L}(V_{in} - V_{TH} - V_{out})R_S}{1 + \mu_n C_{ox}\dfrac{W}{L}(V_{in} - V_{TH} - V_{out})R_S(1+\eta)}. \tag{3.78}$$

Also, note that

$$g_m = \mu_n C_{ox}\frac{W}{L}(V_{in} - V_{TH} - V_{out}). \tag{3.79}$$

Consequently,

$$A_v = \frac{g_m R_S}{1 + (g_m + g_{mb})R_S}. \tag{3.80}$$

The same result is more easily obtained with the aid of a small-signal equivalent circuit. From Fig. 3.28, we have $V_{in} - V_1 = V_{out}$, $V_{bs} = -V_{out}$, and $g_m V_1 - g_{mb} V_{out} = V_{out}/R_S$.

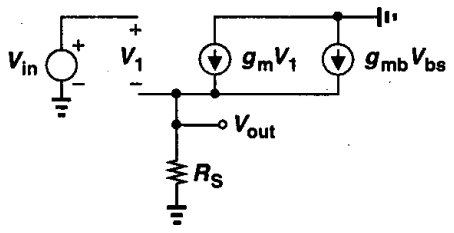


Figure 3.28 Small-signal equivalent circuit of source follower.

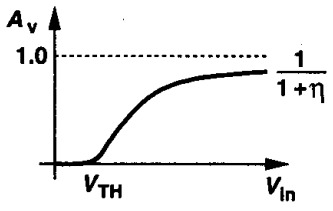


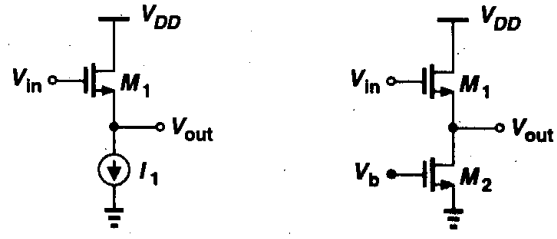
Figure 3.29 Voltage gain of source follower versus input voltage.

Thus, $V_{out}/V_{in} = g_m R_S/[1 + (g_m + g_{mb})R_S]$.

Sketched in Fig. 3.29 vs. V_{in}, the voltage gain begins from zero for $V_{in} \approx V_{TH}$ (that is, $g_m \approx 0$) and monotonically increases. As the drain current and g_m increase, A_v approaches $g_m/(g_m + g_{mb}) = 1/(1 + \eta)$. Since η itself slowly decreases with V_{out}, A_v would eventually become equal to unity, but for typical allowable source-bulk voltages, η remains greater than roughly 0.2.

An important result of (3.80) is that even if $R_S = \infty$, the voltage gain of a source follower is not equal to one. We return to this point later. Note that M_1 in Fig. 3.27 does not enter the triode region if V_{in} remains below V_{DD}.

In the source follower of Fig. 3.27, the drain current of M_1 heavily depends on the input dc level. For example, if V_{in} changes from 1.5 V to 2 V, I_D may increase by a factor of 2 and hence $V_{GS} - V_{TH}$ by $\sqrt{2}$, thereby introducing substantial nonlinearity in the input-output characteristic. To alleviate this issue, the resistor can be replaced by a current source as shown in Fig. 3.30(a). The current source itself is implemented as an NMOS transistor operating in the saturation region [Fig. 3.30(b)].



(a) (b)

Figure 3.30 Source follower using an NMOS transistor as current source.

Example 3.7

Suppose in the source follower of Fig. 3.30(a), $(W/L)_1 = 20/0.5$, $I_1 = 200\ \mu A$, $V_{TH0} = 0.6$ V, $2\Phi_F = 0.7$ V, $\mu_n C_{ox} = 50\ \mu A/V^2$, and $\gamma = 0.4\ V^2$.

(a) Calculate V_{out} for $V_{in} = 1.2$ V.

(b) If I_1 is implemented as M_2 in Fig. 3.30(b), find the minimum value of $(W/L)_2$ for which M_2 remains saturated.

Solution

(a) Since the threshold voltage of M_1 depends on V_{out}, we perform a simple iteration. Noting that

$$(V_{in} - V_{TH} - V_{out})^2 = \frac{2I_D}{\mu_n C_{ox}\left(\frac{W}{L}\right)_1}, \tag{3.81}$$

we first assume $V_{TH} \approx 0.6$ V, obtaining $V_{out} = 0.153$ V. Now we calculate a new V_{TH} as

$$V_{TH} = V_{TH0} + \gamma(\sqrt{2\Phi_F + V_{SB}} - \sqrt{2\Phi_F}) \tag{3.82}$$

$$= 0.635\ \text{V}. \tag{3.83}$$

This indicates that V_{out} is approximately 35 mV less than that calculated above, i.e., $V_{out} \approx 0.119$ V.

(b) Since the drain-source voltage of M_2 is equal to 0.119 V, the device is saturated only if $(V_{GS} - V_{TH})_2 \leq 0.119$ V. With $I_D = 200\ \mu A$, this gives $(W/L)_2 \geq 283/0.5$. Note the substantial drain junction and overlap capacitance contributed by M_2 to the output node.

To gain a better understanding of source followers, let us calculate the small-signal output resistance of the circuit in Fig. 3.31(a). Using the equivalent circuit of Fig. 3.31(b) and noting that $V_1 = -V_X$, we write

$$I_X - g_m V_X - g_{mb} V_X = 0. \tag{3.84}$$

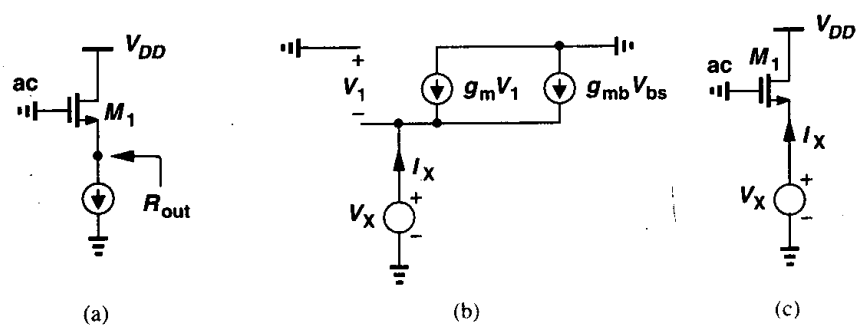


Figure 3.31 Calculation of the output impedance of a source follower.

It follows that

$$R_{out} = \frac{1}{g_m + g_{mb}}. \tag{3.85}$$

Interestingly, body effect decreases the output resistance of source followers. To understand why, suppose in Fig. 3.31(c), V_X decreases by ΔV so that the drain current increases. With no body effect, only the gate-source voltage of M_1 would increase by ΔV. With body effect, on the other hand, the threshold voltage of the device decreases as well. Thus, in $(V_{GS} - V_{TH})^2$ the first term increases and the second decreases, resulting in a greater change in the drain current and hence a lower output impedance.

The above phenomenon can also be studied with the aid of the small-signal model shown in Fig. 3.32(a). It is important to note that the magnitude of the current source $g_{mb}V_{bs}$ is

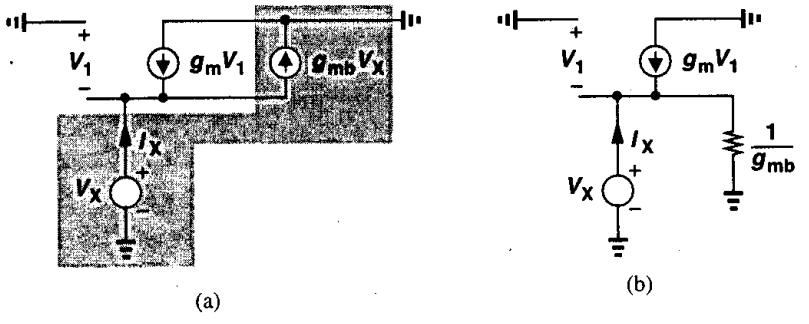


Figure 3.32 Source follower including body effect.

linearly proportional to the voltage across it. Such behavior is that of a simple resistor equal to $1/g_{mb}$, yielding the small-signal model shown in Fig. 3.32(b). The equivalent resistor simply appears in parallel with the output, thereby lowering the overall output resistance. The reader can show that, without $1/g_{mb}$, the output resistance equals $1/g_m$, concluding that

$$R_{out} = \frac{1}{g_m} \| \frac{1}{g_{mb}} \tag{3.86}$$

$$= \frac{1}{g_m + g_{mb}}. \tag{3.87}$$

Modeling the effect of g_{mb} by a resistor—which is only valid for source followers—also helps explain the less-than-unity voltage gain implied by (3.80) for $R_S = \infty$. As shown in

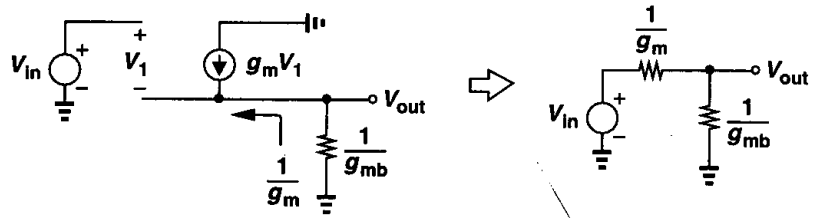


Figure 3.33 Representation of intrinsic source follower by a Thevenin equivalent.

the Thevenin equivalent of Fig. 3.33,

$$A_v = \frac{\dfrac{1}{g_{mb}}}{\dfrac{1}{g_m} + \dfrac{1}{g_{mb}}} \tag{3.88}$$

$$= \frac{g_m}{g_m + g_{mb}}. \tag{3.89}$$

For completeness, we also study the source follower of Fig. 3.34(a) with finite channel-length modulation in M_1 and M_2. From the equivalent circuit in Fig. 3.34(b), we have

$$A_v = \frac{\dfrac{1}{g_{mb}} \| r_{O1} \| r_{o2} \| R_L}{\dfrac{1}{g_{mb}} \| r_{O1} \| r_{o2} \| R_L + \dfrac{1}{g_m}}. \tag{3.90}$$

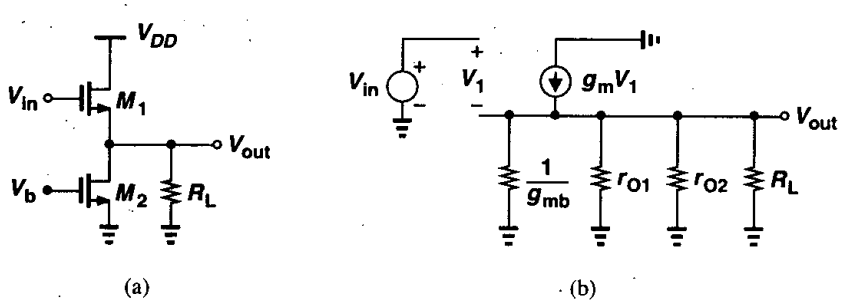


Figure 3.34 (a) Source follower driving load resistance, (b) small-signal equivalent circuit.

Example 3.8

Calculate the voltage gain of the circuit shown in Fig. 3.35.

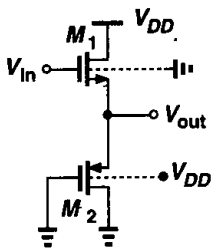


Figure 3.35

Solution

The impedance seen looking into the source of M_2 is equal to $[1/(g_{m2} + g_{mb2})] \| r_{O2}$. Thus,

$$A_v = \frac{\dfrac{1}{g_{m2}+g_{mb2}} \| r_{O2} \| r_{O1} \| \dfrac{1}{g_{mb1}}}{\dfrac{1}{g_{m2}+g_{mb2}} \| r_{O2} \| r_{O1} \| \dfrac{1}{g_{mb1}} + \dfrac{1}{g_{m1}}}. \tag{3.91}$$

Source followers exhibit a high input impedance and a moderate output impedance, but at the cost of two drawbacks: nonlinearity and voltage headroom limitation. We consider these issues in detail.

As mentioned in relation to Fig. 3.27(a), even if a source follower is biased by an ideal current source, its input-output characteristic displays some nonlinearity due to the nonlinear dependence of V_{TH} upon the source potential. In submicron technologies, r_O of the transistor also changes substantially with V_{DS}, thus introducing additional variation in the small-signal gain of the circuit (Chapter 16). For this reason, typical source followers suffer from several percent of nonlinearity.

The nonlinearity due to body effect can be eliminated if the bulk is tied to the source. This is usually possible only for PFETs because all NFETs share the same substrate. Fig. 3.36 shows a PMOS source follower employing two separate n-wells so as to eliminate the body effect of M_1. The lower mobility of PFETs, however, yields a higher output impedance in this case than that available in an NMOS counterpart.

Source followers also shift the dc level of the signal by V_{GS}, thereby consuming voltage headroom and limiting the voltage swings. To understand this point, consider the example illustrated in Fig. 3.37, a cascade of a common-source stage and a source follower. Without the source follower, the minimum allowable value of V_X would be equal to $V_{GS1} - V_{TH1}$ (for M_1 to remain in saturation). With the source follower, on the other hand, V_X must be greater than $V_{GS2} + (V_{GS3} - V_{TH3})$ so that M_3 is saturated. For comparable overdrive voltages in M_1 and M_3, this means the allowable swing at X is reduced by V_{GS2}, a substantial amount.

It is also instructive to compare the gain of source followers and common-source stages when the load impedance is relatively low. A practical example is the need to drive an external 50-Ω termination in a high-frequency setup. As shown in Fig. 3.38(a), the load can

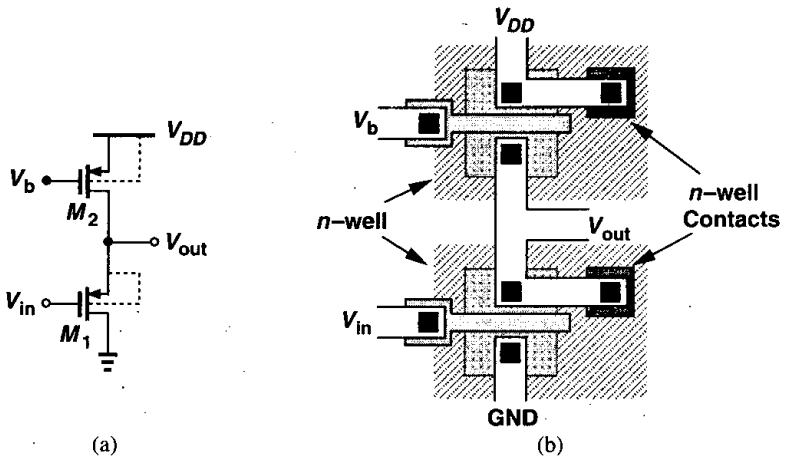


Figure 3.36 PMOS source follower with no body effect.

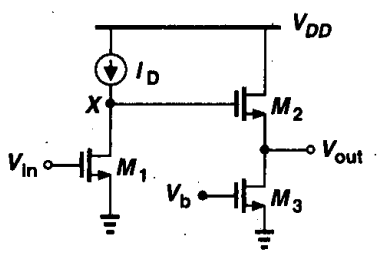


Figure 3.37 Cascade of source follower and CS stage.

be driven by a source follower with an overall voltage gain of

$$\frac{V_{out}}{V_{in}}|_{SF} \approx \frac{R_L}{R_L + 1/g_{m1}}. \tag{3.92}$$

On the other hand, as depicted in Fig. 3.38(b), the load can be included as part of a common-

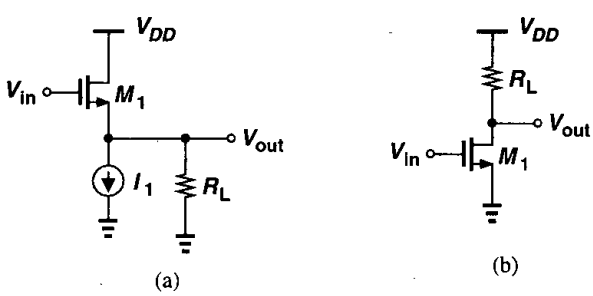


Figure 3.38 (a) Source follower and (b) CS stage driving a load resistance.

source stage, providing a gain of

$$\frac{V_{out}}{V_{in}}|_{CS} \approx -g_{m1}R_L. \tag{3.93}$$

The key difference between these two topologies is the achievable voltage gain for a given bias current. For example, if $1/g_{m1} \approx R_L$, then the source follower exhibits a gain of at most 0.5 whereas the common-source stage provides a gain close to unity. Thus, source followers are not necessarily efficient drivers.

The drawbacks of source followers, namely, nonlinearity due to body effect, voltage headroom consumption due to level shift, and poor driving capability, limit the use of this topology. Perhaps the most common application of source followers is in performing voltage level shift.

Example 3.9

(a) In the circuit of Fig. 3.39(a), calculate the voltage gain if C_1 acts as an ac short at the frequency of interest. What is the maximum dc level of the input signal for which M_1 remains saturated?

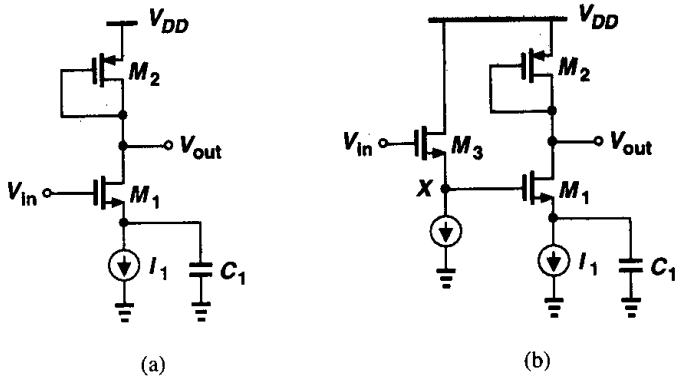


Figure 3.39

(b) To accommodate an input dc level close to V_{DD}, the circuit is modified as shown in Fig. 3.39(b). What relationship among the gate-source voltages of M_1-M_3 guarantees that M_1 is saturated?

Solution

(a) The gain is given by

$$A_v = -g_{m1}[r_{O1} \| r_{O2} \| (1/g_{m2})]. \tag{3.94}$$

Since $V_{out} = V_{DD} - |V_{GS2}|$, the maximum allowable dc level of V_{in} is equal to $V_{DD} - |V_{GS2}| + V_{TH1}$.

(b) If $V_{in} = V_{DD}$, then $V_X = V_{DD} - V_{GS3}$. For M_1 to be saturated, $V_{DD} - V_{GS3} - V_{TH1} \leq V_{DD} - |V_{GS2}|$ and hence $V_{GS3} + V_{TH1} \geq |V_{GS2}|$.

As explained in Chapter 7, source followers also introduce substantial noise. For this reason, the circuit of Fig. 3.39(b) is ill-suited to low-noise applications.

3.4 Common-Gate Stage

In common-source amplifiers and source followers, the input signal is applied to the gate of a MOSFET. It is also possible to apply the signal to the source terminal. Shown in Fig. 3.40(a), a common-gate (CG) stage senses the input at the source and produces the output at the drain. The gate is connected to a dc voltage to establish proper operating conditions. Note that the bias current of M_1 flows through the input signal source. Alternatively, as depicted in Fig. 3.40(b), M_1 can be biased by a constant current source, with the signal capacitively coupled to the circuit.

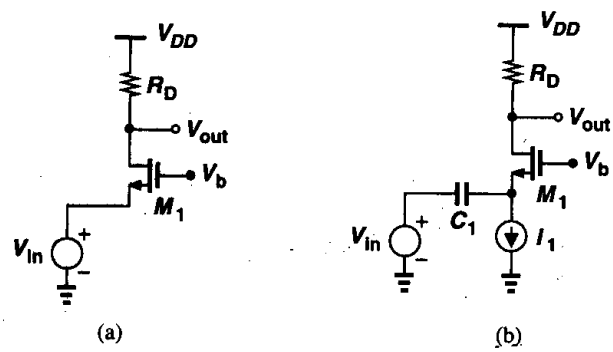


Figure 3.40 (a) Common-gate stage with direct coupling at input, (b) CG stage with capacitive coupling at input.

We first study the large-signal behavior of the circuit in Fig. 3.40(a). For simplicity, let us assume that V_{in} decreases from a large positive value. For $V_{in} \geq V_b - V_{TH}$, M_1 is off and $V_{out} = V_{DD}$. For lower values of V_{in}, we can write

$$I_D = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_b - V_{in} - V_{TH})^2, \tag{3.95}$$

if M_1 is in saturation. As V_{in} decreases, so does V_{out}, eventually driving M_1 into the triode region if

$$V_{DD} - \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_b - V_{in} - V_{TH})^2 R_D = V_b - V_{TH}. \tag{3.96}$$

The input-output characteristic is shown in Fig. 3.41. If M_1 is saturated, we can express the output voltage as

$$V_{out} = V_{DD} - \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_b - V_{in} - V_{TH})^2 R_D, \tag{3.97}$$

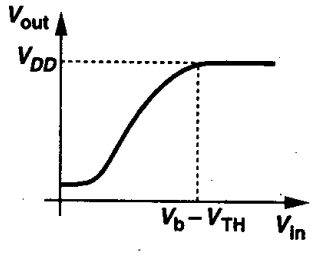


Figure 3.41 Common-gate input-output characteristic.

obtaining a small-signal gain of

$$\frac{\partial V_{out}}{\partial V_{in}} = -\mu_n C_{ox}\frac{W}{L}(V_b - V_{in} - V_{TH})\left(-1 - \frac{\partial V_{TH}}{\partial V_{in}}\right)R_D. \tag{3.98}$$

Since $\partial V_{TH}/\partial V_{in} = \partial V_{TH}/\partial V_{SB} = \eta$, we have

$$\frac{\partial V_{out}}{\partial V_{in}} = \mu_n C_{ox}\frac{W}{L}R_D(V_b - V_{in} - V_{TH})(1+\eta) \tag{3.99}$$

$$= g_m(1+\eta)R_D. \tag{3.100}$$

Note that the gain is positive. Interestingly, body effect increases the equivalent transconductance of the stage.

The input impedance of the circuit is also important. We note that, for $\lambda = 0$, the impedance seen at the source of M_1 in Fig. 3.40(a) is the same as that at the source of M_1 in Fig. 3.31, namely, $1/(g_m + g_{mb}) = 1/[g_m(1+\eta)]$. Thus, the body effect decreases the input impedance of the common-gate stage. The relatively low input impedance of the common-gate stage proves useful in some applications.

Example 3.10

In Fig. 3.42, transistor M_1 senses ΔV and delivers a proportional current to a 50-Ω transmission line. The other end of the line is terminated by a 50-Ω resistor in Fig. 3.42(a) and a common-gate stage in Fig. 3.42(b). Assume $\lambda = \gamma = 0$.

(a) Calculate V_{out}/V_{in} at low frequencies for both arrangements.

(b) What condition is necessary to minimize wave reflection at node X?

Solution

(a) For small signals applied to the gate of M_1, the drain current experiences a change equal to $g_{m1}\Delta V_X$. This current is drawn from R_D in Fig. 3.42(a) and M_2 in Fig. 3.42(b), producing an output voltage swing equal to $-g_{m1}\Delta V_X R_D$. Thus, $A_v = -g_m R_D$ for both cases.

(b) To minimize reflection at node X, the resistance seen at the source of M_2 must equal 50 Ω and the reactance must be small. Thus, $1/(g_m + g_{mb}) = 50\ \Omega$, which can be ensured by proper sizing and biasing of M_2. To minimize the capacitances of the transistor, it is desirable to use a small device biased at a large current. (Recall that $g_m = \sqrt{2\mu_n C_{ox}(W/L)I_D}$.) In addition to higher power dissipation, this remedy also requires a large V_{GS} for M_2.

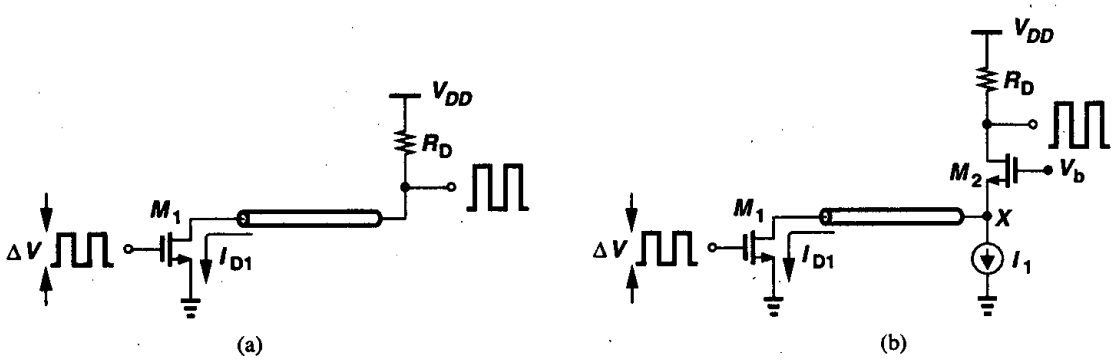


Figure 3.42

The key point in this example is that, while the overall voltage gain in both arrangements equals $-g_{m1}R_D$, the value of R_D in Fig. 3.42(b) can be much greater than 50 Ω without introducing reflections at point X. Thus, the common-gate circuit can provide a much higher voltage gain than that in Fig. 3.42(a).

Now let us study the common-gate topology in a more general case, taking into account both the output impedance of the transistor and the impedance of the signal source. Depicted in Fig. 3.43(a), the circuit can be analyzed with the aid of its equivalent shown

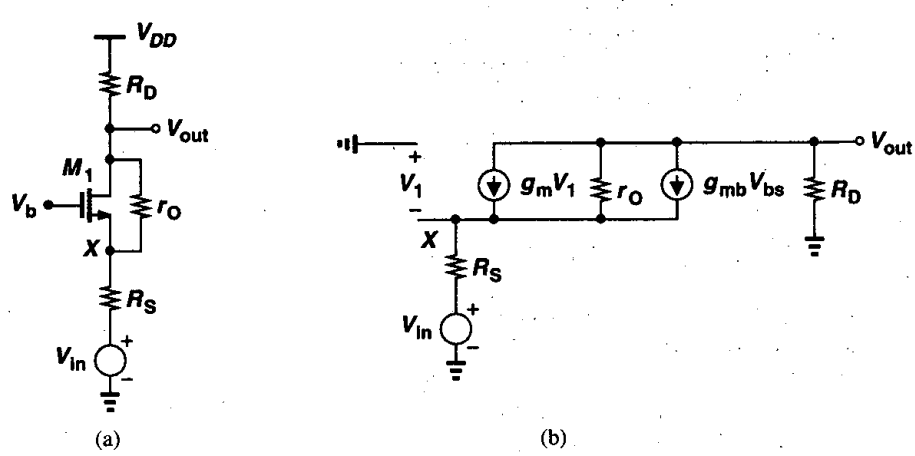


Figure 3.43 (a) CG stage with finite output resistance, (b) small-signal equivalent circuit.

in Fig. 3.43(b). Noting that the current flowing through R_S is equal to $-V_{out}/R_D$, we have:

$$V_1 - \frac{V_{out}}{R_D}R_S + V_{in} = 0. \tag{3.101}$$

Moreover, since the current through r_O is equal to $-V_{out}/R_D - g_m V_1 - g_{mb} V_1$, we can write

$$r_O\left(\frac{-V_{out}}{R_D} - g_m V_1 - g_{mb} V_1\right) - \frac{V_{out}}{R_D} R_S + V_{in} = V_{out}. \tag{3.102}$$

Upon substitution for V_1 from (3.102), (3.101) reduces to

$$r_O\left[\frac{-V_{out}}{R_D} - (g_m + g_{mb})\left(V_{out}\frac{R_S}{R_D} - V_{in}\right)\right] - \frac{V_{out} R_S}{R_D} + V_{in} = V_{out}. \tag{3.103}$$

It follows that

$$\frac{V_{out}}{V_{in}} = \frac{(g_m + g_{mb})r_O + 1}{r_O + (g_m + g_{mb})r_O R_S + R_S + R_D} R_D. \tag{3.104}$$

Note the similarity between (3.104) and (3.71). The gain of the common-gate stage is slightly higher due to body effect.

Example 3.11

Calculate the voltage gain of the circuit shown in Fig. 3.44(a) if $\lambda \neq 0$ and $\gamma \neq 0$.

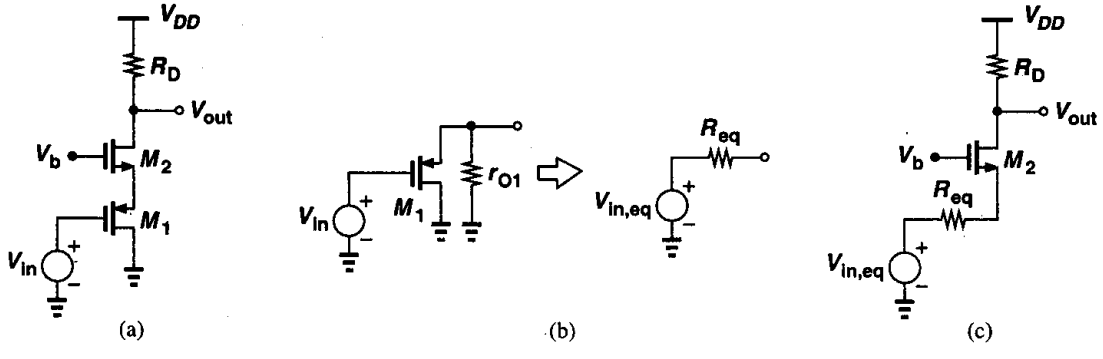


Figure 3.44

Solution

We first find the Thevenin equivalent of M_1. As shown in Fig. 3.44(b), M_1 operates as a source follower and the equivalent Thevenin voltage is given by

$$V_{in,eq} = \frac{r_{O1} \left\| \frac{1}{g_{mb1}} \right.}{r_{O1} \left\| \frac{1}{g_{mb1}} + \frac{1}{g_{m1}} \right.} V_{in}, \tag{3.105}$$

and the equivalent Thevenin resistance is

$$R_{eq} = r_{O1} \left\| \frac{1}{g_{mb1}} \right\| \frac{1}{g_{m1}}. \tag{3.106}$$

Redrawing the circuit as in Fig. 3.44(c), we use (3.104) to write

$$\frac{V_{out}}{V_{in}} = \frac{(g_{m2} + g_{mb2})r_{O2} + 1}{r_{O2} + [1 + (g_{m2} + g_{mb2})r_{O2}]\left(r_{O1} \left\| \frac{1}{g_{mb1}} \right\| \frac{1}{g_{m1}} \right) + R_D} R_D \frac{r_{O1} \left\| \frac{1}{g_{mb1}} \right.}{r_{O1} \left\| \frac{1}{g_{mb1}} + \frac{1}{g_{m2}} \right.}. \tag{3.107}$$

The input and output impedances of the common-gate topology are also of interest. To obtain the impedance seen at the source [Fig. 3.45(a)], we use the equivalent circuit in

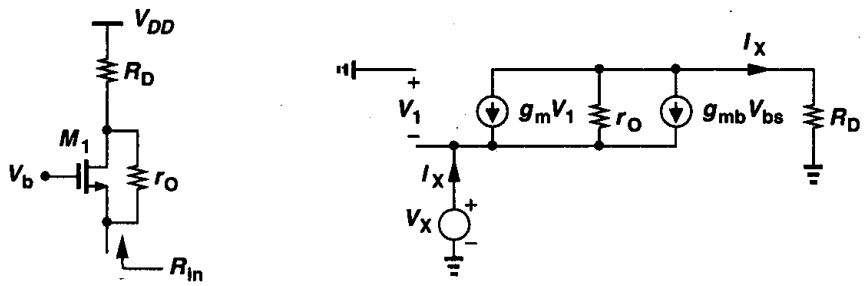


Figure 3.45 (a) Input resistance of a CG stage, (b) small-signal equivalent circuit.

Fig. 3.45(b). Since $V_1 = -V_X$ and the current through r_O is equal to $I_X + g_m V_1 + g_{mb} V_1 = I_X - (g_m + g_{mb})V_X$, we can add up the voltages across r_O and R_D as

$$R_D I_X + r_O[I_X - (g_m + g_{mb})V_X] = V_X. \tag{3.108}$$

Thus,

$$\frac{V_X}{I_X} = \frac{R_D + r_O}{1 + (g_m + g_{mb})r_O} \tag{3.109}$$

$$\approx \frac{R_D}{(g_m + g_{mb})r_O} + \frac{1}{g_m + g_{mb}}, \tag{3.110}$$

if $(g_m + g_{mb})r_O \gg 1$. This result reveals that the drain impedance is divided by $(g_m + g_{mb})r_O$ when seen at the source. This is particularly important in short-channel devices because of their low intrinsic gain. Two special cases of (3.109) are worth studying. First, suppose

$R_D = 0$. Then,

$$\frac{V_X}{I_X} = \frac{r_O}{1 + (g_m + g_{mb})r_O} \tag{3.111}$$

$$= \frac{1}{\dfrac{1}{r_O} + g_m + g_{mb}}, \tag{3.112}$$

which is simply the impedance seen at the source of a source follower, a predictable result because if $R_D = 0$, the circuit configuration is the same as in Fig. 3.31(a).

Second, let us replace R_D with an ideal current source. Equation (3.110) predicts that the input impedance approaches *infinity*. While somewhat surprising, this result can be explained with the aid of Fig. 3.46. Since the total current through the transistor is fixed and equal to I_1, a change in the source potential cannot change the device current, and hence $I_X = 0$. In other words, the input impedance of a common-gate stage is relatively low *only* if the load impedance connected to the drain is small.

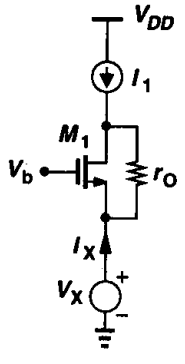


Figure 3.46 Input resistance of a CG stage with ideal current source load.

Example 3.12

Calculate the voltage gain of a common-gate stage with a current-source load [Fig. 3.47(a)].

Solution

Letting R_D approach infinity in (3.104), we have

$$A_v = (g_m + g_{mb})r_O + 1. \tag{3.113}$$

Interestingly, the gain does not depend on R_S. From our foregoing discussion, we recognize that if $R_D \to \infty$, so does the impedance seen at the source of M_1, and the small-signal voltage at node X becomes *equal* to V_{in}. We can therefore simplify the circuit as shown in Fig. 3.47(b), readily arriving at (3.113).

In order to calculate the output impedance of the common-gate stage, we use the circuit

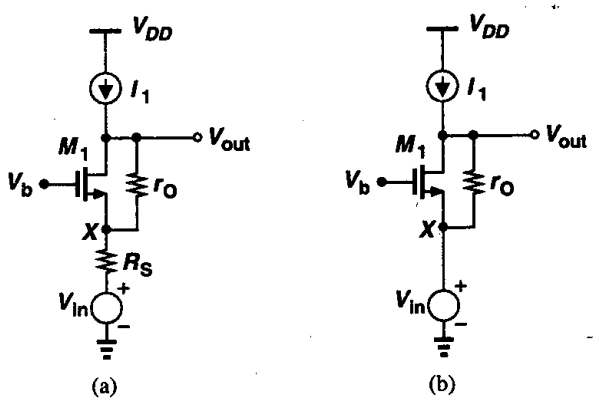


Figure 3.47

in Fig. 3.48. We note that the result is similar to that in Fig. 3.22 and hence

$$R_{out} = \{[1 + (g_m + g_{mb})r_O]R_S + r_O\} \| R_D. \tag{3.114}$$

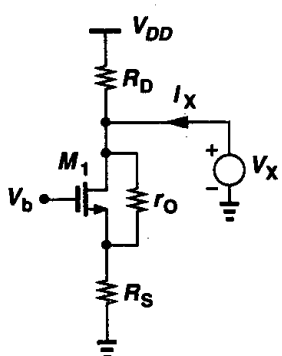


Figure 3.48 Calculation of output resistance of a CG stage.

Example 3.13

As seen in Example 3.10 the input signal of a common-gate stage may be a current rather than a voltage. Shown in Fig. 3.49 is such an arrangement. Calculate V_{out}/I_{in} and the output impedance of the circuit if the input current source exhibits an output impedance equal to R_P.

Solution

To find V_{out}/I_{in}, we replace I_{in} and R_P with a Thevenin equivalent and use (3.104) to write

$$\frac{V_{out}}{I_{in}} = \frac{(g_m + g_{mb})r_O + 1}{r_O + (g_m + g_{mb})r_O R_P + R_P + R_D} R_D R_P. \tag{3.115}$$

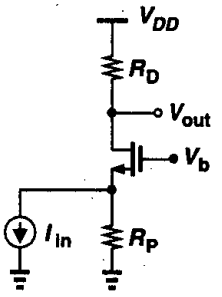


Figure 3.49

The output impedance is simply equal to

$$R_{out} = \{[1 + (g_m + g_{mb})r_O]R_P + r_O\} \| R_D. \tag{3.116}$$

3.5 Cascode Stage

As mentioned in Example 3.10 the input signal of a common-gate stage may be a current. We also know that a transistor in a common-source arrangement converts a voltage signal to a current signal. The cascade of a CS stage and a CG stage is called a "cascode"[1] topology, providing many useful properties. Fig. 3.50 shows the basic configuration: M_1 generates a small-signal drain current proportional to V_{in} and M_2 simply routes the current to R_D.

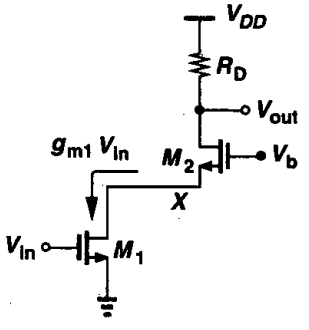


Figure 3.50 Cascode stage.

We call M_1 the input device and M_2 the cascode device. Note that in this example, M_1 and M_2 carry equal currents. As we describe the attributes of the circuit in this section, many advantages of the cascode topology over a simple common-source stage become evident.

First, let us study the bias conditions of the cascode. For M_1 to operate in saturation, $V_X \geq V_{in} - V_{TH1}$. If M_1 and M_2 are both in saturation, then V_X is determined primarily by

[1]The term *cascode* is believed to be the acronym for "cascaded triodes," possibly invented in vacuum tube days.

V_b: $V_X = V_b - V_{GS2}$. Thus, $V_b - V_{GS2} \geq V_{in} - V_{TH1}$ and hence $V_b > V_{in} + V_{GS2} - V_{TH1}$ (Fig. 3.51). For M_2 to be saturated, $V_{out} \geq V_b - V_{TH2}$, that is, $V_{out} \geq V_{in} - V_{TH1} + V_{GS2} -$

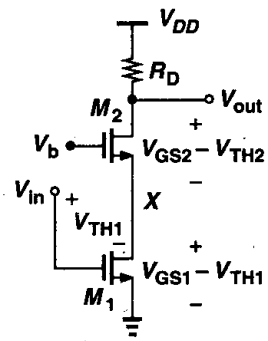


Figure 3.51 Allowable voltages in cascode stage.

V_{TH2} if V_b is chosen to place M_1 at the edge of saturation. Consequently, the minimum output level for which both transistors operate in saturation is equal to the overdrive voltage of M_1 plus that of M_2. In other words, addition of M_2 to the circuit reduces the output voltage swing by at least the overdrive voltage of M_2. We also say M_2 is "stacked" on top of M_1.

We now analyze the large-signal behavior of the cascode stage shown in Fig. 3.50 as V_{in} goes from zero to V_{DD}. For $V_{in} \leq V_{TH1}$, M_1 and M_2 are off, $V_{out} = V_{DD}$, and $V_X \approx V_b - V_{TH2}$ (if subthreshold conduction is neglected) (Fig. 3.52). As V_{in} exceeds V_{TH1}, M_1 begins to draw current, and V_{out} drops. Since I_{D2} increases, V_{GS2} must increase

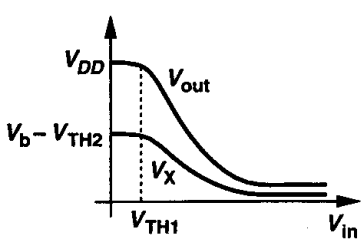


Figure 3.52 Input-output characteristic of a cascode stage.

as well, causing V_X to fall. As V_{in} assumes sufficiently large values, two effects occur: (1) V_X drops below V_{in} by V_{TH1}, forcing M_1 into the triode region; (2) V_{out} drops below V_b by V_{TH2}, driving M_2 into the triode region. Depending on the device dimensions and the values of R_D and V_b, one effect may occur before the other. For example, if V_b is relatively low, M_1 may enter the triode region first. Note that if M_2 goes into deep triode region, V_X and V_{out} become nearly equal.

Let us now consider the small-signal characteristics of a cascode stage, assuming both transistors operate in saturation. If $\lambda = 0$, the voltage gain is equal to that of a common-source stage because the drain current produced by the input device must flow through the cascode device. Illustrated in the equivalent circuit of Fig. 3.53, this result is independent of the transconductance and body effect of M_2.

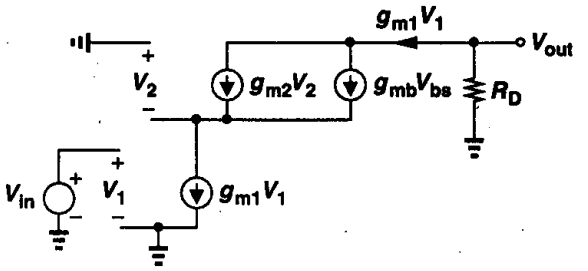


Figure 3.53 Small-signal equivalent circuit of cascode stage.

Example 3.14

Calculate the voltage gain of the circuit shown in Fig. 3.54 if $\lambda = 0$.

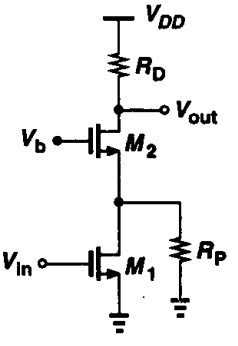


Figure 3.54

Solution

The small-signal drain current of M_1, $g_{m1}V_{in}$, is divided between R_P and the impedance seen looking into the source of M_2, $1/(g_{m2} + g_{mb2})$. Thus, the current flowing through M_2 is

$$I_{D2} = g_{m1}V_{in}\frac{(g_{m2} + g_{mb2})R_P}{1 + (g_{m2} + g_{mb2})R_P}. \tag{3.117}$$

The voltage gain is therefore given by

$$A_v = -\frac{g_{m1}(g_{m2} + g_{mb2})R_P R_D}{1 + (g_{m2} + g_{mb2})R_P}. \tag{3.118}$$

An important property of the cascode structure is its high output impedance. As illustrated in Fig. 3.55, for calculation of R_{out}, the circuit can be viewed as a common-source stage with a degeneration resistor equal to r_{O1}. Thus, from (3.60),

$$R_{out} = [1 + (g_{m2} + g_{mb2})r_{O2}]r_{O1} + r_{O2}. \tag{3.119}$$

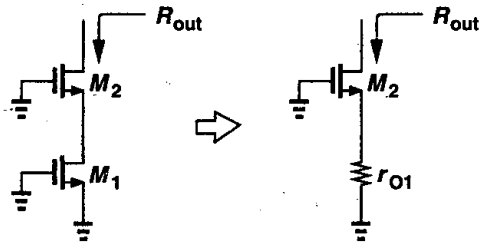


Figure 3.55 Calculation of output resistance of cascode stage.

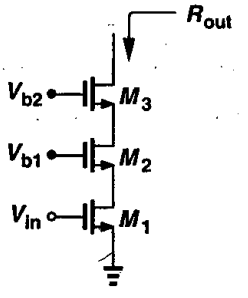


Figure 3.56 Triple cascode.

Assuming $g_m r_O \gg 1$, we have $R_{out} \approx (g_{m2} + g_{mb2}) r_{O2} r_{O1}$. That is, M_2 boosts the output impedance of M_1 by a factor of $(g_{m2} + g_{mb2}) r_{O2}$. As shown in Fig. 3.56, cascoding can be extended to three or more stacked devices to achieve a higher output impedance, but the required additional voltage headroom makes such configurations less attractive. For example, the minimum output voltage of a triple cascode is equal to the sum of three overdrive voltages.

To appreciate the usefulness of a high output impedance, recall from the lemma in Section 3.2.3 that the voltage gain can be written as $G_m R_{out}$. Since G_m is typically determined by the transconductance of a transistor, e.g., M_1 in Fig. 3.50, and hence bears trade-offs with the bias current and device capacitances, it is desirable to increase the voltage gain by maximizing R_{out}. Shown in Fig. 3.57 is an example. If both M_1 and M_2 operate in saturation,

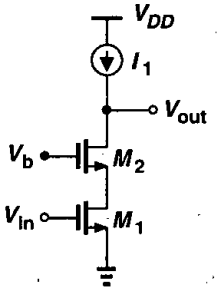


Figure 3.57 Cascode stage with current-source load.

then $G_m \approx g_{m1}$ and $R_{out} \approx (g_{m2} + g_{mb2}) r_{O2} r_{O1}$, yielding $A_v = (g_{m2} + g_{mb2}) r_{O2} g_{m1} r_{O1}$.

Thus, the maximum voltage gain is roughly equal to the *square* of the intrinsic gain of the transistors.

Example 3.15

Calculate the exact voltage gain of the circuit shown in Fig. 3.57.

Solution

The actual G_m of the stage is slightly less than g_{m1} because a fraction of the small-signal current produced by M_1 is shunted to ground by r_{O1}. As depicted in Fig. 3.58:

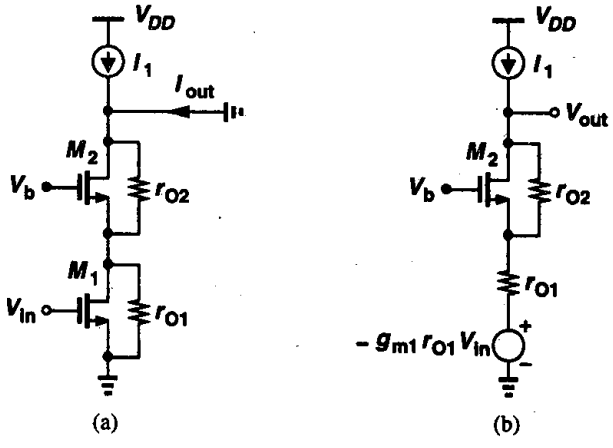


Figure 3.58

$$I_{out} = g_{m1}V_{in}\frac{r_{O1}}{r_{O1} + \dfrac{1}{g_{m2}+g_{mb2}} \Big\| r_{O2}}. \tag{3.120}$$

It follows that the overall transconductance is equal to

$$G_m = \frac{g_{m1}r_{O1}[r_{O2}(g_{m2}+g_{mb2})+1]}{r_{O1}r_{O2}(g_{m2}+g_{mb2}) + r_{O1} + r_{O2}}, \tag{3.121}$$

and hence the voltage gain is given by

$$|A_v| = G_m R_{out} \tag{3.122}$$

$$= g_{m1}r_{O1}[(g_{m2}+g_{mb2})r_{O2}+1]. \tag{3.123}$$

If we had assumed $G_m \approx g_m$, then $|A_v| \approx g_{m1}\{[1+(g_{m2}+g_{mb2})r_{O2}]r_{O1} + r_{O2}\}$.

Another approach to calculating the voltage gain is to replace V_{in} and M_1 by a Thevenin equivalent, reducing the circuit to a common-gate stage. Illustrated in Fig. 3.58(b), this method in conjunction with (3.104) gives the same result as (3.123).

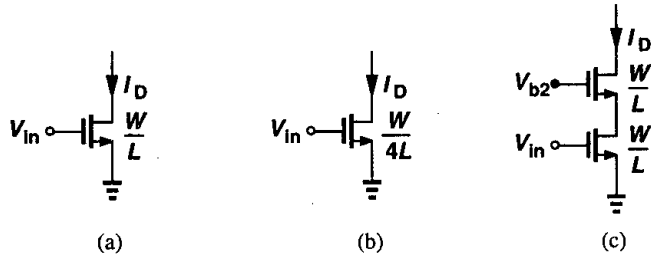


Figure 3.59 Increasing output impedance by increasing the device length or cascoding.

It is also interesting to compare the increase in gain due to cascoding with that due to increasing the length of the input transistor for a given bias current (Fig. 3.59). Suppose, for example, that the length of the input transistor of a CS stage is quadrupled while the width remains constant. Then, since $I_D = (1/2)\mu_n C_{ox}(W/L)(V_{GS} - V_{TH})^2$, the overdrive voltage is doubled, and the transistor consumes the same amount of voltage headroom as does a cascode stage. That is, the circuits of Figs. 3.59(b) and (c) impose equal voltage swing constraints.

Now consider the output impedance achieved in each case. Since

$$g_m r_O = \sqrt{2\mu_n C_{ox}\frac{W}{L}I_D}\frac{1}{\lambda I_D}, \tag{3.124}$$

and $\lambda \propto 1/L$, quadrupling L only doubles the value of $g_m r_O$ while cascoding results in an output impedance of roughly $(g_m r_O)^2$. Note that the transconductance of M_1 in Fig. 3.59(b) is half that in Fig. 3.59(c), leading to higher noise (Chapter 7).

A cascode structure need not operate as an amplifier. Another popular application of this topology is in building constant current sources. The high output impedance yields a current source closer to the ideal, but at the cost of voltage headroom. For example, current source I_1 in Fig. 3.57 can be implemented with a PMOS cascode (Fig. 3.60), exhibiting an impedance equal to $[1 + (g_{m3} + g_{mb3})r_{O3}]r_{O4} + r_{O3}$. If the gate bias voltages are chosen

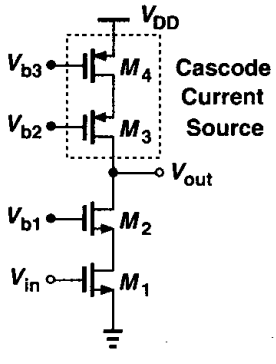


Figure 3.60 NMOS cascode amplifier with PMOS cascode load.

properly, the maximum output swing is equal to $V_{DD} - (V_{GS1} - V_{TH1}) - (V_{GS2} - V_{TH2}) - |V_{GS3} - V_{TH3}| - |V_{GS4} - V_{TH4}|$.

We calculate the voltage gain with the aid of the lemma illustrated in Fig. 3.25. Writing $G_m \approx g_{m1}$ and

$$R_{out} = \{[1 + (g_{m2} + g_{mb2})r_{O2}]r_{O1} + r_{O2}\} \| \{[1 + (g_{m3} + g_{mb3})r_{O3}]r_{O4} + r_{O3}\}, \quad (3.125)$$

we have $|A_v| \approx g_{m1} R_{out}$. For typical values, we approximate the voltage gain as

$$|A_v| \approx g_{m1}[(g_{m2}r_{O2}r_{O1}) \| (g_{m3}r_{O3}r_{O4})]. \quad (3.126)$$

Shielding Property Recall from Fig. 3.23 that the high output impedance arises from the fact that if the output node voltage is changed by ΔV, the resulting change at the source of the cascode device is much less. In a sense, the cascode transistor "shields" the input device from voltage variations at the output. The shielding property of cascodes proves useful in many circuits.

Example 3.16

Two identical NMOS transistors are used as constant current sources in a system [Fig. 3.61(a)]. However, due to internal circuitry of the system, V_X is higher than V_Y by ΔV.

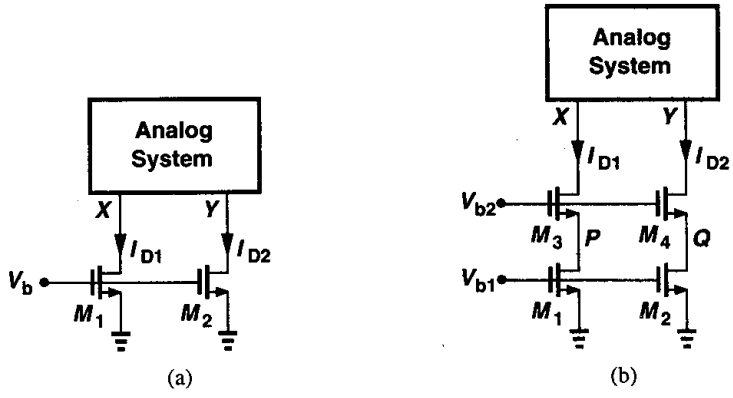


Figure 3.61

(a) Calculate the resulting difference between I_{D1} and I_{D2} if $\lambda \neq 0$.
(b) Add cascode devices to M_1 and M_2 and repeat part (a).

Solution

(a) We have

$$I_{D1} - I_{D2} = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_b - V_{TH})^2(\lambda V_{DS1} - \lambda V_{DS2}) \quad (3.127)$$

$$= \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_b - V_{TH})^2(\lambda \Delta V). \quad (3.128)$$

(b) As shown in Fig. 3.61(b), cascoding reduces the effect of V_X and V_Y upon I_{D1} and I_{D2}, respectively. As depicted in Fig. 3.23 and implied by Eq. (3.63), a difference ΔV between V_X and V_Y translates to a difference ΔV_{PQ} between P and Q equal to

$$\Delta V_{PQ} = \Delta V \frac{r_{O1}}{[1 + (g_{m3} + g_{mb3})r_{O3}]r_{O1} + r_{O3}} \tag{3.129}$$

$$\approx \frac{\Delta V}{(g_{m3} + g_{mb3})r_{O3}}. \tag{3.130}$$

Thus,

$$I_{D1} - I_{D2} = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_b - V_{TH})^2 \frac{\lambda \Delta V}{(g_{m3} + g_{mb3})r_{O3}}. \tag{3.131}$$

In other words, cascoding reduces the mismatch between I_{D1} and I_{D2} by $(g_{m3} + g_{mb3})r_{O3}$.

The shielding property of cascodes diminishes if the cascode device enters the triode region. To understand why, let us consider the circuit in Fig. 3.62, assuming V_X decreases from a large positive value. As V_X falls below $V_{b2} - V_{TH2}$, M_2 requires a greater gate-source

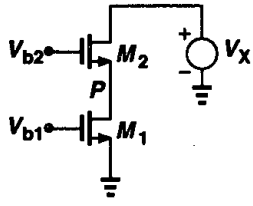


Figure 3.62 Output swing of cascode stage.

overdrive so as to sustain the current drawn by M_1. We can write

$$I_{D2} = \frac{1}{2}\mu_n C_{ox}\left(\frac{W}{L}\right)_2 [2(V_{b2} - V_P - V_{TH2})(V_X - V_P) - (V_X - V_P)^2], \tag{3.132}$$

concluding that as V_X decreases, V_P also drops so that I_{D2} remains constant. In other words, variation of V_X is less attenuated as it appears at P. If V_X falls sufficiently, V_P goes below $V_{b1} - V_{TH1}$, driving M_1 into the triode region.

3.5.1 Folded Cascode

The idea behind the cascode structure is to convert the input voltage to a current and apply the result to a common-gate stage. However, the input device and the cascode device need not be of the same type. For example, as depicted in Fig. 3.63(a), a PMOS-NMOS combination performs the same function. In order to bias M_1 and M_2, a current source must be added as in Fig. 3.63(b). The small-signal operation is as follows. If V_{in} becomes more positive, $|I_{D1}|$ decreases, forcing I_{D2} to increase and hence V_{out} to drop. The voltage gain and output impedance of the circuit can be obtained as calculated for the NMOS-NMOS

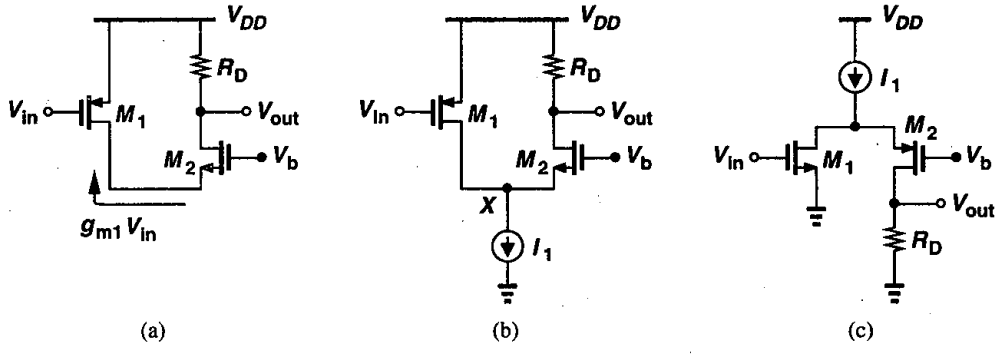


Figure 3.63 (a) Simple folded cascode, (b) folded cascode with proper biasing, (c) folded cascode with NMOS input.

cascode of Fig. 3.50. Shown in Fig. 3.63(c) is an NMOS-PMOS cascode. The advantages and disadvantages of these types will be explained later.

The structures of Figs. 3.63(b) and (c) are called "folded cascode" stages because the small-signal current is "folded" up [in Fig. 3.63(b)] or down [in Fig. 3.63(c)]. Note that the total bias current in this case must be higher than that in Fig. 3.50 to achieve comparable performance.

It is instructive to examine the large-signal behavior of a folded-cascode stage. Suppose in Fig. 3.63(b), V_{in} decreases from V_{DD} to zero. For $V_{in} > V_{DD} - |V_{TH1}|$, M_1 is off and M_2 carries all of I_1,[2] yielding $V_{out} = V_{DD} - I_1 R_D$. For $V_{in} < V_{DD} - |V_{TH1}|$, M_1 turns on in saturation, giving

$$I_{D2} = I_1 - \frac{1}{2}\mu_p C_{ox}\left(\frac{W}{L}\right)_1 (V_{DD} - V_{in} - |V_{TH1}|)^2. \tag{3.133}$$

As V_{in} drops, I_{D2} decreases further, falling to zero if $I_{D1} = I_1$. For this to occur:

$$\frac{1}{2}\mu_p C_{ox}\left(\frac{W}{L}\right)_1 (V_{DD} - V_{in1} - |V_{TH1}|)^2 = I_1. \tag{3.134}$$

Thus,

$$V_{in1} = V_{DD} - \sqrt{\frac{2I_1}{\mu_p C_{ox}(W/L)_1}} - |V_{TH1}|. \tag{3.135}$$

If V_{in} falls below this level, I_{D1} tends to be greater than I_1 and M_1 enters the triode region so as to allow $I_{D1} = I_1$. The result is plotted in Fig. 3.64.

What happens to V_X in the above test? As I_{D2} drops, V_X rises, reaching $V_b - V_{TH2}$ for $I_{D2} = 0$. As M_1 enters the triode region, V_X approaches V_{DD}.

[2] If I_1 is excessively large, M_2 may enter deep triode region, possibly driving I_1 into the triode region as well.

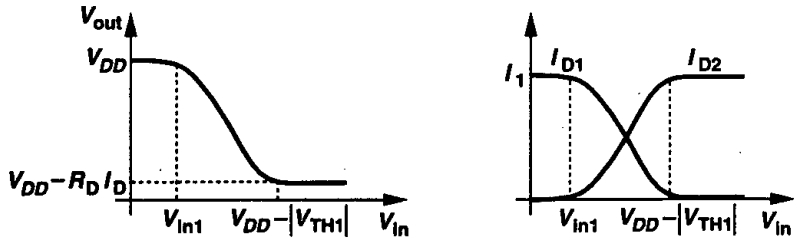


Figure 3.64 Large-signal characteristics of folded cascode.

Example 3.17

Calculate the output impedance of the folded cascode shown in Fig. 3.65 where M_3 operates as a current source.

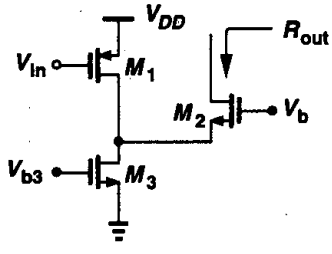


Figure 3.65

Solution

Using (3.60), we have

$$R_{out} = [1 + (g_{m2} + g_{mb2})r_{O2}](r_{O1} \| r_{O3}) + r_{O2}. \tag{3.136}$$

Thus, the circuit exhibits an output impedance lower than that of a nonfolded cascode.

In order to achieve a high voltage gain, the load of a folded cascode can be implemented as a cascode itself (Fig. 3.66). This structure is studied more extensively in Chapter 9.

Throughout this chapter, we have attempted to *increase* the output resistance of voltage amplifiers so as to obtain a high gain. This may seem to make the speed of the circuit quite susceptible to the load capacitance. However, as explained in Chapter 8, a high output impedance per se does not pose a serious issue if the amplifier is placed in a proper feedback loop.

3.6 Choice of Device Models

In this chapter, we have developed various expressions for the properties of single-stage amplifiers. For example, the voltage gain of a degenerated common-source stage can be as simple as $-R_D/(R_S + g_m^{-1})$ or as complex as Eq. (3.71). How does one choose a sufficiently accurate device model or expression?

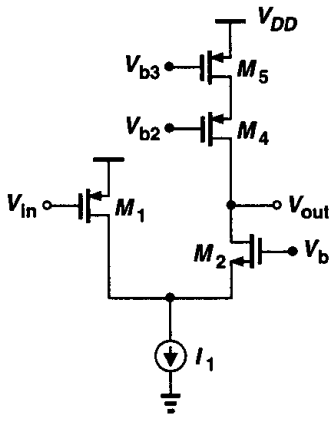


Figure 3.66 Folded cascode with cascode load.

The proper choice is not always straightforward and it is a skill gained by practice, experience, and intuition. However, some general principles in choosing the model for each transistor can be followed. First, break the circuit down into a number of familiar topologies. Next, concentrate on each subcircuit and use the simplest transistor model (a single voltage-dependent current source for FETs operating in saturation) for all transistors. If the drain of a device is connected to a high impedance (e.g., the drain of another), then add r_O to its model. At this point, the basic properties of most circuits can be determined by inspection. In a second, more accurate iteration, the body effect of devices whose source or bulk is not at ac ground can be included as well.

For bias calculations, it is usually adequate to neglect channel-length modulation and body effect in the first pass. These effects do introduce some error but they can be included in the next iteration step—after the basic properties are understood.

In today's analog design, simulation of circuits is essential because the behavior of short-channel MOSFETs cannot be predicted accurately by hand calculations. Nonetheless, if the designer avoids a simple and intuitive analysis of the circuit and hence skips the task of gaining insight, then he/she cannot interpret the simulation results intelligently. For this reason, we say, "Don't let the computer think for you."

Problems

Unless otherwise stated, in the following problems, use the device data shown in Table 2.1 and assume $V_{DD} = 3$ V where necessary. All device dimensions are effective values and in microns.

3.1. For the circuit of Fig. 3.9, calculate the small-signal voltage gain if $(W/L)_1 = 50/0.5$, $(W/L)_2 = 10/0.5$, and $I_{D1} = I_{D2} = 0.5$ mA. What is the gain if M_2 is implemented as a diode-connected PMOS device (Fig. 3.12)?

3.2. In the circuit of Fig. 3.14, assume $(W/L)_1 = 50/0.5$, $(W/L)_2 = 50/2$, and $I_{D1} = I_{D2} = 0.5$ mA when both devices are in saturation. Recall that $\lambda \propto 1/L$.

(a) Calculate the small-signal voltage gain.

(b) Calculate the maximum output voltage swing while both devices are saturated.

3.3. In the circuit of Fig. 3.3(a), assume $(W/L)_1 = 50/0.5$, $R_D = 2\,\text{k}\Omega$, and $\lambda = 0$.
- **(a)** What is the small-signal gain if M_1 is in saturation and $I_D = 1$ mA?
- **(b)** What input voltage places M_1 at the edge of the triode region? What is the small-signal gain under this condition?
- **(c)** What input voltage drives M_1 into the triode region by 50 mV? What is the small-signal gain under this condition?

3.4. Suppose the common-source stage of Fig. 3.3(a) is to provide an output swing from 1 V to 2.5 V. Assume $(W/L)_1 = 50/0.5$, $R_D = 2\,\text{k}\Omega$, and $\lambda = 0$.
- **(a)** Calculate the input voltages that yield $V_{out} = 1$ V and $V_{out} = 2.5$ V.
- **(b)** Calculate the drain current and the transconductance of M_1 for both cases.
- **(c)** How much does the small-signal gain, $g_m R_D$, vary as the output goes from 1 V to 2.5 V? (Variation of small-signal gain can be viewed as nonlinearity.)

3.5. Calculate the intrinsic gain of an NMOS device and a PMOS device operating in saturation with $W/L = 50/0.5$ and $|I_D| = 0.5$ mA. Repeat these calculations if $W/L = 100/1$.

3.6. Plot the intrinsic gain of a satuated device versus the gate-source voltage if **(a)** the drain current is constant, **(b)** W and L are constant.

3.7. Plot the intrinsic gain of a saturated device versus W/L if **(a)** the gate-source voltage is constant, **(b)** the drain current is constant.

3.8. An NMOS transistor with $W/L = 50/0.5$ is biased with $V_G = +1.2$ V and $V_S = 0$. The drain voltage is varied from 0 to 3 V.
- **(a)** Assuming the bulk voltage is zero, plot the intrinsic gain versus V_{DS}.
- **(b)** Repeat part **(a)** for a bulk voltage of -1 V.

3.9. For an NMOS device operating in saturation, plot g_m, r_O, and $g_m r_O$ as the bulk voltage goes from 0 to $-\infty$ while other terminal voltages remain constant.

3.10. Consider the circuit of Fig. 3.9 with $(W/L)_1 = 50/0.5$ and $(W/L)_2 = 10/0.5$. Assume $\lambda = \gamma = 0$.
- **(a)** At what input voltage is M_1 at the edge of the triode region? What is the small-signal gain under this condition?
- **(b)** What input voltage drives M_1 into the triode region by 50 mV? What is the small-signal gain under this condition?

3.11. Repeat Problem 3.10 if body effect is not neglected.

3.12. In the circuit of Fig. 3.13, $(W/L)_1 = 20/0.5$, $I_1 = 1$ mA, and $I_S = 0.75$ mA. Assuming $\lambda = 0$, calculate $(W/L)_2$ such that M_1 is at the edge of the triode region. What is the small-signal voltage gain under this condition?

3.13. Plot the small-signal gain of the circuit shown in Fig. 3.13 as I_S goes from 0 to $0.75I_1$. Assume M_1 is always saturated and neglect channel-length modulation and body effect.

3.14. The circuit of Fig. 3.14 is designed to provide an output voltage swing of 2.2 V with a bias current of 1 mA and a small-signal voltage gain of 100. Calculate the dimensions of M_1 and M_2.

3.15. Sketch V_{out} versus V_{in} for the circuits of Fig. 3.67 as V_{in} varies from 0 to V_{DD}. Identify important transition points.

3.16. Sketch V_{out} versus V_{in} for the circuits of Fig. 3.68 as V_{in} varies from 0 to V_{DD}. Identify important transition points.

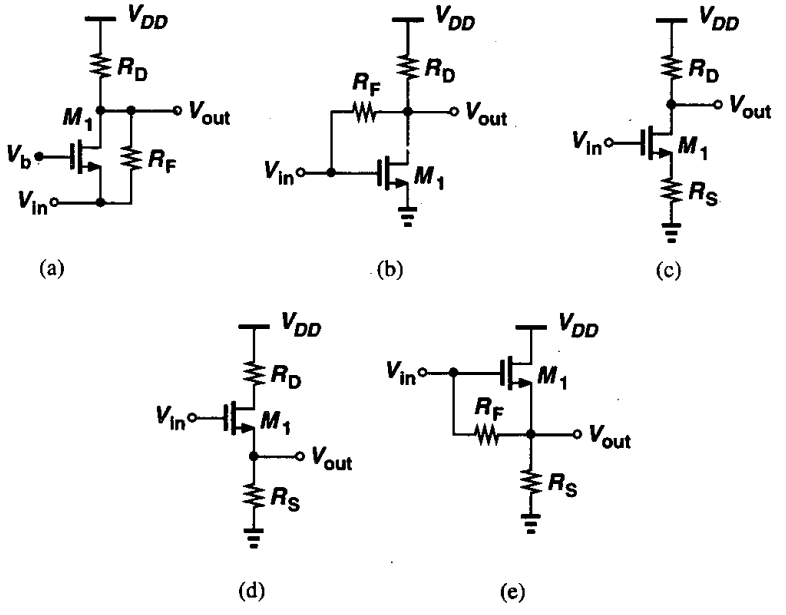


Figure 3.67

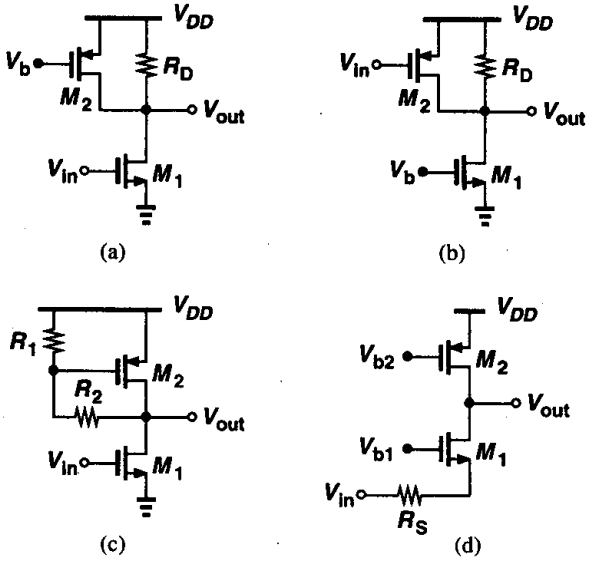


Figure 3.68

3.17. Sketch V_{out} versus V_{in} for the circuits of Fig. 3.69 as V_{in} varies from 0 to V_{DD}. Identify important transition points.

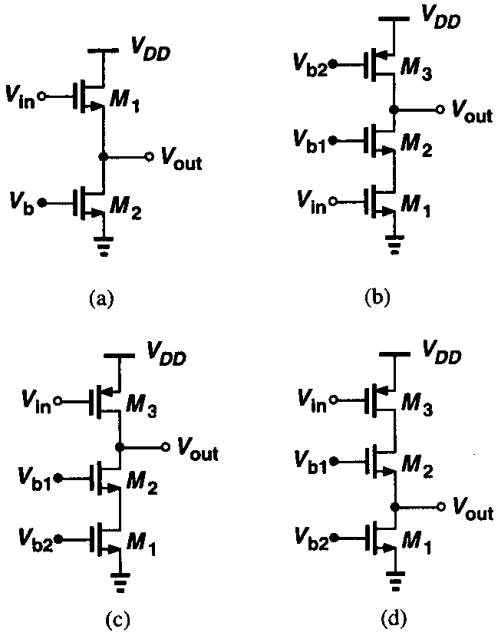


Figure 3.69

3.18. Sketch I_X versus V_X for the circuits of Fig. 3.70 as V_X varies from 0 to V_{DD}. Identify important transition points.

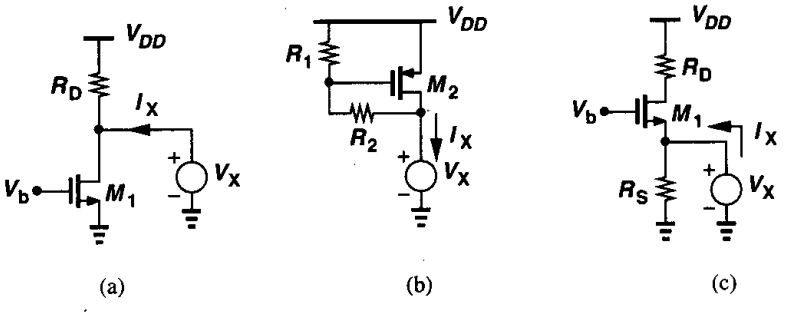


Figure 3.70

3.19. Sketch I_X versus V_X for the circuits of Fig. 3.71 as V_X varies from 0 to V_{DD}. Identify important transition points.

3.20. Assuming all MOSFETs are in saturation, calculate the small-signal voltage gain of each circuit in Fig. 3.72 ($\lambda \neq 0, \gamma = 0$).

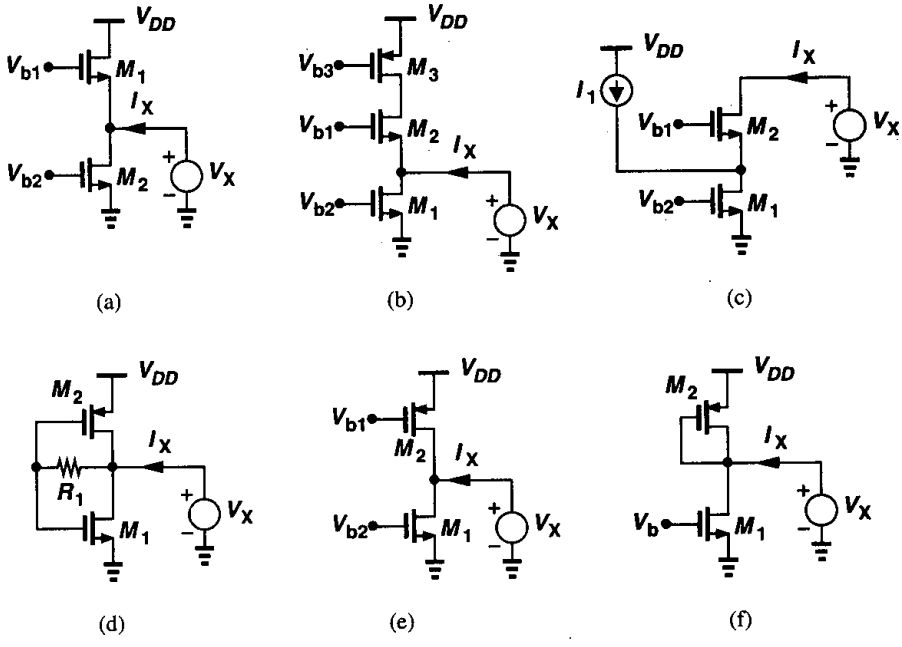


Figure 3.71

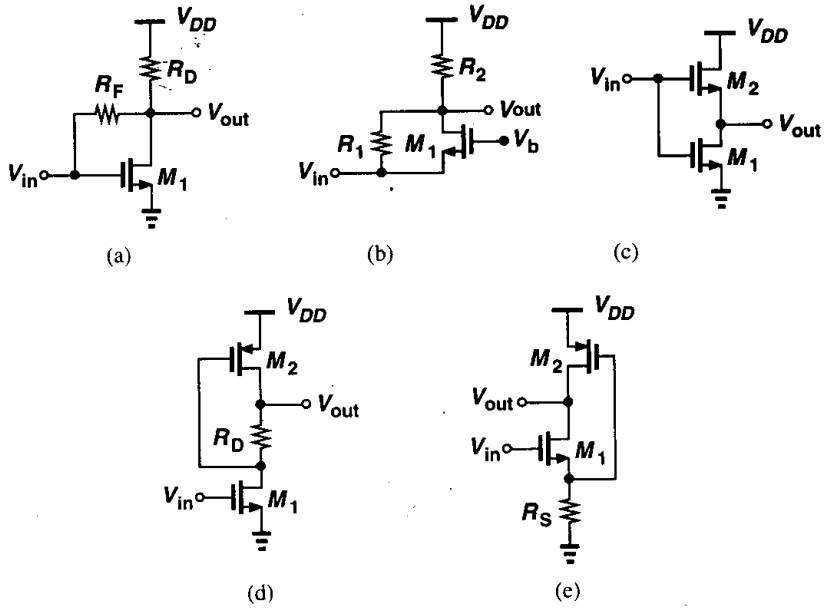


Figure 3.72

3.21. Assuming all MOSFETs are in saturation, calculate the small-signal voltage gain of each circuit in Fig. 3.73 ($\lambda \neq 0, \gamma = 0$).

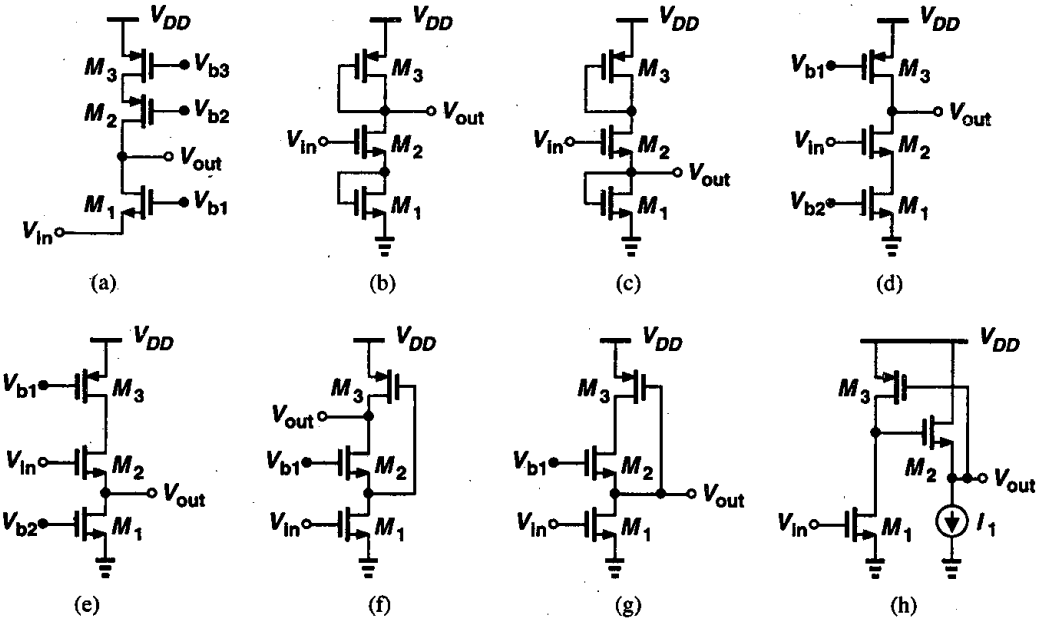


Figure 3.73

3.22. Sketch V_X and V_Y as a function of time for each circuit in Fig. 3.74. The initial voltage across C_1 is equal to V_{DD}.

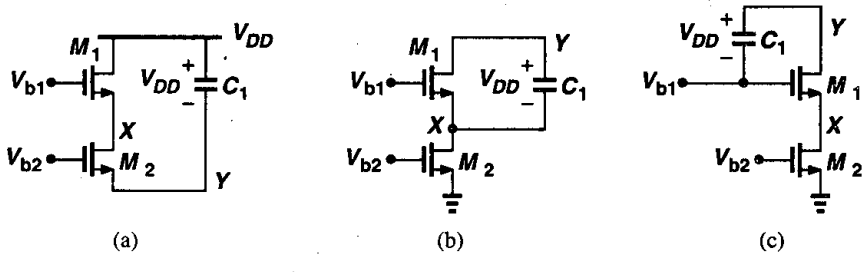


Figure 3.74

3.23. In the cascode stage of Fig. 3.50, assume $(W/L)_1 = 50/0.5$, $(W/L)_2 = 10/0.5$, $I_{D1} = I_{D2} = 0.5$ mA, and $R_D = 1$ kΩ.

(a) Choose V_b such that M_1 is 50 mV away from the triode region.

(b) Calculate the small-signal voltage gain.

(c) Using the value of V_b found in part (a), calculate the maximum output voltage swing. Which device enters the triode region first as V_{out} falls?

(d) Calculate the swing at node X for the maximum output swing obtained above.

3.24. Consider the circuit of Fig. 3.16 with $(W/L)_1 = 50/0.5$, $R_D = 2\ \text{k}\Omega$, and $R_S = 200\ \Omega$.
(a) Calculate the small-signal voltage gain if $I_D = 0.5$ mA.
(b) Assuming $\lambda = \gamma = 0$, calculate the input voltage that places M_1 at the edge of the triode region. What is the gain under this condition?

3.25. Suppose the circuit of Fig. 3.15 is designed for a voltage gain of 5. If $(W/L)_1 = 20/0.5$, $I_{D1} = 0.5$ mA, and $V_b = 0$ V.
(a) Calculate the aspect ratio of M_2.
(b) What input level places M_1 at the edge of the triode region. What is the small-signal gain under this condition?
(c) What input level places M_2 at the edge of the saturation region? What is the small-signal gain under this condition?

3.26. Sketch the small-signal voltage gain of the circuit shown in Fig. 3.15 as V_b varies from 0 to V_{DD}. Consider two cases: (a) M_1 enters the triode region before M_2 is saturated; (b) M_1 enters the triode region after M_2 is saturated.

3.27. A source follower can operate as a level shifter. Suppose the circuit of Fig. 3.30(b) is designed to shift the voltage level by 1 V, i.e., $V_{in} - V_{out} = 1$ V.
(a) Calculate the dimensions of M_1 and M_2 if $I_{D1} = I_{D2} = 0.5$ mA, $V_{GS2} - V_{GS1} = 0.5$ V, and $\lambda = \gamma = 0$.
(b) Repeat part (a) if $\gamma = 0.45\ \text{V}^{-1}$ and $V_{in} = 2.5$ V. What is the minimum input voltage for which M_2 remains saturated?

3.28. Sketch the small-signal gain, V_{out}/V_{in}, of the cascode stage shown in Fig. 3.50 as V_b goes from 0 to V_{DD}. Assume $\lambda = \gamma = 0$.

3.29. The cascode of Fig. 3.60 is designed to provide an output swing of 1.9 V with a bias current of 0.5 mA. If $\gamma = 0$ and $(W/L)_{1-4} = W/L$, calculate V_{b1}, V_{b2}, and W/L. What is the voltage gain if $L = 0.5\ \mu$m?

Chapter 4

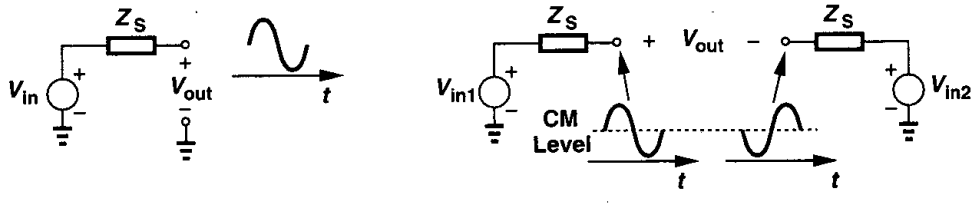
Differential Amplifiers

The differential amplifier is among the most important circuit inventions, dating back to the vacuum tube era. Offering many useful properties, differential operation has become the dominant choice in today's high-performance analog and mixed-signal circuits.

This chapter deals with the analysis and design of CMOS differential amplifiers. Following a review of single-ended and differential operation, we describe the basic differential pair, and analyze both the large-signal and the small-signal behavior. Next, we introduce the concept of common-mode rejection and formulate it for differential amplifiers. We then study differential pairs with diode-connected and current-source loads as well as differential cascode stages. Finally, we describe the Gibert cell.

4.1 Single-Ended and Differential Operation

A single-ended signal is defined as one that is measured with respect to a fixed potential, usually the ground. A differential signal is defined as one that is measured between two nodes that have *equal* and *opposite* signal excursions around a fixed potential. In the strict sense, the two nodes must also exhibit equal impedances to that potential. Fig. 4.1 illustrates the two types of signals conceptually. The "center" potential in differential signaling is called the "common-mode" (CM) level.



(a) (b)

Figure 4.1 (a) Single-ended and (b) differential signals.

An important advantage of differential operation over single-ended signaling is higher immunity to "environmental" noise. Consider the example depicted in Fig. 4.2, where two adjacent lines in a circuit carry a small, sensitive signal and a large clock waveform. Due to capacitive coupling between the lines, transitions on line L_2 corrupt the signal on line L_1. Now suppose, as shown in Fig. 4.2(b), the sensitive signal is distributed as two equal and opposite phases. If the clock line is placed midway between the two, the transitions disturb the differential phases by equal amounts, leaving the *difference* intact. Since the common-mode level of the two phases is disturbed but the differential output is not corrupted, we say this arrangement "rejects" common-mode noise.

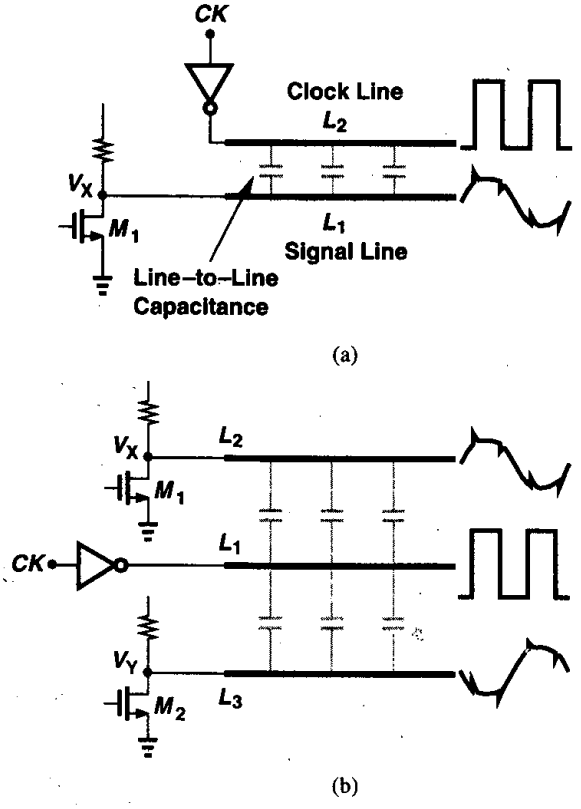


Figure 4.2 (a) Corruption of a signal due to coupling, (b) reduction of coupling by differential operation.

Another example of common-mode rejection occurs with noisy supply voltages. In Fig. 4.3(a), if V_{DD} varies by ΔV, then V_{out} changes by approximately the same amount, i.e., the output is quite susceptible to noise on V_{DD}. Now consider the circuit in Fig. 4.3(b). Here, if the circuit is symmetric, noise on V_{DD} affects V_X and V_Y but not $V_X - V_Y = V_{out}$. Thus, the circuit of Fig. 4.3(b) is much more robust to supply noise.

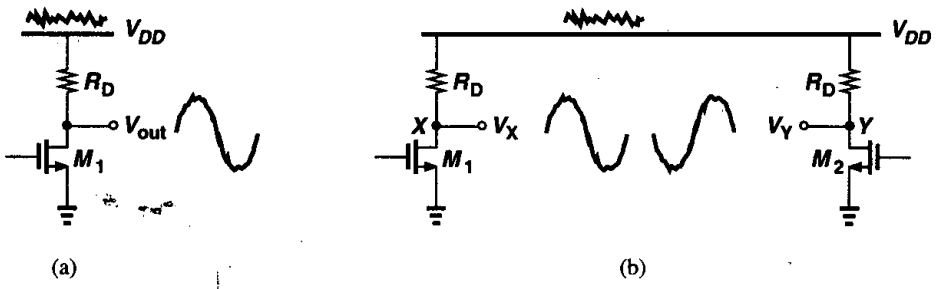


Figure 4.3 Effect of supply noise on (a) a single-ended circuit, (b) a differential circuit.

Thus far, we have seen the importance of employing differential paths for sensitive signals. It is also beneficial to employ differential distribution for *noisy lines*. For example, suppose the clock signal of Fig. 4.2 is distributed in differential form on two lines (Fig. 4.4). Then, with perfect symmetry, the components coupled from CK and $\overline{CK}$ to the signal line cancel each other.

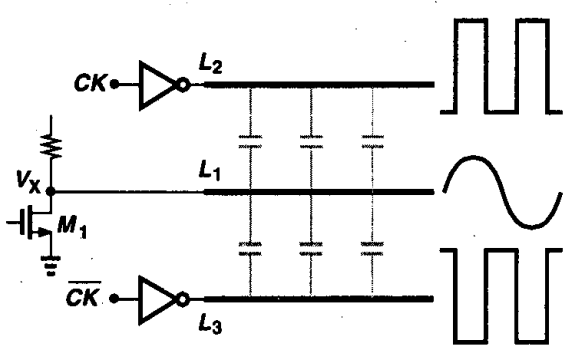


Figure 4.4 Reduction of coupled noise by differential operation.

Another useful property of differential signaling is the increase in maximum achievable voltage swings. In the circuit of Fig. 4.3, for example, the maximum output swing at X or Y is equal to $V_{DD} - (V_{GS} - V_{TH})$, whereas for $V_X - V_Y$, the peak-to-peak swing is equal to $2[V_{DD} - (V_{GS} - V_{TH})]$.

Other advantages of differential circuits over single-ended counterparts include simpler biasing and higher linearity (Chapter 13).

While it may seem that differential circuits occupy twice as much area as single-ended alternatives, in practice this is a minor drawback. Also, the suppression of nonideal effects by differential operation often results in a *smaller* area than that of a brute-force single-ended design. Furthermore, the numerous advantages of differential operation by far outweigh the possible increase in the area.

4.2 Basic Differential Pair

How do we amplify a differential signal? As suggested by the observations in the previous section, we may incorporate two identical single-ended signal paths to process the two phases [Fig. 4.5(a)]. Such a circuit indeed offers some of the advantages of differential

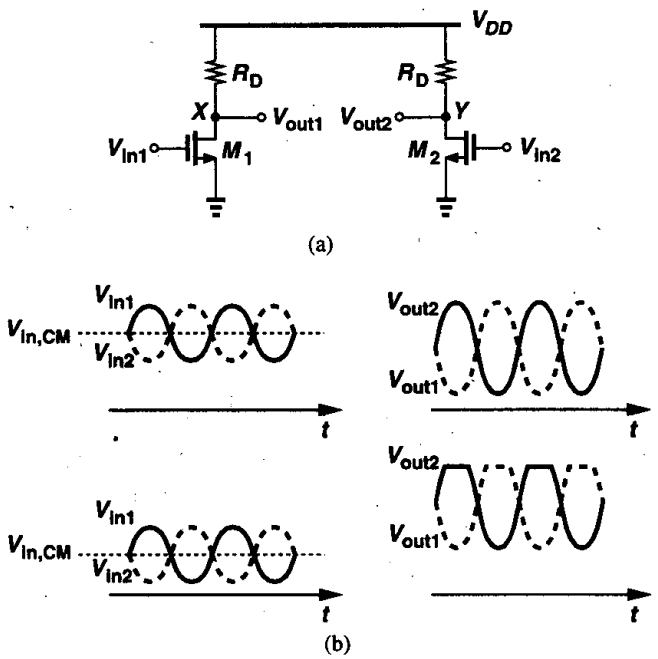


Figure 4.5 (a) Simple differential circuit, (b) illustration of sensitivity to the input common-mode level.

signaling: high rejection of supply noise, higher output swings, etc. But what happens if V_{in1} and V_{in2} experience a large common-mode disturbance or simply do not have a well-defined common-mode dc level? As the input CM level, $V_{in,CM}$, changes, so do the bias currents of M_1 and M_2, thus varying both the transconductance of the devices and the output CM level. The variation of the transconductance in turn leads to a change in the small-signal gain while the departure of the output CM level from its ideal value lowers the maximum allowable output swings. For example, as shown in Fig. 4.5(b), if the input CM level is excessively low, the minimum values of V_{in1} and V_{in2} may in fact turn off M_1 and M_2, leading to severe clipping at the output. Thus, it is important that the bias currents of the devices have minimal dependence on the input CM level.

A simple modification can resolve the above issue. Shown in Fig. 4.6, the "differential pair"[1] employs a current source I_{SS} to make $I_{D1} + I_{D2}$ independent of $V_{in,CM}$. Thus, if $V_{in1} = V_{in2}$, the bias current of each transistor equals $I_{SS}/2$ and the output common-mode

[1] Also called a source-coupled pair or (in the British literature) a long-tailed pair.

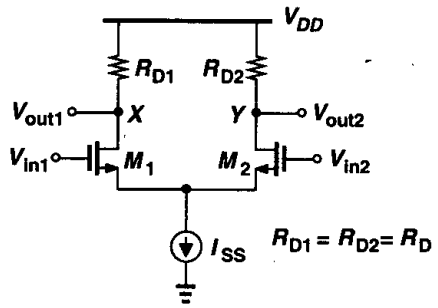


Figure 4.6 Basic differential pair.

level is $V_{DD} - R_D I_{SS}/2$. It is instructive to study the large-signal behavior of the circuit for both differential and common-mode input variations.

4.2.1 Qualitative Analysis

Let us assume that in Fig. 4.6, $V_{in1} - V_{in2}$ varies from $-\infty$ to $+\infty$. If V_{in1} is much more negative than V_{in2}, M_1 is off, M_2 is on, and $I_{D2} = I_{SS}$. Thus, $V_{out1} = V_{DD}$ and $V_{out2} = V_{DD} - R_D I_{SS}$. As V_{in1} is brought closer to V_{in2}, M_1 gradually turns on, drawing a fraction of I_{SS} from R_{D1} and hence lowering V_{out1}. Since $I_{D1} + I_{D2} = I_{SS}$, the drain current of M_2 decreases and V_{out2} rises. As shown in Fig. 4.7(a), for $V_{in1} = V_{in2}$, we have $V_{out1} = V_{out2} = V_{DD} - R_D I_{SS}/2$. As V_{in1} becomes more positive than V_{in2}, M_1 carries a greater current than does M_2 and V_{out1} drops below V_{out2}. For sufficiently large $V_{in1} - V_{in2}$, M_1 "hogs" all of I_{SS}, turning M_2 off. As a result, $V_{out1} = V_{DD} - R_D I_{SS}$ and $V_{out2} = V_{DD}$. Fig. 4.7 also plots $V_{out1} - V_{out2}$ versus $V_{in1} - V_{in2}$.

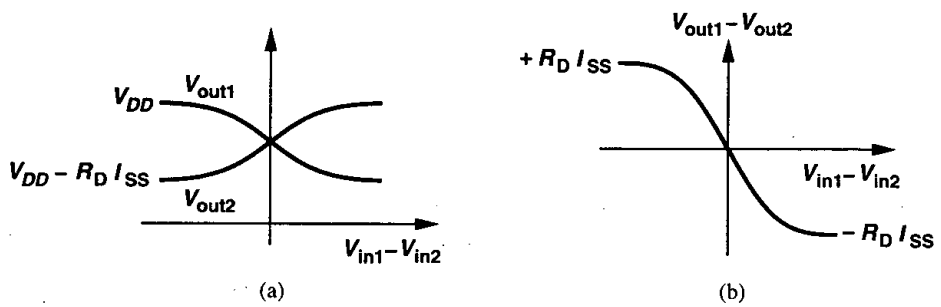


Figure 4.7 Input-output characteristics of a differential pair.

The foregoing analysis reveals two important attributes of the differential pair. First, the maximum and minimum levels at the output are well-defined (V_{DD} and $V_{DD} - R_D I_{SS}$, respectively) and independent of the input CM level. Second, the small-signal gain (the slope of $V_{out1} - V_{out2}$ versus $V_{in1} - V_{in2}$) is maximum for $V_{in1} = V_{in2}$, gradually falling to zero as $|V_{in1} - V_{in2}|$ increases. In other words, the circuit becomes more nonlinear as the input voltage swing increases. For $V_{in1} = V_{in2}$, we say the circuit is in equilibrium.

Now let us consider the common-mode behavior of the circuit. As mentioned earlier, the role of the tail current source is to suppress the effect of input CM level variations on the operation of M_1 and M_2 and the output level. Does this mean that $V_{in,CM}$ can assume arbitrarily low or high values? To answer this question, we set $V_{in1} = V_{in2} = V_{in,CM}$ and vary $V_{in,CM}$ from 0 to V_{DD}. Fig. 4.8(a) shows the circuit with I_{SS} implemented by an NFET. Note that the symmetry of the pair requires that $V_{out1} = V_{out2}$.

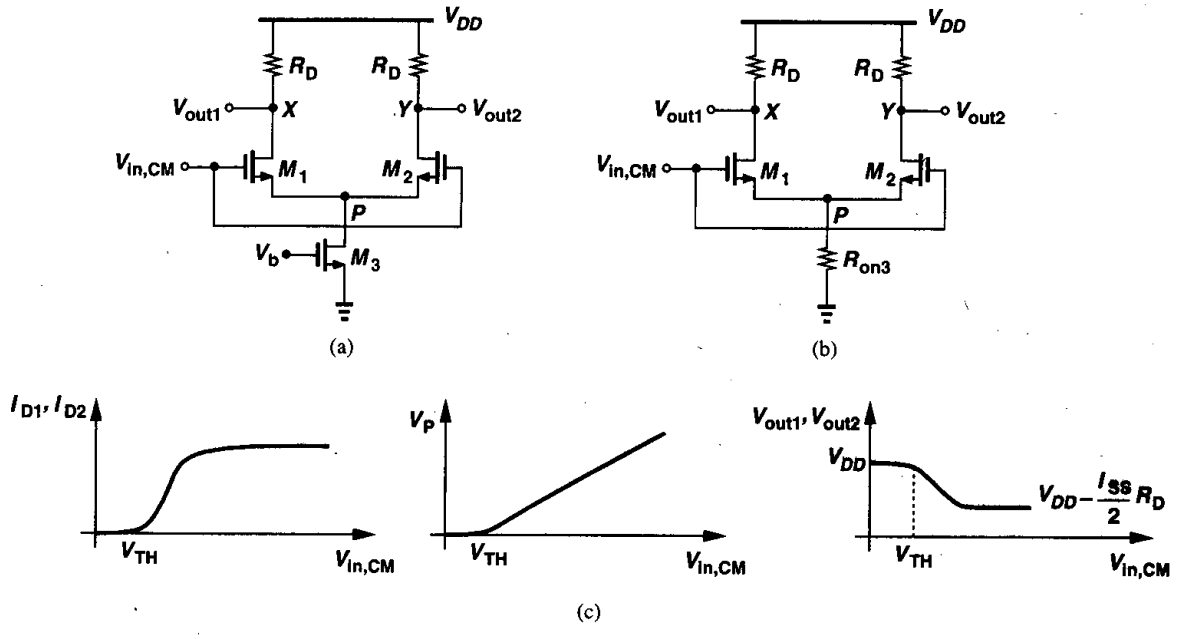


Figure 4.8 (a) Differential pair sensing an input common-mode change, (b) equivalent circuit if M_3 operates in deep triode region, (c) common-mode input-output characteristics.

What happens if $V_{in,CM} = 0$? Since the gate potential of M_1 and M_2 is not more positive than their source potential, both devices are off, yielding $I_{D3} = 0$. This indicates that M_3 is in deep triode region because V_b is high enough to create an inversion layer in the transistor. With $I_{D1} = I_{D2} = 0$, the circuit is incapable of signal amplification, and $V_{out1} = V_{out2} = V_{DD}$.

Now suppose $V_{in,CM}$ becomes more positive. Modeling M_3 by a resistor as in Fig. 4.8(b), we note that M_1 and M_2 turn on if $V_{in,CM} \geq V_{TH}$. Beyond this point, I_{D1} and I_{D2} continue to increase and V_P also rises [Fig. 4.8(c)]. In a sense, M_1 and M_2 constitute a source follower, forcing V_P to track $V_{in,CM}$. For a sufficiently high $V_{in,CM}$, the drain-source voltage of M_3 exceeds $V_{GS3} - V_{TH3}$, allowing the device to operate in saturation. The total current through M_1 and M_2 then remains constant. We conclude that for proper operation, $V_{in,CM} \geq V_{GS1} + (V_{GS3} - V_{TH3})$.

What happens if $V_{in,CM}$ rises further? Since V_{out1} and V_{out2} are relatively constant, we expect that M_1 and M_2 enter the triode region if $V_{in,CM} > V_{out1} + V_{TH} = V_{DD} - R_D I_{SS}/2 + V_{TH}$. This sets an upper limit on the input CM level. In summary, the allowable value of

$V_{in,CM}$ is bounded as follows:

$$V_{GS1} + (V_{GS3} - V_{TH3}) \leq V_{in,CM} \leq \min\left[V_{DD} - R_D\frac{I_{SS}}{2} + V_{TH},\ V_{DD}\right]. \quad (4.1)$$

Example 4.1

Sketch the small-signal differential gain of a differential pair as a function of the input CM level.

Solution

As shown in Fig. 4.9, the gain begins to increase as $V_{in,CM}$ exceeds V_{TH}. After the tail current source

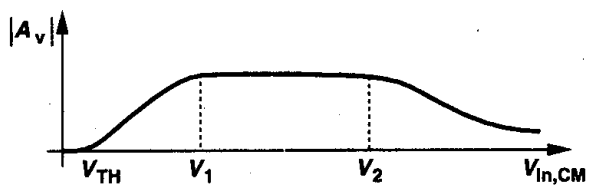


Figure 4.9

enters saturation ($V_{in,CM} = V_1$), the gain remains relatively constant. Finally, if $V_{in,CM}$ is so high that the input transistors enter the triode region ($V_{in,CM} = V_2$), the gain begins to fall.

With our understanding of differential and common-mode behavior of the differential pair, we can now answer another important question: How large can the output voltage swings of a differential pair be? As illustrated in Fig. 4.10, for M_1 and M_2 to be saturated, each output can go as high as V_{DD} but as low as approximately $V_{in,CM} - V_{TH}$. In other

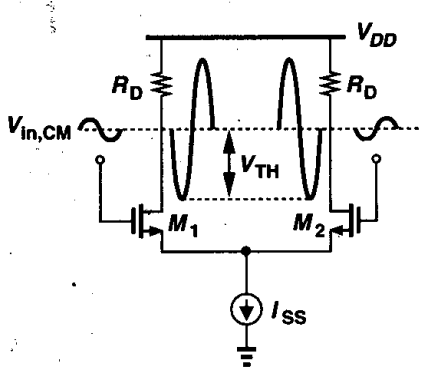


Figure 4.10 Maximum allowable output swings in a differential pair.

words, the higher the input CM level, the smaller the allowable output swings. For this reason, it is desirable to choose a relatively low $V_{in,CM}$, but the preceding stage may not provide such a level easily.

An interesting trade-off exists in the circuit of Fig. 4.10 between the maximum value of $V_{in,CM}$ and the differential *gain*. Similar to a simple common-source stage (Chapter 3),

the gain of a differential pair is a function of the dc drop across the load resistors. Thus, if $R_D I_{SS}/2$ is large, $V_{in,CM}$ must remain close to ground potential.

4.2.2 Quantitative Analysis

We now quantify the behavior of a MOS differential pair as a function of the input differential voltage. We begin with large-signal analysis to arrive at an expression for the plots shown in Fig. 4.7.

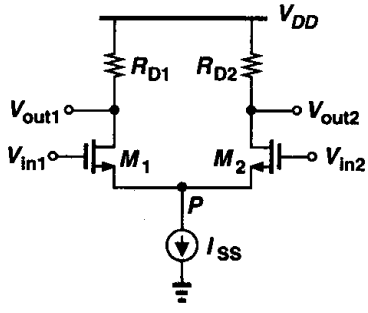


Figure 4.11 Differential pair.

For the differential pair in Fig. 4.11, we have $V_{out1} = V_{DD} - R_{D1} I_{D1}$ and $V_{out2} = V_{DD} - R_{D2} I_{D2}$, i.e., $V_{out1} - V_{out2} = R_{D2} I_{D2} - R_{D1} I_{D1} = R_D(I_{D2} - I_{D1})$ if $R_{D1} = R_{D2} = R_D$. Thus, we simply calculate I_{D1} and I_{D2} in terms of V_{in1} and V_{in2}, assuming the circuit is symmetric, M_1 and M_2 are saturated, and $\lambda = 0$. Since the voltage at node P is equal to $V_{in1} - V_{GS1}$ and $V_{in2} - V_{GS2}$,

$$V_{in1} - V_{in2} = V_{GS1} - V_{GS2}. \tag{4.2}$$

For a square-law device, we have:

$$(V_{GS} - V_{TH})^2 = \frac{I_D}{\frac{1}{2}\mu_n C_{ox} \frac{W}{L}}, \tag{4.3}$$

and, therefore,

$$V_{GS} = \sqrt{\frac{2I_D}{\mu_n C_{ox} \frac{W}{L}}} + V_{TH}. \tag{4.4}$$

It follows from (4.2) and (4.4) that

$$V_{in1} - V_{in2} = \sqrt{\frac{2I_{D1}}{\mu_n C_{ox} \frac{W}{L}}} - \sqrt{\frac{2I_{D2}}{\mu_n C_{ox} \frac{W}{L}}}. \tag{4.5}$$

Our objective is to calculate the differential output current, $I_{D1} - I_{D2}$. Squaring the two sides of (4.5) and recognizing that $I_{D1} + I_{D2} = I_{SS}$, we obtain

$$(V_{in1} - V_{in2})^2 = \frac{2}{\mu_n C_{ox} \dfrac{W}{L}} (I_{SS} - 2\sqrt{I_{D1} I_{D2}}). \tag{4.6}$$

That is,

$$\frac{1}{2} \mu_n C_{ox} \frac{W}{L} (V_{in1} - V_{in2})^2 - I_{SS} = -2\sqrt{I_{D1} I_{D2}}. \tag{4.7}$$

Squaring the two sides again and noting that $4I_{D1}I_{D2} = (I_{D1} + I_{D2})^2 - (I_{D1} - I_{D2})^2 = I_{SS}^2 - (I_{D1} - I_{D2})^2$, we arrive at

$$(I_{D1} - I_{D2})^2 = -\frac{1}{4} \left(\mu_n C_{ox} \frac{W}{L} \right)^2 (V_{in1} - V_{in2})^4 + I_{SS} \mu_n C_{ox} \frac{W}{L} (V_{in1} - V_{in2})^2. \tag{4.8}$$

Thus,

$$I_{D1} - I_{D2} = \frac{1}{2} \mu_n C_{ox} \frac{W}{L} (V_{in1} - V_{in2}) \sqrt{\frac{4I_{SS}}{\mu_n C_{ox} \dfrac{W}{L}} - (V_{in1} - V_{in2})^2}. \tag{4.9}$$

As expected, $I_{D1} - I_{D2}$ is an odd function of $V_{in1} - V_{in2}$, falling to zero for $V_{in1} = V_{in2}$. As $|V_{in1} - V_{in2}|$ increases from zero, $|I_{D1} - I_{D2}|$ also increases because the factor preceding the square root rises more rapidly than the argument in the square root drops.[2]

Before examining (4.9) further, it is instructive to calculate the slope of the characteristic, i.e., the equivalent G_m of M_1 and M_2. Denoting $I_{D1} - I_{D2}$ and $V_{in1} - V_{in2}$ by ΔI_D and ΔV_{in}, respectively, the reader can show that

$$\frac{\partial \Delta I_D}{\partial \Delta V_{in}} = \frac{1}{2} \mu_n C_{ox} \frac{W}{L} \frac{\dfrac{4I_{SS}}{\mu_n C_{ox} W/L} - 2\Delta V_{in}^2}{\sqrt{\dfrac{4I_{SS}}{\mu_n C_{ox} W/L} - \Delta V_{in}^2}}. \tag{4.10}$$

For $\Delta V_{in} = 0$, $G_m = \sqrt{\mu_n C_{ox} (W/L) I_{SS}}$. Moreover, since $V_{out1} - V_{out2} = R_D \Delta I = R_D G_m \Delta V_{in}$, we can write the small-signal differential voltage gain of the circuit in the equilibrium condition as

$$|A_v| = \sqrt{\mu_n C_{ox} \frac{W}{L} I_{SS}} R_D. \tag{4.11}$$

[2]It is interesting to note that, even though I_{D1} and I_{D2} are *even* functions of their respective gate-source voltages, $I_{D1} - I_{D2}$ is an odd function of $V_{in1} - V_{in2}$. This effect is studied in Chapter 13.

Equation (4.10) also suggests that G_m falls to zero for $\Delta V_{in} = \sqrt{2I_{SS}/(\mu_n C_{ox} W/L)}$. As we will see below, this value of ΔV_{in} plays an important role in the operation of the circuit.

Let us now examine Eq. (4.9) more closely. It appears that the argument in the square root drops to zero for $\Delta V_{in} = \sqrt{4I_{SS}/(\mu_n C_{ox} W/L)}$, implying that ΔI_D crosses zero at *two* different values of ΔV_{in}. This was not predicted in our qualitative analysis in Fig. 4.7. This conclusion, however, is incorrect. To understand why, recall that (4.9) was derived with the assumption that both M_1 and M_2 are on. In reality, as ΔV_{in} exceeds a limit, one transistor carries the entire I_{SS}, turning off the other.[3] Denoting this value by ΔV_{in1}, we have $I_{D1} = I_{SS}$ and $\Delta V_{in1} = V_{GS1} - V_{TH}$ because M_2 is nearly off. It follows that

$$\Delta V_{in1} = \sqrt{\frac{2I_{SS}}{\mu_n C_{ox} \dfrac{W}{L}}}. \tag{4.12}$$

For $\Delta V_{in} > \Delta V_{in1}$, M_2 is off and (4.9) does not hold. As mentioned above, G_m falls to zero for $\Delta V_{in} = \Delta V_{in1}$. Figure 4.12 plots the behavior.

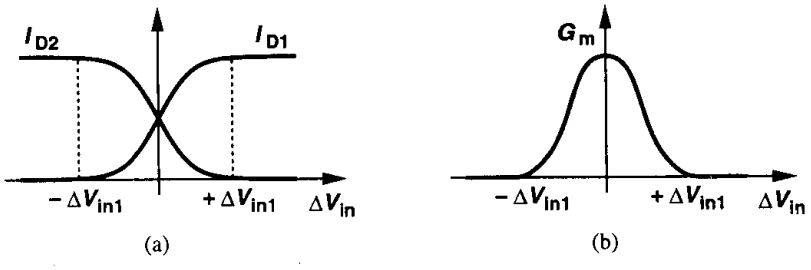


Figure 4.12 Variation of drain currents and overall transconductance of a differential pair versus input voltage.

Example 4.2

Plot the input-output characteristic of a differential pair as the device width and the tail current vary.

Solution

Consider the characteristic shown in Fig. 4.13(a). As W/L increases, ΔV_{in1} decreases, narrowing the input range across which both devices are on [Fig. 4.13(b)]. As I_{SS} increases, both the input range and the output current swing increase [Fig. 4.13(c)]. Intuitively, we expect the circuit to become more linear as I_{SS} increases or W/L decreases.

The value of ΔV_{in1} given by (4.12) in essence represents the maximum differential input that the circuit can "handle." It is possible to relate ΔV_{in1} to the overdrive voltage

[3] We neglect subthreshold conduction here.

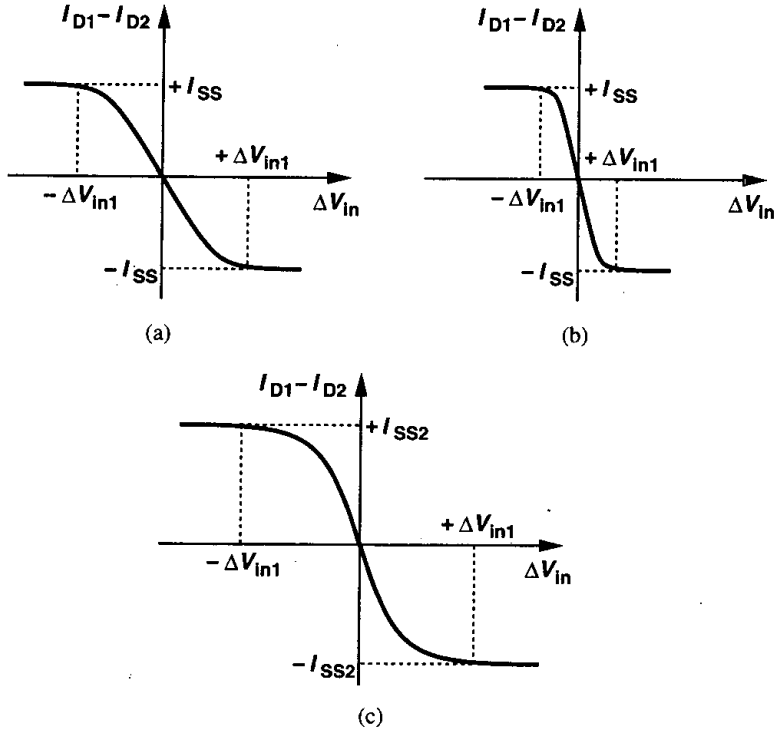


Figure 4.13

of M_1 and M_2 in equilibrium. For a zero differential input, $I_{D1} = I_{D2} = I_{SS}/2$, and hence

$$(V_{GS} - V_{TH})_{1,2} = \sqrt{\frac{I_{SS}}{\mu_n C_{ox} \dfrac{W}{L}}}. \tag{4.13}$$

Thus, the equilibrium overdrive is equal to $\Delta V_{in1}/\sqrt{2}$. The point is that increasing ΔV_{in1} to make the circuit more linear inevitably increases the overdrive voltage of M_1 and M_2. For a given I_{SS}, this is accomplished only by reducing W/L and hence the transconductance of the transistors.

We now study the small-signal behavior of differential pairs. As depicted in Fig. 4.14, we apply small signals V_{in1} and V_{in2} and assume M_1 and M_2 are saturated. What is the differential voltage gain, $V_{out}/(V_{in1} - V_{in2})$? Recall from Eq. (4.11) that this quantity equals $\sqrt{\mu_n C_{ox} I_{SS} W/L} R_D$. Since in the vicinity of equilibrium, each transistor carries approximately $I_{SS}/2$, this expression reduces to $g_m R_D$, where g_m denotes the transconductance of M_1 and M_2. To arrive at the same result by small-signal analysis, we employ two different methods, each providing insight into the circuit's operation. We assume $R_{D1} = R_{D2} = R_D$.

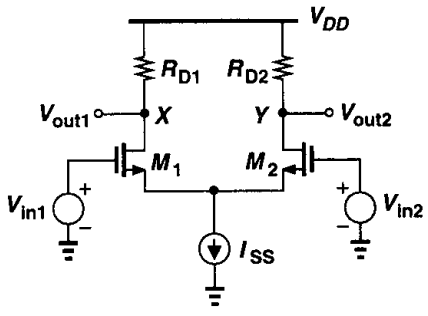
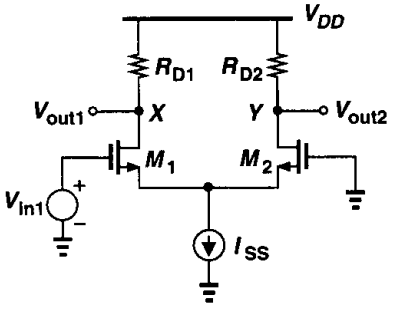
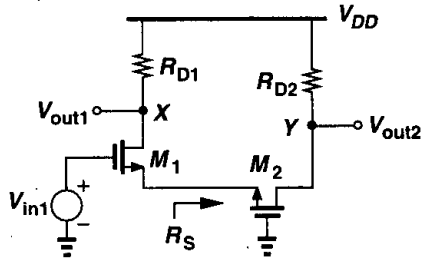


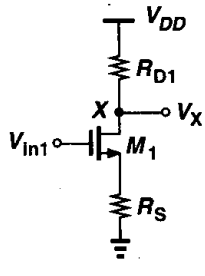
Figure 4.14 Differential pair with small-signal inputs.



(a)



(b)



(c)

Figure 4.15 (a) Differential pair sensing one input signal, (b) circuit of (a) viewed as a CS stage degenerated by M_2, (c) equivalent circuit of (b).

Method I The circuit of Fig. 4.14 is driven by two independent signals. Thus, the output can be computed by superposition.

Let us set V_{in2} to zero and find the effect of V_{in1} at X and Y [Fig. 4.15(a)]. To obtain V_X, we note that M_1 forms a common-source stage with a degeneration resistance equal to the impedance seen looking into the source of M_2 [Fig. 4.15(b)]. Neglecting channel-length

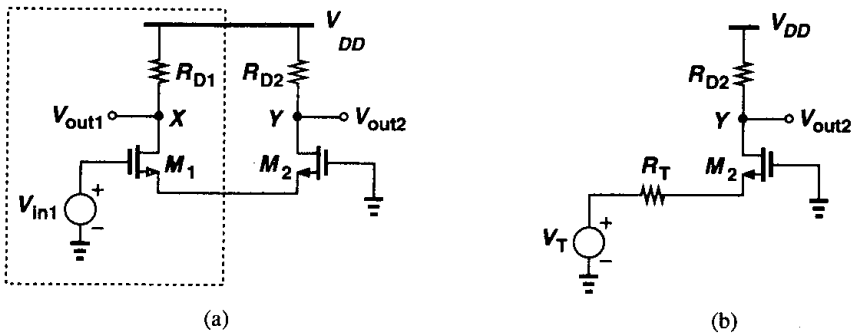


Figure 4.16 Replacing M_1 by a Thevenin equivalent.

modulation and body effect, we have $R_S = 1/g_{m2}$ [Fig. 4.15(c)] and

$$\frac{V_X}{V_{in1}} = \frac{-R_D}{\dfrac{1}{g_{m1}} + \dfrac{1}{g_{m2}}}. \tag{4.14}$$

To calculate V_Y, we note that M_1 drives M_2 as a source follower and replace V_{in1} and M_1 by a Thevenin equivalent (Fig. 4.16): the Thevenin voltage $V_T = V_{in1}$ and the resistance $R_T = 1/g_{m1}$. Here, M_2 operates as a common-gate stage, exhibiting a gain equal to

$$\frac{V_Y}{V_{in1}} = \frac{R_D}{\dfrac{1}{g_{m2}} + \dfrac{1}{g_{m1}}}. \tag{4.15}$$

It follows from (4.14) and (4.15) that the overall voltage gain for V_{in1} is

$$(V_X - V_Y)|_{\text{Due to } V_{in1}} = \frac{-2R_D}{\dfrac{1}{g_{m1}} + \dfrac{1}{g_{m2}}} V_{in1}, \tag{4.16}$$

which, for $g_{m1} = g_{m2} = g_m$ reduces to

$$(V_X - V_Y)|_{\text{Due to } V_{in1}} = -g_m R_D V_{in1}. \tag{4.17}$$

By virtue of symmetry, the effect of V_{in2} at X and Y is identical to that of V_{in1} except for a change in the polarities:

$$(V_X - V_Y)|_{\text{Due to } V_{in2}} = g_m R_D V_{in2}. \tag{4.18}$$

Adding the two sides of (4.17) and (4.18) to perform superposition, we have

$$\frac{(V_X - V_Y)_{tot}}{V_{in1} - V_{in2}} = -g_m R_D. \tag{4.19}$$

Comparison of (4.17), (4.18), and (4.19) indicates that the magnitude of the differential gain is equal to $g_m R_D$ regardless of how the inputs are applied: in Figs. 4.15 and 4.16, the input is applied to only one side whereas in Fig. 4.14 the input is the *difference* between two sources. It is also important to recognize that if the output is single-ended, i.e., it is sensed between X or Y and ground, the gain is halved.

Example 4.3

In the circuit of Fig. 4.17, M_2 is twice as wide as M_1. Calculate the small-signal gain if the bias values of V_{in1} and V_{in2} are equal.

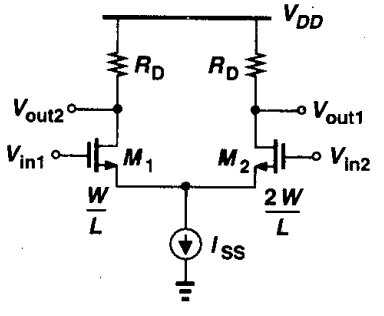


Figure 4.17

Solution

If the gates of M_1 and M_2 are at the same dc potential, then $V_{GS1} = V_{GS2}$ and $I_{D2} = 2I_{D1} = 2I_{SS}/3$. Thus, $g_{m1} = \sqrt{2\mu_n C_{ox}(W/L)I_{SS}/3}$ and $g_{m2} = \sqrt{2\mu_n C_{ox}(2W/L)2I_{SS}/3} = 2g_{m1}$. Following the same procedure as above, the reader can show that

$$|A_v| = \frac{2R_D}{\frac{1}{g_{m1}} + \frac{1}{2g_{m1}}} \tag{4.20}$$

$$= \frac{4}{3} g_{m1} R_D. \tag{4.21}$$

Note that, for a given I_{SS}, this value is lower than the gain of a symmetric differential pair (with $2W/L$ for each device) [Eq. (4.19)] because g_{m1} is smaller.

How does the gain of a differential pair compare with that of a common-source stage? For a given *total* bias current, the value of g_m in (4.19) is $1/\sqrt{2}$ times that of a single transistor biased at I_{SS} with the same dimensions. Thus, the total gain is proportionally less. Equivalently, for given device dimensions and load impedance, a differential pair achieves the same gain as a CS stage at the cost of twice the bias current.

Method II If a fully-symmetric differential pair senses differential inputs (i.e., the two inputs change by equal and opposite amounts from the equilibrium condition), then the concept of "half circuit" can be applied. We first prove a lemma.

Lemma. Consider the symmetric circuit shown in Fig. 4.18(a), where D_1 and D_2 represent

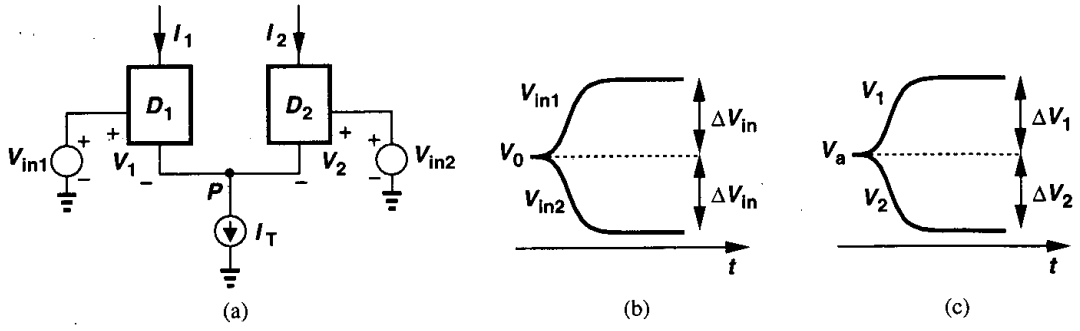


Figure 4.18 Illustration of why node P is a virtual ground.

any three-terminal active device. Suppose V_{in1} changes from V_0 to $V_0 + \Delta V_{in}$ and V_{in2} from V_0 to $V_0 - \Delta V_{in}$ [Fig. 4.18(b)]. Then, if the circuit remains linear, V_P does not change. Assume $\lambda = 0$.

Proof. Let us assume that V_1 and V_2 have an equilibrium value of V_a and change by ΔV_1 and ΔV_2, respectively [Fig. 4.18(c)]. The output currents therefore change by $g_m \Delta V_1$ and $g_m \Delta V_2$. Since $I_1 + I_2 = I_T$, we have $g_m \Delta V_1 + g_m \Delta V_2 = 0$, i.e., $\Delta V_1 = -\Delta V_2$. We also know $V_{in1} - V_1 = V_{in2} - V_2$, and hence $V_0 + \Delta V_{in} - (V_a + \Delta V_1) = V_0 - \Delta V_{in} - (V_a + \Delta V_2)$. Consequently, $2\Delta V_{in} = \Delta V_1 - \Delta V_2 = 2\Delta V_1$. In other words, if V_{in1} and V_{in2} change by $+\Delta V_{in}$ and $-\Delta V_{in}$, respectively, then V_1 and V_2 change by the same values, i.e., a differential change in the inputs is simply "absorbed" by V_1 and V_2. In fact, since $V_P = V_{in1} - V_1$, and since V_1 exhibits the same change as V_{in1}, V_P does not change. ❑

The proof of the foregoing lemma can also be invoked from symmetry. As long as the operation remains linear so that the difference between the bias currents of D_1 and D_2 is negligible, the circuit is symmetric. Thus, V_P cannot "favor" the change at one input and "ignore" the other.

From yet another point of view, the effect of D_1 and D_2 at node P can be represented by Thevenin equivalents (Fig. 4.19). If V_{T1} and V_{T2} change by equal and opposite amounts and R_{T1} and R_{T2} are equal, then V_P remains constant. We emphasize that this is valid if the changes are small such that we can assume $R_{T1} = R_{T2}$.[4]

The above lemma greatly simplifies the small-signal analysis of differential amplifiers. As shown in Fig. 4.20, since V_P experiences no change, node P can be considered "ac ground" and the circuit can be decomposed into two separate halves, hence the term "half-circuit concept" [1]. We can write $V_X / V_{in1} = -g_m R_D$ and $V_Y/(-V_{in1}) = -g_m R_D$, where V_{in1} and $-V_{in1}$ denote the voltage *change* on each side. Thus, $(V_X - V_Y)/(2V_{in1}) = -g_m R_D$.

[4]It is also possible to derive an expression for the large-signal behavior of V_P and prove that for small $V_{in1} - V_{in2}$, V_P remains constant. We defer this calculation to Chapter 14.

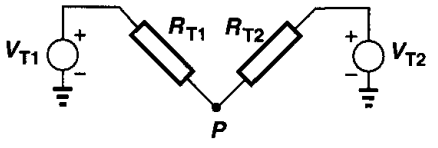


Figure 4.19 Replacing each half of a differential pair by a Thevenin equivalent.

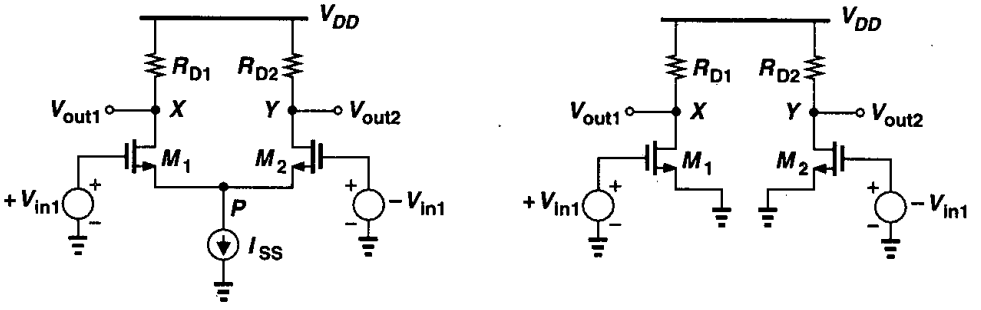


Figure 4.20 Application of the half-circuit concept.

Example 4.4

Calculate the differential gain of the circuit of Fig. 4.20(a) if $\lambda \neq 0$.

Solution

Applying the half-circuit concept as illustrated in Fig. 4.21, we have $V_X/V_{in1} = -g_m(R_D \| r_{O1})$ and $V_Y/(-V_{in1}) = -g_m(R_D \| r_{O2})$, thus arriving at $(V_X - V_Y)/(2V_{in1}) = -g_m(R_D \| r_O)$, where $r_O = r_{O1} = r_{O2}$. Note that Method I would require lengthy calculations here.

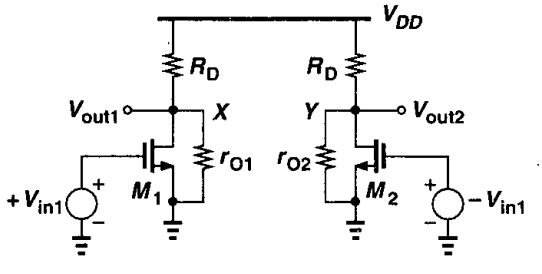


Figure 4.21

The half-circuit concept provides a powerful technique for analyzing symmetric differential pairs with fully differential inputs. But what happens if the two inputs are not fully

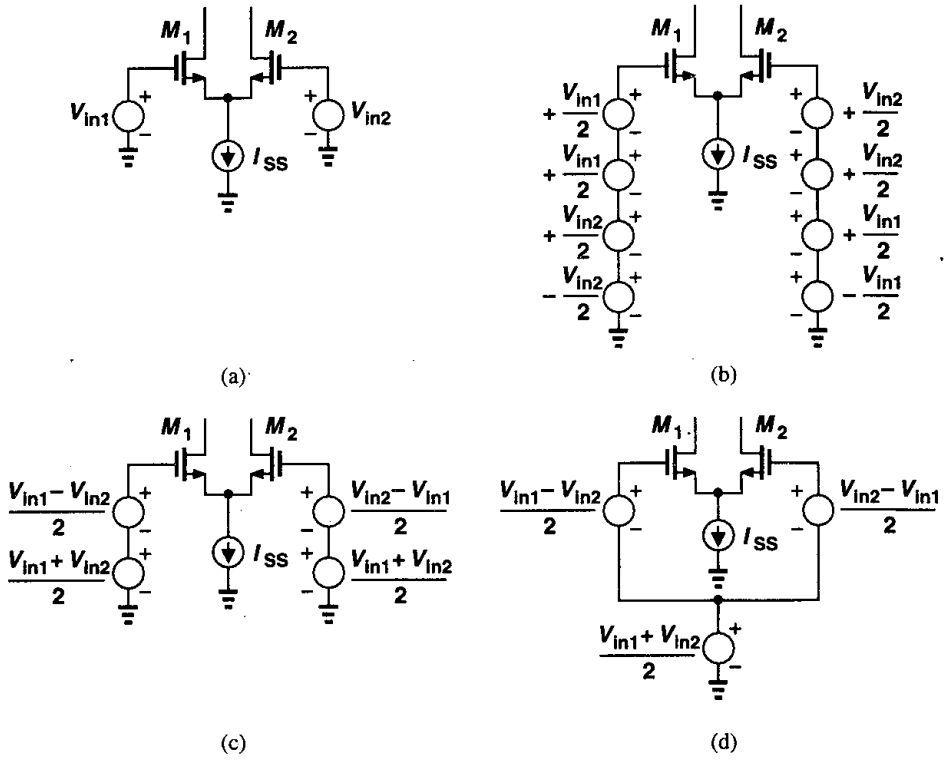


Figure 4.22 Conversion of arbitrary inputs to differential and common-mode components.

differential [Fig. 4.22(a)]? As depicted in Figs. 4.22(b) and (c), the two inputs V_{in1} and V_{in2} can be viewed as

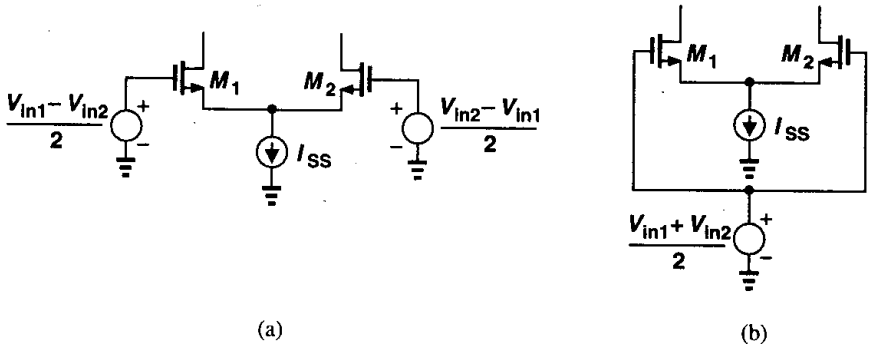
$$V_{in1} = \frac{V_{in1} - V_{in2}}{2} + \frac{V_{in1} + V_{in2}}{2} \tag{4.22}$$

$$V_{in2} = \frac{V_{in2} - V_{in1}}{2} + \frac{V_{in1} + V_{in2}}{2}. \tag{4.23}$$

Since the second term is common to both inputs, we obtain the equivalent circuit in Fig. 4.22(d), recognizing that the circuit senses a combination of a differential input and a common-mode variation. Therefore, as illustrated in Fig. 4.23, the effect of each type of input can be computed by superposition, with the half-circuit concept applied to the differential-mode operation.

Example 4.5

In the circuit of Fig. 4.20(a), calculate V_X and V_Y if $V_{in1} \neq -V_{in2}$ and $\lambda \neq 0$.



(a) (b)

Figure 4.23 Superposition for differential and common-mode signals.

Solution

For differential-mode operation, we have from Fig. 4.24(a)

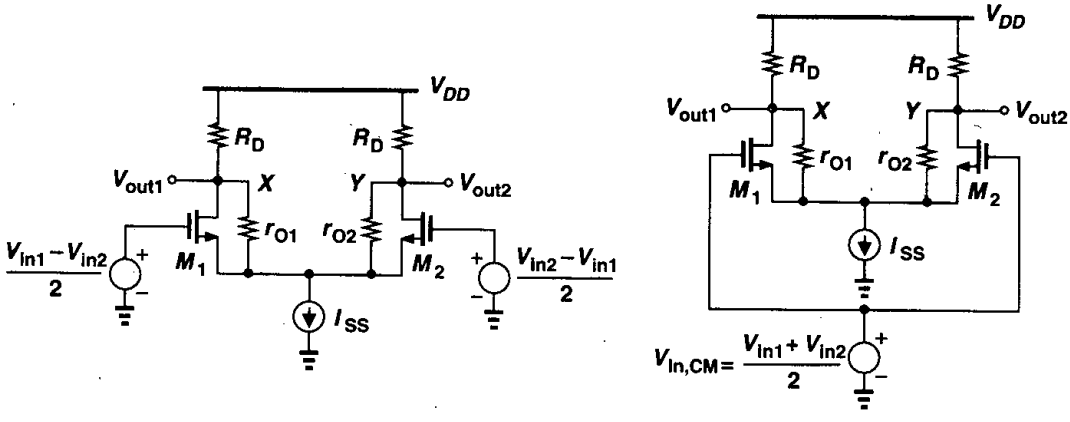
$$V_X = -g_m(R_D \| r_{O1}) \frac{V_{in1} - V_{in2}}{2} \tag{4.24}$$

$$V_Y = -g_m(R_D \| r_{O2}) \frac{V_{in2} - V_{in1}}{2}. \tag{4.25}$$

That is,

$$V_X - V_Y = -g_m(R_D \| r_O)(V_{in1} - V_{in2}), \tag{4.26}$$

which is to be expected.



(a) (b)

Figure 4.24

For common-mode operation, the circuit reduces to that in Fig. 4.24(b). How much do V_X and V_Y change as $V_{in,CM}$ changes? If the circuit is fully symmetric and I_{SS} an ideal current source, the

current drawn by M_1 and M_2 from R_{D1} and R_{D2} is exactly equal to $I_{SS}/2$ and independent of $V_{in,CM}$. Thus, V_X and V_Y experience no change as $V_{in,CM}$ varies. Interestingly, the circuit simply amplifies the difference between V_{in1} and V_{in2} while eliminating the effect of $V_{in,CM}$.

4.3 Common-Mode Response

An important attribute of differential amplifiers is their ability to suppress the effect of common-mode perturbations. Example 4.5 portrays an idealized case of common-mode response. In reality, neither is the circuit fully symmetric nor does the current source exhibit an infinite output impedance. As a result, a fraction of the input CM variation appears at the output.

We first assume the circuit is symmetric but the current source has a finite output impedance, R_{SS} [Fig. 4.25(a)]. As $V_{in,CM}$ changes, so does V_P, thereby increasing the drain currents of M_1 and M_2 and lowering both V_X and V_Y. Owing to symmetry, V_X remains equal to V_Y and, as depicted in Fig. 4.25(b), the two nodes can be shorted together. Since M_1 and M_2 are now "in parallel," i.e., they share all of their respective terminals, the

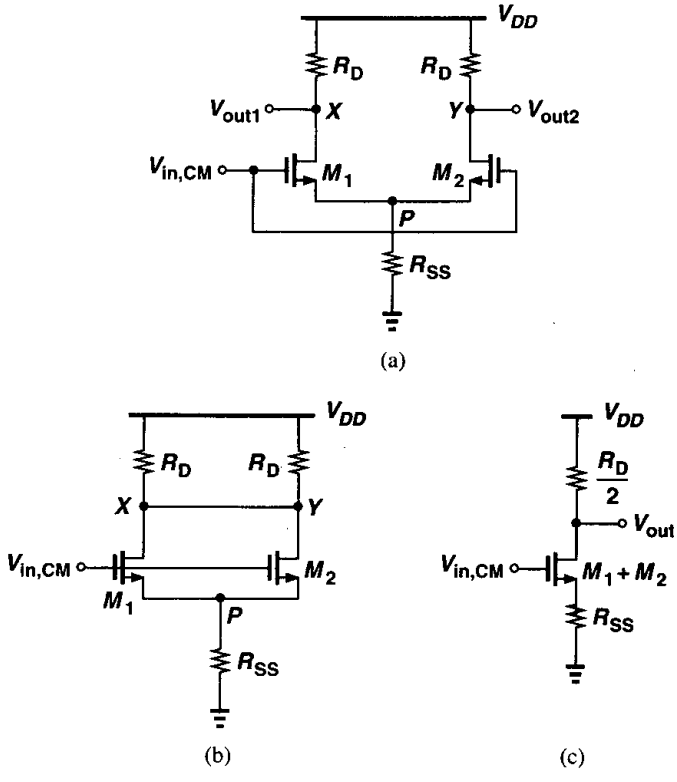


Figure 4.25 (a) Differential pair sensing CM input, (b) simplified version of (a), (c) equivalent circuit of (b).

circuit can be reduced to that in Fig. 4.25(c). Note that the compound device, $M_1 + M_2$, has twice the width and the bias current of each of M_1 and M_2 and, therefore, twice their transconductance. The CM gain of the circuit is thus equal to

$$A_{v,CM} = \frac{V_{out}}{V_{in,CM}} \tag{4.27}$$

$$= -\frac{R_D/2}{1/(2g_m) + R_{SS}}, \tag{4.28}$$

where g_m denotes the transconductance of each of M_1 and M_2 and $\lambda = \gamma = 0$.

What is the significance of this calculation? In a symmetric circuit, input CM variations disturb the bias points, altering the small-signal gain and possibly limiting the output voltage swings. This can be illustrated by an example.

Example 4.6

The circuit of Fig. 4.26 uses a resistor rather than a current source to define a tail current of 1 mA.

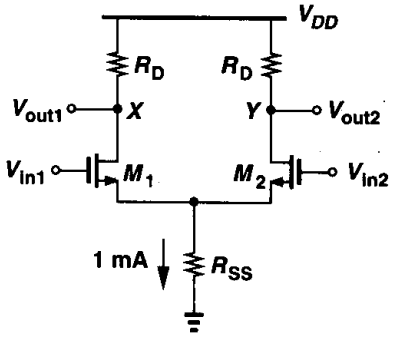


Figure 4.26

Assume $(W/L)_{1,2} = 25/0.5$, $\mu_n C_{ox} = 50\ \mu\text{A/V}^2$, $V_{TH} = 0.6$ V, $\lambda = \gamma = 0$, and $V_{DD} = 3$ V.

(a) What is the required input CM for which R_{SS} sustains 0.5 V?

(b) Calculate R_D for a differential gain of 5.

(c) What happens at the output if the input CM level is 50 mV higher than the value calculated in (a)?

Solution

(a) Since $I_{D1} = I_{D2} = 0.5$ mA, we have

$$V_{GS1} = V_{GS2} = \sqrt{\frac{2I_{D1}}{\mu_n C_{ox}\dfrac{W}{L}}} + V_{TH} \tag{4.29}$$

$$= 1.23\ \text{V}. \tag{4.30}$$

Thus, $V_{in,CM} = V_{GS1} + 0.5$ V$= 1.73$ V. Note that $R_{SS} = 500\ \Omega$.

(b) The transconductance of each device is $g_m = \sqrt{2\mu_n C_{ox}(W/L)I_{D1}} = 1/(632\ \Omega)$, requiring $R_D = 3.16\ \text{k}\Omega$ for a gain of 5.

Note that the output bias level is equal to $V_{DD} - I_{D1}R_D = 1.42$ V. Since $V_{in,CM} = 1.73$ V and $V_{TH} = 0.6$ V, the transistors are 290 mV away from the triode region.

(c) If $V_{in,CM}$ increases by 50 mV, the equivalent circuit of Fig. 4.25(c) suggests that V_X and V_Y drop by

$$|\Delta V_{X,Y}| = \Delta V_{in,CM} \frac{R_D/2}{R_{SS} + 1/(2g_m)} \tag{4.31}$$

$$= 50 \text{ mV} \times 1.94 \tag{4.32}$$

$$= 96.8 \text{ mV}. \tag{4.33}$$

Now, M_1 and M_2 are only 143 mV away from the triode region because the input CM level has increased by 50 mV and the output CM level has decreased by 96.8 mV.

The foregoing discussion indicates that the finite output impedance of the tail current source results in some common-mode gain in a symmetric differential pair. Nonetheless, this is usually a minor concern. More troublesome is the variation of the *differential* output as a result of a change in $V_{in,CM}$, an effect that occurs because in reality the circuit is not fully symmetric, i.e., the two sides suffer from slight mismatches during manufacturing. For example, in Fig. 4.25(a), R_{D1} may not be exactly equal to R_{D2}.

We now study the effect of input common-mode variation if the circuit is asymmetric and the tail current source suffers from a finite output impedance. Suppose, as shown in Fig. 4.27, $R_{D1} = R_D$ and $R_{D2} = R_D + \Delta R_D$, where ΔR_D denotes a small mismatch and

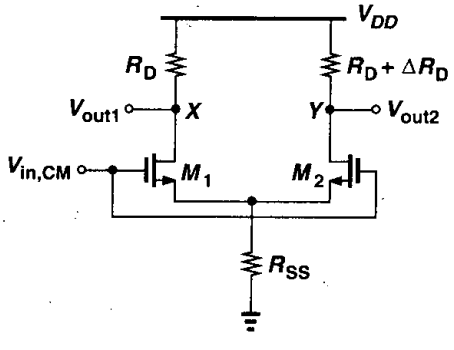


Figure 4.27 Common-mode response in the presence of resistor mismatch.

the circuit is otherwise symmetric. What happens to V_X and V_Y as $V_{in,CM}$ increases? Since M_1 and M_2 are identical, I_{D1} and I_{D2} increase by $[g_m/(1 + 2g_m R_{SS})]\Delta V_{in,CM}$, but V_X and V_Y change by different amounts:

$$\Delta V_X = -\Delta V_{in,CM} \frac{g_m}{1 + 2g_m R_{SS}} R_D \tag{4.34}$$

$$\Delta V_Y = -\Delta V_{in,CM} \frac{g_m}{1 + 2g_m R_{SS}} (R_D + \Delta R_D). \tag{4.35}$$

Thus, a common-mode change at the input introduces a *differential* component at the output. We say the circuit exhibits common-mode to differential conversion. This is a critical problem because if the input of a differential pair includes both a differential signal and

common-mode noise, the circuit corrupts the amplified differential signal by the input CM change. The effect is illustrated in Fig. 4.28.

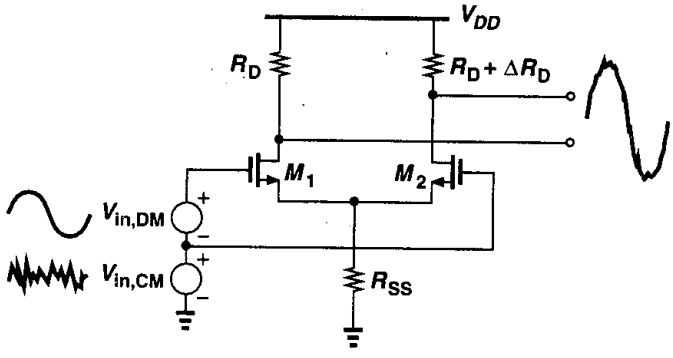


Figure 4.28 Effect of CM noise in the presence of resistor mismatch.

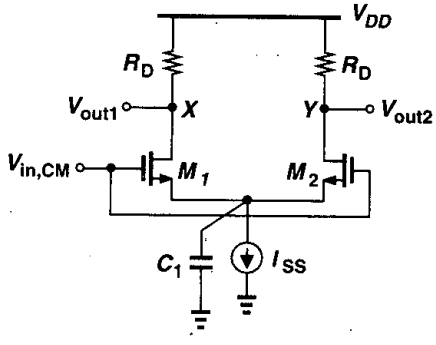


Figure 4.29 CM response with finite tail capacitance.

In summary, the common-mode response of differential pairs depends on the output impedance of the tail current source and asymmetries in the circuit, manifesting itself through two effects: variation of the output CM level (in the absence of mismatches) and conversion of input common-mode variations to differential components at the output. In analog circuits, the latter effect is much more severe than the former. For this reason, the common-mode response should usually be studied with mismatches taken into account.

How significant is common-mode to differential conversion? We make two observations. First, as the *frequency* of the CM disturbance increases, the total capacitance shunting the tail current source introduces larger tail current variations. Thus, even if the output *resistance* of the current source is high, common-mode to differential conversion becomes significant at high frequencies. Shown in Fig. 4.29, this capacitance arises from the parasitics of the current source itself as well as the source-bulk junctions of M_1 and M_2. Second, the asymmetry in the circuit stems from both the load resistors and the input transistors, the latter contributing a typically much greater mismatch.

Let us now study the asymmetry resulting from mismatches between M_1 and M_2 in Fig. 4.30(a). Owing to dimension and threshold voltage mismatches, the two transistors

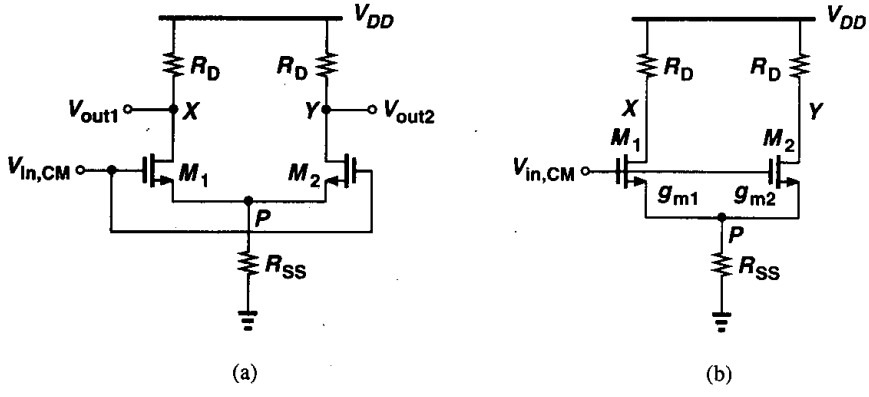


Figure 4.30 (a) Differential pair sensing CM input, (b) equivalent circuit of (a).

carry slightly different currents and exhibit unequal transconductances. To calculate the gain from $V_{in,CM}$ to X and Y, we use the equivalent circuit in Fig. 4.30(b), writing $I_{D1} = g_{m1}(V_{in,CM} - V_P)$ and $I_{D2} = g_{m2}(V_{in,CM} - V_P)$. That is,

$$(g_{m1} + g_{m2})(V_{in,CM} - V_P)R_{SS} = V_P, \tag{4.36}$$

and

$$V_P = \frac{(g_{m1} + g_{m2})R_{SS}}{(g_{m1} + g_{m2})R_{SS} + 1} V_{in,CM}. \tag{4.37}$$

We now obtain the output voltages as

$$V_X = -g_{m1}(V_{in,CM} - V_P)R_D \tag{4.38}$$

$$= \frac{-g_{m1}}{(g_{m1} + g_{m2})R_{SS} + 1} R_D V_{in,CM} \tag{4.39}$$

and

$$V_Y = -g_{m2}(V_{in,CM} - V_P)R_D \tag{4.40}$$

$$= \frac{-g_{m2}}{(g_{m1} + g_{m2})R_{SS} + 1} R_D V_{in,CM}. \tag{4.41}$$

The differential component at the output is therefore given by

$$V_X - V_Y = -\frac{g_{m1} - g_{m2}}{(g_{m1} + g_{m2})R_{SS} + 1} R_D V_{in,CM}. \tag{4.42}$$

In other words, the circuit converts input CM variations to a differential error by a factor

equal to

$$A_{CM-DM} = -\frac{\Delta g_m R_D}{(g_{m1} + g_{m2})R_{SS} + 1}, \tag{4.43}$$

where A_{CM-DM} denotes common-mode to differential-mode conversion and $\Delta g_m = g_{m1} - g_{m2}$.

Example 4.7

Two differential pairs are cascaded as shown in Fig. 4.31. Transistors M_3 and M_4 suffer from a g_m

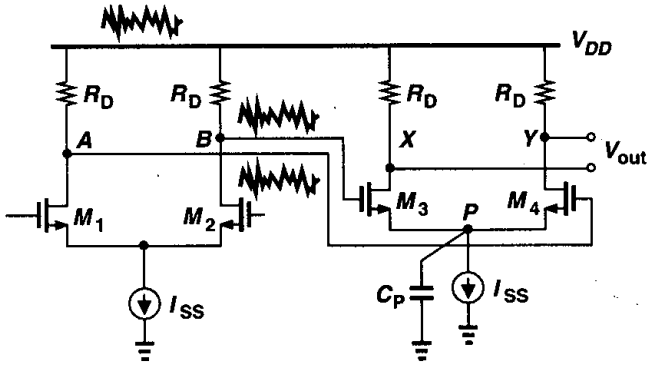


Figure 4.31

mismatch of Δg_m, the total parasitic capacitance at node P is represented by C_P, and the circuit is otherwise symmetric. What fraction of the supply noise appears as a differential component at the output? Assume $\lambda = \gamma = 0$.

Solution

Neglecting the capacitance at nodes A and B, we note that the supply noise appears at these nodes with no attenuation. Substituting $1/(C_P s)$ for R_{SS} in (4.43) and taking the magnitude, we have

$$|A_{CM-DM}| = \frac{\Delta g_m R_D}{\sqrt{1 + (g_{m3} + g_{m4})^2 \left|\frac{1}{C_P \omega}\right|^2}}. \tag{4.44}$$

The key point is that the effect becomes more noticeable as the supply noise frequency, ω, increases.

For meaningful comparison of differential circuits, the undesirable differential component produced by CM variations must be normalized to the wanted differential output resulting from amplification. We define the "common-mode rejection ratio" (CMRR) as

$$CMRR = \left|\frac{A_{DM}}{A_{CM-DM}}\right|. \tag{4.45}$$

If only g_m mismatch is considered, the reader can show from the analysis of Fig. 4.15 that

$$|A_{DM}| = \frac{R_D}{2}\frac{g_{m1} + g_{m2} + 4g_{m1}g_{m2}R_{SS}}{1 + (g_{m1} + g_{m2})R_{SS}}, \tag{4.46}$$

where it is assumed $V_{in1} = -V_{in2}$, and hence

$$CMRR = \frac{g_{m1} + g_{m2} + 4g_{m1}g_{m2}R_{SS}}{2\Delta g_m} \tag{4.47}$$

$$\approx \frac{g_m}{\Delta g_m}(1 + 2g_m R_{SS}), \tag{4.48}$$

where g_m denotes the mean value, i.e., $g_m = (g_{m1} + g_{m2})/2$. In practice, all mismatches must be taken into account.

4.4 Differential Pair with MOS Loads

The load of a differential pair need not be implemented by linear resistors. As with the common-source stages studied in Chapter 3, differential pairs can employ diode-connected or current-source loads (Fig. 4.32). The small-signal differential gain can be derived using

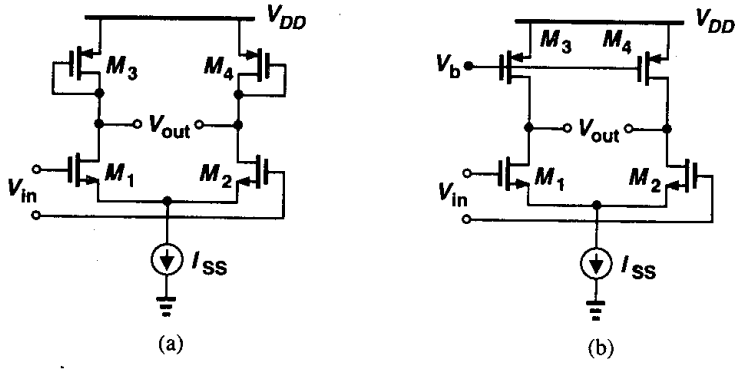


Figure 4.32 Differential pair with (a) diode-connected and (b) current-source loads.

the half-circuit concept. For Fig. 4.32(a),

$$A_v = -g_{mN}\left(g_{mP}^{-1} \| r_{ON} \| r_{OP}\right) \tag{4.49}$$

$$\approx -\frac{g_{mN}}{g_{mP}}, \tag{4.50}$$

where subscripts N and P denote NMOS and PMOS, respectively. Expressing g_{mN} and g_{mP} in terms of device dimensions, we have

$$A_v \approx -\sqrt{\frac{\mu_n(W/L)_N}{\mu_p(W/L)_P}}. \tag{4.51}$$

For Fig. 4.32(b), we have

$$A_v = -g_{mN}(r_{ON} \| r_{OP}). \tag{4.52}$$

In the circuit of Fig. 4.32(a), the diode-connected loads consume voltage headroom, thus creating a trade-off between the output voltage swings, the voltage gain, and the input CM range. Recall from Eq. (3.35) that, for given bias current and input device dimensions, the circuit's gain and the PMOS overdrive voltage scale together. To achieve a higher gain, $(W/L)_P$ must decrease, thereby increasing $|V_{GSP} - V_{THP}|$ and lowering the CM level at nodes X and Y.

In order to alleviate the above difficulty, part of the bias currents of the input transistors can be provided by PMOS current sources. Illustrated in Fig. 4.33, the idea is to lower the g_m of the load devices by reducing their current rather than their aspect ratio. For example,

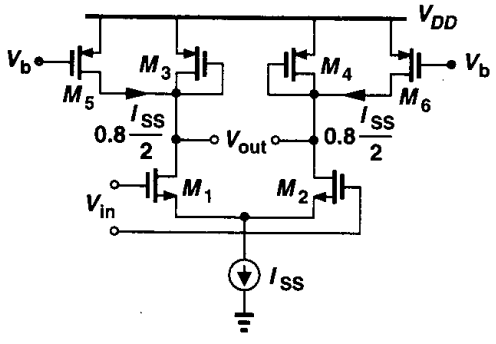


Figure 4.33 Addition of current sources to increase the voltage gain.

if M_5 and M_6 carry 80% of the drain current of M_1 and M_2, the current through M_3 and M_4 is reduced by a factor of five. For a given $|V_{GSP} - V_{THP}|$, this translates to a factor of five reduction in the transconductance of M_3 and M_4 because the aspect ratio of the devices can be lowered by the same factor. Thus, the differential gain is now approximately five times that of the case with no PMOS current sources.

The small-signal gain of the differential pair with current-source loads is relatively low—in the range of 10 to 20 in submicron technologies. How do we increase the voltage gain? Borrowing ideas from the amplifiers in Chapter 3, we increase the output impedance of both PMOS and NMOS devices by cascoding, in essence creating a differential version of the cascode stage introduced in Chapter 3. The result is depicted in Fig. 4.34(a). To calculate the gain, we construct the half circuit of Fig. 4.34(b), which is similar to the cascode stage of Fig. 3.60. Thus,

$$|A_v| \approx g_{m1}[(g_{m3}r_{O3}r_{O1}) \| (g_{m5}r_{O5}r_{O7})]. \tag{4.53}$$

Cascoding therefore increases the differential gain substantially but at the cost of consuming more voltage headroom.

As a final note, we should mention that high-gain fully differential amplifiers require a means of defining the output common-mode level. For example, in Fig. 4.32(b), the output

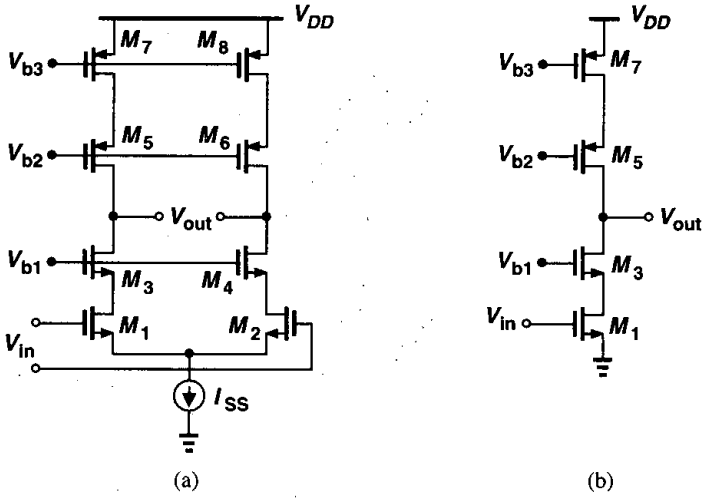


Figure 4.34 (a) Cascode differential pair, (b) half circuit of (a).

common-mode level is not well-defined whereas in Fig. 4.32(a), diode-connected transistors define the output CM level as $V_{DD} - V_{GSP}$. We return to this issue in Chapter 9.

4.5 Gilbert Cell

Our study of differential pairs reveals two important aspects of their operation: (1) the small-signal gain of the circuit is a function of the tail current and (2) the two transistors in a differential pair provide a simple means of steering the tail current to one of two destinations. By combining these two properties, we can develop a versatile building block.

Suppose we wish to construct a differential pair whose gain is varied by a control voltage. This can be accomplished as depicted in Fig. 4.35(a), where the control voltage defines the

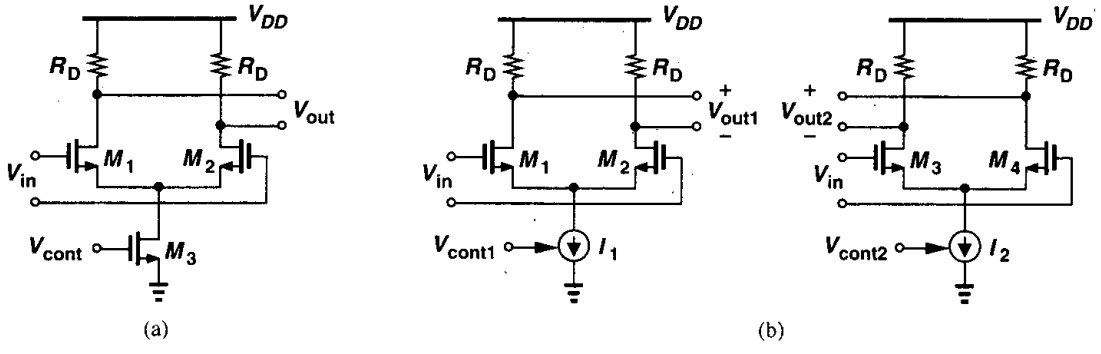


Figure 4.35 (a) Simple VGA, (b) two stages providing variable gain.

tail current and hence the gain. In this topology, $A_v = V_{out}/V_{in}$ varies from zero (if $I_{D3} = 0$) to a maximum value given by voltage headroom limitations and device dimensions. This circuit is a simple example of a "variable-gain amplifier" (VGA). VGAs find application in systems where the signal amplitude may experience large variations and hence requires inverse changes in the gain.

Now suppose we seek an amplifier whose gain can be continuously varied from a negative value to a positive value. Consider two differential pairs that amplify the input by opposite gains [Fig. 4.35(b)]. We now have $V_{out1}/V_{in} = -g_m R_D$ and $V_{out2}/V_{in} = +g_m R_D$, where g_m denotes the transconductance of each transistor in equilibrium. If I_1 and I_2 vary in opposite directions, so do $|V_{out1}/V_{in}|$ and $|V_{out2}/V_{in}|$.

But how should V_{out1} and V_{out2} be combined into a single output? As illustrated in Fig. 4.36(a), the two voltages can be summed, producing $V_{out} = V_{out1} + V_{out2} = A_1 V_{in} + A_2 V_{in}$,

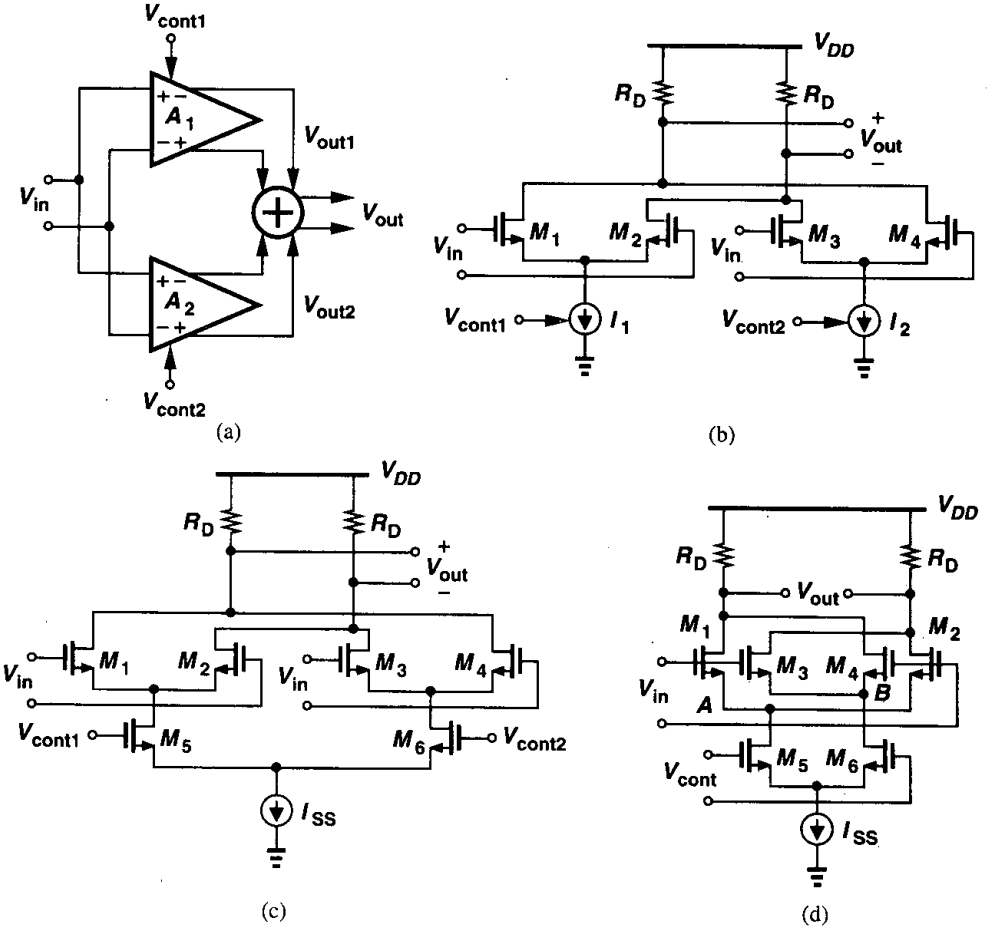


Figure 4.36 (a) Summation of the output voltages of two amplifiers, (b) summation in the current domain, (c) use of M_5-M_6 to control the gain, (d) Gilbert cell.

where A_1 and A_2 are controlled by V_{cont1} and V_{cont2}, respectively. The actual implementation is in fact quite simple: since $V_{out1} = R_D I_{D1} - R_D I_{D2}$ and $V_{out2} = R_D I_{D4} - R_D I_{D3}$, we have $V_{out1} + V_{out2} = R_D(I_{D1} + I_{D4}) - R_D(I_{D2} + I_{D3})$. Thus, rather than add V_{out1} and V_{out2}, we simply short the corresponding drain terminals to sum the currents and subsequently generate the output voltage [Fig. 4.36(b)]. Note that if $I_1 = 0$, then $V_{out} = +g_m R_D V_{in}$ and if $I_2 = 0$, then $V_{out} = -g_m R_D V_{in}$. For $I_1 = I_2$, the gain drops to zero.

In the circuit of Fig. 4.36(b), V_{cont1} and V_{cont2} must vary I_1 and I_2 in opposite directions such that the gain of the amplifier changes monotonically. What circuit can vary two currents in opposite directions? A differential pair provides such a characteristic, leading to the topology of Fig. 4.36(c). Note that for a large $|V_{cont1} - V_{cont2}|$, all of the tail current is steered to one of the top differential pairs and the gain from V_{in} to V_{out} is at its most positive or most negative value. For $V_{cont1} = V_{cont2}$, the gain is zero. For simplicity, we redraw the circuit as shown in Fig. 4.36(d). Called the "Gilbert cell" [2], this circuit is widely used in many analog and communication systems. In a typical design, M_1-M_4 are identical and so are M_5 and M_6.

Example 4.8

Explain why the Gilbert cell can operate as an analog voltage multiplier.

Solution

Since the gain of the circuit is a function of $V_{cont} = V_{cont1} - V_{cont2}$, we have $V_{out} = V_{in} \cdot f(V_{cont})$. Expanding $f(V_{cont})$ in a Taylor series and retaining only the first-order term, αV_{cont}, we have $V_{out} = \alpha V_{in} V_{cont}$. Thus, the circuit can multiply voltages. This property accompanies any voltage-controlled variable-gain amplifier.

As with a cascode structure, the Gilbert cell consumes a greater voltage headroom than a simple differential pair does. This is because the two differential pairs M_1-M_2 and M_3-M_4 are "stacked" on top of the control differential pair. To understand this point, suppose the differential input, V_{in}, in Fig. 4.36(d) has a common-mode level $V_{CM,in}$. Then, $V_A = V_B = V_{CM,in} - V_{GS1}$, where M_1-M_4 are assumed identical. For M_5 and M_6 to operate in saturation, the CM level of V_{cont}, $V_{CM,cont}$, must be such that $V_{CM,cont} \leq V_{CM,in} - V_{GS1} + V_{TH5,6}$. Since $V_{GS1} - V_{TH5,6}$ is roughly equal to one overdrive voltage, we conclude that the control CM level must be lower than the input CM level by at least this value.

In arriving at the Gilbert cell topology, we opted to vary the gain of each differential pair through its tail current, thereby applying the control voltage to the bottom pair and the input signal to the top pairs. Interestingly, the order can be exchanged while still obtaining a VGA. Illustrated in Fig. 4.37(a), the idea is to convert the input voltage to current by means of M_5 and M_6 and route the current through M_1-M_4 to the output nodes. If, as shown in Fig. 4.37(b), V_{cont} is very positive, then only M_1 and M_2 are on and $V_{out} = g_{m5,6} R_D V_{in}$. Similarly, if V_{cont} is very negative [Fig. 4.37(c)], then only M_3 and M_4 are on and $V_{out} = -g_{m5,6} R_D V_{in}$. If the differential control voltage is zero, then $V_{out} = 0$. The input differential pair may incorporate degeneration to provide a linear voltage-to-current conversion.

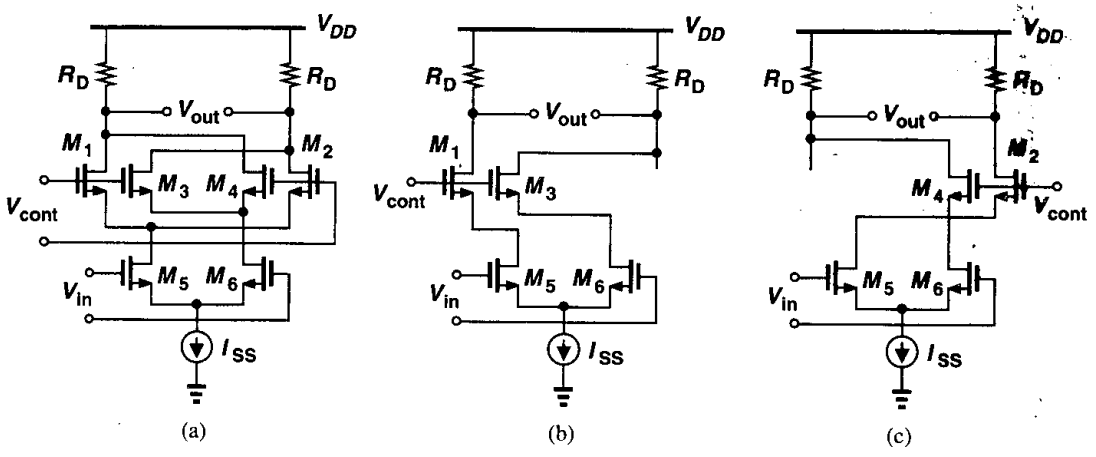


Figure 4.37 (a) Gilbert cell sensing the input voltage by the bottom differential pair, (b) signal path for very positive V_{cont}, (c) signal path for very negative V_{cont}.

Problems

Unless otherwise stated, in the following problems, use the device data shown in Table 2.1 and assume $V_{DD} = 3$ V where necessary. All device dimensions are effective values and in microns.

4.1. Suppose the total capacitance between adjacent lines in Fig. 4.2 is 10 fF and the capacitance from the drains of M_1 and M_2 to ground is 100 fF.
- **(a)** What is the amplitude of the glitches in the analog output in Fig. 4.2(a) for a clock swing of 3 V?
- **(b)** If in Fig. 4.2(b), the capacitance between L_1 and L_2 is 10% less than that between L_1 and L_3, what is the amplitude of the glitches in the differential analog output for a clock swing of 3 V?

4.2. Sketch the small-signal differential voltage gain of the circuit shown in Fig. 4.8(a) if V_{DD} varies from 0 to 3 V. Assume $(W/L)_{1-3} = 50/0.5$, $V_{in,CM} = 1.3$ V, and $V_b = 1$ V.

4.3. Construct the plots of Fig. 4.8(c) for a differential pair using PMOS transistors.

4.4. In the circuit of Fig. 4.10, $(W/L)_{1,2} = 50/0.5$ and $I_{SS} = 0.5$ mA.
- **(a)** What is the maximum allowable output voltage swing if $V_{in,CM} = 1.2$ V?
- **(b)** What is the voltage gain under this condition?

4.5. A differential pair uses input NMOS devices with $W/L = 50/0.5$ and a tail current of 1 mA.
- **(a)** What is the equilibrium overdrive voltage of each transistor?
- **(b)** How is the tail current shared between the two sides if $V_{in1} - V_{in2} = 50$ mA?
- **(c)** What is the equivalent G_m under this condition?
- **(d)** For what value of $V_{in1} - V_{in2}$ does the G_m drop by 10%? By 90%?

4.6. Repeat Problem 4.5 with $W/L = 25/0.5$ and compare the results.

4.7. Repeat Problem 4.5 with a tail current of 2 mA and compare the results.

4.8. Sketch I_{D1} and I_{D2} in Fig. 4.17 versus $V_{in1} - V_{in2}$. For what value of $V_{in1} - V_{in2}$ are the two currents equal?

4.9. Consider the circuit of Fig. 4.28, assuming $(W/L)_{1,2} = 50/0.5$ and $R_D = 2$ kΩ. Suppose R_{SS} represents the output impedance of an NMOS current source with $(W/L)_{SS} = 50/0.5$ and a drain current of 1 mA. The input signal consists of $V_{in,DM} = 10$ mV_{pp} and $V_{in,CM} = 1.5$ V $+ V_n(t)$, where $V_n(t)$ denotes noise with a peak-to-peak amplitude of 100 mV. Assume $\Delta R/R = 0.5\%$.

(a) Calculate the output differential signal-to-noise ratio, defined as the signal amplitude divided by the noise amplitude.

(b) Calculate the CMRR.

4.10. Repeat Problem 4.9 if $\Delta R = 0$ but M_1 and M_2 suffer from a threshold voltage mismatch of 1 mV.

4.11. Suppose the differential pair of Fig. 4.32(a) is designed with $(W/L)_{1,2} = 50/0.5$, $(W/L)_{3,4} = 10/0.5$, and $I_{SS} = 0.5$ mA. Also, I_{SS} is implemented with an NMOS device having $(W/L)_{SS} = 50/0.5$.

(a) What are the minimum and maximum allowable input CM levels if the differential swings at the input and output are small?

(b) For $V_{in,CM} = 1.2$ V, sketch the small-signal differential voltage gain as V_{DD} goes from 0 to 3 V.

4.12. In Problem 4.11, suppose M_1 and M_2 have a threshold voltage mismatch of 1 mV. What is the CMRR?

4.13. In Problem 4.11, suppose $W_3 = 10$ μm but $W_4 = 11$ μm. Calculate the CMRR.

4.14. For the differential pairs of Fig. 4.32(a) and (b), calculate the differential voltage gain if $I_{SS} = 1$ mA, $(W/L)_{1,2} = 50/0.5$, and $(W/L)_{3,4} = 50/1$. What is the minimum allowable input CM level if I_{SS} requires at least 0.4 V across it? Using this value for $V_{in,CM}$, calculate the maximum output voltage swing in each case.

4.15. In the circuit of Fig. 4.33, assume $I_{SS} = 1$ mA and $W/L = 50/0.5$ for all of the transistors.

(a) Determine the voltage gain.

(b) Calculate V_b such that $I_{D5} = I_{D6} = 0.8(I_{SS}/2)$.

(c) If I_{SS} requires a minimum voltage of 0.4 V, what is the maximum differential output swing?

4.16. Assuming all of the circuits shown in Fig. 4.38 are symmetric, sketch V_{out} as **(a)** V_{in1} and V_{in2} vary differentially from zero to V_{DD}, and **(b)** V_{in1} and V_{in2} are equal and they vary from zero to V_{DD}.

4.17. Assuming all of the circuits shown in Fig. 4.39 are symmetric, sketch V_{out} as **(a)** V_{in1} and V_{in2} vary differentially from zero to V_{DD}, and **(b)** V_{in1} and V_{in2} are equal and they vary from zero to V_{DD}.

4.18. Assuming all of the transistors in the circuits of Figs. 4.38 and 4.39 are saturated and $\lambda \neq 0$, calculate the small-signal differential voltage gain of each circuit.

4.19. Consider the circuit shown in Fig. 4.40.

(a) Sketch V_{out} as V_{in1} and V_{in2} vary differentially from zero to V_{DD}.

(b) If $\lambda = 0$, obtain an expression for the voltage gain. What is the voltage gain if $W_{3,4} = 0.8W_{5,6}$?

4.20. For the circuit shown in Fig. 4.41,

(a) Sketch V_{out}, V_X, and V_Y as V_{in1} and V_{in2} vary differentially from zero to V_{DD}.

(b) Calculate the small-signal differential voltage gain.

4.21. Assuming no symmetry in the circuit of Fig. 4.42 and using no equivalent circuits, calculate the small-signal voltage gain $(V_{out})/(V_{in1} - V_{in2})$ if $\lambda = 0$ and $\gamma \neq 0$.

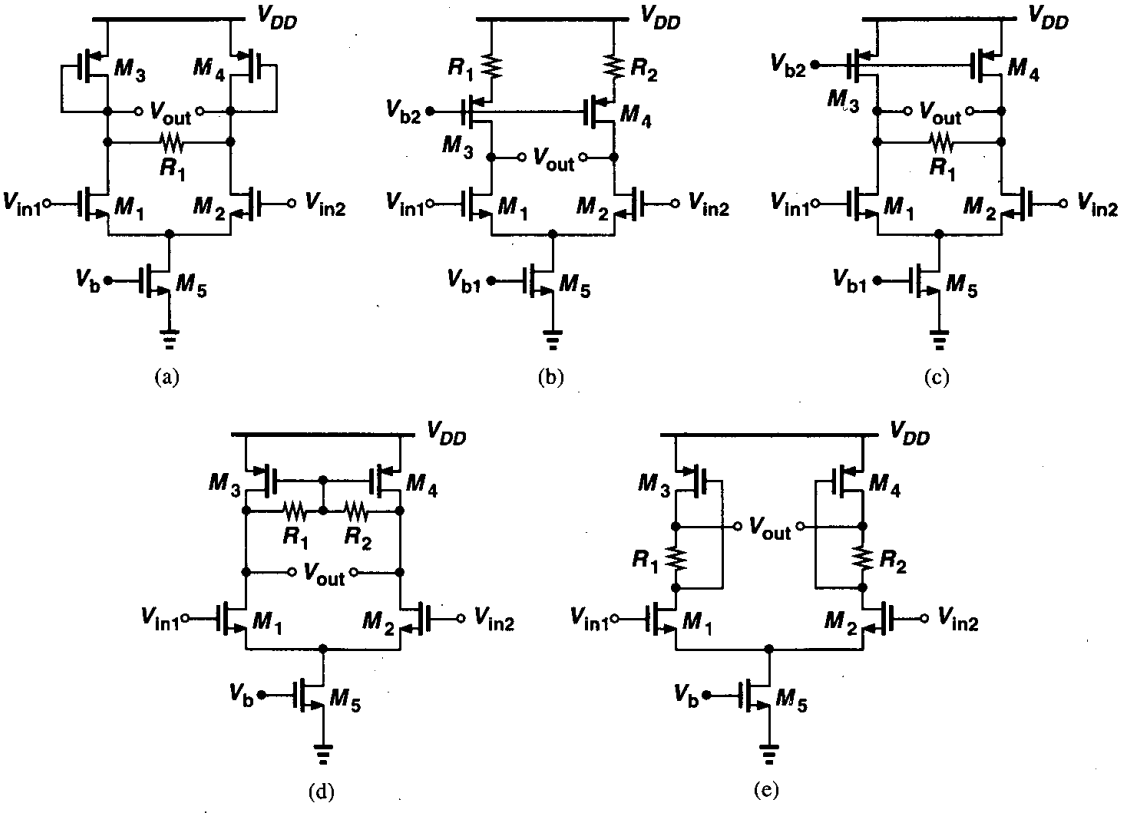


Figure 4.38

4.22. Due to a manufacturing defect, a large parasitic resistance has appeared between the drain and source terminals of M_1 in Fig. 4.43. Assuming $\lambda = \gamma = 0$, calculate the small-signal gain, common-mode gain, and CMRR.

4.23. Due to a manufacturing defect, a large parasitic resistance has appeared between the drains of M_1 and M_4 in the circuit of Fig. 4.44. Assuming $\lambda = \gamma = 0$, calculate the small-signal gain, common-mode gain, and CMRR.

4.24. In the circuit of Fig. 4.45, all of the transistors have a W/L of 50/0.5 and M_3 and M_4 are to operate in deep triode region with an on-resistance of 2 kΩ. Assuming $I_{D5} = 20\ \mu$A and $\lambda = \gamma = 0$, calculate the input common-mode level that yields such resistance. Sketch V_{out1} and V_{out2} as V_{in1} and V_{in2} vary differentially from 0 to V_{DD}.

4.25. In the circuit of Fig. 4.32(b), $(W/L)_{1-4} = 50/0.5$ and $I_{SS} = 1$ mA.
(a) What is the small-signal differential gain?
(b) For $V_{in,CM} = 1.5$ V, what is the maximum allowable output voltage swing?

4.26. In the circuit of Fig. 4.33, assume M_5 and M_6 have a small threshold voltage mismatch of ΔV and I_{SS} has an output impedance R_{SS}. Calculate the CMRR.

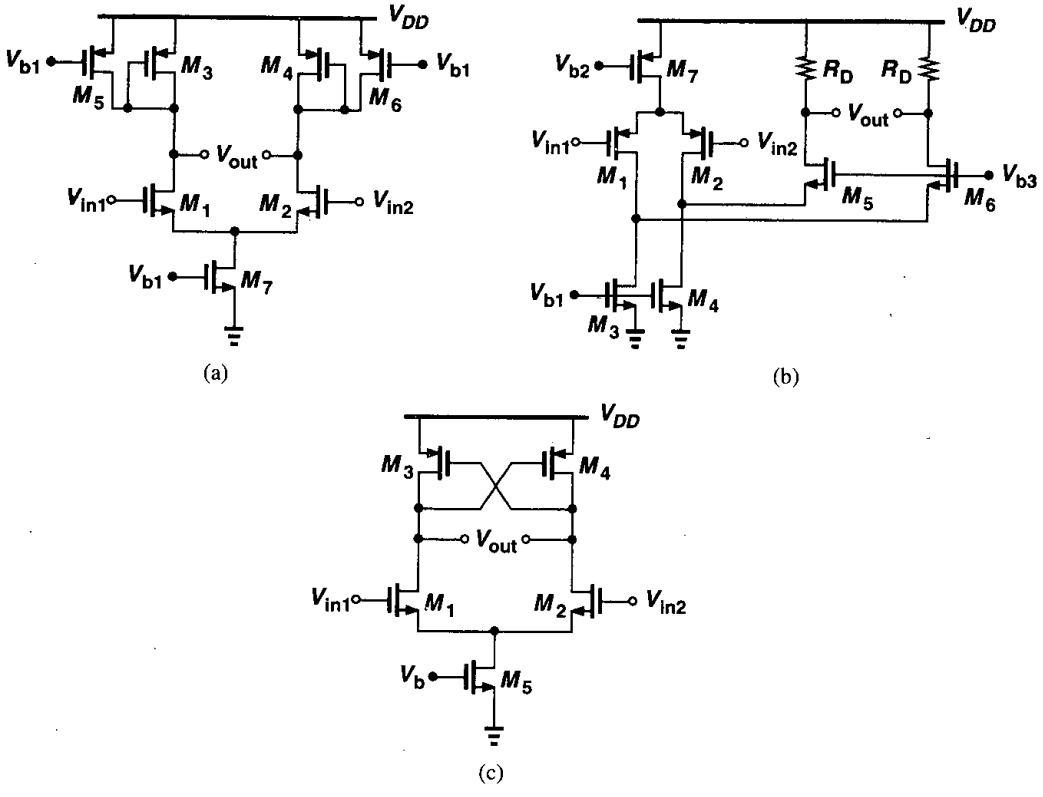


Figure 4.39

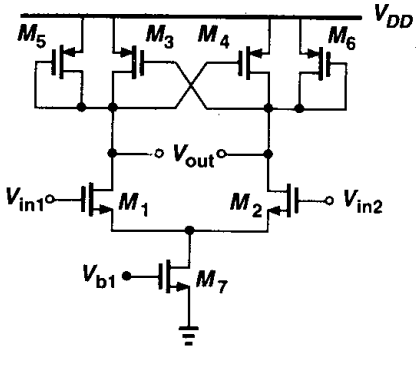


Figure 4.40

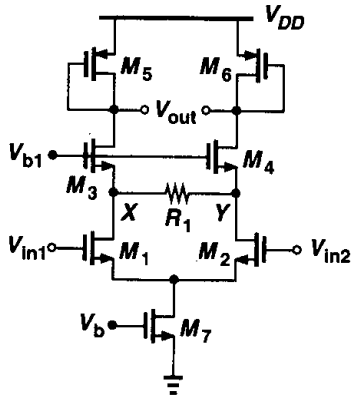


Figure 4.41

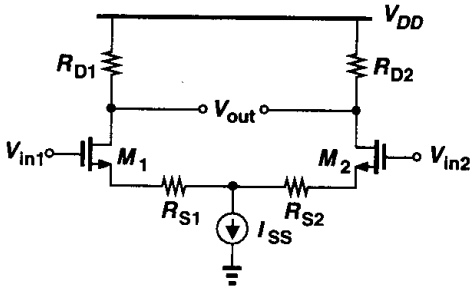


Figure 4.42

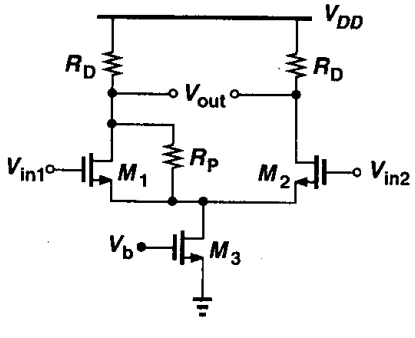


Figure 4.43

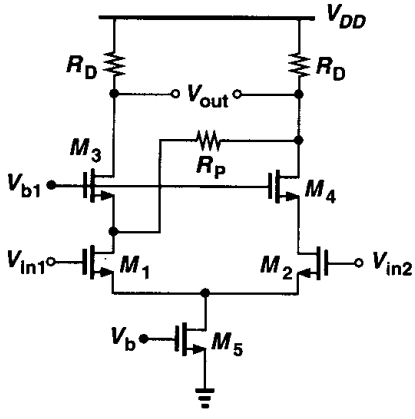


Figure 4.44

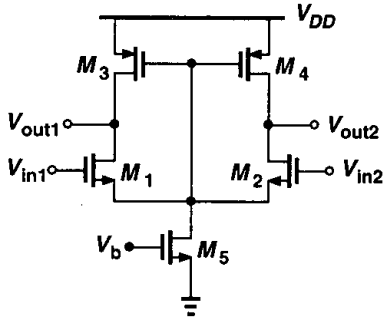


Figure 4.45

References

1. P. R. Gray and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*, Third Ed., New York: Wiley, 1993.
2. B. Gilbert, "A Precise Four-Quadrant Multiplier with Subnanosecond Response," *IEEE J. Solid-State Circuits*, vol. SC-3, pp. 365–373, Dec. 1968.

Chapter 5

Passive and Active Current Mirrors

Our study of single-stage and differential amplifiers in Chapters 3 and 4 points to the wide usage of current sources. In these circuits current sources act as a large resistor without consuming excessive voltage headroom. We also noted that MOS devices operating in saturation can act as a current source.

Current sources find other applications in analog design as well. For example, some digital-to-analog (D/A) converters employ an array of current sources to produce an analog output proportional to the digital input. Also, current sources, in conjunction with "current mirrors," can perform useful functions on analog signals.

This chapter deals with the design of current mirrors as both bias elements and signal processing components. Following a review of basic current mirrors, we study cascode mirror operation. Next, we analyze active current mirrors and describe the properties of differential pairs using such circuits as loads.

5.1 Basic Current Mirrors

Fig. 5.1 illustrates two examples where a current source proves useful. From our study in Chapter 2, recall that the output resistance and capacitance and the voltage headroom of a current source trade with the magnitude of the output current. In addition to these issues, several other aspects of current sources are important: supply, process, and temperature dependence, output noise current, and matching with other current sources. We postpone noise and matching considerations to Chapters 7 and 13, respectively.

How should a MOSFET be biased so as to operate as a stable current source? To gain a better view of the issues, let us consider the simple resistive biasing shown in Fig. 5.2. Assuming M_1 is in saturation, we can write

$$I_{out} \approx \frac{1}{2}\mu_n C_{ox}\frac{W}{L}\left(\frac{R_2}{R_1+R_2}V_{DD} - V_{TH}\right)^2. \tag{5.1}$$

This expression reveals various dependencies of I_{out} upon the supply, process, and temperature. The overdrive voltage is a function of V_{DD} and V_{TH}; the threshold voltage may

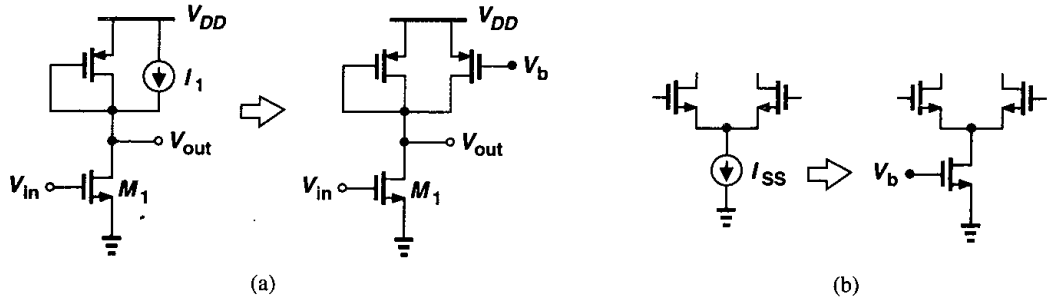


Figure 5.1 Applications of current sources.

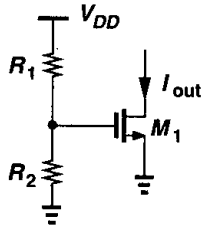


Figure 5.2 Definition of current by resistive divider.

vary by 100 mV from wafer to wafer. Furthermore, both μ_n and V_{TH} exhibit temperature dependence. Thus, I_{out} is poorly defined. The issue becomes more severe as the device is biased with a smaller overdrive voltage, e.g., to consume less headroom. With a nominal overdrive of, say, 200 mV, a 50-mV error in V_{TH} results in a 44% error in the output current.

It is important to note that the above process and temperature dependencies exist even if the gate voltage is not a function of the supply voltage. In other words, if the gate-source *voltage* of a MOSFET is precisely defined, then its drain *current* is not! For this reason, we must seek other methods of biasing MOS current sources.

The design of current sources in analog circuits is based on "copying" currents from a reference, with the assumption that *one* precisely-defined current source is already available. While this method may appear to entail an endless cycle, it is carried out as illustrated in Fig. 5.3. A relatively complex circuit—sometimes requiring external adjustments—is used

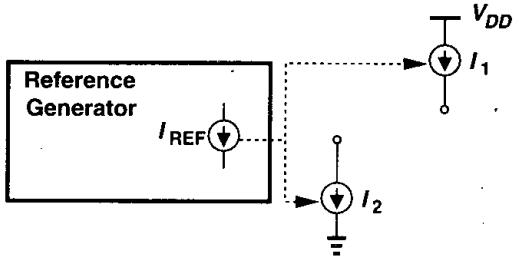


Figure 5.3 Use of a reference to generate various currents.

to generate a stable reference current, I_{REF}, which is then copied to many current sources in the system. We study the copying operation here and the reference generator circuit in Chapter 11.

How do we generate copies of a reference current? For example, in Fig. 5.4, how do we guarantee $I_{out} = I_{REF}$? For a MOSFET, if $I_D = f(V_{GS})$, where $f(\cdot)$ denotes the

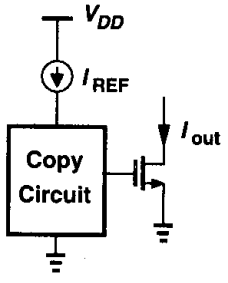


Figure 5.4 Conceptual means of copying currents.

functionality of I_D versus V_{GS}, then $V_{GS} = f^{-1}(I_D)$. That is, if a transistor is biased at I_{REF}, then it produces $V_{GS} = f^{-1}(I_{REF})$ [Fig. 5.5(a)]. Thus, if this voltage is applied to the gate and source terminals of a second MOSFET, the resulting current is $I_{out} = ff^{-1}(I_{REF}) = I_{REF}$ [Fig. 5.5(b)]. From another point of view, two identical MOS devices that have equal gate-source voltages and operate in saturation carry equal currents (if $\lambda = 0$).

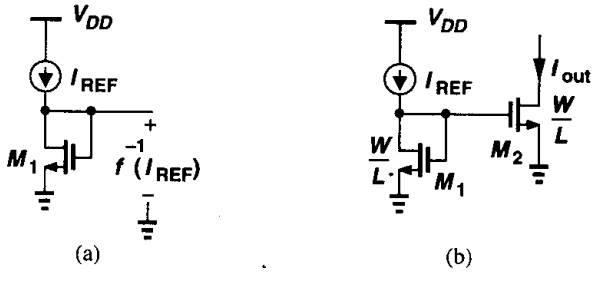


Figure 5.5 (a) Diode-connected device providing inverse function, (b) basic current mirror.

The structure consisting of M_1 and M_2 in Fig. 5.5(b) is called a "current mirror." In the general case, the devices need not be identical. Neglecting channel-length modulation, we can write

$$I_{REF} = \frac{1}{2}\mu_n C_{ox}\left(\frac{W}{L}\right)_1 (V_{GS} - V_{TH})^2 \tag{5.2}$$

$$I_{out} = \frac{1}{2}\mu_n C_{ox}\left(\frac{W}{L}\right)_2 (V_{GS} - V_{TH})^2, \tag{5.3}$$

obtaining

$$I_{out} = \frac{(W/L)_2}{(W/L)_1} I_{REF}. \tag{5.4}$$

The key property of this topology is that it allows precise copying of the current with no dependence on process and temperature. The ratio of I_{out} and I_{REF} is given by the *ratio* of device dimensions, a quantity that can be controlled with reasonable accuracy.

Example 5.1

In Fig. 5.6, find the drain current of M_4 if all of the transistors are in saturation.

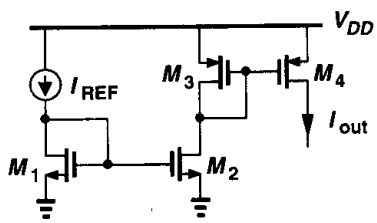


Figure 5.6

Solution

We have $I_{D2} = I_{REF}[(W/L)_2/(W/L)_1]$. Also, $|I_{D3}| = |I_{D2}|$ and $I_{D4} = I_{D3}[(W/L)_4/(W/L)_3]$. Thus, $|I_{D4}| = \alpha\beta I_{REF}$, where $\alpha = (W/L)_2/(W/L)_1$ and $\beta = (W/L)_4/(W/L)_3$. Proper choice of α and β can establish large or small ratios between I_{D4} and I_{REF}. For example, $\alpha = \beta = 5$ yields a magnification factor of 25. Similarly, $\alpha = \beta = 0.2$ can be utilized to generate a small, well-defined current.

Current mirrors find wide application in analog circuits. Fig. 5.7 illustrates a typical case, where a differential pair is biased by means of an NMOS mirror for the tail current source and a PMOS mirror for the load current sources. The device dimensions shown establish a

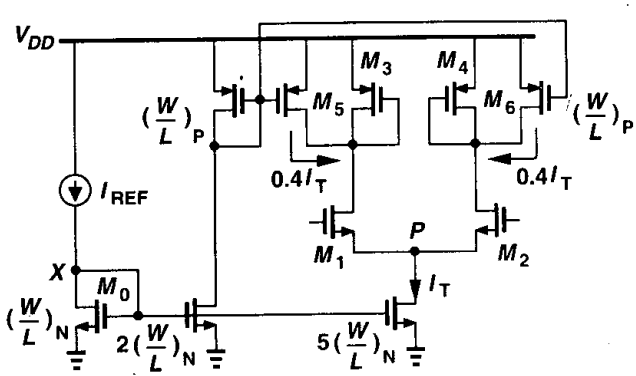


Figure 5.7 Current mirrors used to bias a differential amplifier.

drain current of $0.4I_T$ in M_5 and M_6, reducing the drain current of M_3 and M_4 and hence increasing the gain.

Current mirrors usually employ the same *length* for all of the transistors so as to minimize errors due to the side-diffusion of the source and drain areas (L_D). For example, in Fig. 5.7, the NMOS current sources must have the same channel length as M_0. This is because if, L_{drawn} is, say, doubled, then $L_{eff} = L_{drawn} - 2L_D$ is not. Furthermore, the threshold voltage of short-channel devices exhibits some dependence on the channel length (Chapter 16). Thus, current ratioing is achieved by only scaling the width of transistors.[1]

We should also mention that current mirrors can process *signals* as well. In Fig. 5.5(b), for example, if I_{REF} increases by ΔI, then I_{out} increases by $\Delta I(W/L)_2/(W/L)_1$. That is, the circuit *amplifies* the small-signal current if $(W/L)_2/(W/L)_1 > 1$ (but at the cost of proportional multiplication of the bias current).

Example 5.2

Calculate the small-signal voltage gain of the circuit shown in Fig. 5.8.

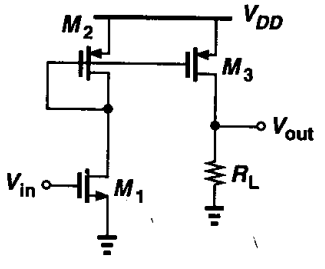


Figure 5.8

Solution

The small-signal drain current of M_1 is equal to $g_{m1}V_{in}$. Since $I_{D2} = I_{D1}$ and $I_{D3} = I_{D2}(W/L)_3/(W/L)_2$, the small-signal drain current of M_3 is equal to $g_{m1}V_{in}(W/L)_3/(W/L)_2$, yielding a voltage gain of $g_{m1}R_L(W/L)_3/(W/L)_2$.

5.2 Cascode Current Mirrors

In our discussion of current mirrors thus far, we have neglected channel length modulation. In practice, this effect results in significant error in copying currents, especially if minimum-length transistors are used so as to minimize the width and hence the output capacitance of the current source. For the simple mirror of Fig. 5.5(b), we can write

$$I_{D1} = \frac{1}{2}\mu_n C_{ox}\left(\frac{W}{L}\right)_1 (V_{GS} - V_{TH})^2(1 + \lambda V_{DS1}) \tag{5.5}$$

[1] As explained in Chapter 18, the widths are actually scaled by placing multiple unit transistors in parallel rather than making a device wider.

$$I_{D2} = \frac{1}{2}\mu_n C_{ox}\left(\frac{W}{L}\right)_2 (V_{GS} - V_{TH})^2(1 + \lambda V_{DS2}), \tag{5.6}$$

and hence

$$\frac{I_{D2}}{I_{D1}} = \frac{(W/L)_2}{(W/L)_1} \cdot \frac{1 + \lambda V_{DS2}}{1 + \lambda V_{DS1}}. \tag{5.7}$$

While $V_{DS1} = V_{GS1} = V_{GS2}$, V_{DS2} may not equal V_{GS2} because of the circuitry fed by M_2. For example, in Fig. 5.7, the potential at node P is determined by the input common-mode level and the gate-source voltage of M_1 and M_2, and it may not equal V_X.

In order to suppress the effect of channel-length modulation, a cascode current source can be used. As shown in Fig. 5.9(a), if V_b is chosen such that $V_Y = V_X$, then I_{out} closely tracks I_{REF}. This is because, as described in conjunction with Fig. 3.61, the cascode device "shields" the bottom transistor from variations in V_P. With the aid of Fig. 3.23, the reader can prove that $\Delta V_Y \approx \Delta V_P/[(g_{m3} + g_{mb3})r_{O3}]$. Thus, we say that V_Y remains close to V_X and hence $I_{D2} \approx I_{D1}$ with high accuracy. Such accuracy is obtained at the cost of the voltage headroom consumed by M_3. Note that, while L_1 must be equal to L_2, the length of M_3 need not be equal to L_1 and L_2.

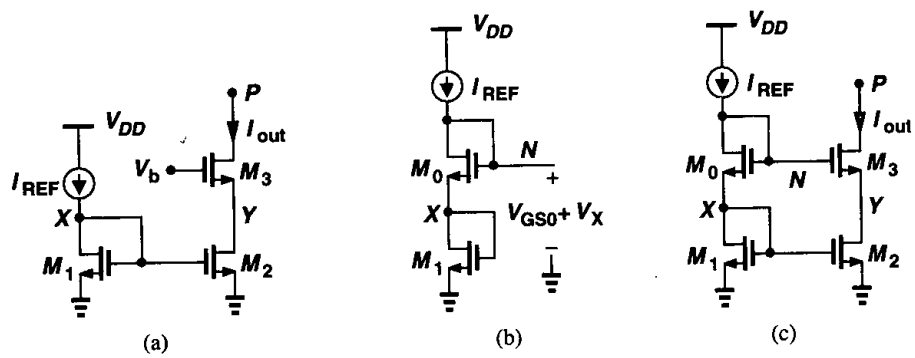


Figure 5.9 (a) Cascode current source, (b) modification of mirror circuit to generate the cascode bias voltage, (c) cascode current mirror.

How do we generate V_b in Fig. 5.9(a)? Since the objective is to ensure $V_Y = V_X$, we must guarantee $V_b - V_{GS3} = V_X$ or $V_b = V_{GS3} + V_X$. This result suggests that if a gate-source voltage is added to V_X, the required value of V_b can be obtained. Depicted in Fig. 5.9(b), the idea is to place another diode-connected device, M_0, in series with M_1, thereby generating a voltage $V_N = V_{GS0} + V_X$. Proper choice of the dimensions of M_0 with respect to those of M_3 yields $V_{GS0} = V_{GS3}$. Connecting node N to the gate of M_3 as shown in Fig. 5.9(c), we have $V_{GS0} + V_X = V_{GS3} + V_Y$. Thus, if $(W/L)_3/(W/L)_0 = (W/L)_2/(W/L)_1$, then $V_{GS3} = V_{GS0}$ and $V_X = V_Y$. Note that this result holds even if M_0 and M_3 suffer from body effect.

Example 5.3

In Fig. 5.10, sketch V_X and V_Y as a function of I_{REF}. If I_{REF} requires 0.5 V to operate as a current source, what is its maximum value?

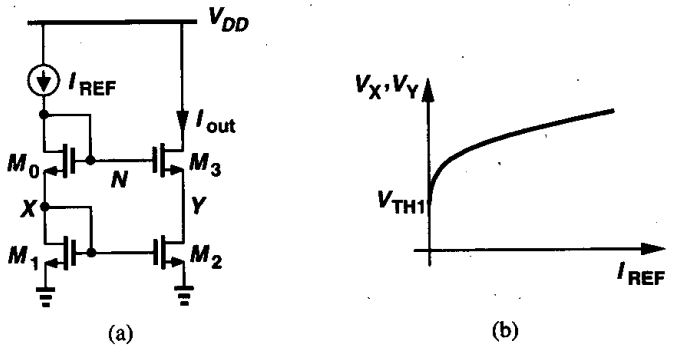


Figure 5.10

Solution

Since M_2 and M_3 are properly ratioed with respect to M_1 and M_0, we have $V_Y = V_X \approx \sqrt{2I_{REF}/[\mu_n C_{ox}(W/L)_1]} + V_{TH1}$. The behavior is plotted in Fig. 5.10(b).

To find the maximum value of I_{REF}, we note that

$$V_N = V_{GS0} + V_{GS1} \tag{5.8}$$

$$= \sqrt{\frac{2I_{REF}}{\mu_n C_{ox}}}\left[\sqrt{\left(\frac{L}{W}\right)_0} + \sqrt{\left(\frac{L}{W}\right)_1}\right] + V_{TH0} + V_{TH1}. \tag{5.9}$$

Thus,

$$V_{DD} - \sqrt{\frac{2I_{REF}}{\mu_n C_{ox}}}\left[\sqrt{\left(\frac{L}{W}\right)_0} + \sqrt{\left(\frac{L}{W}\right)_1}\right] - V_{TH0} - V_{TH1} = 0.5\ \text{V}. \tag{5.10}$$

and hence

$$I_{REF,max} = \frac{\mu_n C_{ox}}{2}\frac{(V_{DD} - 0.5\ \text{V} - V_{TH0} - V_{TH1})^2}{(\sqrt{(L/W)_0} + \sqrt{(L/W)_1})^2}. \tag{5.11}$$

While operating as a current source with high output impedance and accurate value, the topology of Fig. 5.9(c) nonetheless consumes substantial voltage headroom. For simplicity, let us neglect the body effect and assume all of the transistors are identical. Then, the

minimum allowable voltage at node P is equal to

$$V_N - V_{TH} = V_{GS0} + V_{GS1} - V_{TH} \tag{5.12}$$

$$= (V_{GS0} - V_{TH}) + (V_{GS1} - V_{TH}) + V_{TH}, \tag{5.13}$$

i.e., two overdrive voltages plus one threshold voltage. How does this value compare with that in Fig. 5.9(a) if V_b could be chosen more arbitrarily? As shown in Fig. 3.51, V_b could be so low that the minimum allowable voltage at P is merely two overdrive voltages. Thus, the cascode mirror of Fig. 5.9(c) "wastes" one threshold voltage in the headroom. This is because $V_{DS2} = V_{GS2}$, whereas V_{DS2} could be as low as $V_{GS2} - V_{TH}$ while maintaining M_2 in saturation.

Fig. 5.11 summarizes our discussion. In Fig. 5.11(a), V_b is chosen to allow the lowest possible value of V_P but the output current does not accurately track I_{REF} because M_1 and M_2 sustain unequal drain-source voltages. In Fig. 5.11(b), higher accuracy is achieved but the minimum level at P is higher by one threshold voltage.

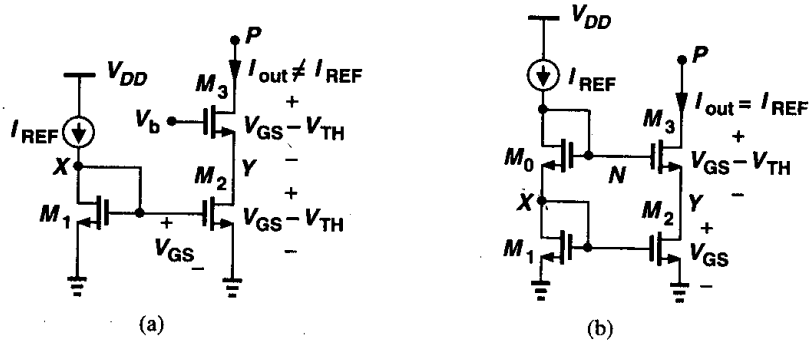


Figure 5.11 (a) Cascode current source with minimum headroom voltage, (b) headroom consumed by a cascode mirror.

Before resolving this issue, it is instructive to examine the large-signal behavior of a cascode current source.

Example 5.4

In Fig. 5.12(a), assuming all of the transistors are identical, sketch I_X and V_B as V_X drops from a large positive value.

Solution

For $V_X \geq V_N - V_{TH}$, both M_2 and M_3 are in saturation, $I_X = I_{REF}$ and $V_B = V_A$. As V_X drops, which transistor enters the triode region first, M_3 or M_2? Suppose M_2 enters the triode region before M_3 does. For this to occur, V_{DS2} must drop and, since V_{GS2} is constant, so must I_{D2}. This means V_{GS3} increases while I_{D3} decreases, which is not possible if M_3 is still in saturation. Thus, M_3 enters the triode region first.

As V_X falls below $V_N - V_{TH}$, M_3 enters the triode region, requiring a greater gate-source overdrive to carry the same current. Thus, as shown in Fig. 5.12(b), V_B begins to drop, causing I_{D2} and hence

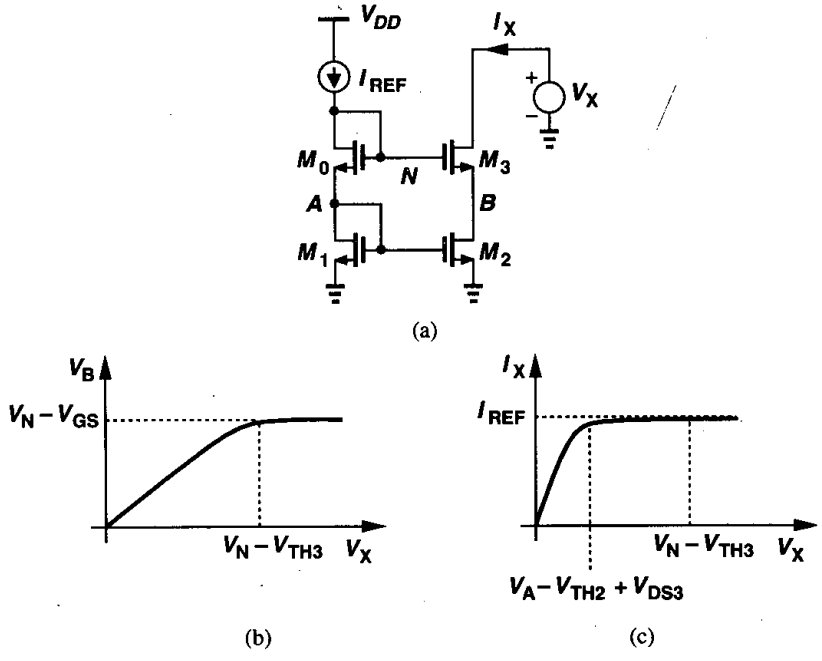


Figure 5.12

I_X to decrease slightly. As V_X and V_B decrease further, eventually we have $V_B < V_A - V_{TH}$, and M_2 enters the triode region. At this point, I_{D2} begins to drop sharply. For $V_X = 0$, $I_X = 0$, and M_2 and M_3 operate in deep triode region. Note that as V_X drops below $V_N - V_{TH3}$, the output impedance of the cascode falls rapidly because g_{m3} degrades in the triode region.

In order to eliminate the accuracy-headroom trade-off described above, we first study the modification depicted in Fig. 5.13(a). Note that this circuit is in fact a cascode topology with its output shorted to its input. How can we choose V_b so that both M_1 and M_2 are in saturation? We must have $V_b - V_{TH2} \leq V_X (= V_{GS1})$ for M_2 to be saturated and $V_{GS1} - V_{TH1} \leq V_A (= V_b - V_{GS2})$ for M_1 to be saturated. Thus,

$$V_{GS2} + (V_{GS1} - V_{TH1}) \leq V_b \leq V_{GS1} + V_{TH2}. \tag{5.14}$$

A solution exists if $V_{GS2} + (V_{GS1} - V_{TH1}) \leq V_{GS1} + V_{TH2}$, i.e., if $V_{GS2} - V_{TH2} \leq V_{TH1}$. We must therefore size M_2 such that its overdrive voltage remains less than one threshold voltage.

Now consider the circuit shown in Fig. 5.13(b), where all of the transistors are in saturation and proper ratioing ensures that $V_{GS2} = V_{GS4}$. If $V_b = V_{GS2} + (V_{GS1} - V_{TH1}) = V_{GS4} + (V_{GS3} - V_{TH3})$, then the cascode current source M_3-M_4 consumes minimum headroom (the overdrive of M_3 plus that of M_4) while M_1 and M_3 sustain equal

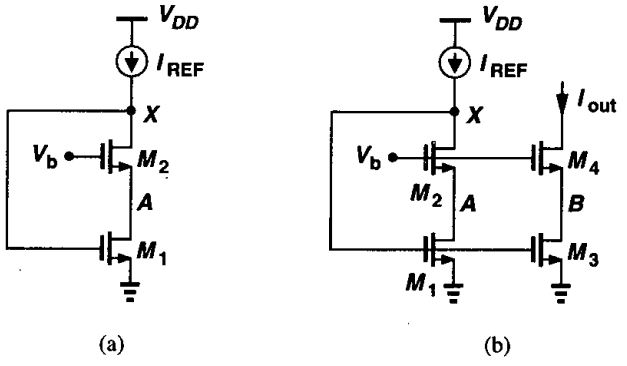


Figure 5.13 Modification of cascode mirror for low-voltage operation.

drain-source voltages, allowing accurate copying of I_{REF}. We call this a "low-voltage cascode."

We must still generate V_b. For minimal voltage headroom consumption, $V_A = V_{GS1} - V_{TH1}$ and hence V_b must be equal to (or slightly greater than) $V_{GS2} + (V_{GS1} - V_{TH1})$. Fig. 5.14(a) depicts an example, where M_5 generates $V_{GS5} \approx V_{GS2}$ and M_6 together with R_b produces $V_{DS6} = V_{GS6} - R_b I_1 \approx V_{GS1} - V_{TH1}$. Some inaccuracy nevertheless arises because M_5 does not suffer from body effect whereas M_2 does. Also, the magnitude of $R_b I_1$ is not well-controlled.

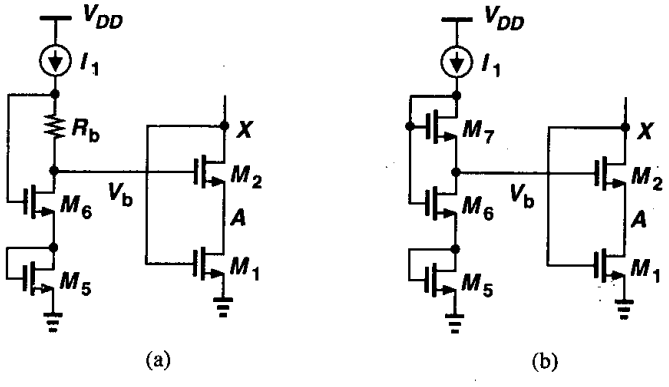


Figure 5.14 Generation of gate voltage V_b for cascode mirrors.

An alternative circuit is shown in Fig. 5.14(b), where the diode-connected transistor M_7 has a large W/L so that $V_{GS7} \approx V_{TH7}$. That is, $V_{DS6} \approx V_{GS6} - V_{TH7}$ and hence $V_b = V_{GS5} + V_{GS6} - V_{TH7}$. While requiring no resistors, this circuit nonetheless suffers from similar errors due to body effect. Some margin is therefore necessary to ensure M_1 and M_2 remain in saturation.

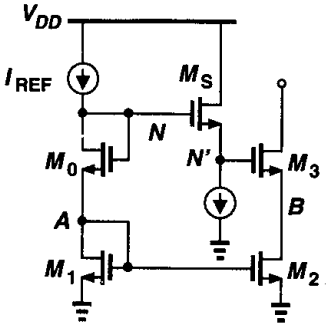


Figure 5.15 Low-voltage cascode using a source follower level shifter.

We should mention that low-voltage cascodes can also be biased using source followers. Shown in Fig. 5.15, the idea is to shift the gate voltage of M_3 down with respect to V_N by interposing a source follower. If M_S is biased at a very low current density, $I_D/(W/L)$, then its gate-source voltage is approximately equal to V_{TH3}, i.e., $V_{N'} \approx V_N - V_{TH3}$, and

$$V_B = V_{GS1} + V_{GS0} - V_{TH3} - V_{GS3} \tag{5.15}$$

$$= V_{GS1} - V_{TH3}, \tag{5.16}$$

implying that M_2 is at the edge of the triode region. In this topology, however, $V_{DS2} \neq V_{DS1}$, introducing substantial mismatch. Also, if the body effect is considered for M_0, M_S, and M_3, it is difficult to guarantee that M_2 operates in saturation. We should mention that, in addition to reducing the systematic mismatch due to channel-length modulation, the cascode structure also provides a high output impedance.

5.3 Active Current Mirrors

As mentioned earlier and exemplified by the circuit of Fig. 5.8, current mirrors can also process signals, i.e., operate as "active" elements. Particularly useful is a type of mirror topology used in conjunction with differential pairs. In this section, we study this circuit and its properties.

First, let us examine the circuit shown in Fig. 5.16, where M_1 and M_2 are identical. Neglecting channel-length modulation, we have $I_{out} = I_{in}$, i.e., with the direction shown for I_{in} and I_{out}, the circuit performs no *inversion*. From the small-signal point of view, if I_{in} increases by ΔI, so does I_{out}.

Now consider the differential amplifier of Fig. 5.17(a), where a current source in a mirror arrangement serves as the load and the output is single-ended. What is the small-signal gain, $A_v = V_{out}/V_{in}$, of this circuit? We calculate A_v using two different approaches,[2] assuming $\gamma = 0$ for simplicity.

[2]Note that, owing to the lack of symmetry, the half-circuit concept cannot be applied here.

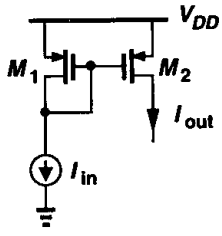
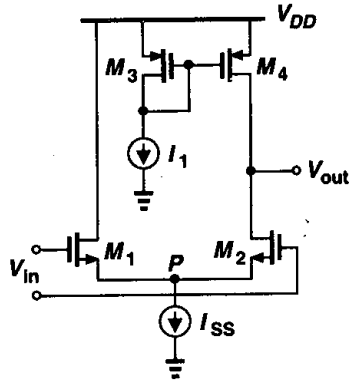
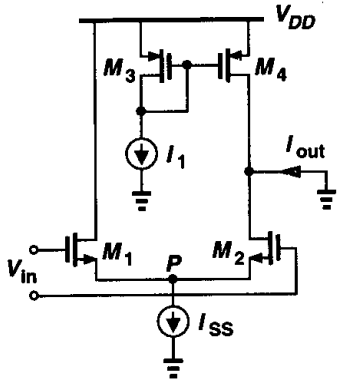


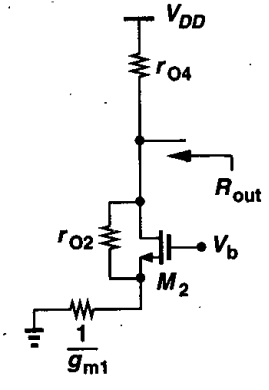
Figure 5.16 Current mirror processing a signal.



(a)



(b)



(c)

Figure 5.17 (a) Differential pair with current-source load, (b) circuit for calculation of G_m, (c) circuit for calculation of R_{out}.

Writing $|A_v| = G_m R_{out}$ and recognizing from Fig. 5.17(b) that $G_m = I_{out}/V_{in} = (g_{m1}V_{in}/2)/V_{in} = g_{m1}/2$, we simply need to compute R_{out}. As illustrated in Fig. 5.17(c), for this calculation, M_2 is degenerated by the source output impedance, $1/g_{m1}$, of M_1, thereby exhibiting an output impedance equal to $(1 + g_{m2}r_{O2})(1/g_{m1,2}) + r_{O2} = 2r_{O2} + 1/g_{m1} \approx$

$2r_{O2}$. Thus, $R_{out} \approx (2r_{O2})\|r_{O4}$, and

$$|A_v| \approx \frac{g_{m1}}{2}[(2r_{O2})\|r_{O4}]. \tag{5.17}$$

Interestingly, if $r_{O4} \to \infty$, then $A_v \to g_{m1}r_{O2}$. This can be explained by the second approach.

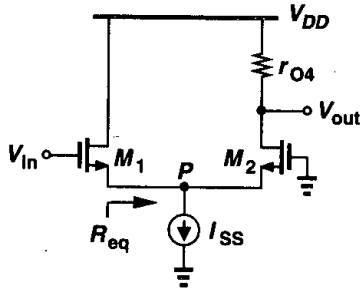


Figure 5.18 Circuit for calculation of V_P/V_{in}.

In our second approach, we calculate V_P/V_{in} and V_{out}/V_P and multiply the results to obtain V_{out}/V_{in}. With the aid of Fig. 5.18,

$$\frac{V_P}{V_{in}} = \frac{R_{eq}}{R_{eq} + \dfrac{1}{g_{m1}}}, \tag{5.18}$$

where R_{eq} denotes the resistance seen looking into the source of M_2. Since the drain of M_2 is terminated by a relatively large resistance, r_{O4}, the value of R_{eq} must be obtained from Eq. (3.110):

$$R_{eq} \approx \frac{1}{g_{m2}} + \frac{r_{O4}}{g_{m2}r_{O2}} \tag{5.19}$$

$$= \frac{1}{g_{m2}}\left(1 + \frac{r_{O4}}{r_{O2}}\right). \tag{5.20}$$

It follows that

$$\frac{V_P}{V_{in}} = \frac{1 + \dfrac{r_{O4}}{r_{O2}}}{2 + \dfrac{r_{O4}}{r_{O2}}}. \tag{5.21}$$

Note that if $r_{O4} \to 0$, $V_P/V_{in} \to 1/2$ and if $r_{O4} \to \infty$, then $V_P/V_{in} \to 1$.

We now calculate V_{out}/V_P while taking r_{O2} into account. From Fig. 5.19,

$$\frac{V_{out}}{V_P} = \frac{1 + g_{m2}r_{O2}}{1 + \dfrac{r_{O2}}{r_{O4}}} \tag{5.22}$$

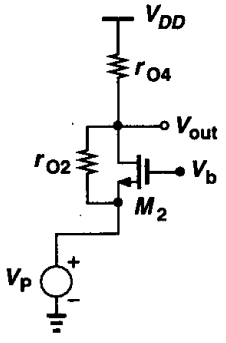


Figure 5.19 Circuit for calculation of V_{out}/V_P.

$$\approx \frac{g_{m2}r_{O2}}{1+\dfrac{r_{O2}}{r_{O4}}}. \tag{5.23}$$

From (5.21) and (5.23), we have

$$\frac{V_{out}}{V_{in}} = \frac{1+\dfrac{r_{O4}}{r_{O2}}}{2+\dfrac{r_{O4}}{r_{O2}}} \cdot \frac{g_{m2}r_{O2}}{1+\dfrac{r_{O2}}{r_{O4}}} \tag{5.24}$$

$$= \frac{g_{m2}r_{O2}r_{O4}}{2r_{O2}+r_{O4}} \tag{5.25}$$

$$= \frac{g_{m2}}{2}[(2r_{O2})\|r_{O4}]. \tag{5.26}$$

In the circuit of Fig. 5.17, the small-signal drain current of M_1 is "wasted." As conceptually shown in Fig. 5.20(a), it is desirable to utilize this current with proper polarity at the output. This can be accomplished as depicted in Fig. 5.20(b), where M_3 and M_4 are identical. To see how M_3 enhances the gain, suppose the gate voltage of M_1 increases by a small amount, increasing I_{D1} by ΔI and decreasing I_{D2} by ΔI. Since $|I_{D3}|$ and hence $|I_{D4}|$ also increase by ΔI, we observe that the output voltage tends to increase through two mechanisms: the drain current of M_2 drops *and* the drain current of M_4 rises.[3] In contrast to the circuit of Fig. 5.17, here M_4 assists M_2 with the voltage change at the output. This configuration is called a differential pair with active current mirror.[4] An important property of this circuit is that it converts a differential input to a single-ended output.

[3]The reader may wonder how this is possible if KCL requires that $I_{D2} = |I_{D4}|$. The explanation in Example 3.2 clarifies this issue.

[4]It is also called a differential pair with active load.

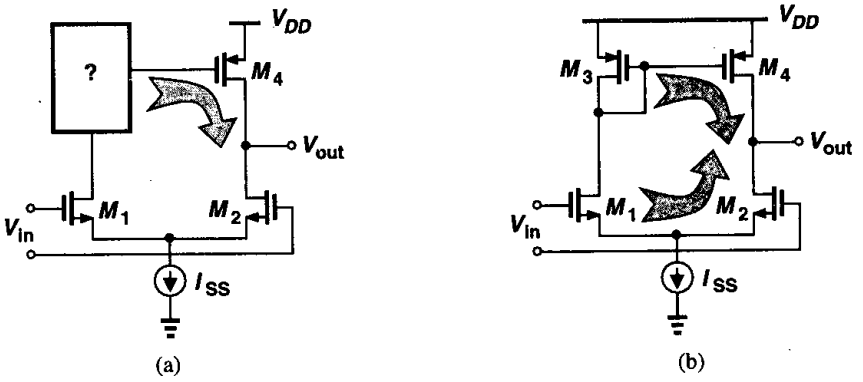


Figure 5.20 (a) Concept of combining the drain currents of M_1 and M_2, (b) realization of (a).

5.3.1 Large-Signal Analysis

Let us study the large-signal behavior of the circuit. To this end, we replace the ideal tail current source by a MOSFET as shown in Fig. 5.21(a). If V_{in1} is much more negative than V_{in2}, M_1 is off and so are M_3 and M_4. Since no current can flow from V_{DD}, both M_2 and M_5 operate in deep triode region, carrying zero current. Thus, $V_{out} = 0$.[5] As V_{in1} approaches V_{in2}, M_1 turns on, drawing part of I_{D5} from M_3 and turning M_4 on. The output voltage then depends on the difference between I_{D4} and I_{D2}. For a small difference between V_{in1} and V_{in2}, both M_2 and M_4 are saturated, providing a high gain [Fig. 5.21(b)]. As V_{in1} becomes more positive than V_{in2}, I_{D1}, $|I_{D3}|$, and $|I_{D4}|$ increase and I_{D2} decreases, eventually driving M_4 into the triode region. If $V_{in1} - V_{in2}$ is sufficiently large, M_2 turns off, M_4 operates in deep triode region with zero current, and $V_{out} = V_{DD}$. Note that if $V_{in1} > V_F + V_{TH}$, then M_1 enters the triode region.

The choice of the input common-mode voltage of the circuit is also important. For M_2 to be saturated, the output voltage cannot be less than $V_{in,CM} - V_{TH}$. Thus, to allow maximum output swings, the input CM level must be as low as possible, with the minimum given by $V_{GS1,2} + V_{DS5,min}$. The direct relationship between the input CM level and the output swing in this circuit is a critical drawback.

What is the output voltage of the circuit when $V_{in1} = V_{in2}$? With perfect symmetry, $V_{out} = V_F = V_{DD} - |V_{GS3}|$. This can be proved by contradiction as well. Suppose, for example, that $V_{out} < V_F$. Then, due to channel-length modulation, M_1 must carry a greater current than M_2 (and M_4 a greater current than M_3). In other words, the total current through M_1 is greater than half of I_{SS}. But this means that the total current through M_3 also exceeds $I_{SS}/2$, violating the assumption that M_4 carries more current than M_3. In reality, however, asymmetries in the circuit may result in a large deviation in V_{out}, possibly driving M_2 or M_4 into the triode region. For example, if the threshold voltage of M_2 is slightly smaller

[5] If V_{in1} is greater than one threshold voltage with respect to ground, M_5 may draw a small current from M_1, raising V_{out} slightly.

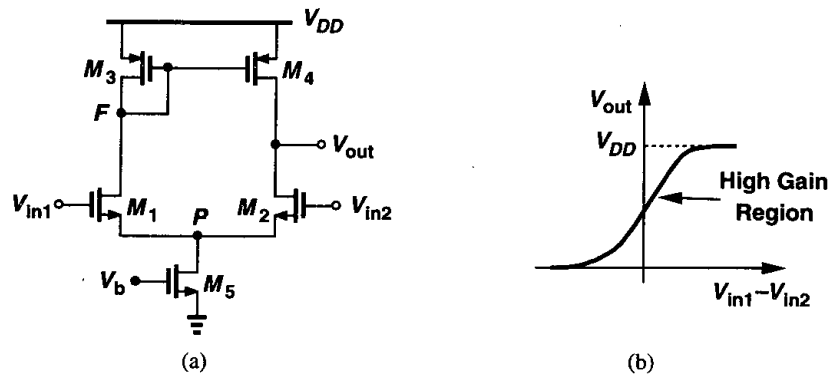


Figure 5.21 (a) Differential pair with active current mirror and realistic current source, (b) large-signal input-output characteristic.

than that of M_1, the former carries a greater current than the latter even with $V_{in1} = V_{in2}$, causing V_{out} to drop significantly. For this reason, the circuit is rarely used in an open-loop configuration to amplify small signals.

Example 5.5

Assuming perfect symmetry, sketch the output voltage of the circuit in Fig. 5.22(a) as V_{DD} varies from 3 V to zero. Assume that for $V_{DD} = 3$ V all of the devices are saturated.

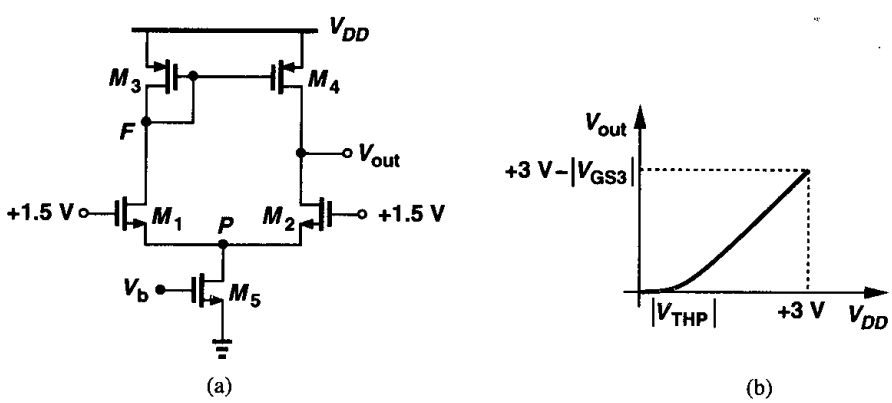


Figure 5.22

Solution

For $V_{DD} = 3$ V, symmetry requires that $V_{out} = V_F$. As V_{DD} drops, so do V_F and V_{out} with a slope close to unity [Fig. 5.22(b)]. As V_F and V_{out} fall below $+1.5\text{ V} - V_{THN}$, M_1 and M_2 enter the triode region, but their drain currents are constant if M_5 is saturated. Further decrease in V_{DD} and hence V_F and V_{out} causes V_{GS1} and V_{GS2} to increase, eventually driving M_5 into the triode region.

Thereafter, the bias current of all of the transistors drops, lowering the rate at which V_{out} decreases. For $V_{DD} < |V_{THP}|$, we have $V_{out} = 0$.

5.3.2 Small-Signal Analysis

We now analyze the small-signal properties of the circuit shown in Fig. 5.21(a), assuming $\gamma = 0$ for simplicity. Can we apply the half-circuit concept to calculate the differential gain here? As illustrated in Fig. 5.23, with small differential inputs, the voltage swings at nodes

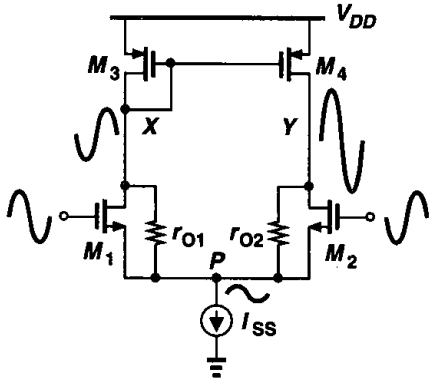


Figure 5.23 Asymmetric swings in a differential pair with active current mirror.

X and Y are vastly different. This is because the diode-connected device M_3 yields a much lower voltage gain from the input to node X than that from the input to node Y. As a result, the effects of V_X and V_Y at node P (through r_{O1} and r_{O2}, respectively) do not cancel each other and this node cannot necessarily be considered a virtual ground. We compute the gain using two different approaches.

In the first approach, we write $|A_v| = G_m R_{out}$ and obtain G_m and R_{out} separately. For the calculation of G_m, consider Fig. 5.24(a). The circuit is not quite symmetric but

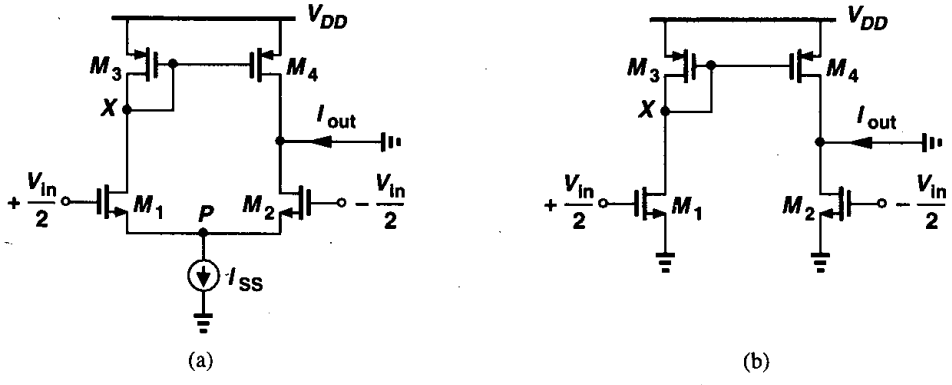


Figure 5.24 (a) Circuit for calculation of G_m, (b) circuit of (a) with node P grounded.

because the impedance seen at node X is relatively low and the swing at this node small, the current returning from X to P through r_{O1} is negligible and node P can be viewed as a virtual ground [Fig. 5.24(b)]. Thus, $I_{D1} = |I_{D3}| = |I_{D4}| = g_{m1,2}V_{in}/2$ and $I_{D2} = -g_{m1,2}V_{in}/2$, yielding $I_{out} = -g_{m1,2}V_{in}$ and hence $|G_m| = g_{m1,2}$. Note that, by virtue of active current mirror operation, this value is twice the transconductance of the circuit of Fig. 5.17(b).

Calculation of R_{out} is less straightforward. We may surmise that the output resistance of this circuit is equal to that of the circuit in Fig. 5.17(c), namely, $(2r_{O2})\|r_{O4}$. In reality, however, the active mirror operation yields a different value because when a voltage is applied to the output to measure R_{out}, the gate voltage of M_4 does not remain constant. Rather than draw the entire equivalent circuit, we observe that, for small signals, I_{SS} is open [Fig. 5.25(a)], any current flowing into M_1 must flow out of M_2, and the role of the two transistors can be

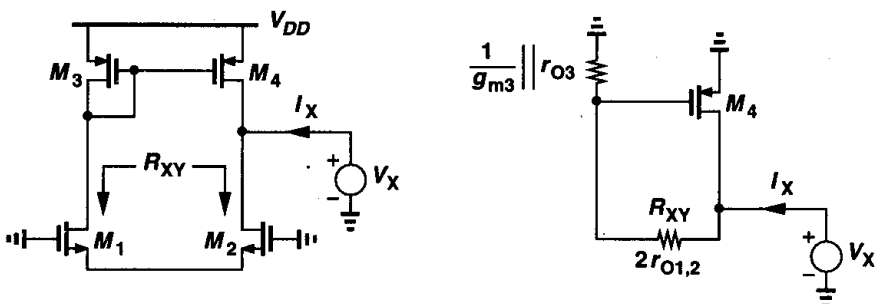


Figure 5.25 (a) Circuit for calculating R_{out}, (b) substitution of M_1 and M_2 by a resistor.

represented by a resistor $R_{XY} = 2r_{O1,2}$ [Fig. 5.25(b)]. As a consequence, the current drawn from V_X by R_{XY} is mirrored by M_3 into M_4 with unity gain. We can therefore write:

$$I_X = 2\frac{V_X}{2r_{O1,2} + \dfrac{1}{g_{m3}}\Big\|r_{O3}} + \frac{V_X}{r_{O4}}, \tag{5.27}$$

where the factor 2 accounts for current copying action of M_3 and M_4. For $2r_{O1,2} \gg (1/g_{m3})\|r_{O3}$, we have

$$R_{out} \approx r_{O2}\|r_{O4}. \tag{5.28}$$

The overall voltage gain is thus equal to $|A_v| = G_m R_{out} = g_{m1,2}(r_{O2}\|r_{O4})$, somewhat higher than that of the circuit in Fig. 5.17(a).

The second approach to calculating the voltage gain of the circuit is illustrated in Fig. 5.26, providing more insight into the operation. We substitute the input source and M_1 and M_2 by a Thevenin equivalent. As illustrated in Fig. 5.27(a), for the Thevenin voltage calculation, node P is a virtual ground because of symmetry, and a half-circuit equivalent yields $V_{eq} = g_{m1,2}r_{O1,2}V_{in}$. Moreover, the output resistance is $R_{eq} = 2r_{O1,2}$. From Fig. 5.27(b),

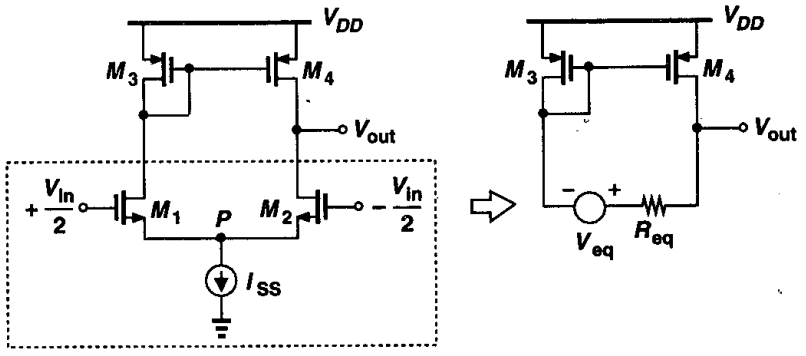
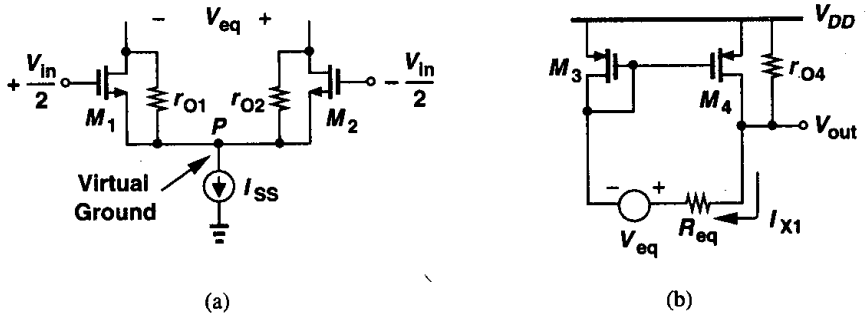


Figure 5.26 Substitution of the input differential pair by a Thevenin equivalent.



(a) (b)

Figure 5.27 (a) Calculation of the Thevenin equivalent voltage, (b) simplified circuit.

we note that the current through R_{eq} is

$$I_{X1} = \frac{V_{out} - g_{m1,2} r_{O1,2} V_{in}}{2r_{O1,2} + \dfrac{1}{g_{m3}} \| r_{O3}}. \tag{5.29}$$

The fraction of this current that flows through $1/g_{m3}$ is mirrored into M_4 with unity gain. That is,

$$2\frac{V_{out} - g_{m1,2} r_{O1,2} V_{in}}{2r_{O1,2} + \dfrac{1}{g_{m3}} \| r_{O3}} \cdot \frac{r_{O3}}{r_{O3} + 1/g_{m3}} = -\frac{V_{out}}{r_{O4}}. \tag{5.30}$$

Assuming $2r_{O1,2} \gg (1/g_{m3,4}) \| r_{O3,4}$, we obtain

$$\frac{V_{out}}{V_{in}} = \frac{g_{m1,2} r_{O3,4} r_{O1,2}}{r_{O1,2} + r_{O3,4}} \tag{5.31}$$

$$= g_{m1,2}(r_{O1,2} \| r_{O3,4}). \tag{5.32}$$

Example 5.6

Calculate the small-signal voltage gain of the circuit shown in Fig. 5.28. How does the performance of this circuit compare with that of a differential pair with active mirror?

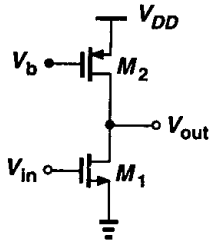


Figure 5.28

Solution

We have $A_v = g_{m1}(r_{O1} \| r_{O2})$, similar to the value derived above. For given device dimensions, this circuit requires half of the bias current to achieve the same gain as a differential pair. However, advantages of differential operation often outweigh the power penalty.

The above calculations of the gain have assumed an ideal tail current source. In reality, the output impedance of this source affects the gain, but the error with respect to $g_{m1,2}(r_{O1,2} \| r_{O3,4})$ is relatively small.

5.3.3 Common-Mode Properties

Let us now study the common-mode properties of the differential pair with active current mirror. We assume $\gamma = 0$ for simplicity and leave a more general analysis including body effect for the reader. Our objective is to predict the consequences of a finite output impedance in the tail current source. As depicted in Fig. 5.29, a change in the input CM level leads to

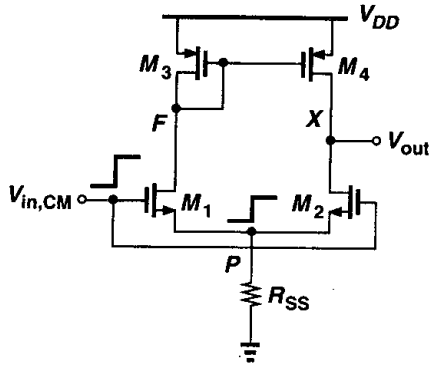


Figure 5.29 Differential pair with active current mirror sensing a common-mode change.

a change in the bias current of all of the transistors. How do we define the common-mode gain here? Recall from Chapter 4 that the CM gain represents the *corruption* of the output

signal of interest due to variations of the input CM level. In the circuits of Chapter 3, the output signal was sensed differentially and hence the CM gain was defined in terms of the output differential component generated by the input CM change. In the circuit of Fig. 5.29, on the other hand, the output signal of interest is sensed with respect to ground. Thus, we define the CM gain in terms of the single-ended output component produced by the input CM change:

$$A_{CM} = \frac{\Delta V_{out}}{\Delta V_{in,CM}}. \tag{5.33}$$

To determine A_{CM}, we observe that if the circuit is symmetric, $V_{out} = V_F$ for any input CM level. For example, as $V_{in,CM}$ increases, V_F drops and so does V_{out}. In other words, nodes F and X can be shorted [Fig. 5.30(a)], resulting in the equivalent circuit shown

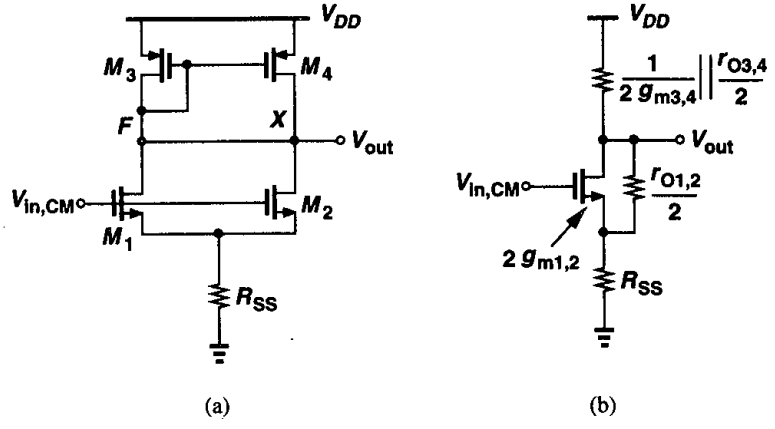


Figure 5.30 (a) Simplified circuit of Fig. 5.29, (b) equivalent circuit of (a).

in Fig. 5.30(b). Here, M_1 and M_2 appear in parallel and so do M_3 and M_4. It follows that

$$A_{CM} \approx \frac{-\dfrac{1}{2g_{m3,4}} \Big\| \dfrac{r_{O3,4}}{2}}{\dfrac{1}{2g_{m1,2}} + R_{SS}} \tag{5.34}$$

$$= \frac{-1}{1 + 2g_{m1,2}R_{SS}} \frac{g_{m1,2}}{g_{m3,4}}, \tag{5.35}$$

where we have assumed $1/(2g_{m3,4}) \ll r_{O3,4}$ and neglected the effect of $r_{O1,2}/2$. The CMRR is then given by

$$CMRR = \left|\frac{A_{DM}}{A_{CM}}\right| \tag{5.36}$$

$$= g_{m1,2}(r_{O1,2} \| r_{O3,4}) \frac{g_{m3,4}(1 + 2g_{m1,2}R_{SS})}{g_{m1,2}} \tag{5.37}$$

$$= (1 + 2g_{m1,2}R_{SS}) g_{m3,4}(r_{O1,2} \| r_{O3,4}). \tag{5.38}$$

Equation (5.35) indicates that, even with perfect symmetry, the output signal is corrupted by input CM variations, a drawback that does not exist in the fully differential circuits of Chapter 3. High-frequency common-mode noise therefore degrades the performance considerably as the capacitance shunting the tail current source exhibits a lower impedance.

Example 5.7

The CM gain of the circuit of Fig. 5.29 can be shown to be *zero* by a (flawed) argument. As shown in Fig. 5.31(a), if $V_{in,CM}$ introduces a change of ΔI in the drain current of each input transistor, then

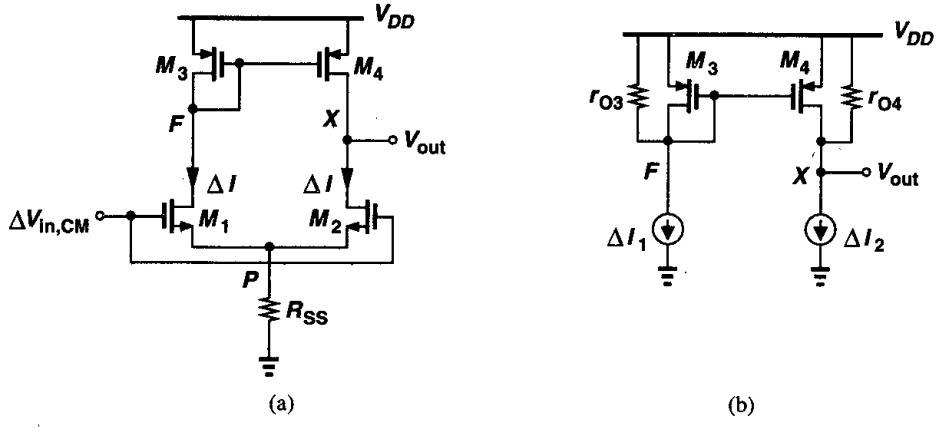


Figure 5.31

I_{D3} also experiences the same change and so does I_{D4}. Thus, M_4 seemingly provides the additional current required by M_2, and the output voltage need not change, i.e., $A_{CM} = 0$. Explain the flaw in this proof.

Solution

The assumption that ΔI_{D4} completely cancels the effect of ΔI_{D2} is incorrect. Consider the equivalent circuit shown in Fig. 5.31(b). Since

$$\Delta V_F = \Delta I_1 \left(\frac{1}{g_{m3}} \| r_{O3}\right), \tag{5.39}$$

we have

$$|\Delta I_{D4}| = g_{m4}\Delta V_F \tag{5.40}$$

$$= g_{m4}\Delta I_1 \frac{r_{O3}}{1 + g_{m3}r_{O3}}. \tag{5.41}$$

This current and ΔI_2 ($= \Delta I_1 = \Delta I$) give a net voltage change equal to

$$\Delta V_{out} = (\Delta I_1 g_{m4} \frac{r_{O3}}{1 + g_{m3}r_{O3}} - \Delta I_2) r_{O4} \tag{5.42}$$

$$= -\Delta I \frac{1}{g_{m3}r_{O3} + 1} r_{O4}, \tag{5.43}$$

which is equal to the voltage change at node F.

It is also instructive to calculate the common-mode gain in the presence of mismatches. As an example, we consider the case where the input transistors exhibit slightly different transconductances [Fig. 5.32(a)]. How does V_{out} depend on $V_{in,CM}$? Since the change at

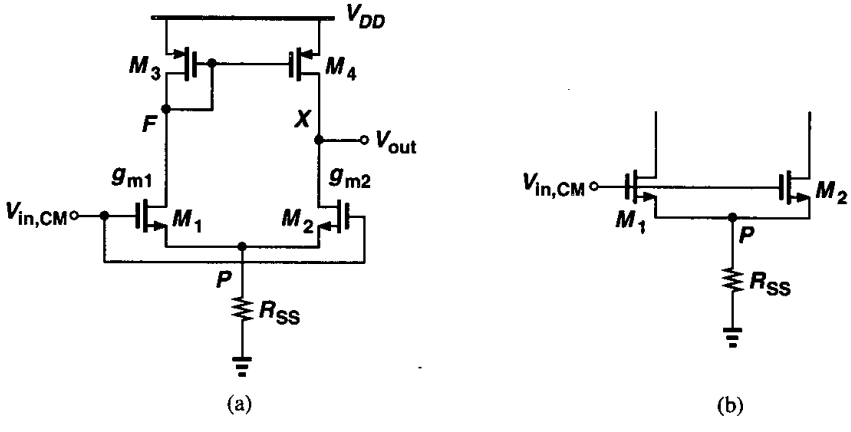


Figure 5.32 Differential pair with g_m mismatch.

nodes F and X is relatively small, we can compute the change in I_{D1} and I_{D2} while neglecting the effect of r_{O1} and r_{O2}. As shown in Fig. 5.32(b), the voltage change at P can be obtained by considering M_1 and M_2 as a single transistor (in a source follower configuration) with a transconductance equal to $g_{m1} + g_{m2}$, i.e.,

$$\Delta V_P = \Delta V_{in,CM} \frac{R_{SS}}{R_{SS} + \dfrac{1}{g_{m1} + g_{m2}}}, \tag{5.44}$$

where body effect is neglected. The changes in the drain currents of M_1 and M_2 are therefore given by

$$\Delta I_{D1} = g_{m1}(\Delta V_{in,CM} - \Delta V_P) \tag{5.45}$$

$$= \frac{\Delta V_{in,CM}}{R_{SS} + \dfrac{1}{g_{m1} + g_{m2}}} \frac{g_{m1}}{g_{m1} + g_{m2}} \tag{5.46}$$

$$\Delta I_{D2} = g_{m2}(\Delta V_{in,CM} - \Delta V_P) \tag{5.47}$$

$$= \frac{\Delta V_{in,CM}}{R_{SS} + \dfrac{1}{g_{m1} + g_{m2}}} \frac{g_{m2}}{g_{m1} + g_{m2}}. \tag{5.48}$$

The change ΔI_{D1} multiplied by $(1/g_{m3})\|r_{O3}$ yields $|\Delta I_{D4}| = g_{m4}[(1/g_{m3})\|r_{O3}]\Delta I_{D1}$. The difference between this current and ΔI_{D2} flows through the output impedance of the circuit, which is equal to r_{O4} because we have neglected the effect of r_{O1} and r_{O2}:

$$\Delta V_{out} = \left[\frac{g_{m1}\Delta V_{in,CM}}{1 + (g_{m1} + g_{m2})R_{SS}} \frac{r_{O3}}{r_{O3} + \dfrac{1}{g_{m3}}} - \frac{g_{m2}\Delta V_{in,CM}}{1 + (g_{m1} + g_{m2})R_{SS}} \right] r_{O4} \tag{5.49}$$

$$= \frac{\Delta V_{in,CM}}{1 + (g_{m1} + g_{m2})R_{SS}} \frac{(g_{m1} - g_{m2})r_{O3} - g_{m2}/g_{m3}}{r_{O3} + \dfrac{1}{g_{m3}}} r_{O4}. \tag{5.50}$$

If $r_{O3} \gg 1/g_{m3}$, we have

$$\frac{\Delta V_{out}}{\Delta V_{in,CM}} \approx \frac{(g_{m1} - g_{m2})r_{O3} - g_{m2}/g_{m3}}{1 + (g_{m1} + g_{m2})R_{SS}}. \tag{5.51}$$

Compared to Eq. (5.35), this result contains the additional term $(g_{m1} - g_{m2})r_{O3}$ in the numerator, revealing the effect of transconductance mismatch on the common-mode gain.

Problems

Unless otherwise stated, in the following problems, use the device data shown in Table 2.1 and assume $V_{DD} = 3$ V where necessary. All device dimensions are effective values and in microns.

5.1. In Fig. 5.2, assume $(W/L)_1 = 50/0.5$, $\lambda = 0$, $I_{out} = 0.5$ mA, and M_1 is saturated.

(a) Determine R_2/R_1.

(b) Calculate the sensitivity of I_{out} to V_{DD}, defined as $\partial I_{out}/\partial V_{DD}$ and normalized to I_{out}.

(c) How much does I_{out} change if V_{TH} changes by 50 mV?

(d) If the temperature dependence of μ_n is expressed as $\mu_n \propto T^{-3/2}$ but V_{TH} is independent of temperature, how much does I_{out} vary if T changes from 300°K to 370°K?

(e) What is the worst-case change in I_{out} if V_{DD} changes by 10%, V_{TH} by 50 mV, and T from 300°K to 370°K?

5.2. Consider the circuit of Fig. 5.6. Assuming I_{REF} is *ideal*, sketch I_{out} versus V_{DD} as V_{DD} varies from 0 to 3 V.

5.3. In the circuit of Fig. 5.7, $(W/L)_N = 10/0.5$, $(W/L)_P = 10/0.5$, and $I_{REF} = 100\ \mu$A. The input CM level applied to the gates of M_1 and M_2 is equal to 1.3 V.

(a) Assuming $\lambda = 0$, calculate V_P and the drain voltage of the PMOS diode-connected transistors.

(b) Now take channel-length modulation into account to determine I_T and the drain current of the PMOS diode-connected transistors more accurately.

5.4. Consider the circuit of Fig. 5.8; sketch V_{out} versus V_{DD} as V_{DD} varies from 0 to 3 V.

5.5. Consider the circuit of Fig. 5.9(a), assuming $(W/L)_{1-3} = 40/0.5$, $I_{REF} = 0.3$ mA, and $\gamma = 0$.

(a) Determine V_b such that $V_X = V_Y$.

(b) If V_b deviates from the value calculated in part (a) by 100 mV, what is the mismatch between I_{out} and I_{REF}?

(c) If the circuit fed by the cascode current source changes V_P by 1 V, how much does V_Y change?

5.6. The circuit of Fig. 5.13 is designed with $(W/L)_{1,2} = 20/0.5$, $(W/L)_{3,4} = 60/0.5$, and $I_{REF} = 100\ \mu$A.

(a) Determine V_X and the acceptable range of V_b.

(b) Estimate the deviation of I_{out} from 300 μA if the drain voltage of M_4 is higher than V_X by 1 V.

5.7. The circuit of Fig. 5.17(a) is designed with $(W/L)_{1-4} = 50/0.5$ and $I_{SS} = 2I_1 = 0.5$ mA.

(a) Calculate the small-signal voltage gain.

(b) Determine the maximum output voltage swing if the input CM level is 1.3 V.

5.8. Consider the circuit of Fig. 5.22(a) with $(W/L)_{1-5} = 50/0.5$ and $I_{D5} = 0.5$ mA.

(a) Calculate the deviation of V_{out} from V_F if $|V_{TH3}|$ is 1 mV less than $|V_{TH4}|$.

(b) Determine the CMRR of the amplifier.

5.9. Sketch V_X and V_Y as a function of V_{DD} for each circuit in Fig. 5.33. Assume the transistors in each circuit are identical.

5.10. Sketch V_X and V_Y as a function of V_{DD} for each circuit in Fig. 5.34. Assume the transistors in each circuit are identical.

5.11. For each circuit in Fig. 5.35, sketch V_X and V_Y as a function of V_1 for $0 < V_1 < V_{DD}$. Assume the transistors in each circuit are identical.

5.12. For each circuit in Fig. 5.36, sketch V_X and V_Y as a function of V_1 for $0 < V_1 < V_{DD}$. Assume the transistors in each circuit are identical.

5.13. For each circuit in Fig. 5.37, sketch V_X and V_Y as a function of I_{REF}.

5.14. For the circuit of Fig. 5.38, sketch I_{out}, V_X, V_A, and V_B as a function of **(a)** I_{REF}, **(b)** V_b.

5.15. In the circuit shown in Fig. 5.39, a source follower using a wide transistor and a small bias current is inserted in series with the gate of M_3 so as to bias M_2 at the edge of saturation. Assuming M_0-M_3 are identical and $\lambda \neq 0$, estimate the mismatch between I_{out} and I_{REF} if **(a)** $\gamma = 0$, **(b)** $\gamma \neq 0$.

5.16. Sketch V_X and V_Y as a function of time for each circuit in Fig. 5.40. Assume the transistors in each circuit are identical.

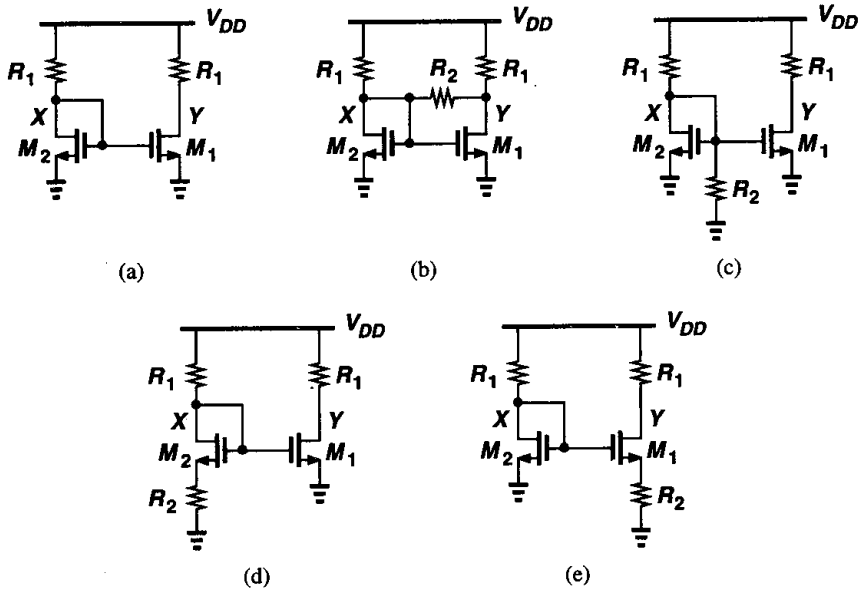


Figure 5.33

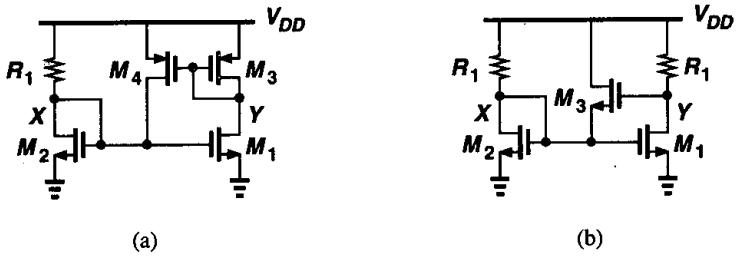


Figure 5.34

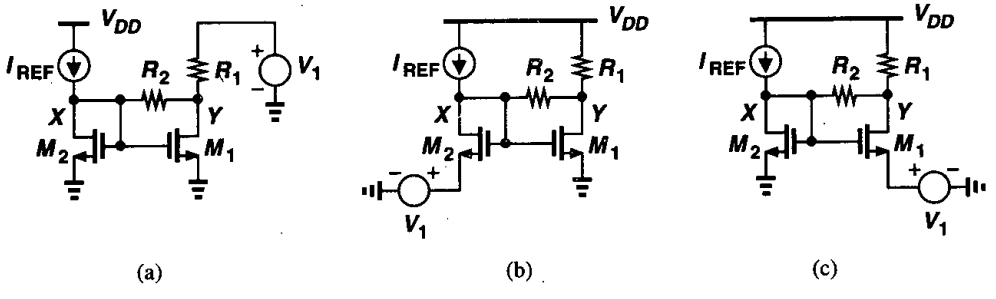


Figure 5.35

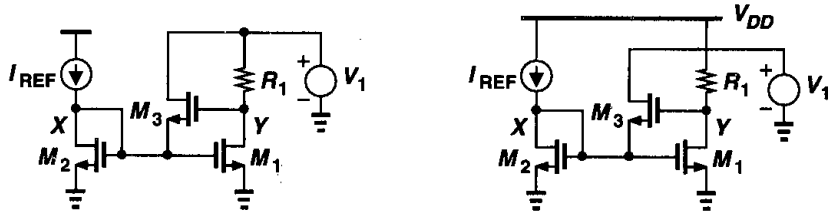


Figure 5.36

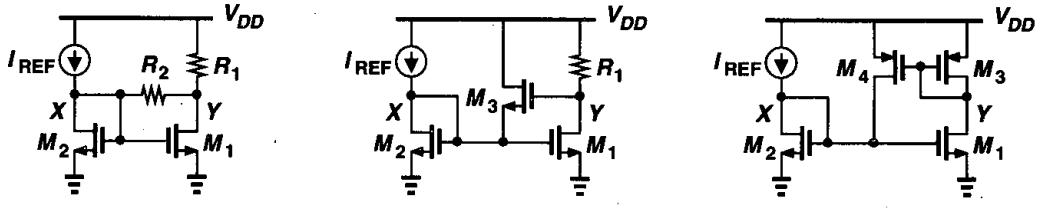


Figure 5.37

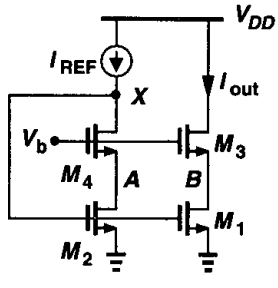


Figure 5.38

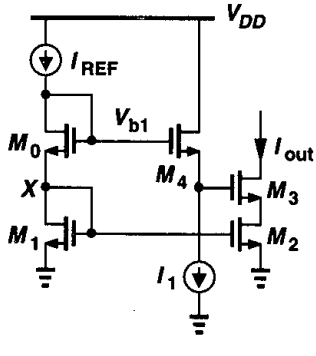


Figure 5.39

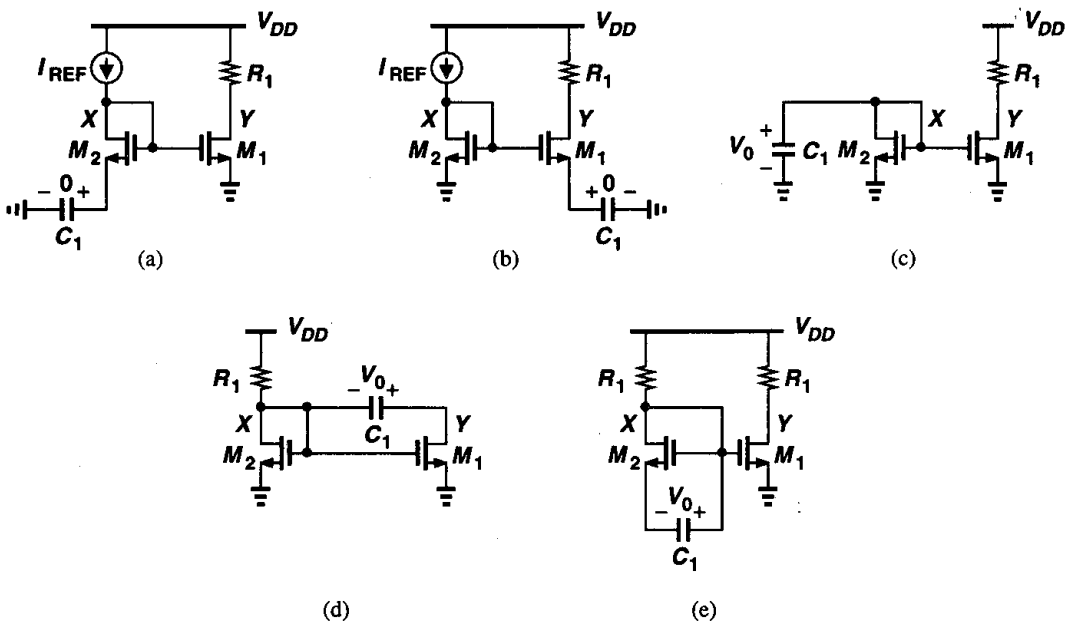


Figure 5.40

5.17. Sketch V_X and V_Y as a function of time for each circuit in Fig. 5.41. Assume the transistors in each circuit are identical.

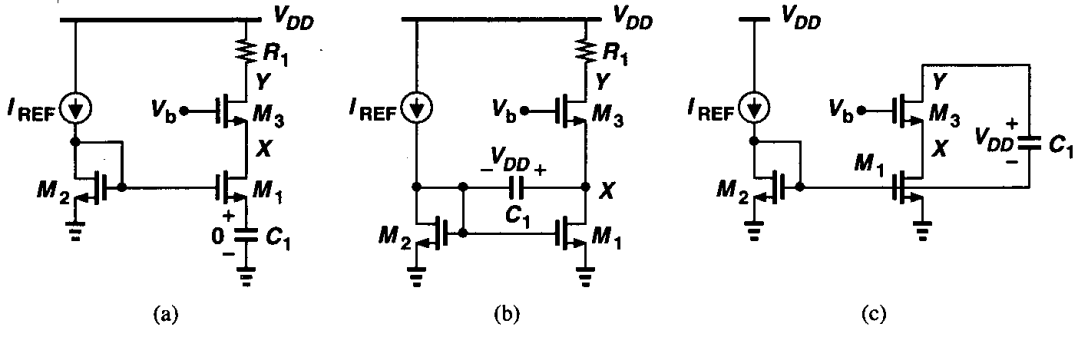


Figure 5.41

5.18. Sketch V_X and V_Y as a function of time for each circuit in Fig. 5.42. Assume the transistors in each circuit are identical.

5.19. The circuit shown in Fig. 5.43 exhibits a *negative* input capacitance. Calculate the input impedance of the circuit and identify the capacitive component.

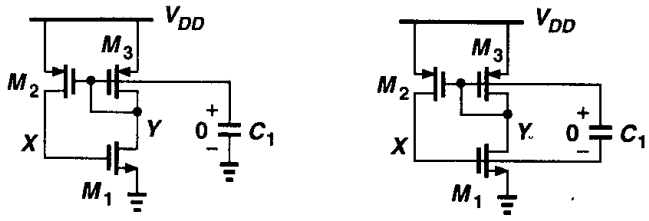


Figure 5.42

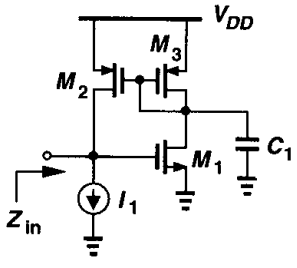


Figure 5.43

5.20. Due to a manufacturing defect, a large parasitic resistance, R_1, has appeared in the circuits of Fig. 5.44. Calculate the gain of each circuit.

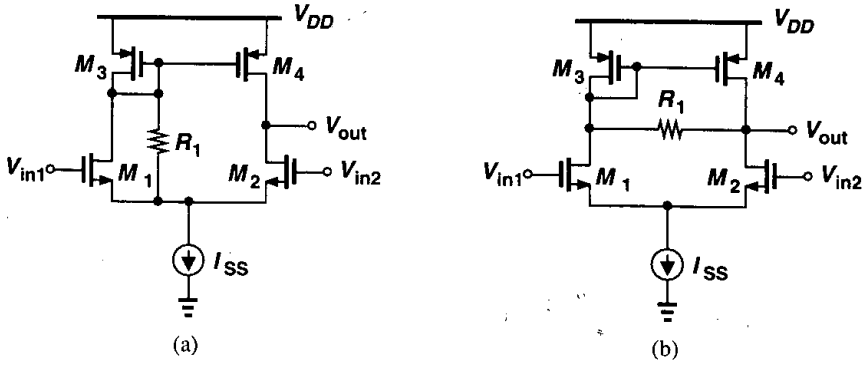


Figure 5.44

5.21. In digital circuits such as memories, a differential pair with active current mirror is used to convert a small differential signal to a large single-ended swing (Fig. 5.45). In such applications, it is desirable that the output levels be as close to the supply rails as possible. Assuming moderate differential input swings (e.g., $\Delta V = 0.1$ V) around a common-mode level $V_{in,CM}$ and a high gain in the circuit, explain why V_{min} depends on $V_{in,CM}$.

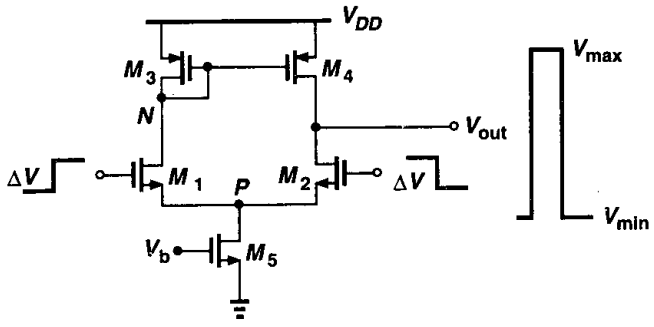


Figure 5.45

5.22. Sketch V_X and V_Y for each circuit in Fig. 5.46 as a function of time. The initial voltage across C_1 is shown.

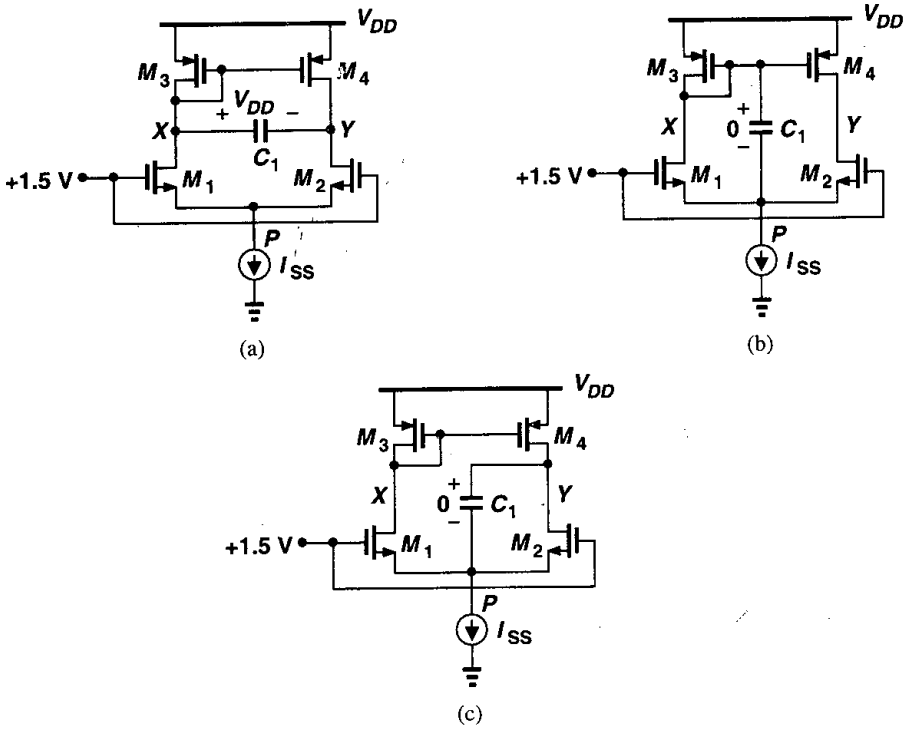


Figure 5.46

5.23. If in Fig. 5.47, ΔV is small enough that all of the transistors remain in saturation, determine the time constant and the initial and final values of V_{out}.

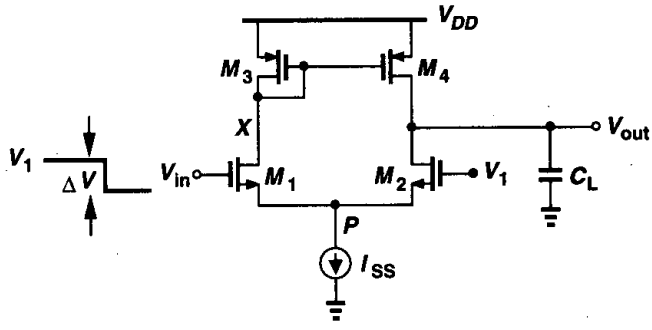


Figure 5.47

Chapter 6

Frequency Response of Amplifiers

Our analysis of simple amplifiers has thus far focussed on low-frequency characteristics, neglecting the effect of device and load capacitances. In most analog circuits, however, the speed trades with many other parameters such as gain, power dissipation, and noise. It is therefore necessary to understand the frequency response limitations of each circuit.

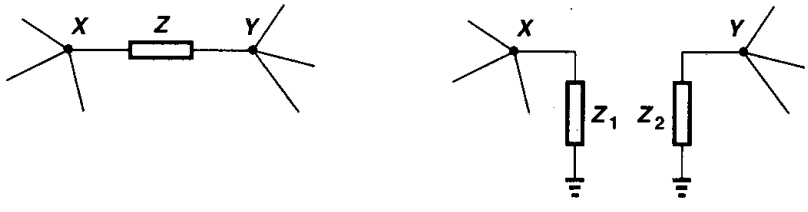
In this chapter, we study the response of single-stage and differential amplifiers in the frequency domain. Following a review of basic concepts, we analyze the high-frequency behavior of common-source and common-gate stages and source followers. Next, we deal with cascode and differential amplifiers. Finally, we consider the effect of active current mirrors on the frequency response of differential pairs.

6.1 General Considerations

6.1.1 Miller Effect

An important phenomenon that occurs in many analog (and digital) circuits is related to "Miller Effect," as described by Miller in a theorem.

Miller's Theorem. If the circuit of Fig. 6.1(a) can be converted to that of Fig. 6.1(b), then $Z_1 = Z/(1 - A_v)$ and $Z_2 = Z/(1 - A_v^{-1})$, where $A_v = V_Y/V_X$.



(a) (b)

Figure 6.1 Application of Miller effect to a floating impedance.

Proof. The current flowing through Z from X to Y is equal to $(V_X - V_Y)/Z$. For the two circuits to be equivalent, the same current must flow through Z_1. Thus,

$$\frac{V_X - V_Y}{Z} = \frac{V_X}{Z_1}, \tag{6.1}$$

that is,

$$Z_1 = \frac{Z}{1 - \dfrac{V_Y}{V_X}}. \tag{6.2}$$

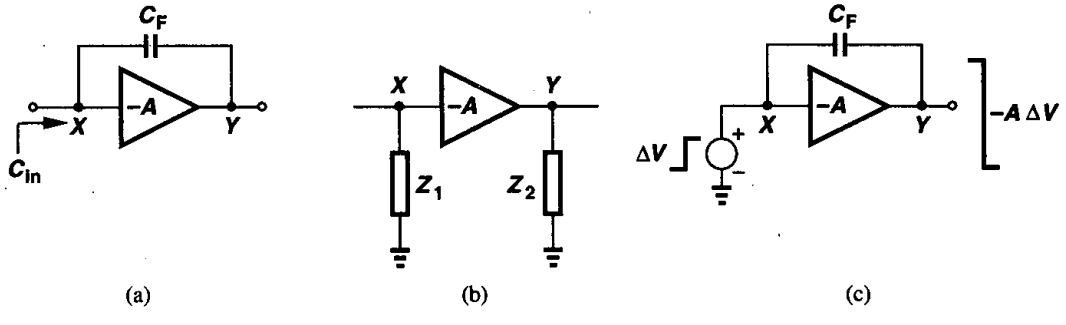
Similarly,

$$Z_2 = \frac{Z}{1 - \dfrac{V_X}{V_Y}}. \tag{6.3}$$

□

Example 6.1

Consider the circuit shown in Fig. 6.2(a), where the voltage amplifier has a negative gain equal to $-A$ and is otherwise ideal. Calculate the input capacitance of the circuit.



(a) (b) (c)

Figure 6.2

Solution

Using Miller's theorem to convert the circuit to that shown in Fig. 6.2(b), we have $Z = 1/(C_F s)$ and $Z_1 = [1/(C_F s)]/(1 + A)$. That is, the input capacitance is equal to $C_F(1 + A)$.

Why is C_F multiplied by $1 + A$? Suppose, as depicted in Fig. 6.2(c), we measure the input capacitance by applying a voltage step at the input and calculating the charge supplied by the voltage source. A step equal to ΔV at X results in a change of $-A\Delta V$ at Y, yielding a total change of $(1+A)\Delta V$ in the voltage across C_F. Thus, the charge drawn by C_F from V_{in} is equal to $(1 + A)C_F\Delta V$ and the equivalent input capacitance equal to $(1 + A)C_F$.

It is important to understand that (6.2) and (6.3) hold *if* we know a priori that the circuit of Fig. 6.1(a) can be converted to that of Fig. 6.1(b). That is, Miller's theorem does not stipulate the conditions under which this conversion is valid. If the impedance Z forms the only signal path between X and Y, then the conversion is often invalid. Illustrated in Fig. 6.3 for a simple resistive divider, the theorem gives a correct input impedance but an incorrect

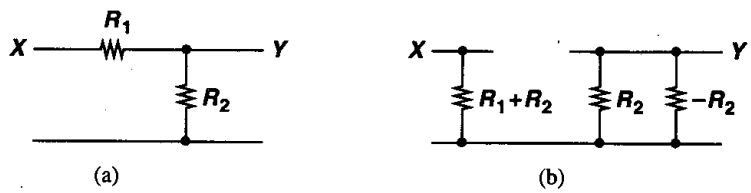


Figure 6.3 Improper application of Miller's theorem.

gain. Nevertheless, Miller's theorem proves useful in cases where the impedance Z appears in parallel with the main signal (Fig. 6.4).

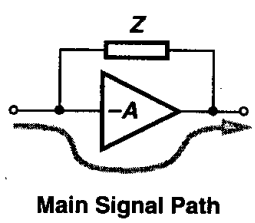


Figure 6.4 Typical case for valid application of Miller's theorem.

Example 6.2

Calculate the input resistance of the circuit shown in Fig. 6.5(a).

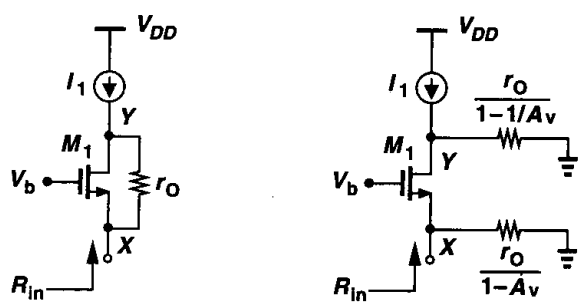


Figure 6.5

Solution

The reader can prove that the voltage gain from X to Y is equal to $1 + (g_m + g_{mb})r_O$. As shown in Fig. 6.5(b), the input resistance is given by the parallel combination of $r_O/(1 - A_v)$ and $1/(g_m + g_{mb})$. Since A_v is usually greater than unity, $r_O/(1 - A_v)$ is a *negative* resistance. We therefore have

$$R_{in} = \frac{r_O}{1 - [1 + (g_m + g_{mb})r_O]} \Bigg\| \frac{1}{g_m + g_{mb}} \tag{6.4}$$

$$= \frac{-1}{g_m + g_{mb}} \Bigg\| \frac{1}{g_m + g_{mb}} \tag{6.5}$$

$$= \infty. \tag{6.6}$$

This is the same result as obtained in Chapter 3 (Fig. 3.46) by direct calculation.

We should also mention that, strictly speaking, the value of $A_v = V_Y/V_X$ in (6.2) and (6.3) must be calculated at the frequency of interest, complicating the algebra significantly. However, in many cases we use the low-frequency value of A_v to gain insight into the behavior of the circuit.

If applied to obtain the input-output transfer function, Miller's theorem cannot be used simultaneously to calculate the output impedance. To derive the transfer function, we apply a voltage source to the *input* of the circuit, obtaining a value for V_Y/V_X in Fig. 6.1(a). On the other hand, to determine the output impedance, we apply a voltage source to the *output* of the circuit, obtaining a value for V_X/V_Y that may not be equal to the inverse of the V_Y/V_X measured in the first test. For example, the circuit of Fig. 6.5(b) may suggest that the output impedance is equal to

$$R_{out} = \frac{r_O}{1 - 1/A_v} \tag{6.7}$$

$$= \frac{r_O}{1 - [1 + (g_m + g_{mb})r_O]^{-1}} \tag{6.8}$$

$$= \frac{1}{g_m + g_{mb}} + r_O, \tag{6.9}$$

whereas the actual value is equal to r_O (if X is grounded). Other subtleties of Miller's theorem are decribed in the appendix.

6.1.2 Association of Poles with Nodes

Consider the simple cascade of amplifiers depicted in Fig. 6.6. Here, A_1 and A_2 are ideal voltage amplifiers, R_1 and R_2 model the output resistance of each stage, C_{in} and C_N represent the input capacitance of each stage, and C_P denotes the load capacitance. The overall transfer function can be written as

$$\frac{V_{out}}{V_{in}}(s) = \frac{A_1}{1 + R_S C_{in} s} \cdot \frac{A_2}{1 + R_1 C_N s} \cdot \frac{1}{1 + R_2 C_P s}. \tag{6.10}$$

The circuit exhibits three poles, each of which is determined by the total capacitance seen from each node to ground multiplied by the total resistance seen at the node to ground. We can therefore associate each pole with one node of the circuit, i.e., $\omega_j = \tau_j^{-1}$, where τ_j is the product of the capacitance and resistance seen at node j to ground. From this perspective, we may say "each node in the circuit contributes one pole to the transfer function."

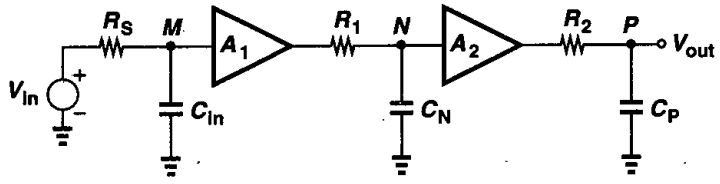


Figure 6.6 Cascade of amplifiers.

The above statement is not valid in general. For example, in the circuit of Fig. 6.7, the location of the poles is difficult to calculate because R_3 and C_3 create interaction between

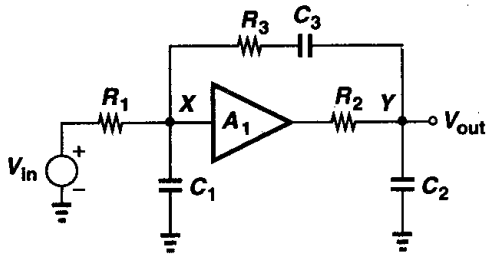


Figure 6.7 Example of interaction between nodes.

X and Y. Nevertheless, in many circuits association of one pole with each node provides an intuitive approach to estimating the transfer function: we simply multiply the total equivalent capacitance by the total incremental resistance (both from the node of interest to ground), thus obtaining an equivalent time constant and hence a pole frequency.

Example 6.3

In Fig. 6.8, calculate the pole associated with node X.

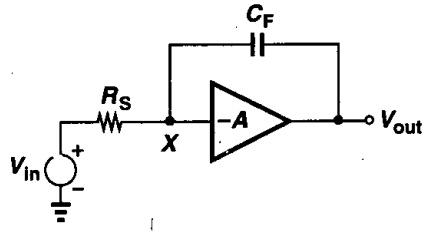


Figure 6.8

Solution

From Fig. 6.2(b), the total equivalent capacitance seen from X to ground equals $(1 + A)C_F$. Since this capacitance is driven by R_S, the pole frequency is equal to $1/[R_S(1 + A)C_F]$ (in rad/s). We call this the "input pole."

The above approach does suffer from some limitations. In particular, the simplification of the circuit through the use of Miller effect often discards the *zeros* of the transfer function. However, the utility of the method becomes apparent in more complex topologies, as described in the following example.

Example 6.4

Neglecting channel-length modulation, compute the transfer function of the common-gate stage shown in Fig. 6.9.

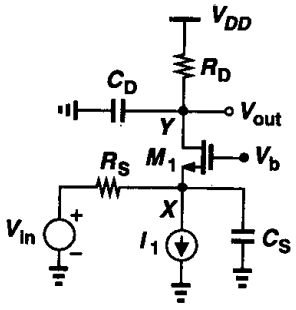


Figure 6.9 Common-gate stage with parasitic capacitances.

Solution

In this circuit, the capacitances contributed by M_1 are connected from the input and output nodes to ground. At node X, $C_S = C_{GS1} + C_{SB1}$, giving a pole frequency

$$\omega_{in} = \left[(C_{GS1} + C_{SB1})\left(R_S \left\| \frac{1}{g_{m1} + g_{mb1}}\right.\right)\right]^{-1} \tag{6.11}$$

Similarly, at node Y, $C_D = C_{DG} + C_{DB}$, yielding a pole frequency

$$\omega_{out} = [(C_{DG} + C_{DB})R_D]^{-1}. \tag{6.12}$$

The overall transfer function is thus given by

$$\frac{V_{out}}{V_{in}}(s) = \frac{(g_m + g_{mb})R_D}{1 + (g_m + g_{mb})R_S} \cdot \frac{1}{\left(1 + \dfrac{s}{\omega_{in}}\right)\left(1 + \dfrac{s}{\omega_{out}}\right)}, \tag{6.13}$$

where the first fraction represents the low-frequency gain of the circuit. Note that if we do not neglect r_{O1}, the input and output nodes interact, making it difficult to calculate the poles.

6.2 Common-Source Stage

The common-source topology exhibits a relatively high input impedance while providing voltage gain and requiring a minimal voltage headroom. As such, it finds wide application in analog circuits and its frequency response is of interest.

Shown in Fig. 6.10 is a common-source stage driven by a finite source resistance, R_S. We identify all of the capacitances in the circuit, noting that C_{GS} and C_{DB} are "grounded" capacitances while C_{GD} appears between the input and the output. Assuming that $\lambda = 0$ and M_1 operates in saturation, let us first estimate the transfer function by associating one pole with each node. The total capacitance seen from X to ground is equal to C_{GS} plus the Miller multiplication of C_{GD}: $C_{GS} + (1 - A_v)C_{GD}$, where $A_v = -g_m R_D$. The magnitude of the input pole is therefore given by

$$\omega_{in} = \frac{1}{R_S[C_{GS} + (1 + g_m R_D)C_{GD}]}. \tag{6.14}$$

At the output node, the total capacitance seen to ground is equal to C_{DB} plus the Miller effect of C_{GD}: $C_{DB} + (1 - A_v^{-1})C_{GD} \approx C_{DB} + C_{GD}$. Thus,

$$\omega_{out} = \frac{1}{R_D(C_{DB} + C_{GD})}. \tag{6.15}$$

Another approximation of the output pole can be obtained if R_S is relatively large. Simplifying the circuit as shown in Fig. 6.11, where the effect of R_S is neglected, the reader can prove that

$$Z_X = \frac{1}{C_{eq}s} \left\| \left(\frac{C_{GD} + C_{GS}}{C_{GD}} \cdot \frac{1}{g_{m1}} \right), \right. \tag{6.16}$$

where $C_{eq} = C_{GD}C_{GS}/(C_{GD} + C_{GS})$. Thus, the output pole is roughly equal to

$$\omega_{out} = \frac{1}{\left[R_D \left\| \left(\frac{C_{GD} + C_{GS}}{C_{GD}} \cdot \frac{1}{g_{m1}} \right) \right. \right] (C_{eq} + C_{DB})}. \tag{6.17}$$

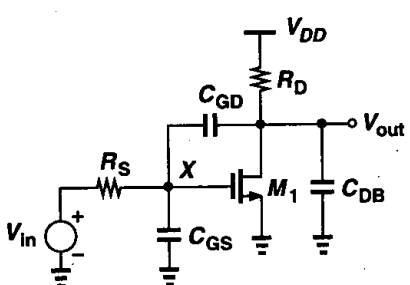


Figure 6.10 High-frequency model of a common-source stage.

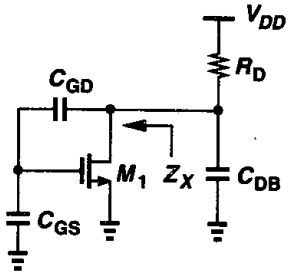


Figure 6.11 Model for calculation of output impedance.

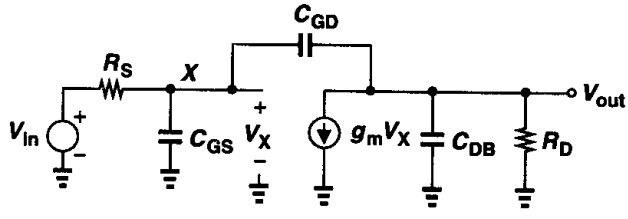


Figure 6.12 Equivalent circuit of Fig. 6.10.

We then surmise that the transfer function is

$$\frac{V_{out}}{V_{in}}(s) = \frac{-g_m R_D}{\left(1 + \frac{s}{\omega_{in}}\right)\left(1 + \frac{s}{\omega_{out}}\right)}. \tag{6.18}$$

Note that r_{O1} and any load capacitance can easily be included here.

The primary error in this estimation is that we have not considered the existence of zeros in the circuit. Another concern stems from approximating the gain of the amplifier by $-g_m R_D$ whereas in reality the gain varies with frequency (for example, due to the capacitance at the output node).

We now obtain the exact transfer function, investigating the validity of the above approach. Using the equivalent circuit depicted in Fig. 6.12, we can sum the currents at each node:

$$\frac{V_X - V_{in}}{R_S} + V_X C_{GS} s + (V_X - V_{out}) C_{GD} s = 0 \tag{6.19}$$

$$(V_{out} - V_X) C_{GD} s + g_m V_X + V_{out}\left(\frac{1}{R_D} + C_{DB} s\right) = 0. \tag{6.20}$$

From (6.20), V_X is obtained as

$$V_X = -\frac{V_{out}\left(C_{GD}s + \frac{1}{R_D} + C_{DB}s\right)}{g_m - C_{GD}s}, \tag{6.21}$$

which, upon substitution in (6.19), yields

$$-V_{out}\frac{[R_S^{-1} + (C_{GS} + C_{GD})s][R_D^{-1} + (C_{GD} + C_{DB})s]}{g_m - C_{GD}s} - V_{out}C_{GD}s = \frac{V_{in}}{R_S}. \tag{6.22}$$

That is,

$$\frac{V_{out}}{V_{in}}(s) = \frac{(C_{GD}s - g_m)R_D}{R_S R_D \xi s^2 + [R_S(1 + g_m R_D)C_{GD} + R_S C_{GS} + R_D(C_{GD} + C_{DB})]s + 1}, \tag{6.23}$$

where $\xi = C_{GS}C_{GD} + C_{GS}C_{DB} + C_{GD}C_{DB}$. Note that the transfer function is of second order even though the circuit contains three capacitors. This is because the capacitors form a "loop," allowing only *two* independent initial conditions in the circuit and hence yielding a second-order differential equation for the time response.

If manipulated judiciously, Eq. (6.23) reveals several interesting points about the circuit. While the denominator appears rather complicated, it can yield intuitive expressions for the two poles, ω_{p1} and ω_{p2}, if we assume $|\omega_{p1}| \ll |\omega_{p2}|$ [1]. Writing the denominator as

$$D = \left(\frac{s}{\omega_{p1}} + 1\right)\left(\frac{s}{\omega_{p2}} + 1\right) \tag{6.24}$$

$$= \frac{s^2}{\omega_{p1}\omega_{p2}} + \left(\frac{1}{\omega_{p1}} + \frac{1}{\omega_{p2}}\right)s + 1, \tag{6.25}$$

we recognize that the coefficient of s is approximately equal to $1/\omega_{p1}$ if ω_{p2} is much farther from the origin. It follows from (6.23) that

$$\omega_{p1} = \frac{1}{R_S(1 + g_m R_D)C_{GD} + R_S C_{GS} + R_D(C_{GD} + C_{DB})}. \tag{6.26}$$

How does this compare with the "input" pole given by (6.14)? The only difference results from the term $R_D(C_{GD} + C_{DB})$, which may be negligible in some cases. The key point here is that the intuitive approach of associating a pole with the input node provides a rough estimate with much less effort. We also note that the Miller multiplication of C_{GD} by the low-frequency gain of the amplifier is relatively accurate in this case.

Example 6.5

For the circuit shown in Fig. 6.13, calculate the transfer function (with $\lambda = 0$) and explain why Miller effect vanishes as C_{DB} increases.

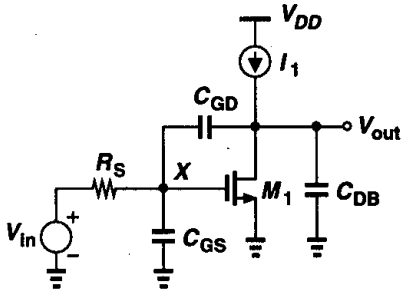


Figure 6.13

Solution

Using (6.23) and letting R_D approach infinity, we have

$$\frac{V_{out}}{V_{in}}(s) = \frac{C_{GD}s - g_m}{R_S\xi s^2 + [g_m R_S C_{GD} + (C_{GD} + C_{DB})]s} \tag{6.27}$$

$$= \frac{C_{GD}s - g_m}{s[R_S(C_{GS}C_{GD} + C_{GS}C_{DB} + C_{GD}C_{DB})s + (g_m R_S + 1)C_{GD} + C_{DB}]}. \tag{6.28}$$

As expected, the circuit exhibits two poles—one at the origin because the dc gain is infinity. The magnitude of the other pole is given by

$$\omega_2 \approx \frac{(1 + g_m R_S)C_{GD} + C_{DB}}{R_S(C_{GD}C_{GS} + C_{GS}C_{DB} + C_{GD}C_{DB})}. \tag{6.29}$$

For large C_{DB}, this expression reduces to

$$\omega_2 \approx \frac{1}{R_S(C_{GS} + C_{GD})}, \tag{6.30}$$

indicating that C_{GD} experiences no Miller multiplication. This can be explained by noting that, for large C_{DB}, the voltage gain from node X to the output begins to drop even at low frequencies. As a result, for frequencies close to $[R_S(C_{GS} + C_{GD})]^{-1}$, the effective gain is quite small and $C_{GD}(1 - A_v) \approx C_{GD}$. Such a case is an example where the application of Miller effect using low-frequency gain does not provide a reasonable estimate.

From (6.23), we can also estimate the second pole of the CS stage of Fig. 6.10. Since the coefficient of s^2 is equal to $(\omega_{p1}\omega_{p2})^{-1}$, we have

$$\omega_{p2} = \frac{1}{\omega_{p1}} \cdot \frac{1}{R_S R_D(C_{GS}C_{GD} + C_{GS}C_{DB} + C_{GD}C_{DB})} \tag{6.31}$$

$$= \frac{R_S(1 + g_m R_D)C_{GD} + R_S C_{GS} + R_D(C_{GD} + C_{DB})}{R_S R_D(C_{GS}C_{GD} + C_{GS}C_{DB} + C_{GD}C_{DB})}. \tag{6.32}$$

If $C_{GS} \gg (1 + g_m R_D)C_{GD} + R_D(C_{GD} + C_{DB})/R_S$, then

$$\omega_{p2} \approx \frac{R_S C_{GS}}{R_S R_D(C_{GS}C_{GD} + C_{GS}C_{DB})} \tag{6.33}$$

$$= \frac{1}{R_D(C_{GD} + C_{DB})}, \tag{6.34}$$

the same as (6.15). Thus, the "output" pole approach is valid only if C_{GS} dominates the response.

The transfer function of (6.23) exhibits a zero given by $\omega_z = +g_m/C_{GD}$, an effect not predicted by the simple approach leading to (6.18). Located in the *right* half plane, the zero arises from direct coupling of the input to the output through C_{GD}. As illustrated in Fig. 6.14, C_{GD} provides a feedforward path that conducts the input signal to the output at very high frequencies, resulting in a slope in the frequency response that is less negative than −40 dB/dec. As explained in Chapter 10, a zero in the right half plane introduces stability issues in feedback amplifiers.

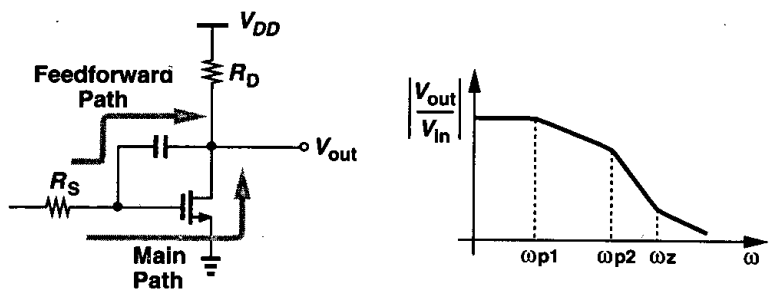


Figure 6.14 Feedforward path through C_{GD} (log-log scale).

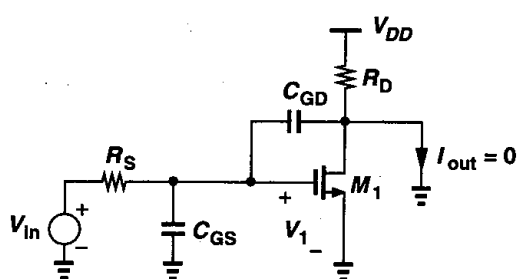


Figure 6.15 Calculation of the zero in a CS stage.

The zero, s_z, can also be computed by noting that the transfer function $V_{out}(s)/V_{in}(s)$ must drop to zero for $s = s_z$. For a finite V_{in}, this means that $V_{out}(s_z) = 0$ and hence the output can be *shorted* to ground at this (possibly complex) frequency with no current flowing through the short (Fig. 6.15). Therefore, the currents through C_{GD} and M_1 are equal

and opposite:

$$V_1 C_{GD} s_z = g_m V_1. \tag{6.35}$$

That is, $s_z = +g_m/C_{GD}$.[1]

In high-speed applications, the input impedance of the common-source stage is also important. As a first-order approximation, we have from Fig. 6.16(a)

$$Z_{in} = \frac{1}{[C_{GS} + (1 + g_m R_D) C_{GD}]s}. \tag{6.36}$$

But at high frequencies, the effect of the output node must be taken into account. Ignoring C_{GS} for the moment and using the circuit of Fig. 6.16(b), we write

$$(I_X - g_m V_X)\frac{R_D}{1 + R_D C_{DB} s} + \frac{I_X}{C_{GD} s} = V_X, \tag{6.37}$$

and hence

$$\frac{V_X}{I_X} = \frac{1 + R_D(C_{GD} + C_{DB})s}{C_{GD} s(1 + g_m R_D + R_D C_{DB} s)}. \tag{6.38}$$

The actual input impedance consists of the parallel combination of (6.38) and $1/(C_{GS}s)$.

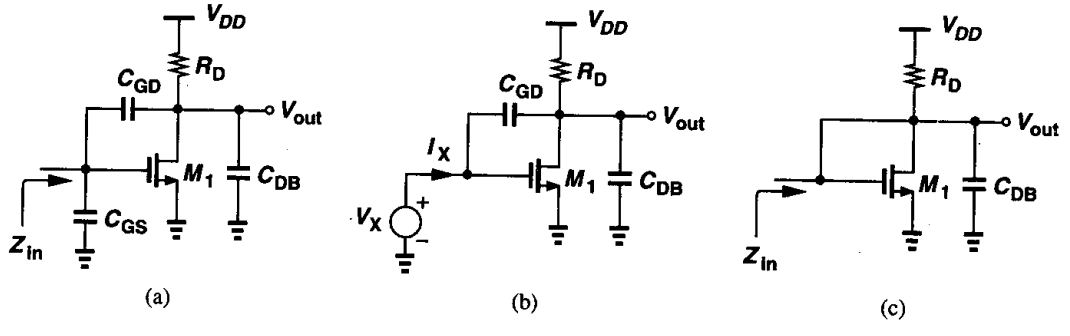


Figure 6.16 Calculation of input impedance of a CS stage.

At frequencies where $|R_D(C_{GD} + C_{DB})s| \ll 1$ and $|R_D C_{DB} s| \ll 1 + g_m R_D$, (6.38) reduces to $[(1 + g_m R_D)C_{GD}s]^{-1}$ (as expected), indicating that the input impedance is primarily capacitive. At higher frequencies, however, (6.38) contains both real and imaginary parts. In fact, if C_{GD} is large, it provides a low-impedance path between the gate and drain of M_1, yielding the equivalent circuit of Fig. 6.16(c) and suggesting that $1/g_{m1}$ and R_D appear in parallel with the input.

[1] This approach is similar to expressing the transfer function as $G_m Z_{out}$ and finding the zeros of G_m and Z_{out}.

6.3 Source Followers

Source followers are occasionally employed as level shifters or buffers, impacting the overall frequency response. Consider the circuit depicted in Fig. 6.17(a), where C_L represents the total capacitance seen at the output node to ground, including C_{SB1}. The strong inter-

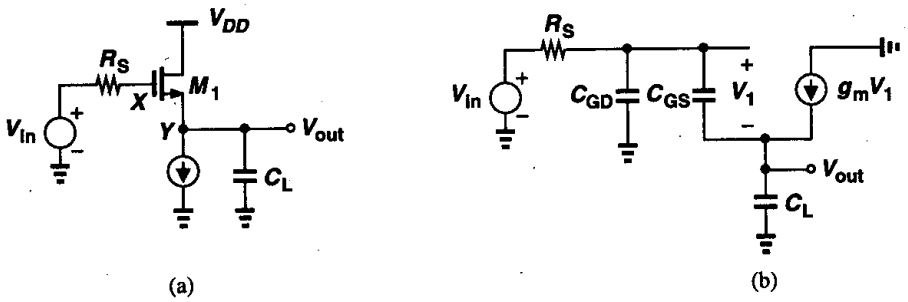


Figure 6.17 (a) Source follower, (b) high-frequency equivalent circuit.

action between nodes X and Y through C_{GS} in Fig. 6.17(a) makes it difficult to associate a pole with each node in a source follower. Neglecting body effect for simplicity and using the equivalent circuit shown in Fig. 6.17(b), we can sum the currents at the output node:

$$V_1 C_{GS} s + g_m V_1 = V_{out} C_L s, \tag{6.39}$$

obtaining

$$V_1 = \frac{C_L s}{g_m + C_{GS} s} V_{out}. \tag{6.40}$$

Also, beginning from V_{in}, we can add up all of the voltages:

$$V_{in} = R_S[V_1 C_{GS} s + (V_1 + V_{out}) C_{GD} s] + V_1 + V_{out}. \tag{6.41}$$

Substituting for V_1 from (6.40), we have

$$\frac{V_{out}}{V_{in}}(s) = \frac{g_m + C_{GS} s}{R_S(C_{GS}C_L + C_{GS}C_{GD} + C_{GD}C_L)s^2 + (g_m R_S C_{GD} + C_L + C_{GS})s + g_m}. \tag{6.42}$$

Interestingly, the transfer function contains a zero in the *left* half plane. This is because the signal conducted by C_{GS} at high frequencies adds with the same polarity to the signal produced by the intrinsic transistor.

If the two poles of (6.42) are assumed far apart, then the more significant one has a magnitude of

$$\omega_{p1} \approx \frac{g_m}{g_m R_S C_{GD} + C_L + C_{GS}} \tag{6.43}$$

$$= \frac{1}{R_S C_{GD} + \dfrac{C_L + C_{GS}}{g_m}}. \tag{6.44}$$

Also, if $R_S = 0$, then $\omega_{p1} = g_m/(C_L + C_{GS})$.

Let us now calculate the input impedance of the circuit, noting that C_{GD} simply shunts the input and can be ignored initially. From the equivalent shown in Fig. 6.18, the small-signal

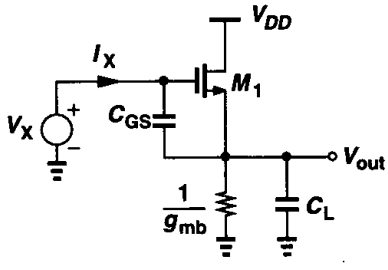


Figure 6.18 Calculation of source follower input impedance.

gate-source voltage of M_1 is equal to $I_X/(C_{GS}s)$, giving a source current of $g_m I_X/(C_{GS}s)$. Starting from the input and adding the voltages, we have

$$V_X = \frac{I_X}{C_{GS}s} + \left(I_X + \frac{g_m I_X}{C_{GS}s}\right)\left(\frac{1}{g_{mb}} \Big\| \frac{1}{C_L s}\right), \tag{6.45}$$

that is,

$$Z_{in} = \frac{1}{C_{GS}s} + \left(1 + \frac{g_m}{C_{GS}s}\right)\frac{1}{g_{mb} + C_L s}. \tag{6.46}$$

At relatively low frequencies, $g_{mb} \gg |C_L s|$ and

$$Z_{in} \approx \frac{1}{C_{GS}s}\left(1 + \frac{g_m}{g_{mb}}\right) + \frac{1}{g_{mb}}, \tag{6.47}$$

indicating that the equivalent input capacitance is equal to $C_{GS}g_{mb}/(g_m + g_{mb})$. This result can also be obtained by Miller approximation. Since the low-frequency gain from the input to the output equals $g_m/(g_m + g_{mb})$, the effect of C_{GS} at the input can be expressed as $C_{GS}[1 - g_m/(g_m + g_{mb})] = C_{GS}g_{mb}(g_m + g_{mb})$. In other words, the overall input capacitance is equal to C_{GD} plus a *fraction* of C_{GS}.

At high frequencies, $g_{mb} \ll |C_L s|$ and

$$Z_{in} \approx \frac{1}{C_{GS}s} + \frac{1}{C_L s} + \frac{g_m}{C_{GS}C_L s^2}. \tag{6.48}$$

For a given $s = j\omega$, the input impedance consists of the series combination of capacitors C_{GS} and C_L and a *negative* resistance equal to $-g_m/(C_{GS}C_L\omega^2)$. The negative resistance property can be utilized in oscillators [2].

Example 6.6

Calculate the transfer function of the circuit shown in Fig. 6.19(a).

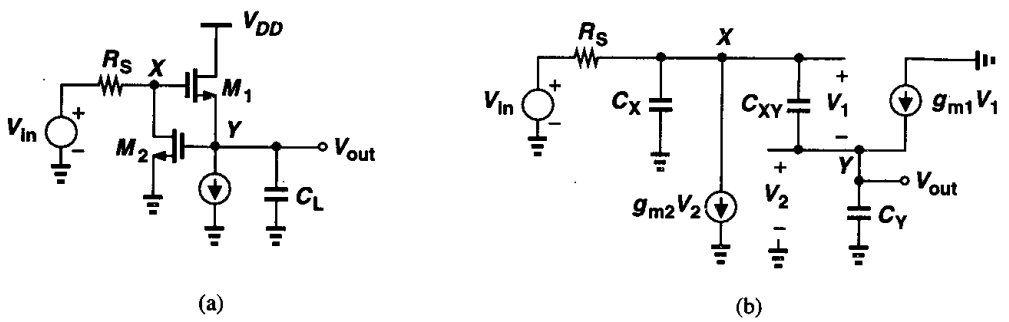


Figure 6.19

Solution

Let us first identify all of the capacitances in the circuit. At node X, C_{GD1} and C_{DB2} are connected to ground and C_{GS1} and C_{GD2} to Y. At node Y, C_{SB1}, C_{GS2}, and C_L are connected to ground. Similar to the source follower of Fig. 6.17(b), this circuit has three capacitances in a loop and hence a second-order transfer function. Using the equivalent circuit shown in Fig. 6.19(b), where $C_X = C_{GD1}+C_{DB2}$, $C_{XY} = C_{GS1} + C_{GD2}$, and $C_Y = C_{SB1} + C_{GS2} + C_L$, we have $V_1C_{XY}s + g_{m1}V_1 = V_{out}C_Ys$ and hence $V_1 = V_{out}C_Ys/(C_{XY}s + g_{m1})$. Also, since $V_2 = V_{out}$, the summation of currents at node X gives

$$(V_1 + V_{out})C_Xs + g_{m2}V_{out} + V_1C_{XY}s = \frac{V_{in} - V_1 - V_{out}}{R_S}. \tag{6.49}$$

Substituting for V_1 and simplifying the result, we obtain

$$\frac{V_{out}}{V_{in}}(s) = \frac{g_{m1} + C_{XY}s}{R_S\xi s^2 + [C_Y + g_{m1}R_SC_X + (1 + g_{m2}R_S)C_{XY}]s + g_{m1}(1 + g_{m2}R_S)}, \tag{6.50}$$

where $\xi = C_XC_Y + C_XC_{XY} + C_YC_{XY}$. As expected, (6.50) reduces to a form similar to (6.42) for $g_{m2} = 0$.

The output impedance of source followers is also of interest. In Fig. 6.17(a), the body effect and C_{SB} simply yield an impedance in parallel with the output. Ignoring this impedance and neglecting C_{GD}, we note from the equivalent circuit of Fig. 6.20(a) that $V_1C_{GS}s +$

$g_m V_1 = -I_X$. Also, $V_1 C_{GS} s R_S + V_1 = -V_X$. Dividing both sides of these equations gives

$$Z_{out} = \frac{V_X}{I_X} \tag{6.51}$$

$$= \frac{R_S C_{GS} s + 1}{g_m + C_{GS} s}. \tag{6.52}$$

It is instructive to examine the magnitude of this impedance as a function of frequency. At low frequencies, $Z_{out} \approx 1/g_m$, as expected. At very high frequencies, $Z_{out} \approx R_S$ (because C_{GS} shorts the gate and the source). We therefore surmise that $|Z_{out}|$ varies as shown in Figs. 6.20(b) or (c). Which one of these variations is more realistic? Operating as buffers, source followers must lower the output impedance, i.e., $1/g_m < R_S$. For this reason, the characteristic shown in Fig. 6.20(c) occurs more commonly than that in Fig. 6.20(b).

The behavior illustrated in Fig. 6.20(c) reveals an important attribute of source followers. Since the output impedance *increases* with frequency, we postulate that it contains an *inductive* component. To confirm this guess, we represent Z_{out} by a first-order passive network, noting that Z_{out} equals $1/g_m$ at $\omega = 0$ and R_S at $\omega = \infty$. The network can therefore be assumed as shown in Fig. 6.21 because Z_1 equals R_2 at $\omega = 0$ and $R_1 + R_2$ at $\omega = \infty$. In other words, $Z_1 = Z_{out}$ if $R_2 = 1/g_m$, $R_1 = R_S - 1/g_m$, and L is chosen properly.

To calculate L, we can simply obtain an expression for Z_1 in terms of the three components in Fig. 6.21 and equate the result to Z_{out} found above. Alternatively, since R_2 is a series component of Z_1, we can subtract its value from Z_{out}, thereby obtaining an expression

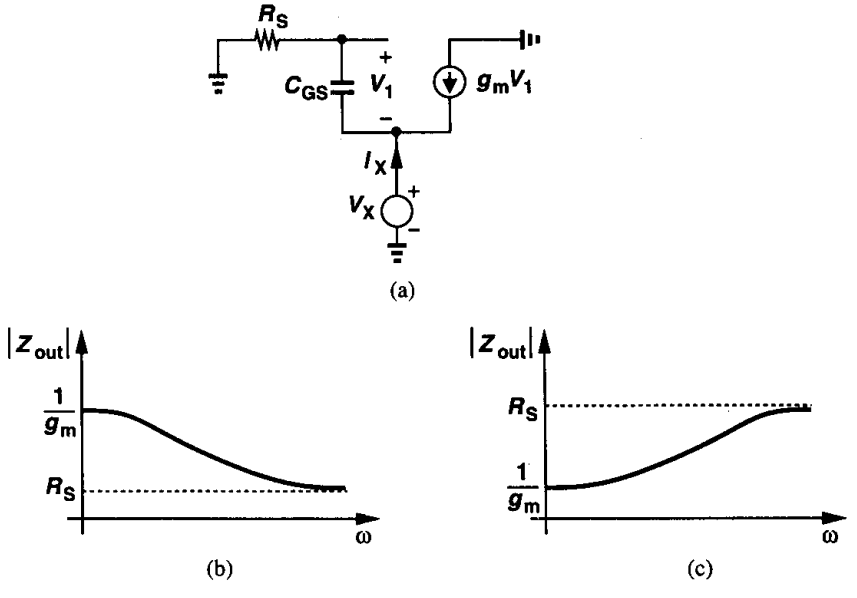


Figure 6.20 Calculation of source follower output impedance.

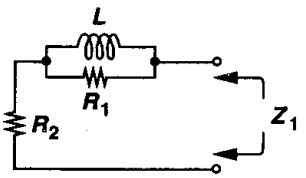


Figure 6.21 Equivalent output impedance of a source follower.

for the parallel combination of R_1 and L:

$$Z_{out} - \frac{1}{g_m} = \frac{C_{GS}s\left(R_S - \dfrac{1}{g_m}\right)}{g_m + C_{GS}s}. \tag{6.53}$$

Inverting the result to obtain the admittance of the parallel circuit, we have

$$\frac{1}{Z_{out} - \dfrac{1}{g_m}} = \frac{1}{R_S - \dfrac{1}{g_m}} + \frac{1}{\dfrac{C_{GS}s}{g_m}\left(R_S - \dfrac{1}{g_m}\right)}. \tag{6.54}$$

We can thus identify the first term on the right hand side as the inverse of R_1 and the second term as the inverse of an impedance equal to $(C_{GS}s/g_m)(R_S - 1/g_m)$, i.e., an inductor with the value

$$L = \frac{C_{GS}}{g_m}\left(R_S - \frac{1}{g_m}\right). \tag{6.55}$$

The dependence of L upon R_S implies that if a source follower is driven by a large resistance, then it exhibits substantial inductive behavior. As depicted in Fig. 6.22, this effect manifests itself as "ringing" in the step response if the circuit drives a large load capacitance.

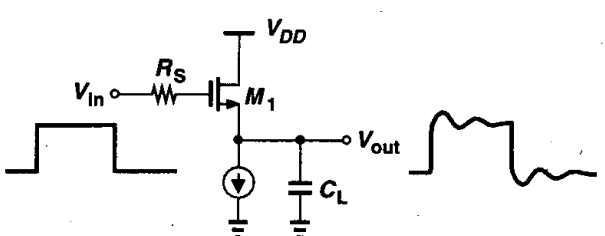


Figure 6.22 Ringing in step response of a source follower with heavy capacitive load.

6.4 Common-Gate Stage

As explained in Example 6.4, in a common-gate stage the input and output nodes are "isolated" if channel-length modulation is neglected. For a common-gate stage such as that in Fig. 6.23, the calculation of Example 6.4 suggested a transfer function

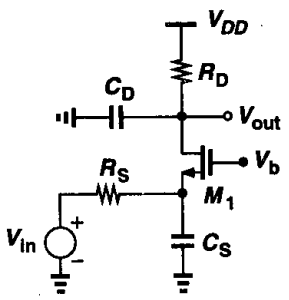


Figure 6.23 Common-gate stage at high frequencies.

$$\frac{V_{out}}{V_{in}}(s) = \frac{(g_m + g_{mb})R_D}{1 + (g_m + g_{mb})R_S} \frac{1}{\left(1 + \dfrac{C_S}{g_m + g_{mb} + R_S^{-1}}s\right)(1 + R_D C_D s)}. \tag{6.56}$$

An important property of this circuit is that it exhibits no Miller multiplication of capacitances, potentially achieving a wide band. Note, however, that the low input impedance may load the preceding stage. Furthermore, since the voltage drop across R_D is typically maximized to obtain a reasonable gain, the dc level of the input signal must be quite low.

If channel-length modulation is not negligible, the calculations become quite complex. Recall from Chapter 3 that the input impedance of a common-gate topology does depend on the drain load if $\lambda \neq 0$. From Eq. (3.110), we can express the impedance seen looking into the source of M_1 in Fig. 6.23 as

$$Z_{in} \approx \frac{Z_L}{(g_m + g_{mb})r_O} + \frac{1}{g_m + g_{mb}}, \tag{6.57}$$

where $Z_L = R_D \| [1/(C_D s)]$. Since Z_{in} now depends on Z_L, it is difficult to associate a pole with the input node.

Example 6.7

For the common-gate stage shown in Fig. 6.24(a), calculate the transfer function and the input impedance, Z_{in}. Explain why Z_{in} becomes independent of C_L as this capacitance increases.

Solution

Using the equivalent circuit shown in Fig. 6.24(b), we can write the current through R_S as $-V_{out}C_L s + V_1 C_{in} s$. Noting that the voltage across R_S plus V_{in} must equal $-V_1$, we have

$$(-V_{out}C_L s + V_1 C_{in} s)R_S + V_{in} = -V_1, \tag{6.58}$$

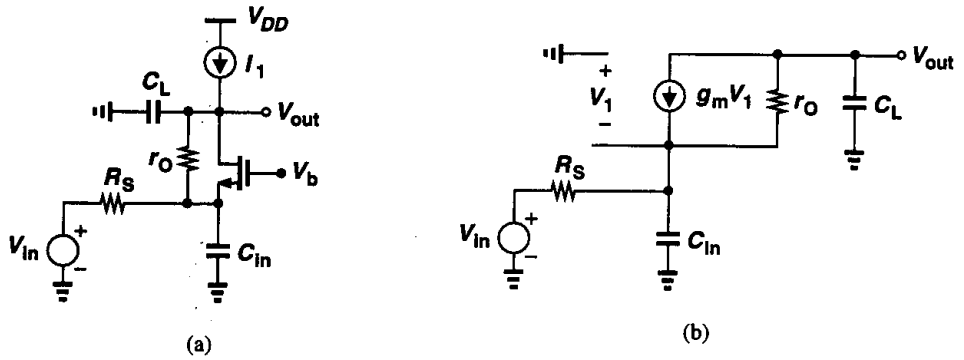


Figure 6.24

that is,

$$V_1 = -\frac{-V_{out} C_L s R_S + V_{in}}{1 + C_{in} R_S s}. \tag{6.59}$$

We also observe that the voltage across r_O minus V_1 equals V_{out}:

$$r_O(-V_{out} C_L s - g_m V_1) - V_1 = V_{out}. \tag{6.60}$$

Substituting for V_1 from (6.59), we obtain the transfer function:

$$\frac{V_{out}}{V_{in}}(s) = \frac{1 + g_m r_O}{r_O C_L C_{in} R_S s^2 + [r_O C_L + C_{in} R_S + (1 + g_m r_O) C_L R_S] s + 1}. \tag{6.61}$$

The reader can prove that body effect can be included by simply replacing g_m with $g_m + g_{mb}$. As expected, the gain at very low frequencies is equal to $1 + g_m r_O$. For Z_{in}, we can use (6.57) by replacing Z_L with $1/(C_L s)$, obtaining

$$Z_{in} = \frac{1}{g_m + g_{mb}} + \frac{1}{C_L s} \cdot \frac{1}{(g_m + g_{mb}) r_O}. \tag{6.62}$$

We note that as C_L or s increases, Z_{in} approaches $1/(g_m + g_{mb})$ and hence the input pole can be defined as

$$\omega_{p,in} = \frac{1}{\left(R_S \left\| \frac{1}{g_m + g_{mb}} \right. \right) C_{in}}. \tag{6.63}$$

Why does Z_{in} become independent of C_L at high frequencies? This is because C_L lowers the voltage gain of the circuit, thereby suppressing the effect of the negative resistance introduced by Miller effect through r_O (Fig. 6.5). In the limit, C_L shorts the output node to ground, and r_O affects the input impedance negligibly.

If a common-gate stage is driven by a relatively large source impedance, then the output impedance of the circuit drops at high frequencies. This effect is better described in the context of cascode circuits.

6.5 Cascode Stage

As explained in Chapter 3, cascoding proves beneficial in increasing the voltage gain of amplifiers and the output impedance of current sources while providing shielding as well. The invention of the cascode (in the vacuum tube era), however, was motivated by the need for high-frequency amplifiers with relatively high input impedance. Viewed as a cascade of a common-source stage and a common-gate stage, a cascode circuit offers the speed of the latter—by suppressing the Miller effect—and the input impedance of the former.

Let us consider the cascode shown in Fig. 6.25, first identifying all of the device capacitances. At node A, C_{GS1} is connected to ground and C_{GD1} to node X. At node X, C_{DB1}, C_{SB2}, and C_{GS2} are tied to ground, and at node Y, C_{DB2}, C_{GD2}, and C_L are connected to ground. The Miller effect of C_{GD1} is determined by the gain from A to X. As an approximation, we use the low-frequency value of this gain, which for low values of R_D (or negligible channel-length modulation) is equal to $-g_{m1}/(g_{m2} + g_{mb2})$. Thus, if M_1 and M_2 have roughly equal dimensions, C_{GD1} is multiplied by approximately 2 rather than the large voltage gain in a simple common-source stage. We therefore say Miller effect is less significant in cascode amplifiers than in common-source stages. The pole associated with node A is estimated as

$$\omega_{p,A} = \frac{1}{R_S\left[C_{GS1} + \left(1 + \frac{g_{m1}}{g_{m2} + g_{mb2}}\right) C_{GD1}\right]}. \tag{6.64}$$

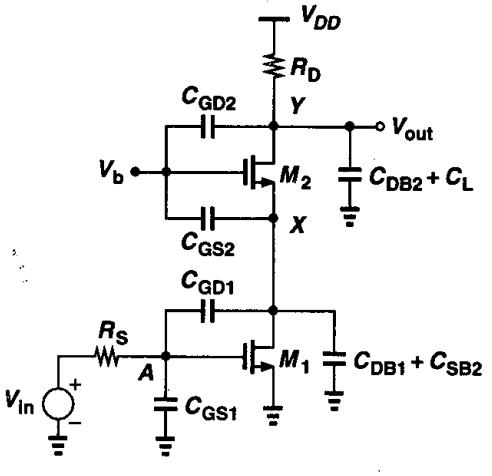


Figure 6.25 High-frequency model of a cascode stage.

We can also attribute a pole to node X. The total capacitance at this node is roughly equal to $2C_{GD1} + C_{DB1} + C_{SB2} + C_{GS2}$, giving a pole

$$\omega_{p,X} = \frac{g_{m2} + g_{mb2}}{2C_{GD1} + C_{DB1} + C_{SB2} + C_{GS2}}. \tag{6.65}$$

Finally, the output node yields a third pole:

$$\omega_{p,Y} = \frac{1}{R_D(C_{DB2} + C_L + C_{GD2})}. \tag{6.66}$$

The relative magnitudes of the three poles in a cascode circuit depend on the actual design parameters, but $\omega_{p,X}$ is typically chosen to be farther from the origin than the other two. As explained in Chapter 10, this choice plays an important role in the stability of op amps.

But what if R_D in Fig. 6.25 is replaced by a current source so as to achieve a higher dc gain? We know from Chapter 3 that the impedance seen at node X reaches high values if the load impedance at the drain of M_2 is large. For example, Eq. (3.110) predicts that the pole at node X may be quite lower than $(g_{m2} + g_{mb2})/C_X$ if R_D itself is the output impedance of a PMOS cascode current source. Interestingly, however, the overall transfer function is negligibly affected by this phenomenon. This can be better seen by an example.

Example 6.8

Consider the cascode stage shown in Fig. 6.26(a), where the load resistor is replaced by an ideal

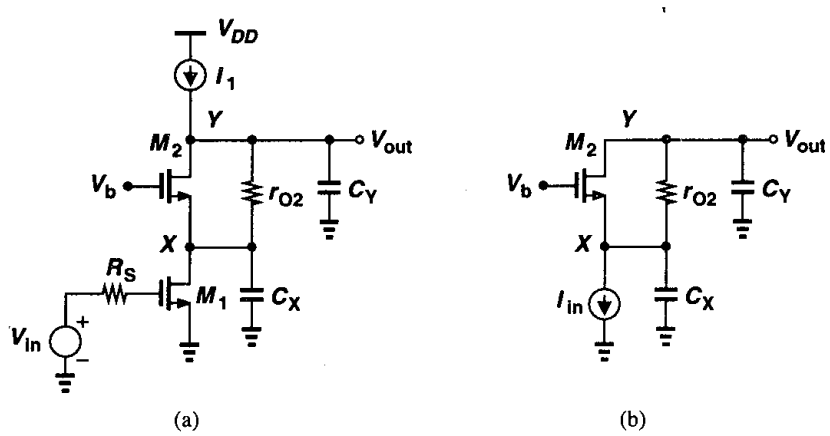


Figure 6.26 Simplified model of a cascode stage.

current source. Neglecting the capacitances associated with M_1, representing V_{in} and M_1 by a Norton equivalent as in Fig. 6.26(b), and assuming $\gamma = 0$, compute the transfer function.

Solution

Since the current through C_X is equal to $-V_{out}C_Y s - I_{in}$, we have $V_X = -(V_{out}C_Y s + I_{in})/(C_X s)$, and the small-signal drain current of M_2 is $-g_{m2}(-V_{out}C_Y s - I_{in})/(C_X s)$. The current through r_{O2}

is then equal to $-V_{out}C_Y s - g_{m2}(V_{out}C_Y s + I_{in})/(C_X s)$. Noting that V_X plus the voltage drop across r_{O2} is equal to V_{out}, we write

$$-r_{O2}\left[(V_{out}C_Y s + I_{in})\frac{g_{m2}}{C_X s} + V_{out}C_Y s\right] - (V_{out}C_Y s + I_{in})\frac{1}{C_X s} = V_{out}. \tag{6.67}$$

That is,

$$\frac{V_{out}}{I_{in}} = -\frac{g_{m2}r_{O2}+1}{C_X s}\cdot\frac{1}{1+(1+g_{m2}r_{O2})\dfrac{C_Y}{C_X}+C_Y r_{O2}s}, \tag{6.68}$$

which, for $g_{m2}r_{O2} \gg 1$ and $g_{m2}r_{O2}C_Y/C_X \gg 1$ (i.e., $C_Y > C_X$), reduces to

$$\frac{V_{out}}{I_{in}} \approx -\frac{g_{m2}}{C_X s}\frac{1}{\dfrac{C_Y}{C_X}g_{m2}+C_Y s}. \tag{6.69}$$

and hence

$$\frac{V_{out}}{V_{in}} = -\frac{g_{m1}g_{m2}}{C_Y C_X s}\frac{1}{g_{m2}/C_X + s}. \tag{6.70}$$

The magnitude of the pole at node X is still given by g_{m2}/C_X. This is because at high frequencies (as we approach this pole) C_Y shunts the output node, dropping the gain and suppressing the Miller effect of r_{O2}.

If a cascode structure is used as a current source, then the variation of its output impedance with frequency is of interest. Neglecting C_{GD1} and C_Y in Fig. 6.26(a), we have

$$Z_{out} = (1 + g_{m2}r_{O2})Z_X + r_{O2}, \tag{6.71}$$

where $Z_X = r_{O1}||(C_X s)^{-1}$. Thus, Z_{out} contains a pole at $(r_{O1}C_X)^{-1}$ and falls at frequencies higher than this value.

6.6 Differential Pair

The versatility of differential pairs and their extensive use in analog systems motivate us to characterize their frequency response for both differential and common-mode signals.

Consider the simple differential pair shown in Fig. 6.27(a), with the differential half circuit and the common-mode equivalent circuit depicted in Figs. 6.27(b) and (c), respectively. For differential signals, the response is identical to that of a common-source stage, exhibiting Miller multiplication of C_{GD}. Note that since $+V_{in2}/2$ and $-V_{in2}/2$ are multiplied by the same transfer function, the number of poles in V_{out}/V_{in} is equal to that of each path (rather than the sum of the number of the poles in the two paths).

For common-mode signals, the total capacitance at node P in Fig. 6.27(c) determines the high-frequency gain. Arising from C_{GD3}, C_{DB3}, C_{SB1}, and C_{SB2}, this capacitance can be

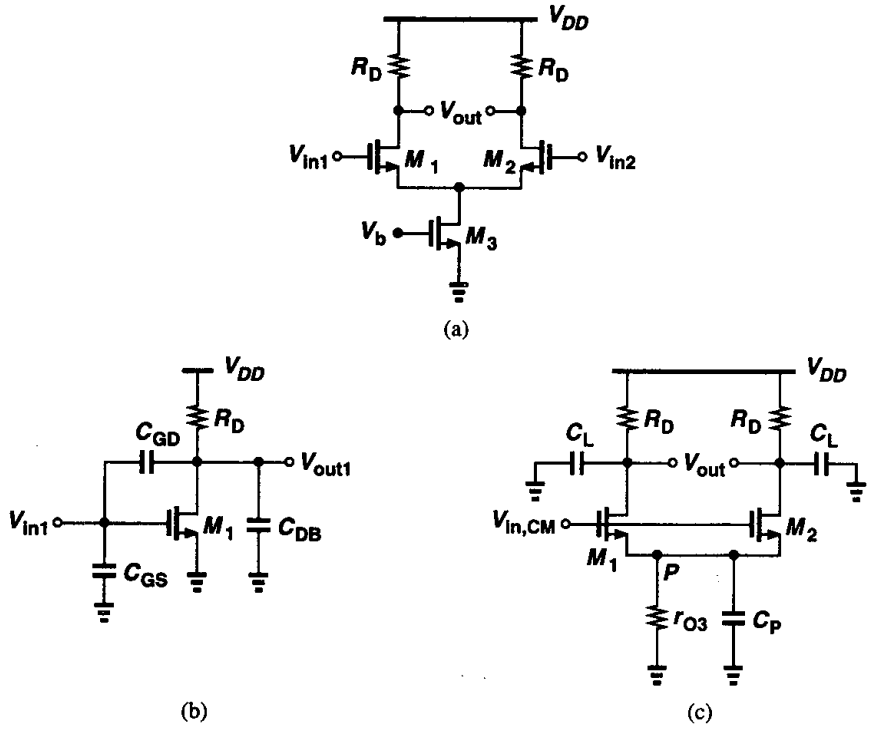


Figure 6.27 (a) Differential pair, (b) half-circuit equivalent, (c) equivalent circuit for common-mode inputs.

quite substantial if M_1-M_3 are wide transistors. For example, limited voltage headroom often necessitates that W_3 be so large that M_3 does not require a large drain-source voltage for operating in the saturation region. If only the mismatch between M_1 and M_2 is considered, the high-frequency common-mode gain can be calculated with the aid of Eq. (4.43). We replace r_{O3} with $r_{O3} \| [1/(C_P s)]$ and R_D by $R_D \| [1/(C_L s)]$, where C_L denotes the total capacitance seen at each output node. Thus,

$$A_{v,CM} = -\frac{\Delta g_m \left[R_D \left\| \left(\frac{1}{C_L s} \right) \right. \right]}{(g_{m1} + g_{m2}) \left[r_{O3} \left\| \left(\frac{1}{C_P s} \right) \right. \right] + 1}, \tag{6.72}$$

where other capacitances in the circuit are neglected.

This result suggests that, if the output pole is much farther from the origin than is the pole at node P, the common-mode rejection of the circuit degrades considerably at high frequencies. For example, as illustrated in Fig. 6.28, if the supply voltage contains high-frequency noise and the circuit exhibits mismatches, the resulting common-mode disturbance at node P leads to a differential noise component at the output.

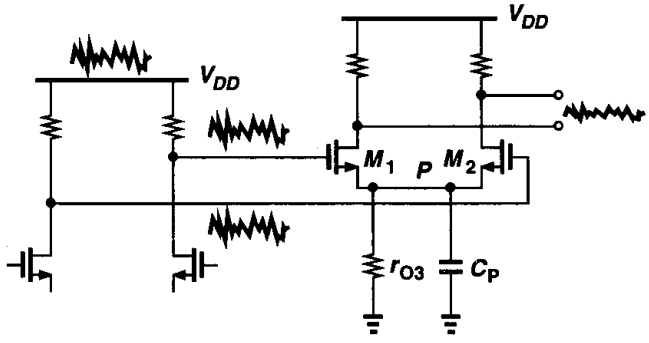


Figure 6.28 Effect of high-frequency supply noise in differential pairs.

We should emphasize that the circuit of Fig. 6.27(a) suffers from a trade-off between voltage headroom and $CMRR$. To minimize the headroom consumed by M_3, its width is maximized, introducing substantial capacitance at the sources of M_1 and M_2 and degrading the high-frequency $CMRR$. The issue becomes more serious at low supply voltages.

We now study the frequency response of differential pairs with high-impedance loads. Shown in Fig. 6.29(a) is a fully differential implementation. As with the topology of Fig. 6.27, this circuit can be analyzed for differential and common-mode signals separately. Note that here C_L includes the drain junction capacitance and the gate-drain overlap capacitance of each PMOS transistor as well. Also, as depicted in Fig. 6.29(b) for differential output signals, C_{GD3} and C_{GD4} conduct equal and opposite currents to node G, making this node an ac ground. (In practice, node G is nonetheless bypassed to ground by means of a capacitor.)

The differential half circuit is depicted in Fig. 6.29(c), with the output resistance of M_1 and M_3 shown explicitly. This topology implies that Eq. (6.23) can be applied to this circuit if R_L is replaced by $r_{O1} \| r_{O3}$. In practice, the relatively high value of this resistance makes the output pole, given by $[(r_{O1} \| r_{O3})C_L]^{-1}$, the "dominant" pole. We return to this observation in Chapter 10. The common-mode behavior of the circuit is similar to that of Fig. 6.27(c).

Let us now consider a differential pair with active current mirror (Fig. 6.30). How many poles does this circuit have? In contrast to the fully differential configuration of Fig. 6.29(a), this topology contains two signal paths with *different* transfer functions. The path consisting of M_3 and M_4 includes a pole at node E, approximately given by g_{m3}/C_E, where C_E denotes the total capacitance at E to ground. This capacitance arises from C_{GS3}, C_{GS4}, C_{DB3}, C_{DB1}, and the Miller effect of C_{GD1} and C_{GD4}. Even if only C_{GS3} and C_{GS4} are considered, the severe trade-off between g_m and C_{GS} of PMOS devices results in a pole that greatly impacts

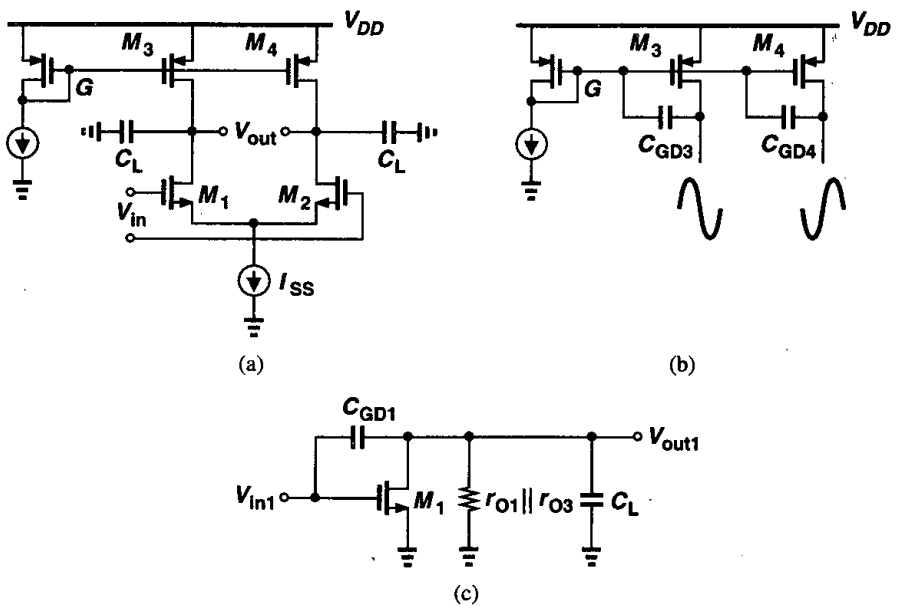


Figure 6.29 (a) Differential pair with current-source loads, (b) effect of differential swings at node G, (c) half-circuit equivalent.

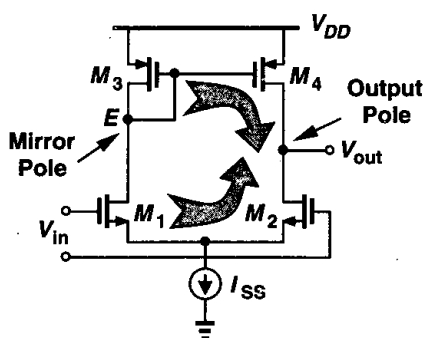


Figure 6.30 High-frequency behavior of differential pair with active current mirror.

the performance of the circuit. The pole associated with node E is called a "mirror pole." Note that, as with the circuit of Fig. 6.29(a), both signal paths shown in Fig. 6.30 contain a pole at the output node.

In order to estimate the frequency response of the differential pair with active current mirror, we construct the simplified model depicted in Fig. 6.31(a), where all other capacitances are neglected. Replacing V_{in}, M_1, and M_2 by a Thevenin equivalent, we arrive at the circuit of Fig. 6.31(b), where, from the analysis of Fig. 5.26, $V_X = g_{mN} r_{ON} V_{in}$ and $R_X = 2r_{ON}$. Here, the subscripts P and N refer to PMOS and NMOS devices, respectively,

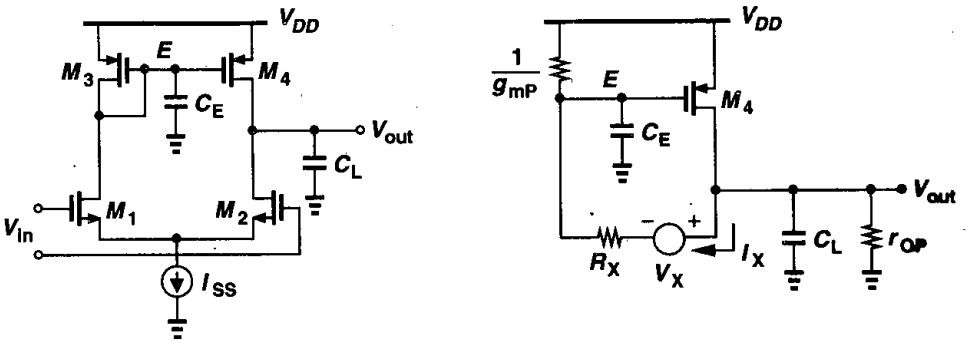


Figure 6.31 (a) Simplified high-frequency model of differential pair with active current mirror, (b) circuit of (a) with a Thevenin equivalent.

and we have assumed $1/g_{mP} \ll r_{OP}$. The small-signal voltage at E is equal to

$$V_E = (V_{out} - V_X)\frac{\dfrac{1}{C_E s + g_{mP}}}{\dfrac{1}{C_E s + g_{mP}} + R_X}, \tag{6.73}$$

and the small-signal drain current of M_4 is $g_{m4}V_E$. Noting that $-g_{m4}V_E - I_X = V_{out}(C_L s + r_{OP}^{-1})$, we have

$$\frac{V_{out}}{V_{in}} = \frac{g_{mN}r_{ON}(2g_{mP} + C_E s)}{2r_{OP}r_{ON}C_E C_L s^2 + [(2r_{ON} + r_{OP})C_E + r_{OP}(1 + 2g_{mP}r_{ON})C_L]s + 2g_{mP}(r_{ON} + r_{OP})}. \tag{6.74}$$

Since the mirror pole is typically quite higher in magnitude than the output pole, we can utilize the results of Eq. (6.25) to write

$$\omega_{p1} \approx \frac{2g_{mP}(r_{ON} + r_{OP})}{(2r_{ON} + r_{OP})C_E + r_{OP}(1 + 2g_{mP}r_{ON})C_L}. \tag{6.75}$$

Neglecting the first term in the denominator and assuming $2g_{mP}r_{ON} \gg 1$, we have

$$\omega_{p1} \approx \frac{1}{(r_{ON} \| r_{OP})C_L}, \tag{6.76}$$

an expected result. The second pole is then given by

$$\omega_{p2} \approx \frac{g_{mP}}{C_E}, \tag{6.77}$$

which is also expected.

An interesting point revealed by Eq. (6.74) is a zero with a magnitude of $2g_{mP}/C_E$ in the left half plane. The appearance of such a zero can be understood by noting that the circuit consists of a "slow path" (M_1, M_3, and M_4) in parallel with a "fast path" (M_1 and M_2). Representing the two by $A_0/[(1 + s/\omega_{p1})(1 + s/\omega_{p2})]$ and $A_0/(1 + s/\omega_{p1})$, respectively, we have

$$\frac{V_{out}}{V_{in}} = \frac{A_0}{1 + s/\omega_{p1}} \left(\frac{1}{1 + s/\omega_{p2}} + 1 \right) \tag{6.78}$$

$$= \frac{A_0(2 + s/\omega_{p2})}{(1 + s/\omega_{p1})(1 + s/\omega_{p2})}. \tag{6.79}$$

That is, the system exibits a zero at $2\omega_{p2}$. The zero can also be obtained by the method of Fig. 6.15 (Problem 6.15).

Comparing the circuits of Figs. 6.29(a) and 6.30, we conclude that the former entails no mirror pole, another advantage of fully differential circuits over single-ended topologies.

Example 6.9

Not all fully differential circuits are free from mirror poles. Fig. 6.32(a) illustrates an example, where current mirrors M_3-M_5 and M_4-M_6 "fold" the signal current. Estimate the low-frequency gain and the transfer function of this circuit.

Solution

Neglecting channel-length modulation and using the differential half-circuit shown in Fig. 6.32(b), we observe that M_5 multiplies the drain current of M_3 by K, yielding an overall low-frequency voltage gain $A_v = g_{m1} K R_D$.

To obtain the transfer function, we utilize the equivalent circuit depicted in Fig. 6.32(c), including a source resistance R_S for completeness. To simplify calculations, we assume $R_D C_L$ is relatively small so that the Miller multiplication of C_{GD5} can be approximated as $C_{GD5}(1 + g_{m5} R_D)$. The circuit thus reduces to that in Fig. 6.32(d), where $C_X \approx C_{GS3} + C_{GS5} + C_{DB3} + C_{GD5}(1 + g_{m5} R_D) + C_{DB1}$. The overall transfer function is then equal to V_X/V_{in1} multiplied by V_{out1}/V_X. The former is readily obtained from (6.23) by replacing R_D with $1/g_{m3}$ and C_{DB} with C_X, while the latter is

$$\frac{V_{out1}}{V_X}(s) = -g_{m5} R_D \frac{1}{1 + R_D C_L s}. \tag{6.80}$$

Note that we have neglected the zero due to C_{GD5}.

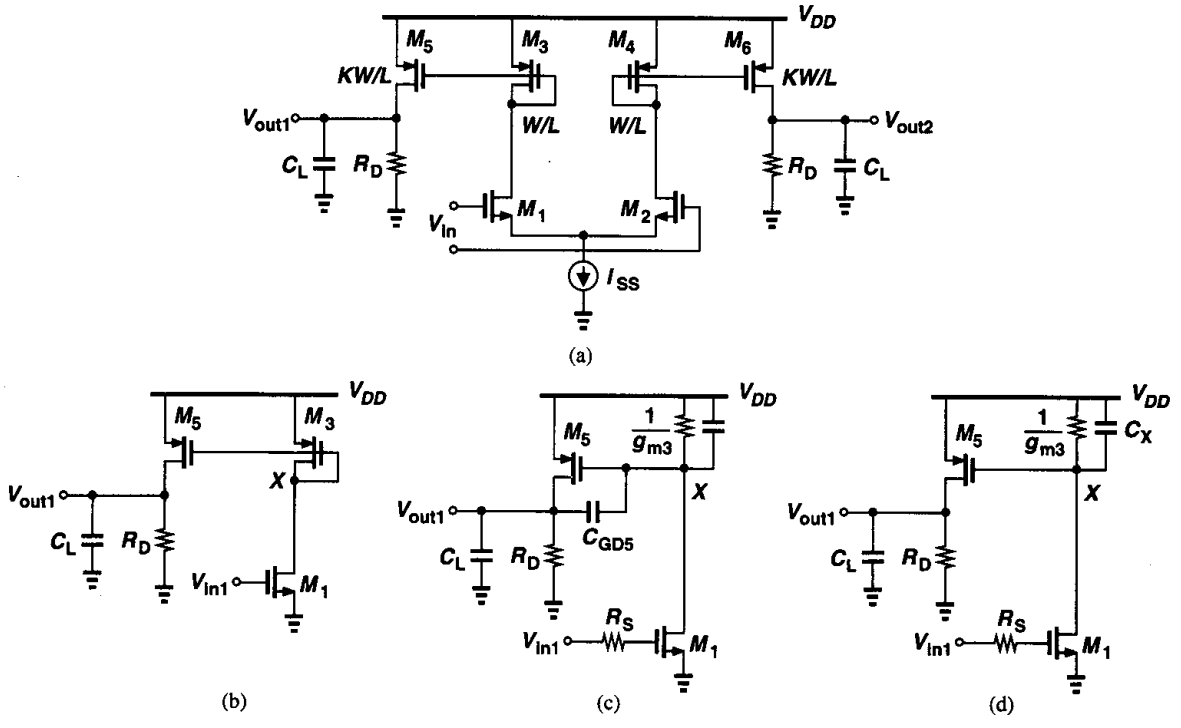


Figure 6.32

Appendix A: Dual of Miller's Theorem

In the Miller's theorem (Fig. 6.1), we readily observe that $Z_1 + Z_2 = Z$. This is no coincidence and it has interesting implications.

Redrawing Fig. 6.1 as shown in Fig. 6.33(a), we surmise that since the point between Z_1 and Z_2 can be grounded, then if we "walk" from X towards Y along the impedance Z, the local potential drops to zero at some intermediate point [Fig. 6.33(b)]. Indeed, for $V_P = 0$, we have

$$\frac{Z_a}{Z_a + Z_b}(V_Y - V_X) + V_X = 0, \tag{6.81}$$

and, since $Z_a + Z_b = Z$,

$$Z_a = \frac{Z}{1 - V_Y/V_X}. \tag{6.82}$$

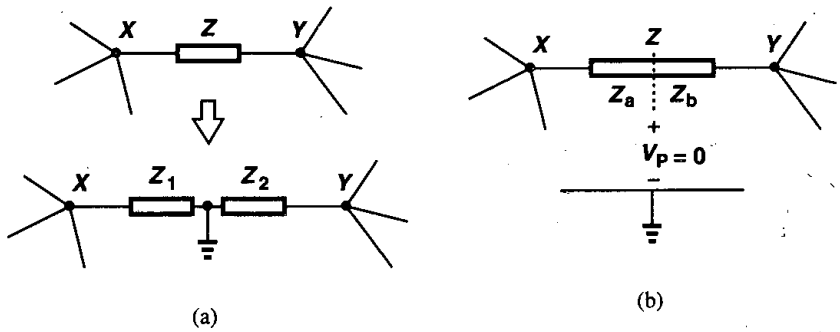


Figure 6.33 Illustration of Miller's theorem identifying a local zero potential along Z.

Similarly,

$$Z_b = \frac{Z}{1 - V_X/V_Y}. \tag{6.83}$$

In other words, $Z_1(= Z_a)$ and $Z_2(= Z_b)$ are such decompositions of Z that provide an intermediate node having a zero potential. For example, since in the common-source stage of Fig. 6.10 V_X and V_Y have opposite polarities, the potential falls to zero at some point "inside" C_{GD}.

The above observation explains the difficulty with the transformation depicted in Fig. 6.3. Drawing Fig. 6.33(b) for this case as in Fig. 6.34(a), we recognize that the circuit is

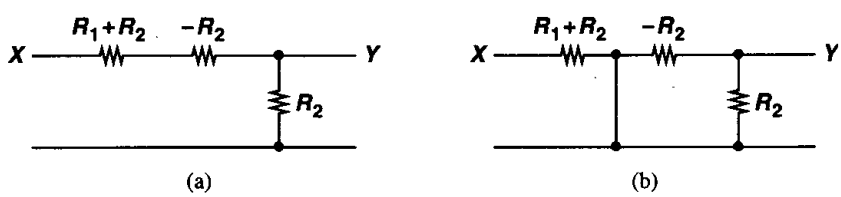


Figure 6.34 Resistive divider with decomposition of R_1.

still valid before node P is grounded because the current through $R_1 + R_2$ must equal that through $-R_2$. However, if, as shown in Fig. 6.34(b), node P is tied to ground, then the only current path between X and Y vanishes.

The concept of a zero local potential along the floating impedance Z also allows us to develop the "dual" of Miller's theorem, i.e., decomposition in terms of admittances and current ratios. Suppose two loops carrying currents I_1 and I_2 share an admittance Y [Fig. 6.35(a)]. Then, if Y is properly decomposed into two *parallel* admittances Y_1 and Y_2, the *current* flowing between the two is zero [Fig. 6.35(b)] and the connection can be broken [Fig. 6.35(c)]. In Fig. 6.35(a), the voltage across Y is equal to $(I_1 - I_2)/Y$ and in Fig. 6.35(c),

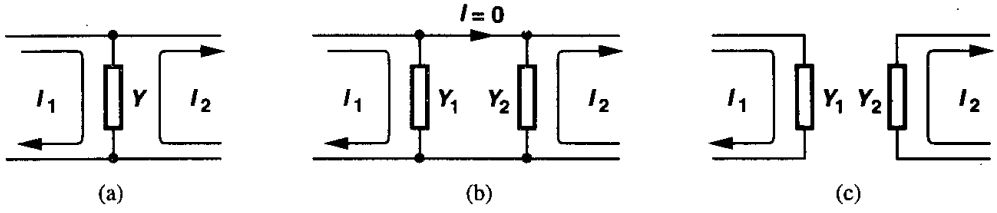


Figure 6.35 (a) Two loops sharing admittance Y, (b) decomposition of Y into Y_1 and Y_2 such that $I = 0$, (c) equivalent circuit.

the voltage across Y_1 is I_1/Y_1. For the two circuits to be equivalent,

$$\frac{I_1 - I_2}{Y} = \frac{I_1}{Y1}, \tag{6.84}$$

and

$$Y_1 = \frac{Y}{1 - I_2/I_1}. \tag{6.85}$$

Note the duality between this expression and $Z_1 = (1 - V_Y/V_X)Z$. We also have

$$Y_2 = \frac{Y}{1 - I_1/I_2}. \tag{6.86}$$

Problems

Unless otherwise stated, in the following problems, use the device data shown in Table 2.1 and assume $V_{DD} = 3$ V where necessary. Also, assume all transistors are in saturation. All device dimensions are effective values and in microns.

6.1. In the circuit of Fig. 6.2(c), suppose the amplifier has a finite output resistance R_{out}.
- **(a)** Explain why the output jumps *up* by ΔV before it begins to go down. This indicates the existence of a zero in the transfer function.
- **(b)** Determine the transfer function and the step response without using Miller's theorem.

6.2. Repeat Problem 6.1 if the amplifier has an output resistance R_{out} and the circuit drives a load capacitance C_L.

6.3. The CS stage of Fig. 6.10 is designed with $(W/L)_1 = 50/0.5$, $R_S = 1\ \text{k}\Omega$ and $R_D = 2\ \text{k}\Omega$. If $I_{D1} = 1$ mA, determine the poles and the zero of the circuit.

6.4. Consider the CS stage of Fig. 6.13, where I_1 is realized by a PMOS device operating in saturation. Assume $(W/L)_1 = 50/0.5$, $I_{D1} = 1$ mA, and $R_S = 1\ \text{k}\Omega$.
- **(a)** Determine the aspect ratio of the PMOS transistor such that the maximum allowable output level is 2.6 V. What is the maximum peak-to-peak swing?
- **(b)** Determine the poles and the zero.

6.5. A source follower employing an NFET with $W/L = 50/0.5$ and a bias current of 1 mA is driven by a source impedance of 10 kΩ. Calculate the equivalent inductance seen at the output.

6.6. Neglecting other capacitances, calculate the input impedance of each circuit shown in Fig. 6.36.

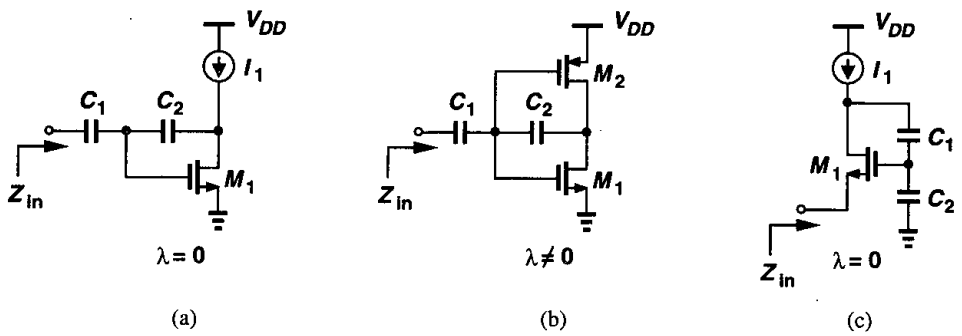


Figure 6.36

6.7. Estimate the poles of each circuit in Fig. 6.37.

6.8. Calculate the input impedance and the transfer function of each circuit in Fig. 6.38.

6.9. Calculate the gain of each circuit in Fig. 6.39 at very low and very high frequencies. Neglect all other capacitances and assume $\lambda = 0$ for circuits (a) and (b) and $\gamma = 0$ for all of the circuits.

6.10. Calculate the gain of each circuit in Fig. 6.40 at very low and very high frequencies. Neglect all other capacitances and assume $\lambda = \gamma = 0$.

6.11. Consider the cascode stage shown in Fig. 6.41. In our analysis of the frequency response of a cascode stage, we assumed that the gate-drain overlap capacitance of M_1 is multiplied by $g_{m1}/(g_{m2} + g_{mb2})$. Recall from Chapter 3, however, that with a high resistance loading the drain of M_2, the resistance seen looking into the source of M_2 can be quite high, suggesting a much higher Miller multiplication factor for C_{GD1}. Explain why C_{GD1} is still multiplied by $1 + g_{m1}/(g_{m2} + g_{mb2})$ if C_L is relatively large.

6.12. Neglecting other capacitances, calculate Z_X in the circuits of Fig. 6.42. Sketch $|Z_X|$ versus frequency.

6.13. The common-gate stage of Fig. 6.23 is designed with $(W/L)_1 = 50/0.5$, $I_{D1} = 1$ mA, $R_D = 2\ \text{k}\Omega$, and $R_S = 1\ \text{k}\Omega$. Assuming $\lambda = 0$, determine the poles and the low-frequency gain. How do these results compare with those obtained in Problem 6.9?

6.14. Suppose in the cascode stage of Fig. 6.25, a resistor R_G appears in series with the gate of M_2. Including only C_{GS2}, neglecting other capacitances, and assuming $\lambda = \gamma = 0$, determine the transfer function.

6.15. Apply the method of Fig. 6.15 to the circuit of Fig. 6.31(b) to determine the zero of the transfer function.

6.16. The circuit of Fig. 6.32(a) is designed with $(W/L)_{1,2} = 50/0.5$ and $(W/L)_{3,4} = 10/0.5$. If $I_{SS} = 100\ \mu\text{A}$, $K = 2$, $C_L = 0$, and R_D is implemented by an NFET having $W/L = 50/0.5$, estimate the poles and zeros of the circuit. Assume the amplifier is driven by an ideal voltage source.

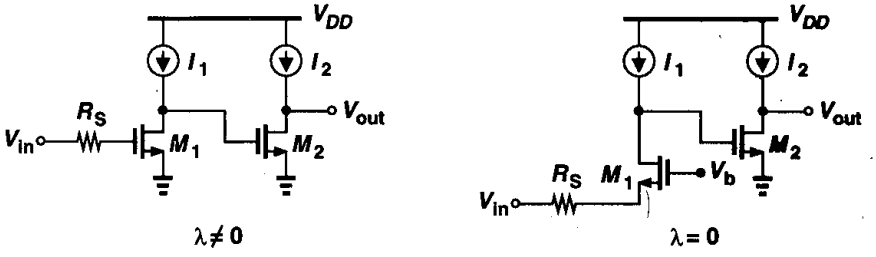


Figure 6.37

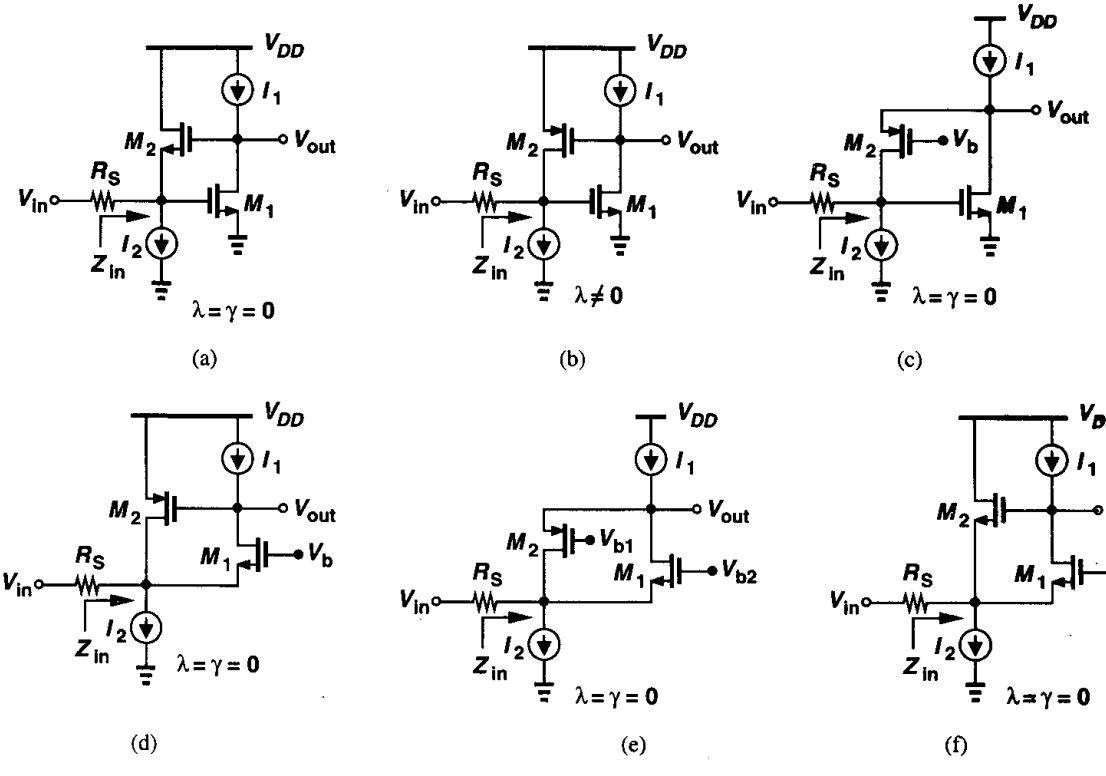


Figure 6.38

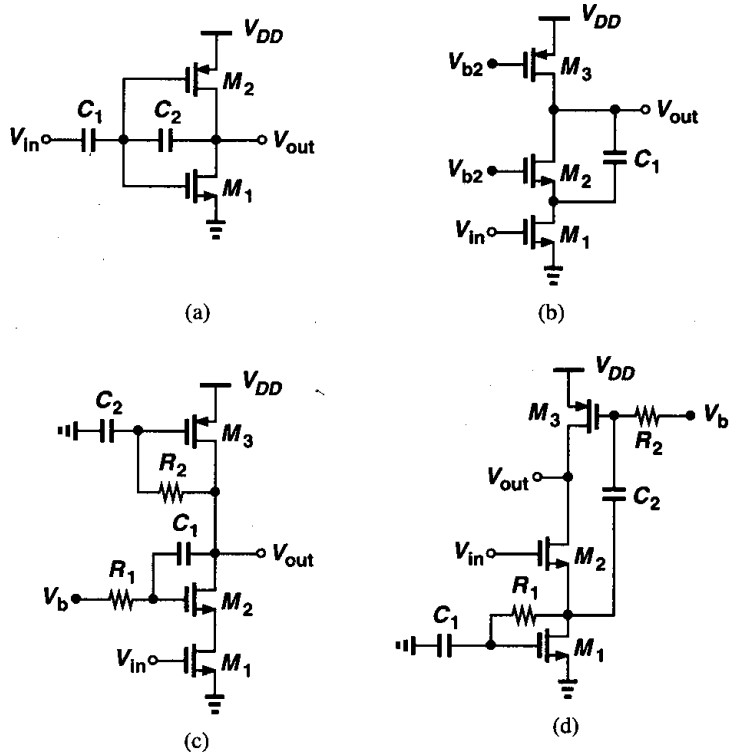


Figure 6.39

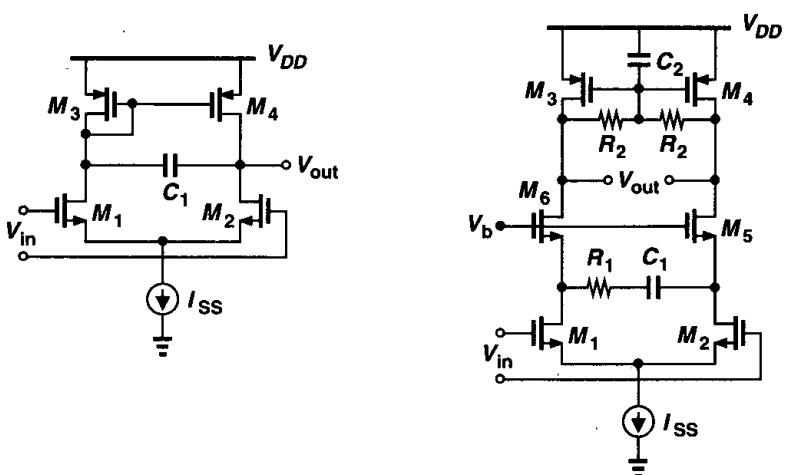


Figure 6.40

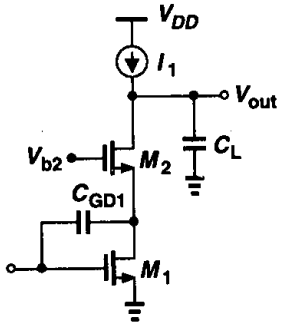


Figure 6.41

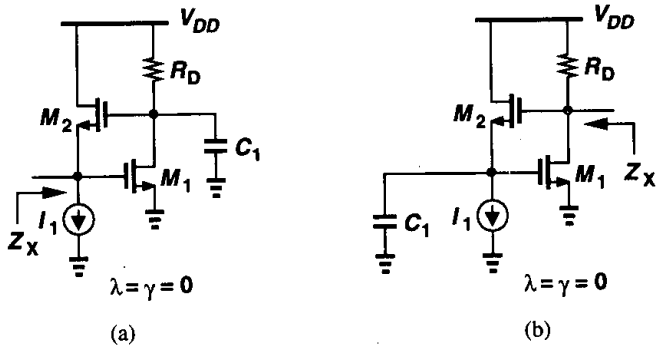


Figure 6.42

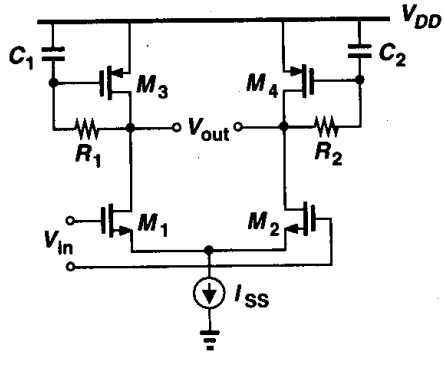


Figure 6.43

6.17. A differential pair driven by an ideal voltage source is required to have a total phase shift of 135° at the frequency where its gain drops to unity.

(a) Explain why a topology in which the load is realized by diode-connected devices or current sources does not satisfy this condition.

(b) Consider the circuit shown in Fig. 6.43. Neglecting other capacitances, determine the transfer function. Explain under what conditions the load exhibits an inductive behavior. Can this circuit provide a total phase shift of 135° at the frequency where its gain drops to unity?

References

1. P. R. Gray and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*, Third Ed., New York: Wiley, 1993.
2. B. Razavi, *RF Microelectronics*, Upper Saddle River, NJ: Prentice Hall, 1998.