Chapter 7

Noise

Noise limits the minimum signal level that a circuit can process with acceptable quality. Today's analog designers constantly deal with the problem of noise because it trades with power dissipation, speed, and linearity.

In this chapter, we describe the phenomenon of noise and its effect on analog circuits. The objective is to provide sufficient understanding of the problem so that further developments of analog circuits in the following chapters take noise into account as readily as other circuit parameters such as gain, input and output impedance, etc. Seemingly a complex subject, noise is introduced at this early stage so as to accompany the reader for the remainder of the book and become more intuitive through various examples.

Following a general description of noise characteristics in the frequency and time domains, we introduce thermal, shot, and flicker noise. Next, we consider methods of representing noise in circuits. Finally, we describe the effect of noise in single-stage and differential amplifiers along with trade-offs with other performance parameters.

7.1 Statistical Characteristics of Noise

Noise is a random process. For our purposes in this book, this statement means the value of noise cannot be predicted at any time even if the past values are known. Compare the output of a sinewave generator with that of a microphone picking up the sound of water flow in a river (Fig. 7.1). While the value of $x_1(t)$ at $t = t_1$ can be predicted from the observed waveform, the value of $x_2(t)$ at $t = t_2$ cannot. This is the principal difference between deterministic and random phenomena.

If the instantaneous value of noise in the time domain cannot be predicted, how can we incorporate noise in circuit analysis? This is accomplished by observing the noise for a long time and using the measured results to construct a "statistical model" for the noise. While the instantaneous *amplitude* of noise cannot be predicted, a statistical model provides knowledge about some other important properties of the noise that prove useful and adequate in circuit analysis.

Which properties of noise *can* be predicted? In many cases, the average power of noise is predictable. For example, if a microphone picking up the sound of a river is brought

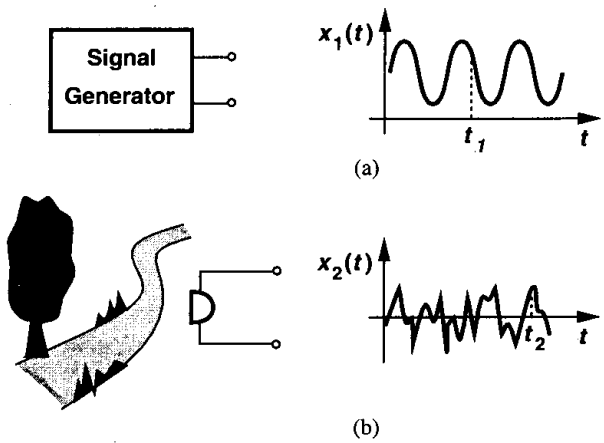


Figure 7.1 Output of a generator and the sound of a river.

closer to the river, the resulting electrical signal displays, on the average, larger excursions and hence higher power (Fig. 7.2). The reader may wonder if a random process can be so random that even its average power is unpredictable. Such processes do exist, but we are fortunate that most sources of noise in circuits exhibit a constant average power.

The concept of average power proves essential in our analysis and must be defined carefully. Recall from basic circuit theory that the average power delivered by a periodic

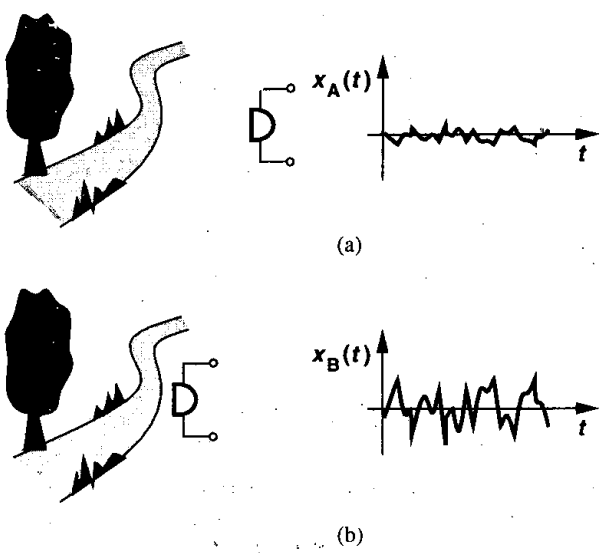


Figure 7.2 Illustration of the average power of a random signal.

voltage $v(t)$ to a load resistance R_L is given by

$$P_{av} = \frac{1}{T}\int_{-T/2}^{+T/2} \frac{v^2(t)}{R_L} dt, \tag{7.1}$$

where T denotes the period.[1] This quantity can be visualized as the average heat produced in R_L by $v(t)$.

How do we define P_{av} for a random signal? In the example of Fig. 7.2, we expect that $x_B(t)$ generates more heat than $x_A(t)$ if the microphone drives a resistive load. However, since the signals are not periodic, the measurement must be carried out over a long time:

$$P_{av} = \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{+T/2} \frac{x^2(t)}{R_L} dt, \tag{7.2}$$

where $x(t)$ is a voltage quantity. Figure 7.3 illustrates the operation on $x_A(t)$ and $x_B(t)$; each signal is squared, the area under the resulting waveform is calculated for a long time T, and the average power is obtained by normalizing the area to T.[2]

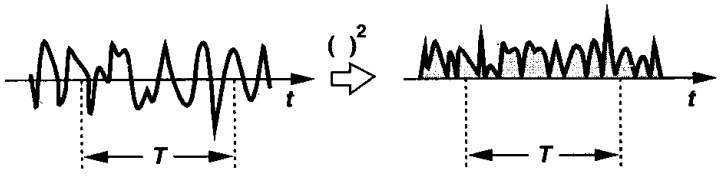


Figure 7.3 Average noise power.

To simplify calculations, we write the definition of P_{av} as

$$P_{av} = \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{+T/2} x^2(t) dt, \tag{7.3}$$

where P_{av} is expressed in V^2 rather than W. The idea is that if we know P_{av} from (7.3), then the actual power delivered to a load R_L can be readily calculated as P_{av}/R_L. In analogy with deterministic signals, we can also define a root-mean-square (rms) voltage for noise as $\sqrt{P_{av}}$ where P_{av} is given by (7.3).

7.1.1 Noise Spectrum

The concept of average power becomes more versatile if defined with regard to the *frequency content* of noise. The noise made by a group of men contains weaker high-frequency components than that made by a group of women, a difference observable from the "spectrum"

[1] To be more rigorous, $v^2(t)$ should be replaced by $v(t)\cdot v^*(t)$, where $v^*(t)$ is the complex conjugate waveform.

[2] Strictly speaking, this definition holds only for "stationary" processes [1].

of each type of noise. Also called the "power spectral density" (PSD), the spectrum shows how much power the signal carries at each frequency. More specifically, the PSD, $S_x(f)$, of a noise waveform $x(t)$ is defined as the average power carried by $x(t)$ in a one-hertz bandwidth around f. That is, as illustrated in Fig. 7.4(a), we apply $x(t)$ to a bandpass filter

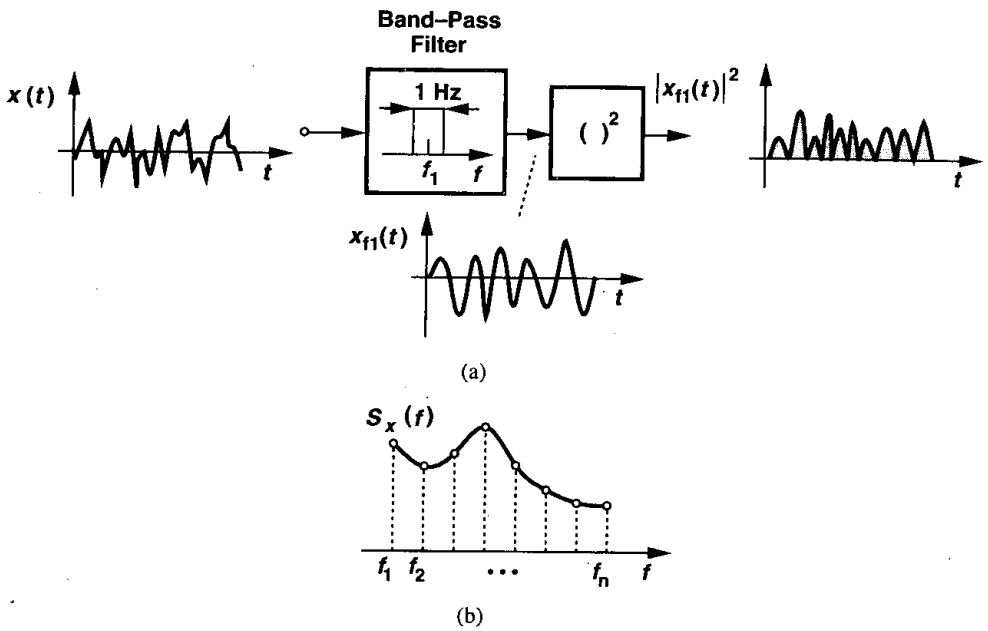


Figure 7.4 Calculation of noise spectrum.

with center frequency f_1 and 1-Hz bandwidth, square the output, and calculate the average over a long time to obtain $S_X(f_1)$. Repeating the procedure with bandpass filters having different center frequencies, we arrive at the overall shape of $S_X(f)$ [Fig. 7.4(b)].[3] While it is possible that the PSD of a random process is random itself, most of the noise sources of interest to us exhibit a predictable spectrum.

As with the definition of P_{av} in (7.3), it is customary to eliminate R_L from $S_X(f)$. Thus, since each value on the plot in Fig. 7.4(b) is measured for a 1-Hz bandwidth, $S_X(f)$ is expressed in V^2/Hz rather than W/Hz. It is also common to take the square root of $S_X(f)$, expressing the result in $V/\sqrt{Hz}$. For example, we say the input noise voltage of an amplifier at 100 MHz is equal to 3 $nV/\sqrt{Hz}$, simply to mean that the average power in a 1-Hz bandwidth at 100 MHz is equal to $(3 \times 10^{-9})^2$ V^2.

An example of a common type of noise PSD is the "white spectrum," also called white noise. Shown in Fig. 7.5, such a PSD displays the same value at all frequencies (similar

[3] In signal processing theory, the PSD is defined as the Fourier transform of the autocorrelation function of the noise. The two definitions are equivalent in most cases of interest to us.

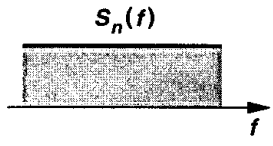


Figure 7.5 White spectrum.

to white light). Strictly speaking, we note that white noise does not exist because the total area under the power spectral density, i.e., the total power carried by the noise, is infinite. In practice, however, any noise spectrum that is flat *in the band of interest* is usually called white.

The PSD is a powerful tool in analyzing the effect of noise in circuits, especially in conjunction with the following theorem.

Theorem If a signal with spectrum $S_X(f)$ is applied to a linear time-invariant system with transfer function $H(s)$, then the output spectrum is given by

$$S_Y(f) = S_X(f)|H(f)|^2, \tag{7.4}$$

where $H(f) = H(s = 2\pi jf)$. The proof can be found in textbooks on signal processing or communications, e.g., [1].

This theorem agrees with our intuition that the spectrum of the signal should be "shaped" by the transfer function of the system (Fig. 7.6). For example, as illustrated in Fig. 7.7,

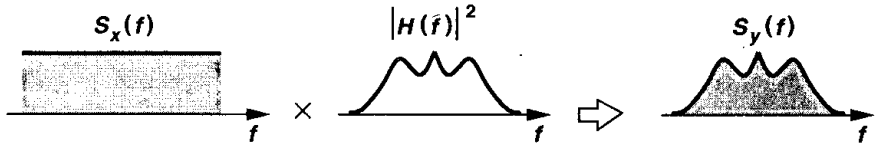


Figure 7.6 Noise shaping by a transfer function.

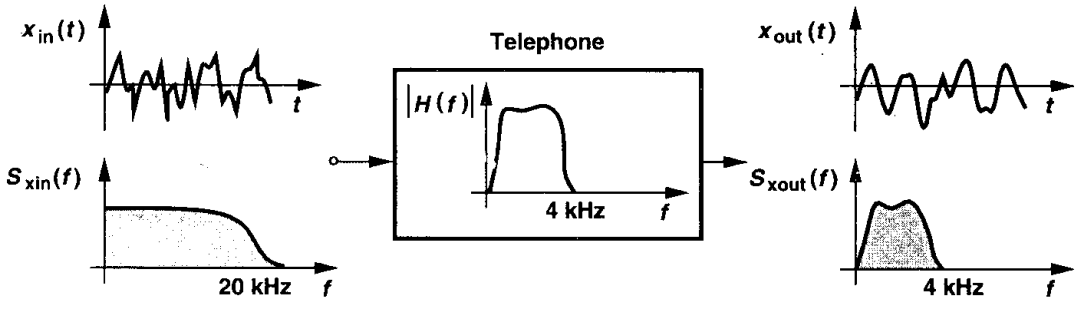


Figure 7.7 Spectral shaping by telephone bandwidth.

since regular telephones have a bandwidth of approximately 4 kHz, they suppress the high-frequency components of the caller's voice. Note that, owing to its limited bandwidth, $x_{out}(t)$ exhibits slower changes than does $x_{in}(t)$.

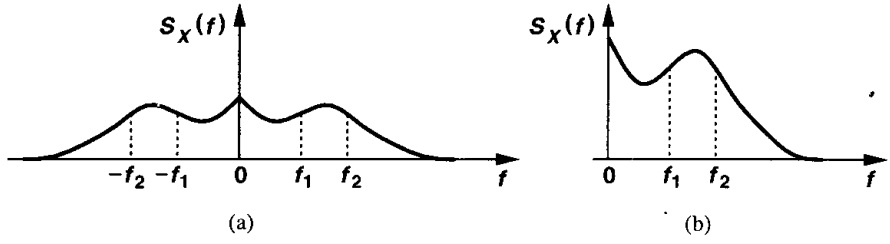


Figure 7.8 (a) Two-sided and (b) one-sided noise spectra.

Since $S_X(f)$ is an even function of f for real $x(t)$ [1], as depicted in Fig. 7.8, the total power carried by $x(t)$ in the frequency range $[f_1\ f_2]$ is equal to

$$P_{f1,f2} = \int_{-f_2}^{-f_1} S_X(f)df + \int_{+f_1}^{+f_2} S_X(f)df \tag{7.5}$$

$$= \int_{+f_1}^{+f_2} 2S_X(f)df. \tag{7.6}$$

In fact, the integral in (7.6) is the quantity measured by a power meter sensing the output of a bandpass filter between f_1 and f_2. That is, the negative-frequency part of the spectrum is folded around the vertical axis and added to the positive-frequency part. We call the representation of Fig. 7.8(a) the "two-sided" spectrum and that of Fig. 7.8(b) the "one-sided" spectrum. For example, the two-sided white spectrum of Fig. 7.5 has the one-sided counterpart shown in Fig. 7.9.

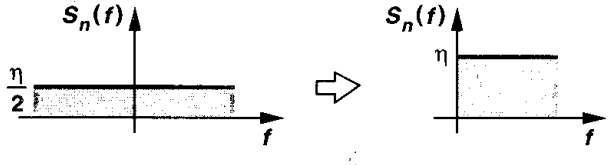


Figure 7.9 Folded white spectrum.

In summary, the spectrum shows the power carried in a small bandwidth at each frequency, revealing how *fast* the waveform is expected to vary in the time domain.

7.1.2 Amplitude Distribution

As mentioned earlier, the instantaneous amplitude of noise is usually unpredictable. However, by observing the noise waveform for a long time, we can construct a "distribution" of

the amplitude, indicating how *often* each value occurs. Also called the "probability density function" (PDF), the distribution of $x(t)$ is defined as

$$p_X(x)dx = \text{probability of } x < X < x + dx, \tag{7.7}$$

where X is the measured value of $x(t)$ at some point in time.

As illustrated in Fig. 7.10, to estimate the distribution, we sample $x(t)$ at many points, construct bins of small width, choose the bin height equal to the number of samples whose value falls between the two edges of the bin, and normalize the bin heights to the total number of samples. Note that the PDF provides no information as to how fast $x(t)$ varies in the time domain. For example, the sound generated by a violin may have the same amplitude distribution as that produced by a drum even though their frequency contents are vastly different.

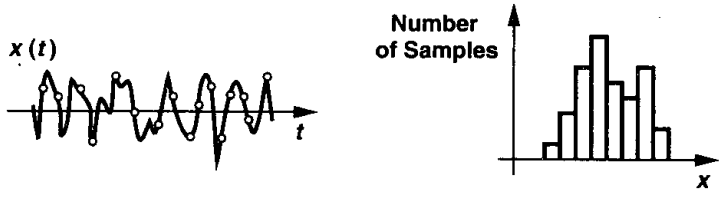


Figure 7.10 Amplitude distribution of noise.

An important example of PDFs is the Gaussian (or normal) distribution. The central limit theorem states that if many independent random processes with arbitrary PDFs are added, the PDF of the sum approaches a Gaussian distribution [1]. It is therefore not surprising that many natural phenomena exhibit Gaussian statistics. For example, since the noise of a resistor results from random "walk" of a very large number of electrons, each having relatively independent statistics, the overall amplitude follows a Gaussian PDF.

In this book, we employ the spectrum and average power of noise to a much greater extent than the amplitude distribution. For completeness, however, we note that the Gaussian PDF is defined as

$$p_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \frac{-(x-m)^2}{2\sigma^2}, \tag{7.8}$$

where σ and m are the standard deviation and mean of the distribution, respectively.

7.1.3 Correlated and Uncorrelated Sources

In analyzing circuits, we often need to add the effect of several sources of noise to obtain the total noise. While for deterministic voltages and currents, we simply use the superposition principle, the procedure is somewhat different for random signals. Since in noise analysis, ultimately the average noise *power* is of interest, we add two noise waveforms and take the

average of the resulting power:

$$P_{av} = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{+T/2} [x_1(t) + x_2(t)]^2 dt \tag{7.9}$$

$$= \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{+T/2} x_1^2(t)dt + \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{+T/2} x_2^2(t)dt$$

$$+ \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{+T/2} 2x_1(t)x_2(t)dt \tag{7.10}$$

$$= P_{av1} + P_{av2} + \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{+T/2} 2x_1(t)x_2(t)dt, \tag{7.11}$$

where P_{av1} and P_{av2} denote the average power of $x_1(t)$ and $x_2(t)$, respectively. Called the "correlation" between $x_1(t)$ and $x_2(t)$,[4] the third term in (7.11) indicates how "similar" these two waveforms are. If generated by independent devices, the noise waveforms are usually "uncorrelated" and the integral in (7.11) vanishes. For example, the noise produced by a resistor has no correlation with that generated by a transistor. In such a case,

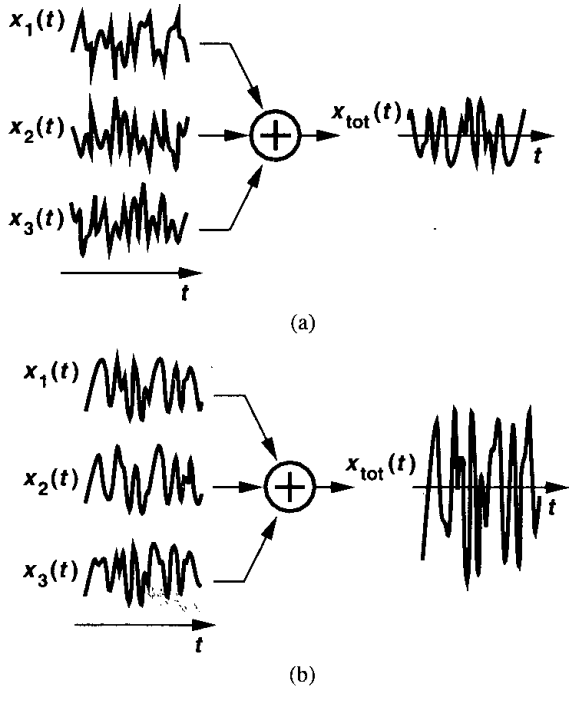


Figure 7.11 (a) Uncorrelated noise and (b) correlated noise generated in a stadium.

[4]This terminology applies only to stationary signals.

$P_{av} = P_{av1} + P_{av2}$. From this observation, we say superposition holds for the *power* of uncorrelated noise sources.

A familiar analogy is that of the spectators in a sports stadium. Before the game begins, many conversations are in progress, generating uncorrelated noise components [Fig. 7.11(a)]. During the game, the spectators applaud (or scream) simultaneously, producing correlated noise at much higher power level [Fig. 7.11(b)].

In most cases studied in this book, the noise sources are uncorrelated. One exception is studied in Section 7.3.

7.2 Types of Noise

Analog signals processed by integrated circuits are corrupted by two different types of noise: device electronic noise and "environmental" noise. The latter refers to (seemingly) random disturbances that a circuit experiences through the supply or ground lines or through the substrate. We focus on device electronic noise here and defer the study of environmental noise to Chapter 18.

7.2.1 Thermal Noise

Resistor Thermal Noise The random motion of electrons in a conductor introduces fluctuations in the voltage measured across the conductor even if the average current is zero. Thus, the spectrum of thermal noise is proportional to the absolute temperature.

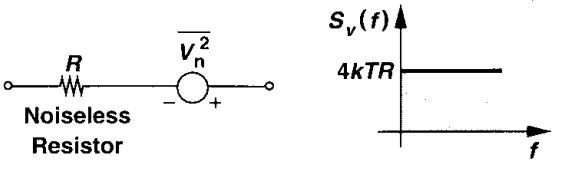


Figure 7.12 Thermal noise of a resistor.

As shown in Fig. 7.12, the thermal noise of a resistor R can be modeled by a series voltage source, with the one-sided spectral density

$$S_v(f) = 4kTR, \quad f \geq 0, \tag{7.12}$$

where $k = 1.38 \times 10^{-23}$ J/K is the Boltzmann constant. Note that $S_v(f)$ is expressed in V^2/Hz. Thus, we write $\overline{V_n^2} = 4kTR$, where the overline indicates averaging.[5] We may even say the noise "voltage" is given by $4kTR$ even though this quantity is in fact the noise voltage squared. For example, a 50-Ω resistor held at $T = 300°K$ exhibits 8.28×10^{-19} V^2/Hz of thermal noise. To convert this number to a more familiar voltage quantity, we take the square root, obtaining 0.91 nV/$\sqrt{\text{Hz}}$. While the square root of hertz may appear strange,

[5] Some books write $\overline{V_n^2} = 4kTR\Delta f$ to emphasize that $4kTR$ is the noise power per unit bandwidth. To simplify the notation, we assume $\Delta f = 1$ Hz, unless otherwise stated.

it is helpful to remember that 0.91 nV/$\sqrt{\text{Hz}}$ has little significance per se and simply means that the power in a 1-Hz bandwidth is equal to $(0.91 \times 10^{-9})^2$ V^2.

The equation $S_v(f) = 4kTR$ suggests that thermal noise is white. In reality, $S_v(f)$ is flat for up to roughly 100 THz, dropping at higher frequencies. For our purposes, the white spectrum is quite accurate.

Since noise is a random quantity, the polarity used for the voltage source in Fig. 7.12 is unimportant. Nevertheless, once a polarity is chosen, it must be retained throughout the analysis of the circuit so as to obtain consistent results.

Example 7.1

Consider the RC circuit shown in Fig. 7.13. Calculate the noise spectrum and the total noise power in V_{out}.

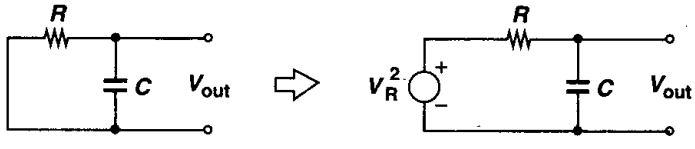


Figure 7.13 Noise generated in a low-pass filter.

Solution

Modeling the noise of R by a series voltage source V_R, we compute the transfer function from V_R to V_{out}:

$$\frac{V_{out}}{V_R}(s) = \frac{1}{RCs + 1}. \tag{7.13}$$

From the theorem in Section 6.1.1, we have

$$S_{out}(f) = S_R(f)\left|\frac{V_{out}}{V_R}(j\omega)\right|^2 \tag{7.14}$$

$$= 4kTR\frac{1}{4\pi^2R^2C^2f^2 + 1}. \tag{7.15}$$

Thus, the white noise spectrum of the resistor is shaped by a low-pass characteristic (Fig. 7.14). To calculate the total noise power at the output, we write

$$P_{n,out} = \int_0^\infty \frac{4kTR}{4\pi^2R^2C^2f^2 + 1}df, \tag{7.16}$$

which, since

$$\int \frac{dx}{x^2 + 1} = \tan^{-1}x, \tag{7.17}$$

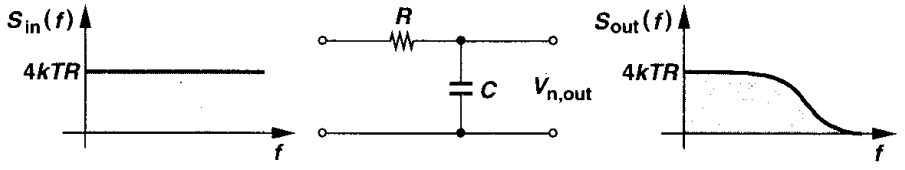


Figure 7.14 Noise spectrum shaping by a low-pass filter.

reduces to

$$P_{n,out} = \frac{2kT}{\pi C} \tan^{-1} u \Big|_{u=0}^{u=\infty} \tag{7.18}$$

$$= \frac{kT}{C}. \tag{7.19}$$

Note that the unit of kT/C is V^2. We may also consider $\sqrt{kT/C}$ as the total rms noise voltage measured at the output. For example, with a 1-pF capacitor, the total noise voltage is equal to 64.3 μV_{rms}.

Equation (7.19) implies that the total noise at the output of the circuit shown in Fig. 7.13 is independent of the value of R. Intuitively, this is because for larger values of R, the associated noise per unit bandwidth increases while the overall bandwidth of the circuit decreases. The fact that kT/C noise can be decreased by only increasing C (if T is fixed) introduces many difficulties in the design of analog circuits (Chapter 12).

The thermal noise of a resistor can be represented by a parallel current source as well (Fig. 7.15). For the representations of Figs. 7.12 and 7.15 to be equivalent, we have $\overline{V_n^2}/R^2 = \overline{I_n^2}$, that is, $\overline{I_n^2} = 4kT/R$. Note that $\overline{I_n^2}$ is expressed in A^2/Hz. Depending on the circuit topology, one model may lead to simpler calculations than the other.

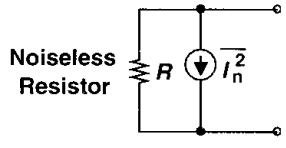
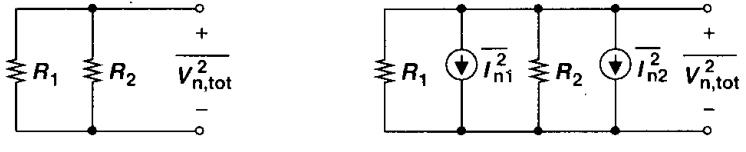


Figure 7.15 Representation of resistor thermal noise by a current source.

Example 7.2

Calculate the equivalent noise voltage of two parallel resistors R_1 and R_2 [Fig. 7.16(a)].



(a) (b)

Figure 7.16

Solution

As shown in Fig. 7.16(b), each resistor exhibits an equivalent noise current with the spectral density $4kT/R$. Since the two noise sources are uncorrelated, we add the *powers*:

$$\overline{I_{n,tot}^2} = \overline{I_{n1}^2} + \overline{I_{n2}^2} \tag{7.20}$$

$$= 4kT\left(\frac{1}{R_1} + \frac{1}{R_2}\right). \tag{7.21}$$

Thus, the equivalent noise voltage is given by

$$\overline{V_{n,tot}^2} = \overline{I_{n,tot}^2}(R_1 \| R_2)^2 \tag{7.22}$$

$$= 4kT(R_1 \| R_2), \tag{7.23}$$

as intuitively expected. Note that our notation assumes a 1-Hz bandwidth.

The dependence of thermal noise (and some other types of noise) upon T suggests that low-temperature operation can decrease the noise in analog circuits. This approach becomes more attractive with the observation that the mobility of charge carriers in MOS devices increases at low temperatures [2].[6] Nonetheless, the required cooling equipment limits the practicality of low-temperature circuits.

MOSFETs MOS transistors also exhibit thermal noise. The most significant source is the noise generated in the channel. It can be proved [3] that for long-channel MOS devices operating in saturation, the channel noise can be modeled by a current source connected between the drain and source terminals (Fig. 7.17) with a spectral density:[7]

$$\overline{I_n^2} = 4kT\gamma g_m. \tag{7.24}$$

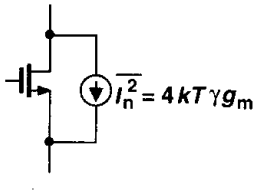


Figure 7.17 Thermal noise of a MOSFET.

The coefficient γ (not to be confused with the body effect coefficient!) is derived to be equal to 2/3 for long-channel transistors and may need to be replaced by a larger value for submicron MOSFETs [4]. For example, γ is about 2.5 in some 0.25-μm MOS devices. It

[6] At extremely low temperatures, the mobility drops due to "carrier freezeout" [2].

[7] The actual equation reads $\overline{I_n^2} = 4kT\gamma g_{ds}$, where g_{ds} is the drain-source conductance with $V_{DS} = 0$, i.e., the same as R_{on}^{-1}. For long-channel devices, g_{ds} with $V_{DS} = 0$ is equal to g_m in saturation.

also varies to some extent with the drain-source voltage. The theoretical determination of γ is still under active research.

Example 7.3

Find the maximum noise voltage that a singie MOSFET can generate.

Solution

As shown in Fig. 7.18, the maximum output noise occurs if the transistor sees only its own output

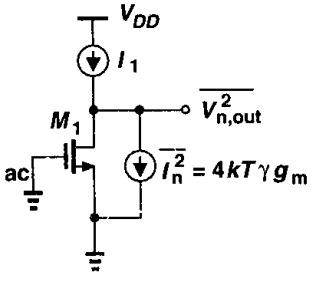


Figure 7.18

impedance as the load, i.e., if the external load is an ideal current source. The output noise voltage is then given by

$$\overline{V_n^2} = \overline{I_n^2} r_O^2 \tag{7.25}$$

$$= 4kT\left(\frac{2}{3}g_m\right) r_O^2. \tag{7.26}$$

Equation (7.26) suggests that the noise *current* of a MOS transistor decreases if the transconductance drops. For example, if the transistor operates as a constant current source, it is desirable to minimize its transconductance.

Another important conclusion is that the noise measured at the output of the circuit does not depend on where the input terminal is because for output noise calculation, the input is set to zero.[8] For example, the circuit of Fig. 7.18 may be a common-source or a common-gate stage, exhibiting the same output noise.

The ohmic sections of a MOSFET also contribute thermal noise. As conceptually illustrated in the top view of Fig. 7.19(a), the gate, source, and drain materials exhibit finite resistivity, thereby introducing noise. For a relatively wide transistor, the source and drain resistance is typically negligible whereas the gate distributed resistance may become noticeable.

In the noise model of Fig. 7.19(b), a lumped resistor R_1 represents the distributed gate resistance. Viewing the overall transistor as the distributed structure shown in Fig. 7.19(c),

[8] Of course, if the input voltage or current source has an output impedance that generates noise, this statement must be interpreted carefully.

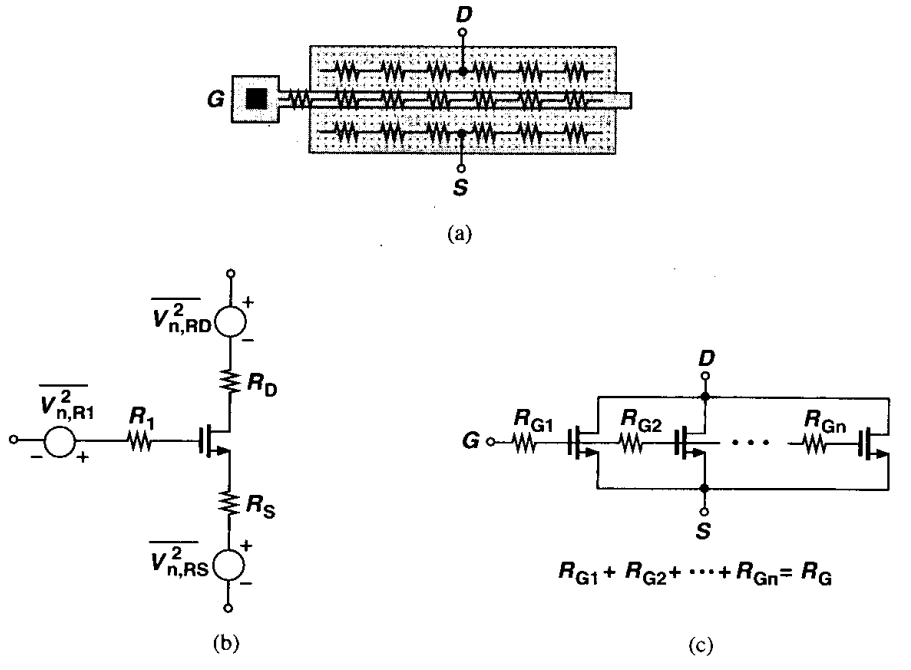


Figure 7.19 L ayout of a MOSFET indicating the terminal resistances, (b) circuit model, (c) distributed gate resistance.

we observe that the unit transistors near the left end see the noise of only a fraction of R_G whereas those near the right end see the noise of most of R_G. We therefore expect the lumped resistor in the noise model to be *less* than R_G. In fact, it can be proved that $R_1 = R_G/3$ (Problem 7.3).

While the thermal noise generated in the channel is controlled by only the transconductance of the device, the effect of R_G can be reduced by proper layout. Shown in Fig. 7.20 are two examples. In Fig. 7.20(a), the gate is contacted on both ends and in Fig. 7.20(b), the

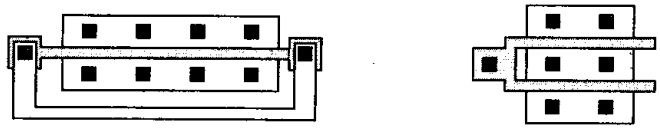


Figure 7.20 Reduction of gate resistance by (a) adding contacts to both sides or (b) folding.

device is folded (Chapter 2), each technique reducing R_G by a factor of 4. We will hereafter neglect the thermal noise due to the ohmic sections of MOS devices.

Example 7.4

Find the maximum thermal noise voltage that the gate resistance of a single MOSFET can generate.

Solution

If the total distributed gate resistance is R_G, then from Fig. 7.18, the output noise voltage due to R_G is given by

$$\overline{V_{n,out}^2} = 4kT\frac{R_G}{3}(g_m r_O)^2. \tag{7.27}$$

7.2.2 Flicker Noise

The interface between the gate oxide and the silicon substrate in a MOSFET entails an interesting phenomenon. Since the silicon crystal reaches an end at this interface, many "dangling" bonds appear, giving rise to extra energy states (Fig.7.21). As charge carriers move at the interface, some are randomly trapped and later released by such energy states, introducing "flicker" noise in the drain current. In addition to trapping, several other mechanisms are believed to generate flicker noise [3].

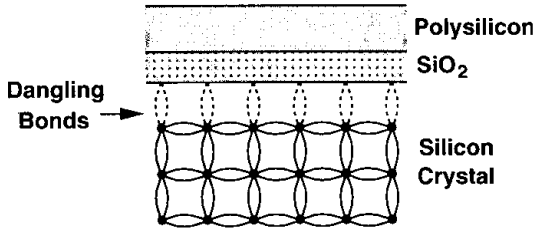


Figure 7.21 Dangling bonds at the oxide-silicon interface.

Unlike thermal noise, the average power of flicker noise cannot be predicted easily. Depending on the "cleanness" of the oxide-silicon interface, flicker noise may assume considerably different values and as such varies from one CMOS technology to another. The flicker noise is more easily modeled as a voltage source in series with the gate and roughly given by

$$\overline{V_n^2} = \frac{K}{C_{ox}WL}\cdot\frac{1}{f}, \tag{7.28}$$

where K is a process-dependent constant on the order of 10^{-25} V^2F. Note that our notation assumes a bandwidth of 1 Hz. Interestingly, as shown in Fig. 7.22, the noise spectral density is inversely proportional to the frequency. For example, the trap-and-release phenomenon associated with the dangling bonds occurs at low frequencies more often. For this reason, flicker noise is also called $1/f$ noise. Note that (7.28) does not depend on the bias current or the temperature. This is only an approximation and in reality, the flicker noise equation is somewhat more complex [3].

The inverse dependence of (7.28) on WL suggests that to decrease $1/f$ noise, the device *area* must be increased. It is therefore not surprising to see devices having areas of several

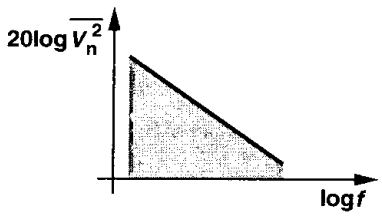


Figure 7.22 Flicker noise spectrum.

thousand square microns in low-noise applications. It is also believed that PMOS devices exhibit less $1/f$ noise than NMOS transistors because the former carry the holes in a "buried channel," i.e., at some distance from the oxide-silicon interface. Nonetheless, this difference between PMOS and NMOS transistors is not consistently observed [3].

Example 7.5

For an NMOS current source, calculate the total thermal and $1/f$ noise in the drain current for a band from 1 kHz to 1 MHz.

Solution

The thermal noise current per unit bandwidth is given by $\overline{I_{n,th}^2} = 4kT(2/3)g_m$. Thus, the total thermal noise integrated across the band of interest is

$$\overline{I_{n,th,tot}^2} = 4kT\left(\frac{2}{3}g_m\right)(10^6 - 10^3) \tag{7.29}$$

$$\approx 4kT\left(\frac{2}{3}g_m\right) \times 10^6 \text{ A}^2. \tag{7.30}$$

For $1/f$ noise, the drain noise current per unit bandwidth is obtained by multiplying the noise voltage at the gate by the device transconductance:

$$\overline{I_{n,1/f}^2} = \frac{K}{C_{ox}WL} \cdot \frac{1}{f} \cdot g_m^2. \tag{7.31}$$

The total $1/f$ noise is then equal to

$$\overline{I_{n,1/f,tot}^2} = \frac{Kg_m^2}{C_{ox}WL}\int_{1\text{ kHz}}^{1\text{ MHz}} \frac{df}{f} \tag{7.32}$$

$$= \frac{Kg_m^2}{C_{ox}WL} \ln 10^3 \tag{7.33}$$

$$= \frac{6.91Kg_m^2}{C_{ox}WL}. \tag{7.34}$$

The above example raises an interesting question. What happens to $\overline{I_{n,1/f,tot}^2}$ if the lower end of the band, f_L, is zero rather than 1 kHz? Equation (7.33) then contains the natural

logarithm of zero, yielding an infinite value for the total noise. To overcome the fear of infinite noise, we make two observations. First, extending f_L to zero means that we are interested in *arbitrarily* slow noise components. A noise component at 0.01 Hz varies significantly in roughly 10 s and one at 10^{-6} in roughly one week. Second, the infinite flicker noise power simply means that if we observe the circuit for a very long time, the very slow noise components can randomly assume a very large power level. At such slow rates, noise becomes indistinguishable from thermal drift or aging of devices.

The foregoing observations lead to the following conclusions. First, since the *signals* encountered in most applications do not contain significant low-frequency components, our observation window need not be very long. For example, voice signals display negligible energy below 20 Hz and if a noise component varies at a lower rate, it does not corrupt the voice significantly. Second, the logarithmic dependence of the flicker noise power upon f_L allows some margin for error in selecting f_L. For example, if the band of interest is so wide that the total integrated thermal noise power is comparable with the flicker noise contribution, then the choice of f_L is quite relaxed.

In order to quantify the significance of $1/f$ noise with respect to thermal noise for a given device, we plot both spectral densities on the same axes (Fig. 7.23). Called the $1/f$ noise "corner frequency," the intersection point serves as a measure of what part of the band

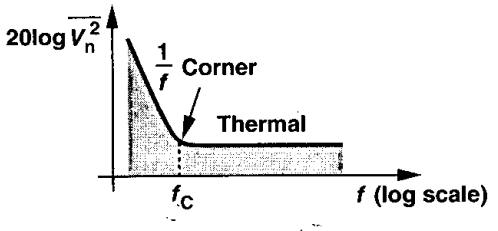


Figure 7.23 Concept of flicker noise corner frequency.

is mostly corrupted by flicker noise. In the above example, the $1/f$ noise corner, f_C, of the output current is determined as

$$4kT\left(\frac{2}{3}g_m\right) = \frac{K}{C_{ox}WL}\cdot\frac{1}{f_C}\cdot g_m^2, \tag{7.35}$$

that is,

$$f_C = \frac{K}{C_{ox}WL}g_m\frac{3}{8kT}. \tag{7.36}$$

This result implies that f_C generally depends on device dimensions and bias current. Nonetheless, since for a given L, the dependence is relatively weak, the $1/f$ noise corner is relatively constant, falling in the vicinity of 500 kHz to 1 MHz for submicron transistors.

Example 7.6

For a 100-μm/0.5-μm MOS device with $g_m = 1/(100\,\Omega)$, the $1/f$ noise corner frequency is measured to be 500 kHz. If $t_{ox} = 90$ Å, what is the flicker noise coefficient, K, in this technology?

Solution

For $t_{ox} = 90$ Å, we have $C_{ox} = 3.84$ fF/μm^2. Using Eq. (7.36), we write

$$500 \text{ kHz} = \frac{K}{3.84 \times 100 \times 0.5 \times 10^{-15}} \cdot \frac{1}{100} \cdot \frac{3}{8 \times 1.38 \times 10^{-23} \times 300}. \tag{7.37}$$

That is, $K = 1.06 \times 10^{-25}$ V^{2}F.

7.3 Representation of Noise in Circuits

Consider a general circuit with one input port and one output port (Fig. 7.24). How do we quantify the effect of noise here? The natural approach would be to set the input to zero and calculate the total noise at the output due to various sources of noise in the circuit. This is indeed how the noise is measured in the laboratory or in simulations.

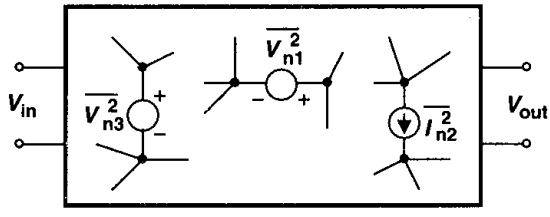
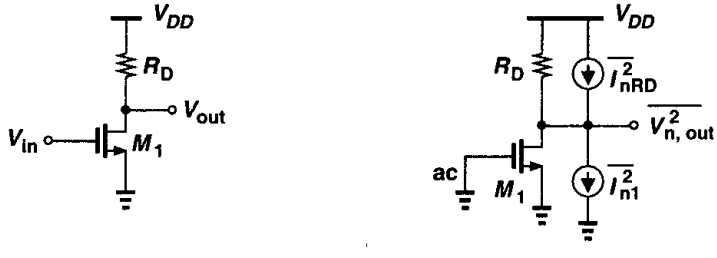


Figure 7.24 Noise sources in a circuit.

Example 7.7

What is the total output noise voltage of the common-source stage shown in Fig. 7.25(a)?



(a) (b)

Figure 7.25 (a) CS stage, (b) circuit including noise sources.

Solution

We model the thermal and flicker noise of M_1 by two current sources: $\overline{I_{n,th}^2} = 4kT(2/3)g_m$ and $\overline{I_{n,1/f}^2} = Kg_m^2/(C_{ox}WLf)$. We also represent the thermal noise of R_D by a current source $\overline{I_{n,RD}^2} = 4kT/R_D$. The output noise voltage per unit bandwidth is therefore equal to

$$\overline{V_{n,out}^2} = \left(4kT\frac{2}{3}g_m + \frac{K}{C_{ox}WL}\cdot\frac{1}{f}\cdot g_m^2 + \frac{4kT}{R_D}\right)R_D^2. \tag{7.38}$$

Note that the noise mechanisms are added as "power" quantities because they are uncorrelated. The value given by (7.38) represents the noise power in 1 Hz at a frequency f. The total output noise can be obtained by integration over the bandwidth of interest.

While intuitively appealing, the output-referred noise does not allow a fair comparison of the performance of different circuits because it depends on the gain. For example, as depicted in Fig. 7.26, if a common-source stage is followed by a noiseless amplifier having

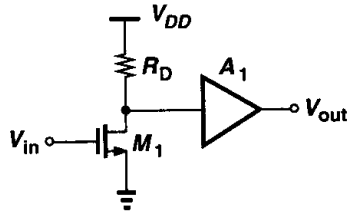


Figure 7.26 Addition of gain stage to a CS stage.

a voltage gain A_1, then the output noise is equal to the expression in (7.38) multiplied by A_1^2. Considering only the output noise, we may conclude that as A_1 increases, the circuit becomes noisier, an incorrect result because a larger A_1 also provides a proportionally higher *signal* level at the output. That is, the output signal-to-noise ratio (SNR) does not depend on A_1.

To overcome the above quandary, we usually specify the "input-referred noise" of circuits. Illustrated conceptually in Fig. 7.27, the idea is to represent the effect of all noise sources in the circuit by a single source, $\overline{V_{n,in}^2}$, at the input such that the output noise

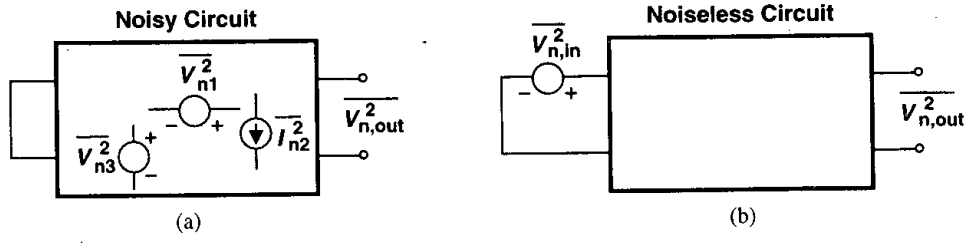


Figure 7.27 Determination of input-referred noise voltage.

in Fig. 7.27(b) equals that in Fig. 7.27(a). If the voltage gain is A_v, then we must have $\overline{V_{n,out}^2} = A_v^2 \overline{V_{n,in}^2}$, that is, the input-referred noise voltage in this simple case is given by the output noise voltage divided by the gain.

Example 7.8

For the circuit of Fig. 7.25, calculate the input-referred noise voltage.

Solution

We have

$$\overline{V_{n,in}^2} = \frac{\overline{V_{n,out}^2}}{A_v^2} \tag{7.39}$$

$$= \left(4kT\frac{2}{3}g_m + \frac{K}{C_{ox}WL}\cdot\frac{1}{f}\cdot g_m^2 + \frac{4kT}{R_D}\right) R_D^2 \frac{1}{g_m^2 R_D^2} \tag{7.40}$$

$$= 4kT\frac{2}{3g_m} + \frac{K}{C_{ox}WL}\cdot\frac{1}{f} + \frac{4kT}{g_m^2 R_D}. \tag{7.41}$$

Note that the first term in (7.41) can be viewed as the thermal noise of a resistor equal to $2/(3g_m)$ placed in series with the gate. Similarly, the third term corresponds to the noise of a resistor equal to $(g_m^2 R_D)^{-1}$. We sometimes say the "equivalent thermal noise resistance" of a circuit is equal to R_T, meaning that the total input-referred thermal noise of the circuit in unit bandwidth is equal to $4kTR_T$.

At this point of our study, we make two observations. First, the input-referred noise and the input signal are both multiplied by the gain as they are processed by the circuit. Thus, the input-referred noise indicates how much the input signal is corrupted by the circuit's noise, i.e., how small an input the circuit can detect with acceptable SNR. For this reason, input-referred noise allows a fair comparison of different circuits. Second, the input-referred noise is a fictitious quantity in that it cannot be *measured* at the input of the circuit. The two circuits of Figs. 7.27(a) and (b) are *mathematically* equivalent but the physical circuit is still that in Fig. 7.27(a).

In the foregoing discussion, we have assumed that the input-referred noise can be modeled by a single voltage source in series with the input. This is generally an incomplete representation if the circuit has a finite input impedance and is driven by a finite source impedance. To understand why, consider the common-source stage of Fig. 7.28(a), where the input capacitance is denoted by C_{in} and $1/f$ noise is neglected for simplicity. From Eq. (7.41), the input-referred noise voltage of the circuit is given by $8kT/(3g_m)+4kT/(g_m^2 R_D)$. Now suppose the preceding stage is modeled by a Thevenin equivalent having an inductive output impedance [Fig. 7.28(b)]. Simplifying the circuit for noise calculations as shown in Fig. 7.28(c), we seek to find the output noise as L_1 increases. Owing to the voltage division between $L_1 s$ and $1/(C_{in}s)$, the effect of $\overline{V_{n,in}^2}$ at the gate of M_1 and hence at the output vanishes as L_1 approaches infinity. This result, however, is incorrect because the output noise of the circuit is equal to $(8kT/3)g_m R_D^2 + 4kTR_D$ and independent of L_1 and C_{in}.

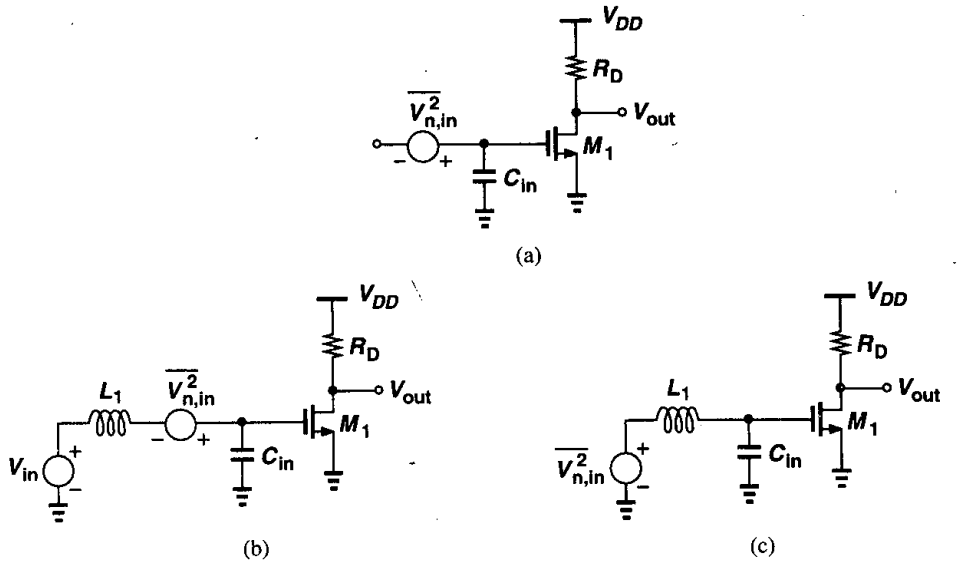


Figure 7.28 CS stage including input capacitance, (b) CS stage stimulated by a finite source impedance, (c) Effect of single noise source.

Let us summarize the problem. If the circuit has a finite input impedance, modeling the input-referred noise by merely a voltage source implies that the output noise vanishes as the source impedance becomes large, an incorrect conclusion. To resolve this issue, we model the input-referred noise by both a series voltage source and a parallel current source (Fig. 7.29) so that if the output impedance of the preceding stage assumes large values—thereby reducing the effect of $\overline{V^2_{n,in}}$—the noise current source still flows through a finite impedance, producing noise at the input. It can be proved that $\overline{V^2_{n,in}}$ and $\overline{I^2_{n,in}}$ are necessary and sufficient to represent the noise of any linear two-port circuit [5].

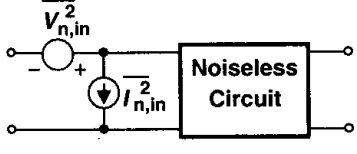


Figure 7.29 Representation of noise by voltage and current sources.

How do we calculate $\overline{V^2_{n,in}}$ and $\overline{I^2_{n,in}}$? Since the model is valid for any source impedance, we consider two extreme cases: zero and infinite source impedances. As shown in Fig. 7.30(a), if the source impedance is zero, $\overline{I^2_{n,in}}$ flows through $\overline{V^2_{n,in}}$ and has no effect on the output. Thus, the output noise measured in this case arises solely from $\overline{V^2_{n,in}}$. Similarly, if the input is open [Fig. 7.30(b)], then $\overline{V^2_{n,in}}$ has no effect and the output noise is due to only $\overline{I^2_{n,in}}$. Let us apply this method to the circuit of Fig. 7.28.

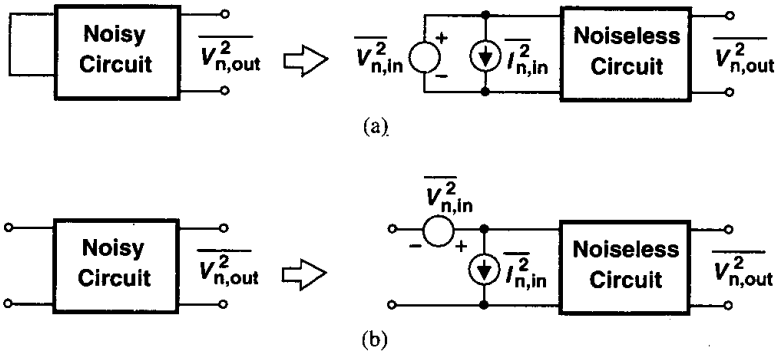


Figure 7.30 Calculation of input-referred noise (a) voltage, and (b) current.

Example 7.9

Calculate the input-referred noise voltage and current of Fig. 7.28.

Solution

From (7.41), the input-referred noise voltage (excluding $1/f$ noise) is simply

$$\overline{V^2_{n,in}} = 4kT\frac{2}{3g_m} + \frac{4kT}{g_m^2 R_D}. \tag{7.42}$$

As depicted in Fig. 7.31(a), this voltage generates the same output noise as the actual circuit if the input is shorted.

To obtain the input-referred noise current, we open the input and find the output noise in terms of $\overline{I^2_{n,in}}$ [Fig. 7.31(b)]. The noise current flows through C_{in}, generating at the output

$$\overline{V^2_{n,out}} = \overline{I^2_{n,in}}\left(\frac{1}{C_{in}\omega}\right)^2 g_m^2 R_D^2. \tag{7.43}$$

This value must be equal to the output of the noisy circuit when its input is open:

$$\overline{V^2_{n,out}} = \left(4kT\frac{2}{3}g_m + \frac{4kT}{R_D}\right)R_D^2. \tag{7.44}$$

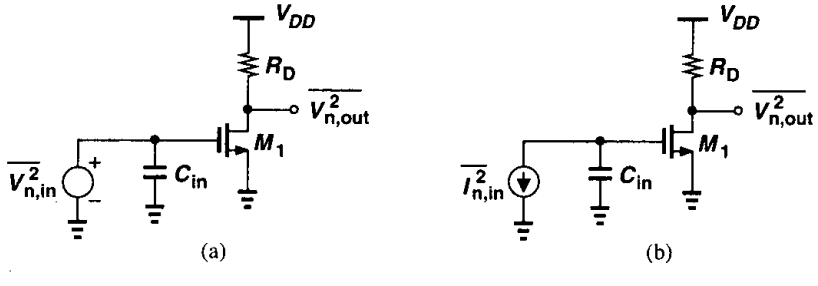


Figure 7.31

From (7.43) and (7.44), it follows that

$$\overline{I_{n,in}^2} = (C_{in}\omega)^2 \frac{4kT}{g_m^2}\left(\frac{2}{3}g_m + \frac{1}{R_D}\right). \tag{7.45}$$

The reader may wonder if the use of both a voltage source and a current source to represent the input-referred noise "counts the noise twice." We utilize the circuit of Fig. 7.28 as an example to demonstrate that this is not so. Considering the environment depicted in Fig. 7.32, we prove that the output noise is correct for any source impedance Z_S. Assuming Z_S is noise-

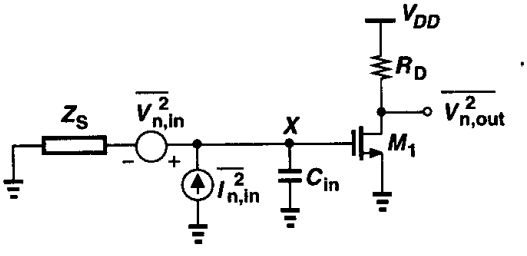


Figure 7.32 CS stage stimulated by a source impedance.

less for simplicity, we first calculate the total noise voltage at the gate of M_1 due to $\overline{V_{n,in}^2}$ and $\overline{I_{n,in}^2}$. How is this voltage obtained: by superposition of voltages or powers? The two sources $\overline{V_{n,in}^2}$ and $\overline{I_{n,in}^2}$ are in general *correlated* simply because they may represent the same noise mechanisms in the circuit. In fact, Eqs. (7.42) and (7.45) can be respectively rewritten as

$$V_{n,in} = V_{n,M1} + \frac{1}{g_m R_D} V_{n,RD} \tag{7.46}$$

$$I_{n,in} = C_{in}sV_{n,M1} + \frac{C_{in}s}{g_m R_D} V_{n,RD}, \tag{7.47}$$

where $V_{n,M1}$ denotes the gate-referred noise voltage of M_1 and $V_{n,RD}$ the noise voltage of R_D. We recognize that $V_{n,M1}$ and $V_{n,RD}$ appear in both $V_{n,in}$ and $I_{n,in}$, creating a strong correlation between the two. Thus, the calculations must use superposition of voltages—as if $V_{n,in}$ and $I_{n,in}$ were deterministic quantities.

Adding the contributions of $V_{n,in}$ and $I_{n,in}$ at node X in Fig. 7.32, we have

$$V_{n,X} = V_{n,in}\frac{\dfrac{1}{C_{in}s}}{\dfrac{1}{C_{in}s} + Z_S} + I_{n,in}\frac{\dfrac{Z_S}{C_{in}s}}{\dfrac{1}{C_{in}s} + Z_S} \tag{7.48}$$

$$= \frac{V_{n,in} + I_{n,in}Z_S}{Z_S C_{in}s + 1}. \tag{7.49}$$

Substituting for $V_{n,in}$ and $I_{n,in}$ from (7.46) and (7.47), respectively, we obtain

$$V_{n,X} = \frac{1}{Z_S C_{in} s + 1}\left[V_{n,M1} + \frac{1}{g_m R_D} V_{n,RD} + C_{in} s Z_S (V_{n,M1} + \frac{1}{g_m R_D} V_{n,RD})\right] \quad (7.50)$$

$$= V_{n,M1} + \frac{1}{g_m R_D} V_{n,RD}. \quad (7.51)$$

Note that $V_{n,X}$ is independent of Z_S and C_{in}. It follows that

$$\overline{V_{n,out}^2} = g_m^2 R_D^2 \overline{V_{n,X}^2} \quad (7.52)$$

$$= 4kT\left(\frac{2}{3} g_m + \frac{1}{R_D}\right) R_D^2, \quad (7.53)$$

the same as (7.44).

7.4 Noise in Single-Stage Amplifiers

Having developed basic mathematical tools and models for noise analysis, we now study the noise performance of single-stage amplifiers at low frequencies. Before considering specific topologies, we describe a lemma that simplifies noise calculations.

Lemma The circuits shown in Fig. 7.33(a) and (b) are equivalent at low frequencies if $\overline{V_n^2} = \overline{I_n^2}/g_m^2$ and the circuits are driven by a finite impedance.

Proof. Since the circuits have equal output impedances, we simply examine the output short-circuit currents [Figs. 7.33(c) and (d)]. It can be proved (Problem 7.4) that the output noise current of the circuit in Fig. 7.33(c) is given by

$$I_{n,out1} = \frac{I_n}{Z_S(g_m + 1/r_O) + 1} \quad (7.54)$$

and that of Fig. 7.33(d) is

$$I_{n,out2} = \frac{g_m V_n}{Z_S(g_m + 1/r_O) + 1}. \quad (7.55)$$

Equating (7.54) and (7.55), we have $V_n = I_n/g_m$. ❑

This lemma suggests that the noise source can be transformed from a drain-source current to a gate series voltage for arbitrary Z_S.

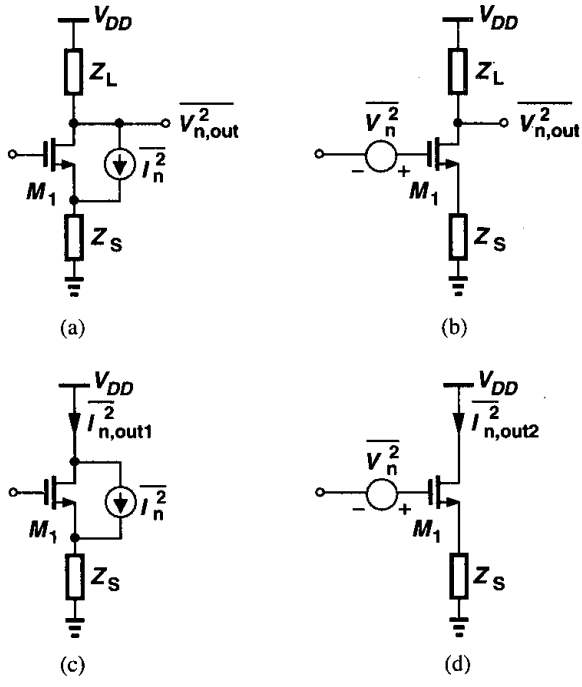


Figure 7.33 Equivalent CS stages.

7.4.1 Common-Source Stage

From Example 7.8, the input-referred noise voltage per unit bandwidth of a simple CS stage is equal to

$$\overline{V_{n,in}^2} = 4kT\left(\frac{2}{3g_m} + \frac{1}{g_m^2 R_D}\right) + \frac{K}{C_{ox}WL}\frac{1}{f}. \tag{7.56}$$

From the above lemma, we recognize that the term $4kT[2/(3g_m)]$ is in fact the thermal noise current of M_1 expressed as a voltage in series with the gate.

How can we reduce the input-referred noise voltage? Equation (7.56) implies that the transconductance of M_1 must be maximized. Thus, the transconductance must be maximized if the transistor is to amplify a voltage signal applied to its gate [Fig. 7.34(a)] whereas it must be minimized if the transistor operates as a current source [Fig. 7.34(b)].

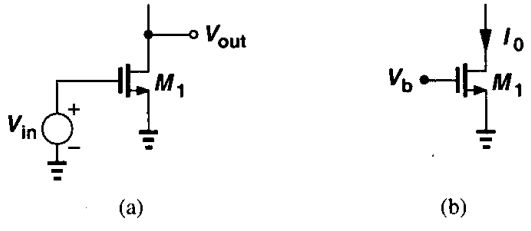


Figure 7.34 Voltage amplification versus current generation.

Example 7.10

Calculate the input-referred thermal noise voltage of the amplifier shown in Fig. 7.35(a), assuming both transistors are in saturation. Also, determine the total output thermal noise if the circuit drives a load capacitance C_L. What is the output signal-to-noise ratio if a low-frequency sinusoid of amplitude V_m is applied to the input?

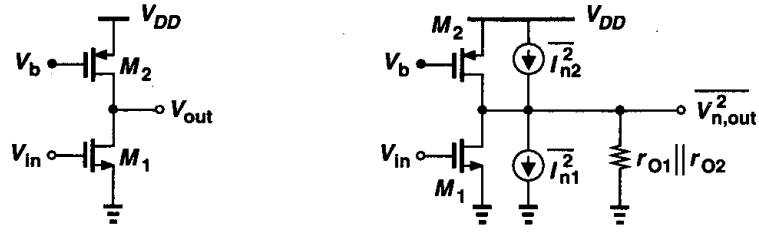


Figure 7.35

Solution

Representing the thermal noise of M_1 and M_2 by current sources [Fig. 7.35(b)] and noting that they are uncorrelated, we write

$$\overline{V_{n,out}^2} = 4kT\left(\frac{2}{3}g_{m1} + \frac{2}{3}g_{m2}\right)(r_{O1}\|r_{O2})^2. \tag{7.57}$$

Since the voltage gain is equal to $g_{m1}(r_{O1}\|r_{O2})$, the total noise voltage referred to the gate of M_1 is

$$\overline{V_n^2} = 4kT\left(\frac{2}{3}g_{m1} + \frac{2}{3}g_{m2}\right)\frac{1}{g_{m1}^2} \tag{7.58}$$

$$= 4kT\left(\frac{2}{3g_{m1}} + \frac{2}{3}\frac{g_{m2}}{g_{m1}^2}\right). \tag{7.59}$$

Equation (7.59) reveals the dependence of $\overline{V_{n,in}^2}$ upon g_{m1} and g_{m2}, confirming that g_{m2} must be minimized because M_2 serves as a current source.

The reader may wonder why M_1 and M_2 in Fig. 7.35 exhibit different noise effects. After all, if the noise currents of both transistors flow through $r_{O1}\|r_{O2}$, why should g_{m1} be maximized and g_{m2} minimized? This is simply because, as g_{m1} increases, the output noise *voltage* rises in proportion to $\sqrt{g_{m1}}$ whereas the *voltage gain* of the stage increases in proportion to g_{m1}. As a result, the input-referred noise voltage decreases.

To compute the total output noise, we integrate (7.57) across the band:

$$\overline{V_{n,out,tot}^2} = \int_0^\infty 4kT\left(\frac{2}{3}g_{m1} + \frac{2}{3}g_{m2}\right)(r_{O1}\|r_{O2})^2\frac{df}{1+(r_{O1}\|r_{O2})^2C_L^2(2\pi f)^2}. \tag{7.60}$$

Using the results of Example 7.1, we have

$$\overline{V_{n,out,tot}^2} = \frac{2}{3}(g_{m1}+g_{m2})(r_{O1}\|r_{O2})\frac{kT}{C_L}. \tag{7.61}$$

An input sinusoid of amplitude V_m yields an output amplitude equal to $g_{m1}(r_{O1}\|r_{O2})V_m$. The output SNR is equal to the ratio of the signal power and the noise power:

$$SNR_{out} = \left[\frac{g_{m1}(r_{O1}\|r_{O2})V_m}{\sqrt{2}}\right]^2 \cdot \frac{1}{(2/3)(g_{m1}+g_{m2})(r_{O1}\|r_{O2})(kT/C_L)} \tag{7.62}$$

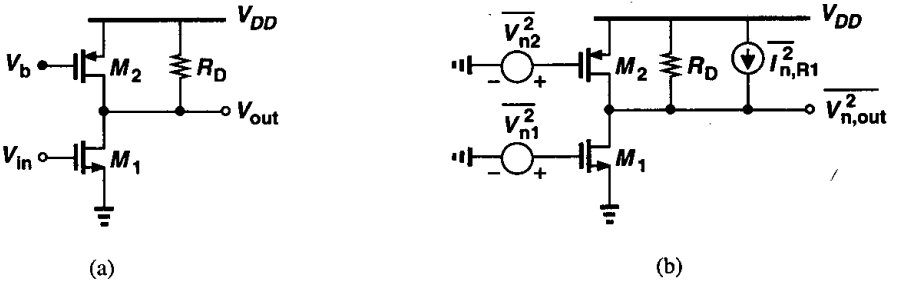
$$= \frac{3C_L}{4kT} \cdot \frac{g_{m1}^2(r_{O1}\|r_{O2})}{g_{m1}+g_{m2}} V_m^2. \tag{7.63}$$

We note that to maximize the output SNR, C_L must be maximized, i.e., the bandwidth must be minimized. Of course, the bandwidth is also dictated by the input signal spectrum. This example indicates that it becomes exceedingly difficult to design broadband circuits while maintaining a low noise.

It is also important to observe from (7.56) that the noise contributed by R_D in Fig. 7.25(a) decreases as R_D increases. This is again because the noise voltage due to R_D at the output is proportional to $\sqrt{R_D}$ while the voltage gain of the circuit is proportional to R_D.

Example 7.11

Calculate the input-referred $1/f$ and thermal noise voltage of the circuit depicted in Fig. 7.36(a) assuming M_1 and M_2 are in saturation.



(a) (b)

Figure 7.36

Solution

We model the $1/f$ and thermal noise of the transistors as voltage sources in series with their gates [Fig. 7.36(b)]. The noise voltage at the gate of M_2 experiences a gain of $g_{m2}(R_D\|r_{O1}\|r_{O2})$ as it appears at the output. The result must then be divided by $g_{m1}(R_D\|r_{O1}\|r_{O2})$ to be referred to the main input. The noise current of R_D is multiplied by $R_D\|r_{O1}\|r_{O2}$ and divided by $g_{m1}(R_D\|r_{O1}\|r_{O2})$. Thus, the overall input-referred noise voltage is given by

$$\overline{V_{n,in}^2} = 4kT\frac{2}{3}\left(\frac{g_{m2}}{g_{m1}^2}+\frac{1}{g_{m1}}\right) + \frac{1}{C_{ox}}\left[\frac{K_P g_{m2}^2}{(WL)_2 g_{m1}^2} + \frac{K_N}{(WL)_1}\right]\frac{1}{f} + \frac{4kT}{g_{m1}^2 R_D}, \tag{7.64}$$

where K_P and K_N denote the flicker noise coefficients of PMOS and NMOS devices, respectively. As expected, the input-referred noise voltage increases if g_{m2} increases.

How do we design a common-source stage for low-noise operation? For thermal noise in the simple topology of Fig. 7.34, we must maximize g_{m1} by increasing the drain current or the device width. A higher I_D translates to greater power dissipation and limited output voltage swings while a wider device leads to larger input and output capacitance. We can also increase R_D, but at the cost of limiting the voltage headroom and lowering the speed.

For $1/f$ noise, the primary approach is to increase the area of the transistor. If WL is increased while W/L remains constant, then the device transconductance and hence its thermal noise do not change but the device capacitances increase. These observations point to the trade-offs between noise, power dissipation, voltage headroom, and speed.

7.4.2 Common-Gate Stage

Consider the common-gate configuration shown in Fig. 7.37(a). Neglecting channel-

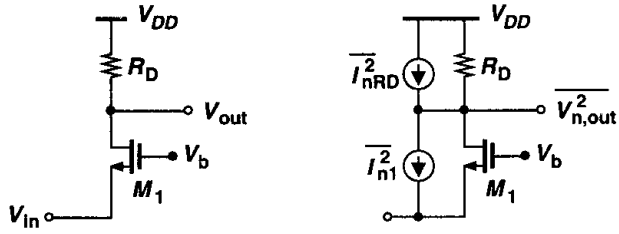


Figure 7.37 (a) CG stage, (b) circuit including noise sources.

length modulation, we represent the thermal noise of M_1 and R_D by two current sources [Fig. 7.37(b)]. Note that, owing to the low input impedance of the circuit, the input-referred noise current is not negligible even at low frequencies. To calculate the input-referred noise voltage, we short the input to ground and equate the output noise of the circuits in Figs. 7.38(a) and (b):

$$\left(4kT\frac{2}{3}g_m + \frac{4kT}{R_D}\right)R_D^2 = \overline{V_{n,in}^2}(g_m + g_{mb})^2 R_D^2. \tag{7.65}$$

That is,

$$\overline{V_{n,in}^2} = \frac{4kT(2g_m/3 + 1/R_D)}{(g_m + g_{mb})^2}. \tag{7.66}$$

Similarly, equating the output noise of the circuits in Figs. 7.38(c) and (d) yields the input-referred noise current. What is the effect of $\overline{I_{n1}^2}$ at the output in Fig. 7.38(c)? Since the sum of the currents at the source of M_1 is zero, $I_{n1} + I_{D1} = 0$. Consequently, I_{n1} creates an equal and opposite current in M_1, producing *no* noise at the output. The output noise voltage of Fig. 7.37(a) is therefore equal to $4kTR_D$ and hence $\overline{I_{n,in}^2}R_D^2 = 4kTR_D$. That is,

$$\overline{I_{n,in}^2} = \frac{4kT}{R_D}. \tag{7.67}$$

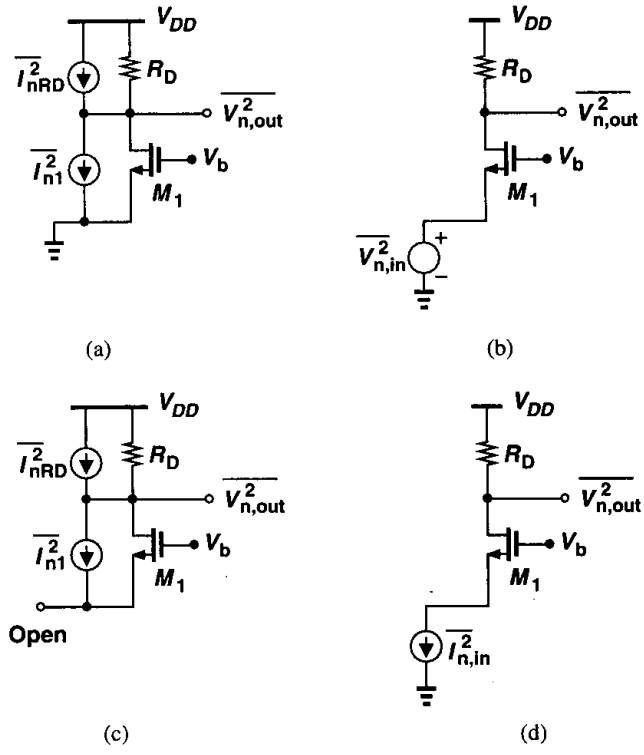


Figure 7.38 Calculation of input-referred noise of a CG stage.

An important drawback of common-gate topologies is that they directly refer the noise current produced by the load to the input. Exemplified by (7.67), this effect arises because such circuits provide no *current* gain, a point of contrast to common-source amplifiers.

We have thus far neglected the noise contributed by the bias current source of a common-gate stage. Shown in Fig. 7.39 is a simple mirror arrangement establishing the bias current of M_1 as a multiple of I_1. Capacitor C_0 shunts the noise generated by M_0 to ground. We

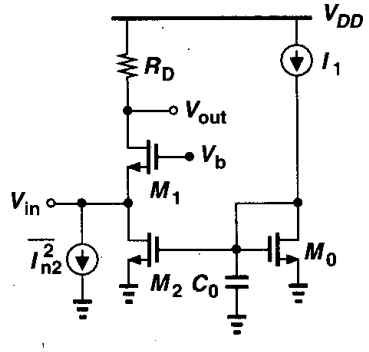


Figure 7.39 Noise contributed by bias current source.

note that if the input of the circuit is shorted to ground, then the drain noise current of M_2 does not flow through R_D, contributing no input-referred noise voltage. On the other hand, if the input is open, all of $\overline{I_{n2}^2}$ flows from M_1 and R_D (at low frequencies), producing an output noise equal to $\overline{I_{n2}^2}R_D^2$ and hence an input-referred noise current of $\overline{I_{n2}^2}$. As a result, the noise current of M_2 directly adds to the input-referred noise current, making it desirable to *minimize* the transconductance of M_2. For a given bias current, however, this translates to a higher drain-source voltage for M_2 because $g_{m2} = 2I_{D2}/(V_{GS2} - V_{TH2})$, requiring a high value for V_b and limiting the voltage swing at the output node.

Example 7.12

Calculate the input-referred thermal noise voltage and current of the circuit shown in Fig. 7.40 assuming all of the transistors are in saturation.

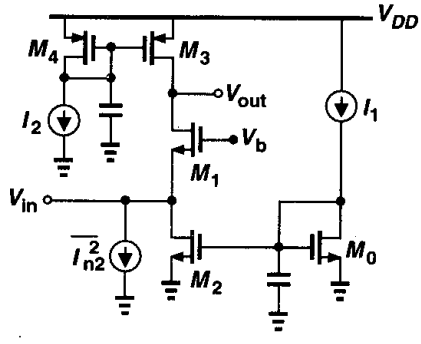


Figure 7.40

Solution

To compute the input-referred noise voltage, we short the input to ground, obtaining

$$\overline{V_{n,out}^2} = 4kT\frac{2}{3}(g_{m1} + g_{m3})(r_{O1} \| r_{O3})^2. \tag{7.68}$$

Thus, the input-referred noise voltage, $V_{n,in}$, must satisfy this relationship:

$$\overline{V_{n,in}^2}(g_{m1} + g_{mb1})^2(r_{O1} \| r_{O3})^2 = 4kT\frac{2}{3}(g_{m1} + g_{m3})(r_{O1} \| r_{O3})^2, \tag{7.69}$$

and hence

$$\overline{V_{n,in}^2} = 4kT\frac{2}{3}\frac{(g_{m1} + g_{m3})}{(g_{m1} + g_{mb1})^2}. \tag{7.70}$$

As expected, the noise is proportional to g_{m3}.

To calculate the input-referred noise current, we open the input and note that the output noise voltage is simply given by $(\overline{I_{n2}^2} + \overline{I_{n3}^2})R_{out}^2$, where $R_{out} \approx r_{O3} \| (g_{m1}r_{O1}r_{O2})$ denotes the output impedance when the input is open. It follows that

$$\overline{I_{n,in}^2} = 4kT\frac{2}{3}(g_{m2} + g_{m3}). \tag{7.71}$$

Again, the noise is proportional to the transconductance of the two current sources.

The effect of $1/f$ noise in a common-gate topology is also of interest. As a typical case, we compute the input-referred $1/f$ noise voltage and current of the circuit shown in Fig. 7.40. Illustrated in Fig. 7.41, each $1/f$ noise generator is modeled by a voltage source in series with the gate of the corresponding transistor. Note that the $1/f$ noise of M_0 and M_4 is neglected. A more realistic case is studied in Problem 7.10.

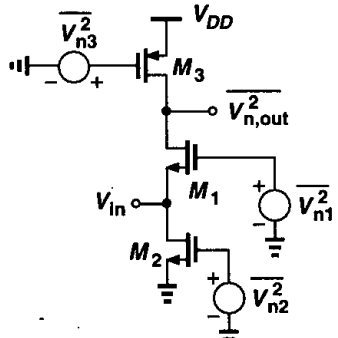


Figure 7.41 Flicker noise in a CG stage.

With the input shorted to ground, we have

$$\overline{V_{n,out}^2} = \frac{1}{C_{ox} f}\left[\frac{g_{m1}^2 K_N}{(WL)_1} + \frac{g_{m3}^2 K_P}{(WL)_3}\right](r_{O1} \| r_{O3})^2, \tag{7.72}$$

where K_N and K_P denote the flicker noise coefficient of NMOS and PMOS devices, respectively. Thus,

$$\overline{V_{n,in}^2} = \frac{1}{C_{ox} f}\left[\frac{g_{m1}^2 K_N}{(WL)_1} + \frac{g_{m3}^2 K_P}{(WL)_3}\right]\frac{1}{(g_{m1} + g_{mb1})^2}. \tag{7.73}$$

With the input open, the output noise voltage is given by

$$\overline{V_{n,out}^2} = \frac{1}{C_{ox} f}\left[\frac{g_{m2}^2 K_N}{(WL)_2} + \frac{g_{m3}^2 K_P}{(WL)_3}\right] R_{out}^2, \tag{7.74}$$

yielding

$$\overline{I_{n,in}^2} = \frac{1}{C_{ox} f}\left[\frac{g_{m2}^2 K_N}{(WL)_2} + \frac{g_{m3}^2 K_P}{(WL)_3}\right]. \tag{7.75}$$

Equations (7.73) and (7.75) describe the $1/f$ noise behavior of the circuit and must be added to (7.70) and (7.71), respectively, to obtain the overall noise per unit bandwidth.

7.4.3 Source Followers

Consider the source follower depicted in Fig. 7.42(a), where M_2 serves as the bias current source. Since the input impedance of the circuit is quite high, even at relatively high frequencies, the input-referred noise current can usually be neglected for moderate driving

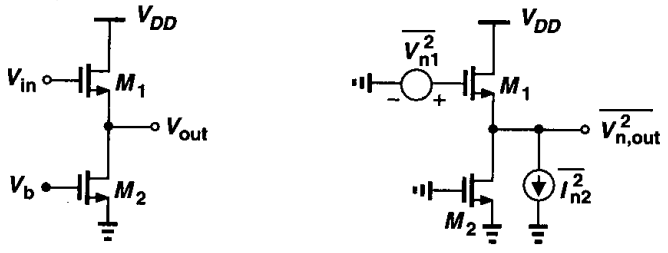


Figure 7.42 (a) Source follower, (b) circuit including noise sources.

source impedances. To compute the input-referred thermal noise voltage, we employ the representation in Fig. 7.42(b), expressing the output noise due to M_2 as

$$\overline{V_{n,out}^2}|_{M2} = \overline{I_{n2}^2}\left(\frac{1}{g_{m1}} \Big\| \frac{1}{g_{mb1}} \Big\| r_{Q1} \| r_{O2}\right)^2. \tag{7.76}$$

From Chapter 3,

$$A_v = \frac{\dfrac{1}{g_{mb1}} \Big\| r_{O1} \| r_{O2}}{\dfrac{1}{g_{mb1}} \Big\| r_{O1} \| r_{O2} + \dfrac{1}{g_{m1}}}. \tag{7.77}$$

Thus, the total input-referred noise voltage is

$$\overline{V_{n,in}^2} = \overline{V_{n1}^2} + \frac{\overline{V_{n,out}^2}|_{M2}}{A_v^2} \tag{7.78}$$

$$= 4kT\frac{2}{3}\left(\frac{1}{g_{m1}} + \frac{g_{m2}}{g_{m1}^2}\right). \tag{7.79}$$

Note the similarity between (7.59) and (7.79).

Since source followers add noise to the input signal while providing a voltage gain less than unity, they are usually avoided in low-noise amplification. The $1/f$ noise performance of source followers is studied in Problem 7.11.

7.4.4 Cascode Stage

Consider the cascode stage of Fig. 7.43(a). Since at low frequencies the noise currents of M_1 and R_D flow through R_D, the noise contributed by these two devices is quantified as in a common-source stage:

$$\overline{V_{n,in}^2}|_{M1,RD} = 4kT\left(\frac{2}{3g_{m1}} + \frac{1}{g_{m1}^2 R_D}\right), \tag{7.80}$$

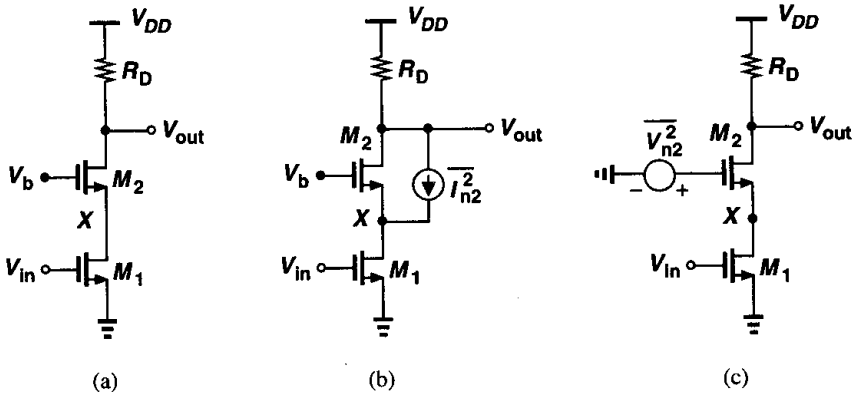


Figure 7.43 (a) Cascode stage, (b) noise of M_2 modeled by a current source, (c) noise of M_2 modeled by a voltage source.

where $1/f$ noise of M_1 is ignored. What is the effect of noise of M_2? Modeled as in Fig. 7.43(b), this noise contributes negligibly to the output, especially at low frequencies. This is because, if channel length modulation in M_1 is neglected, then $I_{n2} + I_{D2} = 0$, and hence M_2 does not affect $V_{n,out}$. From another point of view, using the lemma of Fig. 7.33 to construct the equivalent in Fig. 7.43(c), we note that the voltage gain from V_{n2} to the output is quite small if the impedance at node X is large. At high frequencies, on the other hand, the total capacitance at node X, C_X, gives rise to a gain:

$$\frac{V_{n,out}}{V_{n2}} \approx \frac{-R_D}{1/g_{m2} + 1/(C_X s)}, \tag{7.81}$$

increasing the output noise. This capacitance also decreases the gain from the main input to the output by shunting the signal current produced by M_1 to ground. As a result, the input-referred noise of a cascode stage may rise considerably at high frequencies.

7.5 Noise in Differential Pairs

With our understanding of noise in basic amplifiers, we can now study the noise behavior of differential pairs. Shown in Fig. 7.44(a), a differential pair can be viewed as a two-port circuit. It is therefore possible to model the overall noise as depicted in Fig. 7.44(b). For low-frequency operation, the magnitude of $\overline{I_{n,in}^2}$ is typically negligible.

To calculate the thermal component of $\overline{V_{n,in}^2}$, we first obtain the total output noise with the inputs shorted together [Fig. 7.45(a)], noting that superposition of power quantities is possible because the noise sources in the circuit are uncorrelated. Since I_{n1} and I_{n2} are uncorrelated, node P cannot be considered a virtual ground, making it difficult to use the half-circuit concept. Thus, we simply derive the effect of each source individually. Depicted in Fig. 7.45(b), the contribution of I_{n1} is obtained by first reducing the circuit to that in Fig. 7.45(c). With the aid of this figure and neglecting channel-length modulation, the reader can prove that half of I_{n1} flows through R_{D1} and the other half through M_2 and R_{D2}.

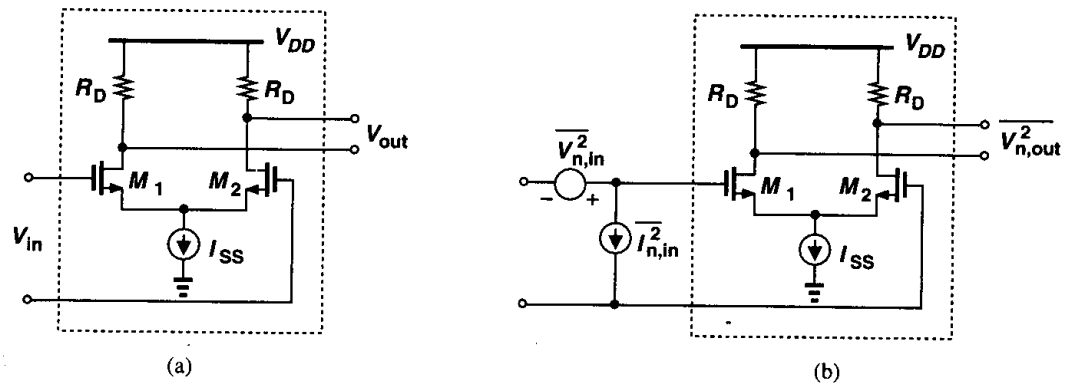


Figure 7.44 (a) Differential pair, (b) circuit including input-referred noise sources.

As shown in Fig. 7.45(d), this can also be proved by decomposing I_{n1} into two (correlated) current sources and calculating their effect at the output. Thus,

$$V_{n,out}|_{M1} = \frac{I_{n1}}{2} R_{D1} + \frac{I_{n1}}{2} R_{D2}. \tag{7.82}$$

Note that the two noise voltages are directly added because they both arise from I_{n1} and are therefore correlated. It follows that, if $R_{D1} = R_{D2} = R_D$,

$$\overline{V^2_{n,out}}|_{M1} = \overline{I^2_{n1}} R_D^2. \tag{7.83}$$

Similarly,

$$\overline{V^2_{n,out}}|_{M2} = \overline{I^2_{n2}} R_D^2, \tag{7.84}$$

yielding

$$\overline{V^2_{n,out}}|_{M1,M2} = (\overline{I^2_{n1}} + \overline{I^2_{n2}}) R_D^2. \tag{7.85}$$

Taking into account the noise of R_{D1} and R_{D2}, we have for the total output noise:

$$\overline{V^2_{n,out}} = \left(\overline{I^2_{n1}} + \overline{I^2_{n2}}\right) R_D^2 + 2(4kTR_D) \tag{7.86}$$

$$= 8kT\left(\frac{2}{3} g_m R_D^2 + R_D\right). \tag{7.87}$$

Dividing the result by the square of the differential gain, $g_m^2 R_D^2$, we have

$$\overline{V^2_{n,in}} = 8kT\left(\frac{2}{3g_m} + \frac{1}{g_m^2 R_D}\right). \tag{7.88}$$

This is simply twice the input noise voltage squared of a common-source stage.

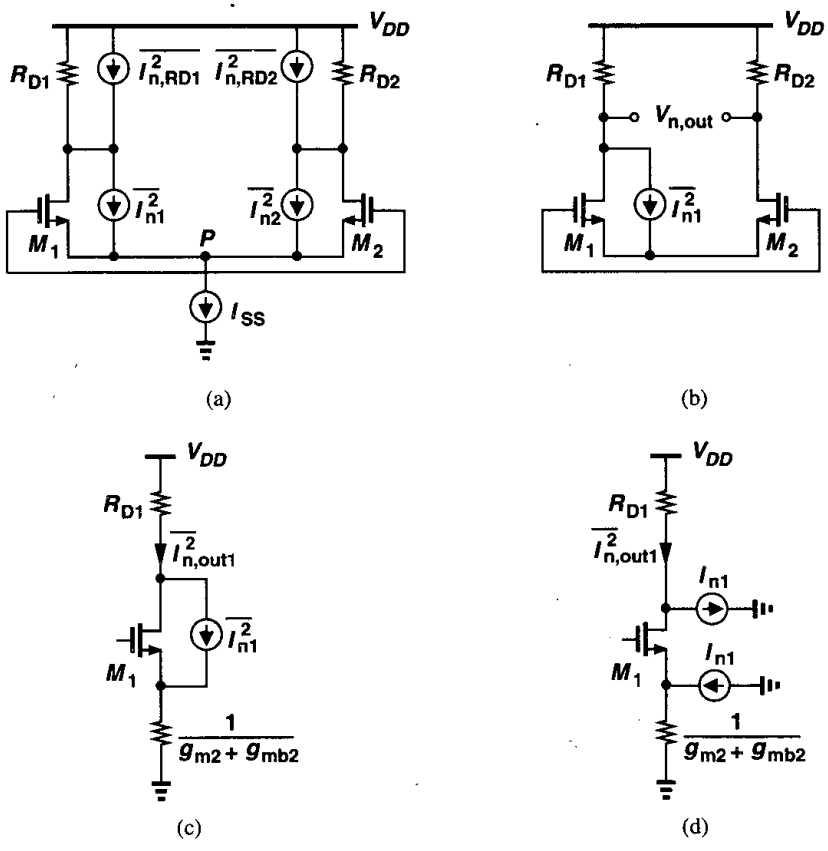


Figure 7.45 Calculation of input-referred noise of a differential pair.

The input-referred noise voltage can also be calculated by exploiting the lemma illustrated in Fig. 7.33. As shown in Fig. 7.46, the noise of M_1 and M_2 is modeled as a voltage source in series with their gates, and the noise of R_{D1} and R_{D2} is divided by $g_m^2 R_D^2$, thereby resulting in (7.88).

It is instructive to compare the noise performance of a differential pair and a common-source stage, as expressed by (7.56) and (7.88). We conclude that, if each transistor has a transconductance g_m, then the input-referred noise *voltage* of a differential pair is $\sqrt{2}$ times that of a common-source stage. This is simply because the former includes twice as many devices in the signal path, as exemplified by the two series voltage sources in Fig. 7.46. (Since the noise sources are uncorrelated, their powers add.) It is also important to recognize that, with the assumption of equal device transconductances, a differential pair consumes twice as much power as a common-source stage if the transistors have the same dimensions.

The noise modeling of Fig. 7.46 can readily account for $1/f$ noise of the transistors as well. Placing the voltage sources given by $K/(C_{ox}WL)$ in series with each gate, we can

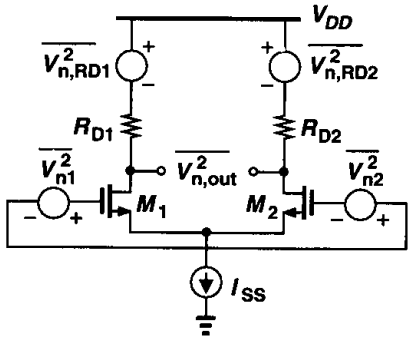


Figure 7.46 Alternative method of calculating the input-referred noise.

rewrite (7.88) as

$$\overline{V^2_{n,in,tot}} = 8kT\left(\frac{2}{3g_m} + \frac{1}{g_m^2 R_D}\right) + \frac{2K}{C_{ox}WL}\frac{1}{f}. \tag{7.89}$$

Does the tail current source in Fig. 7.44 contribute noise? If the differential input signal is zero and the circuit is symmetric, then the noise in I_{SS} divides equally between M_1 and M_2, producing only a common-mode noise voltage at the output. On the other hand, for a small differential input, ΔV_{in}, we have

$$\Delta I_{D1} - \Delta I_{D2} = g_m \Delta V_{in} \tag{7.90}$$

$$= \sqrt{2\mu_n C_{ox}\frac{W}{L}\left(\frac{I_{SS} + I_n}{2}\right)}\Delta V_{in}, \tag{7.91}$$

where I_n denotes the noise in I_{SS} and $I_n \ll I_{SS}$. In essence, the noise modulates the transconductance of each device. Equation (7.91) can be written as

$$\Delta I_{D1} - \Delta I_{D2} \approx \sqrt{2\mu_n C_{ox}\frac{W}{L}\cdot\frac{I_{SS}}{2}}\left(1 + \frac{I_n}{2I_{SS}}\right)\Delta V_{in} \tag{7.92}$$

$$= g_{m0}\left(1 + \frac{I_n}{2I_{SS}}\right)\Delta V_{in}, \tag{7.93}$$

where g_{m0} is the transconductance of the noiseless circuit. Equation (7.93) suggests that as the circuit departs from equilibrium, I_n is more unevenly divided between M_1 and M_2, thereby generating differential noise at the output. This effect is nonetheless usually negligible.

Example 7.13

Assuming the devices in Fig. 7.47(a) operate in saturation and the circuit is symmetric, calculate the input-referred noise voltage.

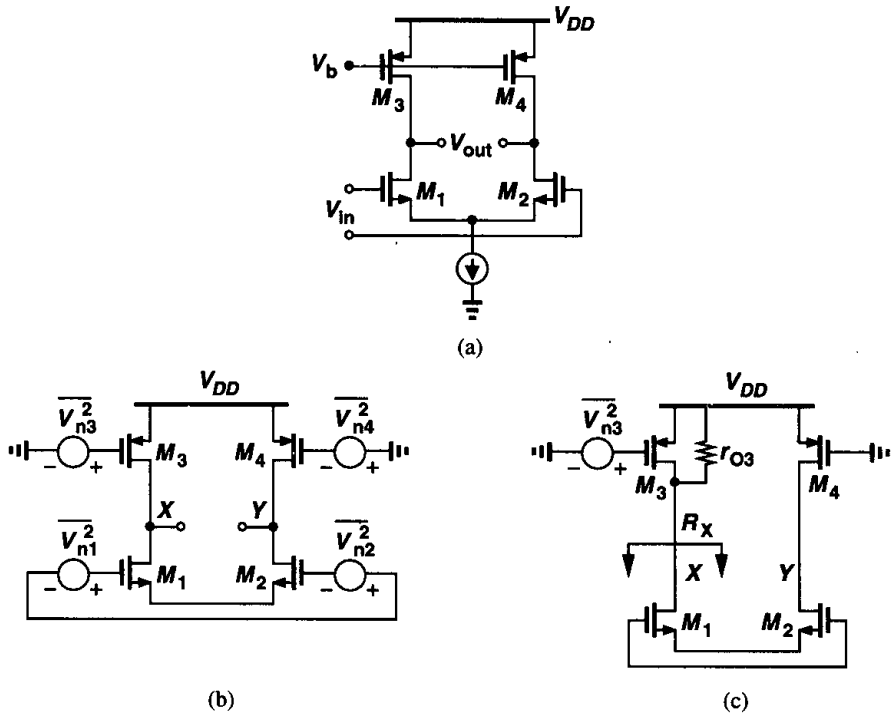


Figure 7.47

Solution

Since the thermal and $1/f$ noise of M_1 and M_2 can be modeled as voltage sources in series with the input, we only need to refer the noise of M_3 and M_4 to the input. Let us calculate the output noise contributed by M_3. The drain noise current of M_3 is divided between r_{O3} and the resistance seen looking into the drain of M_1 [Fig. 7.47(c)]. From Chapter 5, this resistance equals $R_X = r_{O4} + 2r_{O1}$. Denoting the resulting noise currents flowing through r_{O3} and R_X by I_{nA} and I_{nB}, respectively, we have

$$I_{nA} = g_{m3} V_{n3} \frac{r_{O4} + 2r_{O1}}{2r_{O4} + 2r_{O1}} \tag{7.94}$$

and

$$I_{nB} = g_{m3} V_{n3} \frac{r_{O3}}{2r_{O4} + 2r_{O1}}. \tag{7.95}$$

The former produces a noise voltage $g_{m3} V_{n3} r_{O3}(r_{O4} + 2r_{O1})/(2r_{O4} + 2r_{O1})$ at node X with respect to ground whereas the latter flows through M_1, M_2, and r_{O4}, generating $g_{m3} V_{n3} r_{O3} r_{O4}/(2r_{O4} + 2r_{O1})$ at node Y with respect to ground. Thus, the total differential output noise due to M_3 is

equal to

$$V_{nXY} = V_{nX} - V_{nY} \tag{7.96}$$

$$= g_{m3} V_{n3} \frac{r_{O3} r_{O1}}{r_{O3} + r_{O1}}. \tag{7.97}$$

(The reader can verify that V_{nY} must be *subtracted* from V_{nX}.)

Equation (7.97) implies that the noise current of M_3 is simply multiplied by the parallel combination of r_{O1} and r_{O3} to produce the differential output voltage. This is of course not surprising because, as depicted in Fig. 7.48, the effect of V_{n3} at the output can also be derived by decomposing

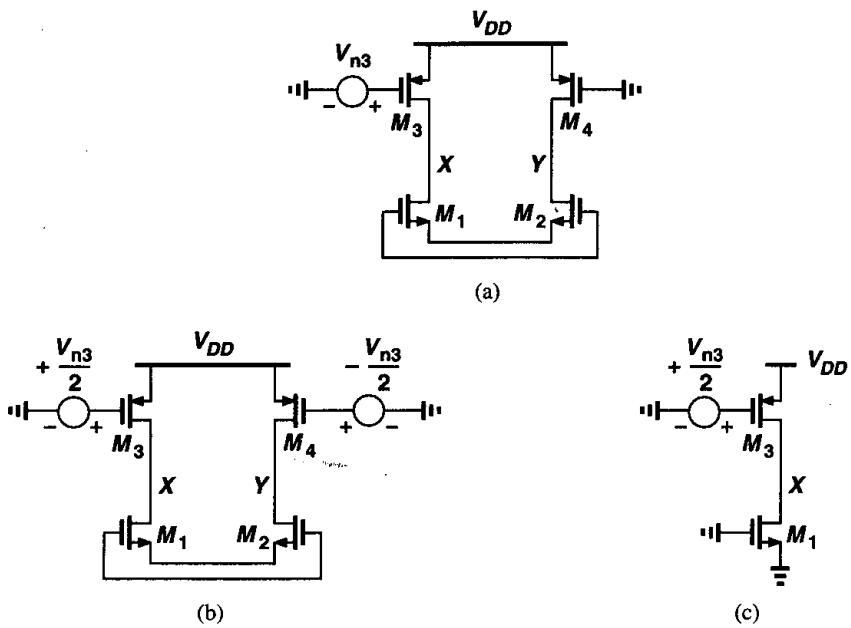


Figure 7.48 Calculation of input-referred noise in a differential pair with current-source loads.

V_{n3} into two differential components applied to the gates of M_3 and M_4 and subsequently using the half-circuit concept. Since this calculation relates to a *single* noise source, we can temporarily ignore the random nature of noise and treat V_{n3} and the circuit as familiar deterministic, linear components.

Applying (7.97) to M_4 as well and adding the resulting powers, we have

$$\overline{V_{n,out}^2}|_{M3,M4} = g_{m3}^2 (r_{O1} \| r_{O3})^2 \overline{V_{n3}^2} + g_{m4}^2 (r_{O2} \| r_{O4})^2 \overline{V_{n4}^2} \tag{7.98}$$

$$= 2 g_{m3}^2 (r_{O1} \| r_{O3})^2 \overline{V_{n3}^2}. \tag{7.99}$$

To refer the noise to the input, we divide (7.99) by $g_{m1}^2 (r_{O1} \| r_{O3})^2$, obtaining the *total* input-referred noise voltage per unit bandwidth as

$$\overline{V_{n,in}^2} = 2\overline{V_{n1}^2} + 2\frac{g_{m3}^2}{g_{m1}^2} \overline{V_{n3}^2}, \tag{7.100}$$

which, upon substitution for $\overline{V_{n1}^2}$ and $\overline{V_{n3}^2}$, reduces to:

$$\overline{V_{n,in}^2} = 8kT\left(\frac{2}{3g_{m1}} + \frac{2g_{m3}}{3g_{m1}^2}\right) + \frac{2K_N}{C_{ox}(WL)_1 f} + \frac{2K_P}{C_{ox}(WL)_3 f}\frac{g_{m3}^2}{g_{m1}^2}. \tag{7.101}$$

The effect of noise must be studied for many other analog circuits as well. For example, feedback systems, op amps, and bandgap references exhibit interesting and important noise characteristics. We return to these topics in other chapters.

7.6 Noise Bandwidth

The total noise corrupting a signal in a circuit results from all of the frequency components that fall in the bandwidth of the circuit. Consider a multipole circuit having the output noise spectrum shown in Fig. 7.49(a). Since the noise components above ω_{p1} are not negligible,

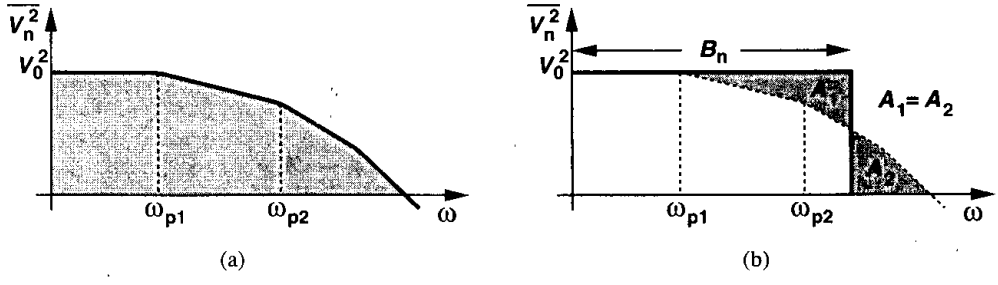


Figure 7.49 (a) Output noise spectrum of a circuit, (b) concept of noise bandwidth.

the total output noise must be evaluated by calculating the total area under the spectral density:

$$\overline{V_{n,out,tot}^2} = \int_0^\infty \overline{V_{n,out}^2} df. \tag{7.102}$$

However, as depicted in Fig. 7.49(b), it is sometimes helpful to represent the total noise simply as $V_0^2 \cdot B_n$, where the bandwidth B_n is chosen such that

$$V_0^2 \cdot B_n = \int_0^\infty \overline{V_{n,out}^2} df. \tag{7.103}$$

Called the "noise bandwidth," B_n allows a fair comparison of circuits that exhibit the same low-frequency noise, V_0^2, but different high-frequency transfer functions. As an exercise, the reader can prove that the noise bandwidth of a one-pole system is equal to $\pi/2$ times the pole frequency.

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.

7.1. A common-source stage incorporates a 50-μm/0.5-μm NMOS device biased at $I_D = 1$ mA along with a load resistor of 2 kΩ. What is the total input-referred thermal noise voltage in a 100-MHz bandwidth?

7.2. Consider the common-source stage of Fig. 7.35. Assume $(W/L)_1 = 50/0.5$, $I_{D1} = I_{D2} = 0.1$ mA, and $V_{DD} = 3$ V. If the contribution of M_2 to the input-referred noise voltage (not voltage squared) must be one-fifth of that of M_1, what is the maximum output voltage swing of the amplifier?

7.3. Using the distributed model of Fig. 7.19(c) and ignoring the channel thermal noise, prove that, for gate noise calculations, a distributed gate resistance of R_G can be replaced by a lumped resistance equal to $R_G/3$. (Hint: model the noise of R_{Gj} by a series voltage source and calculate the total drain noise current. Watch for correlated sources of noise.)

7.4. Prove that the output noise current of Fig. 7.33(c) is given by Eq. (7.54).

7.5. Calculate the input-referred noise voltage of the circuit shown in Fig. 7.50 and compare the result with Eq. (7.59).

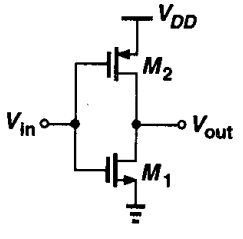


Figure 7.50

7.6. Calculate the input-referred thermal noise voltage of each circuit in Fig. 7.51. Assume $\lambda = \gamma = 0$.

7.7. Calculate the input-referred thermal noise voltage of each circuit in Fig. 7.52. Assume $\lambda = \gamma = 0$.

7.8. Calculate the input-referred thermal noise voltage and current of each circuit in Fig. 7.53. Assume $\lambda = \gamma = 0$.

7.9. Calculate the input-referred thermal noise voltage and current of each circuit in Fig. 7.54. Assume $\lambda = \gamma = 0$.

7.10. Calculate the input-referred $1/f$ noise voltage and current of Fig. 7.40 if the two capacitors are removed.

7.11. Calculate the input-referred $1/f$ noise voltage of the source follower shown in Fig. 7.42.

7.12. Assuming $\lambda = \gamma = 0$, calculate the input-referred thermal noise voltage of each circuit in Fig. 7.55. For part (a), assume $g_{m3,4} = 0.5g_{m5,6}$.

7.13. Consider the degenerated common-source stage shown in Fig. 7.56.

(a) Calculate the input-referred thermal noise voltage if $\lambda = \gamma = 0$.

(b) Suppose linearity requirements necessitate that the dc voltage drop across R_S be equal to the overdrive voltage of M_1. How does the thermal noise contributed by R_S compare with that contributed by M_1?

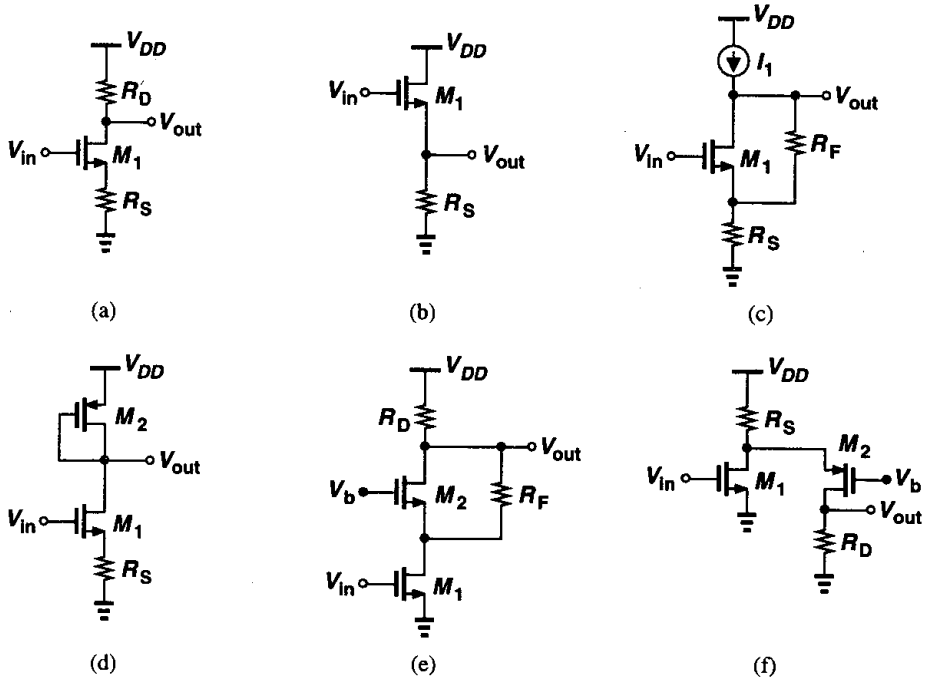


Figure 7.51

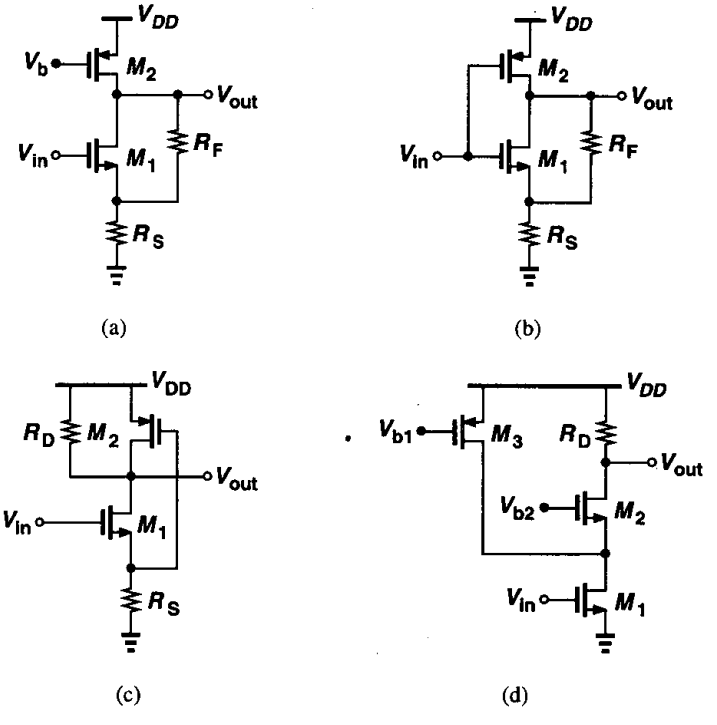


Figure 7.52

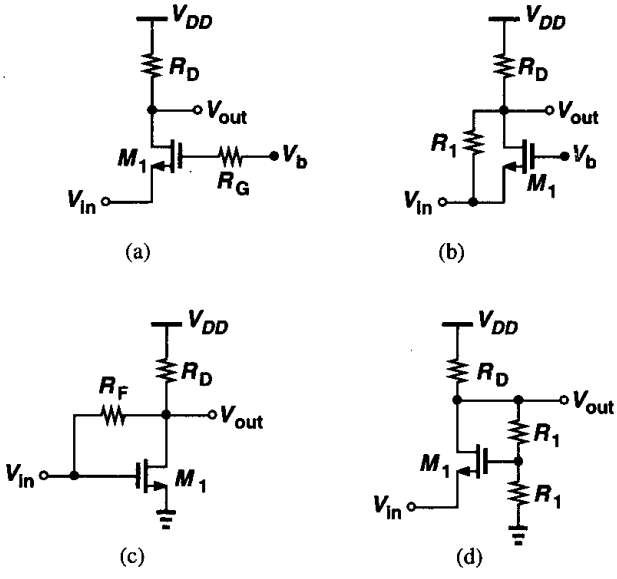


Figure 7.53

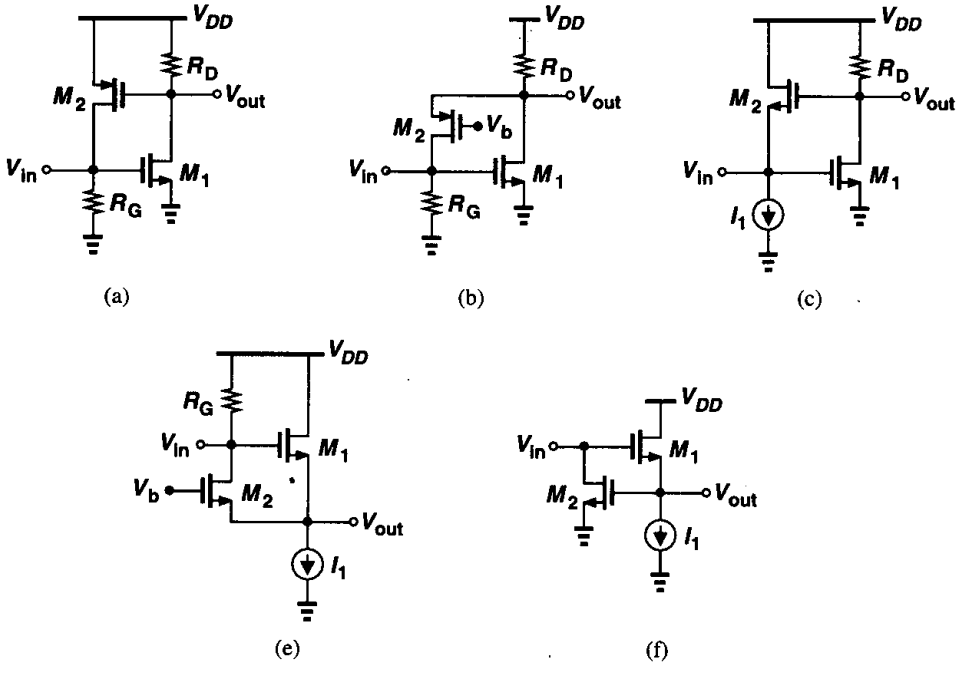


Figure 7.54

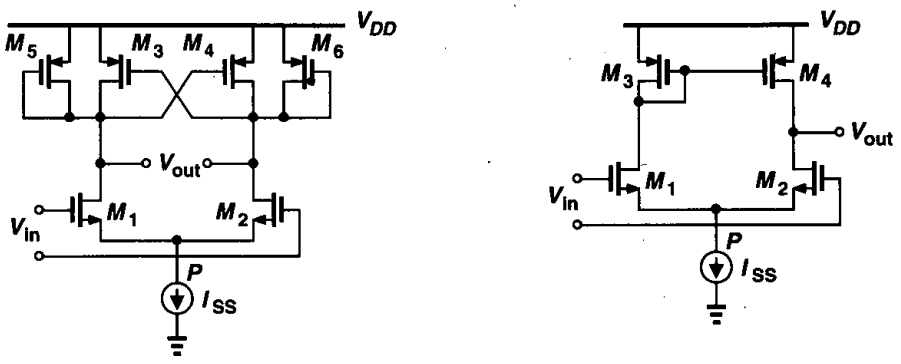


Figure 7.55

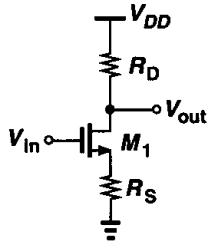


Figure 7.56

7.14. Explain why Miller's theorem cannot be applied to calculate the effect of the thermal noise of a floating resistor.

7.15. The circuit of Fig. 7.18 is designed with $(W/L)_1 = 50/0.5$ and $I_{D1} = 0.05$ mA. Calculate the total rms thermal noise voltage at the output in a 50-MHz bandwidth.

7.16. For the circuit shown in Fig. 7.58, calculate the total output thermal and $1/f$ noise in a bandwidth $[f_L, f_H]$. Assume $\lambda \neq 0$ but neglect other capacitances.

7.17. Suppose in the circuit of Fig. 7.35, $(W/L)_{1,2} = 50/0.5$ and $I_{D1} = |I_{D2}| = 0.5$ mA. What is the input-referred thermal noise voltage?

7.18. The circuit of Fig. 7.35 is modified as depicted in Fig. 7.59.

(a) Calculate the input-referred thermal noise voltage.

(b) For a given bias current and output voltage swing, what value of R_S minimizes the input-referred thermal noise?

7.19. A common-gate stage incorporates an NMOS device with $W/L = 50/0.5$ biased at $I_D = 1$ mA and a load resistor of 1 kΩ. Calculate the input-referred thermal noise voltage and current.

7.20. The circuit of Fig. 7.39 is designed with $(W/L)_1 = 50/0.5$ and $I_{D1} = I_{D2} = 0.05$ mA and $R_D = 1$ kΩ.

(a) Determine $(W/L)_2$ such that the contribution of M_2 to the input-referred thermal noise current (not current squared) is one-fifth of that due to R_D.

(b) Now calculate the minimum value of V_b to place M_2 at the edge of the triode region. What is the maximum allowable output voltage swing?

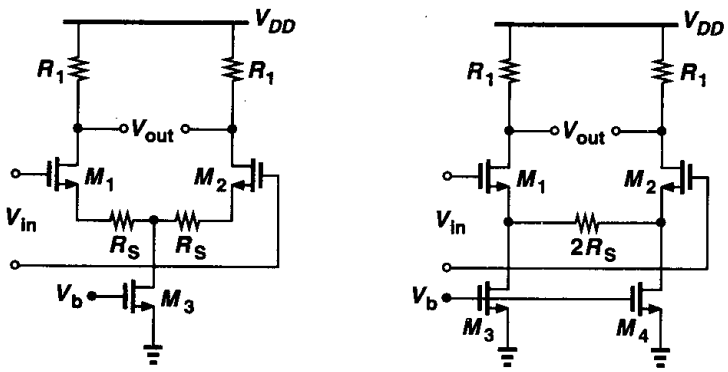


Figure 7.57

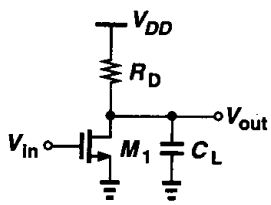


Figure 7.58

7.21. Design the circuit of Fig. 7.39 for an input-referred thermal noise voltage of 3 nV/$\sqrt{\text{Hz}}$ and maximum output swing. Assume $I_{D1} = I_{D2} = 0.5$ mA.

7.22. Consider the circuit of Fig. 7.40. If $(W/L)_{1-3} = 50/0.5$ and $I_{D1-3} = 0.5$ mA, determine the input-referred thermal noise voltage and current.

7.23. The circuit of Fig. 7.40 is designed with $(W/L)_1 = 50/0.5$ and $I_{D1-3} = 0.5$ mA. If an output swing of 2 V is required, estimate by iteration the dimensions of M_2 and M_3 such that the input-referred thermal noise current is minimum.

7.24. The source follower of Fig. 7.42 is to provide an output resistance of 100 Ω with a bias current of 0.1 mA.

(a) Calculate $(W/L)_1$.

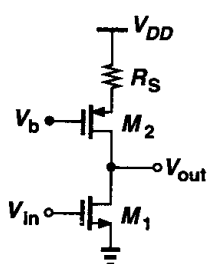


Figure 7.59

(b) Determine $(W/L)_2$ such that the input-referred thermal noise voltage (not voltage squared) contributed by M_2 is one-fifth of that due to M_1. What is the maximum output swing?

7.25. The cascode stage of Fig. 7.43(a) exhibits a capacitance C_X from node X to ground. Neglecting other capacitances, determine the input-referred thermal noise voltage.

7.26. Determine the input-referred thermal and $1/f$ noise voltages of the circuits shown in Fig. 7.57 and compare the results. Assume the circuits draw equal supply currents.

References

1. L. W. Couch, *Digital and Analog Communication Systems,* Fourth Ed., New York: Macmillan Co., 1993.
2. S. M. Sze, *Physics of Semiconductor Devices,* Second Ed., New York: Wiley, 1981.
3. Y. Tsividis, *Operation and Modeling of the MOS Transistor,* Second Ed., Boston: McGraw-Hill, 1999.
4. A. A. Abidi, "High-Frequency Noise Measurements on FETs with Small Dimensions," *IEEE Trans. Electron Devices,* vol. 33, pp. 1801–1805, Nov. 1986.
5. H. A. Haus et al., "Representation of Noise in Linear Twoports," *Proc. IRE,* vol. 48, pp. 69–74, Jan. 1960.

Chapter 8

Feedback

On a mild August morning in 1921, Harold Black was riding the ferry from New York to New Jersey, where he worked at Bell Laboratories. Black and many other researchers had been investigating the problem of nonlinearity in amplifiers used in long-distance telephone networks, seeking a practical solution. While reading the newspaper on the ferry, Black was suddenly struck by an idea and began to draw a diagram on the newspaper, which would later be used as the evidence in his patent application. The idea is known to us as the negative feedback amplifier.

Feedback is a powerful technique that finds wide application in analog circuits. For example, negative feedback allows high-precision signal processing and positive feedback makes it possible to build oscillators. In this chapter, we consider only negative feedback and use the term feedback to mean that.

We begin with a general view of feedback circuits, describing important benefits that result from feedback. Next, we study four feedback topologies and their properties. Finally, we examine the effects of loading in feedback amplifiers.

8.1 General Considerations

Fig. 8.1 shows a negative feedback system, where $H(s)$ and $G(s)$ are called the feedforward and the feedback networks, respectively. Since the output of $G(s)$ is equal to $G(s)Y(s)$, the input to $H(s)$, called the feedback error, is given by $X(s) - G(s)Y(s)$. That is,

$$Y(s) = H(s)[X(s) - G(s)Y(s)]. \tag{8.1}$$

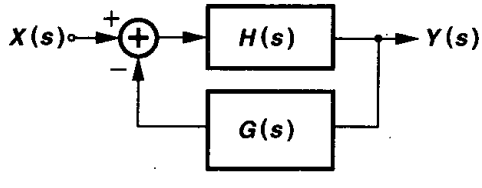


Figure 8.1 General feedback system.

Thus,

$$\frac{Y(s)}{X(s)} = \frac{H(s)}{1 + G(s)H(s)}. \tag{8.2}$$

We call $H(s)$ the "open-loop" transfer function and $Y(s)/X(s)$ the "closed-loop" transfer function. In most cases of interest in this book, $H(s)$ represents an amplifier and $G(s)$ is a frequency-independent quantity. In other words, a fraction of the output signal is sensed and compared with the input, generating an error term. In a well-designed negative feedback system, the error term is minimized, thereby making the output of $G(s)$ an accurate "copy" of the input and hence the output of the system a faithful replica of the input (Fig. 8.2). We also say the input of $H(s)$ is a "virtual ground" because the signal amplitude at this point is very small. In subsequent developments, we replace $G(s)$ by a frequency-independent quantity β and call it the "feedback factor."

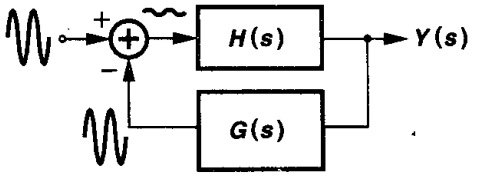


Figure 8.2 Similarity between output of feedback network and input signal.

It is instructive to identify four elements in the feedback system of Fig. 8.1: (1) the feedforward amplifier, (2) a means of sensing the output, (3) the feedback network, (4) a means of generating the feedback error. These elements exist in every feedback system, even though they may not be obvious in cases such as a simple common-source stage with resistive degeneration.

8.1.1 Properties of Feedback Circuits

Before proceeding to the analysis of feedback circuits, we study some simple examples to describe the benefits of negative feedback as well as identify the above four elements in each case.

Gain Desensitization Consider the common-source stage shown in Fig. 8.3(a), where the voltage gain is equal to $g_{m1}r_{O1}$. A critical drawback of this circuit is the poor definition of the gain: both g_{m1} and r_{O1} vary with process and temperature. Now suppose the circuit is configured as in Fig. 8.3(b), where the gate bias of M_1 is set by means not shown here (Chapter 12). Let us calculate the overall voltage gain of the circuit at relatively low frequencies such that C_2 does not load the output node, i.e., $V_{out}/V_X = -g_{m1}r_{O1}$. Since $(V_{out} - V_X)C_2 s = (V_X - V_{in})C_1 s$, we have

$$\frac{V_{out}}{V_{in}} = -\frac{1}{\left(1 + \dfrac{1}{g_{m1}r_{O1}}\right)\dfrac{C_2}{C_1} + \dfrac{1}{g_{m1}r_{O1}}}. \tag{8.3}$$

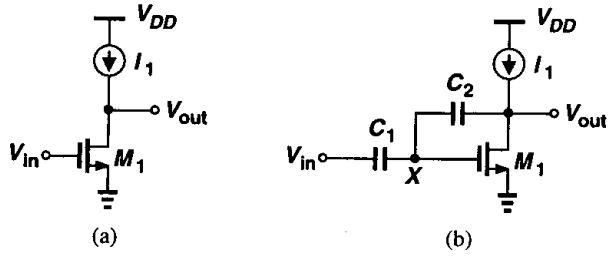


Figure 8.3 (a) Simple common-source stage, (b) circuit of (a) with feedback.

If $g_{m1}r_{O1}$ is sufficiently large, the $1/(g_{m1}r_{O1})$ terms in the denominator can be neglected, yielding

$$\frac{V_{out}}{V_{in}} = -\frac{C_1}{C_2}. \tag{8.4}$$

Compared to $g_{m1}r_{O1}$, this gain can be controlled with much higher accuracy because it is given by the *ratio* of two capacitors. If C_1 and C_2 are made of the same material, then process and temperature variations do not change C_1/C_2.

The above example reveals that negative feedback provides gain "desensitization," i.e., the closed-loop gain is much less sensitive to device parameters than the open-loop gain is. Illustrated for a more general case in Fig. 8.4, this property can be quantified by writing

$$\frac{Y}{X} = \frac{A}{1+\beta A} \tag{8.5}$$

$$\approx \frac{1}{\beta}\left(1 - \frac{1}{\beta A}\right), \tag{8.6}$$

where we have assumed $\beta A \gg 1$. We note that the closed-loop gain is determined, to the first order, by the feedback factor, β. More importantly, even if the open-loop gain, A, varies by a factor of, say, 2, Y/X varies by a small percentage because $1/(\beta A) \ll 1$.

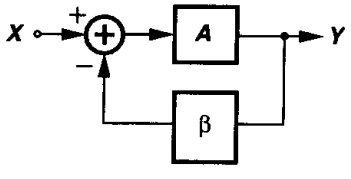


Figure 8.4 Simple feedback system.

Called the "loop gain," the quantity βA plays an important role in feedback systems.[1] We see from (8.6) that the higher βA is, the less sensitive Y/X will be to variations in A.

[1] Loop gain and open-loop gain must not be confused with each other.

From another perspective, the accuracy of the closed-loop gain improves by maximizing A or β. Note that as β increases, the closed-loop gain, $Y/X \approx 1/\beta$, decreases, suggesting a trade-off between precision and the closed-loop gain. In other words, we begin with a high-gain amplifier and apply feedback to obtain a low, but less sensitive closed-loop gain. Another conclusion here is that the output of the feedback network is equal to $X \cdot A/(1 + \beta A)$, approaching A as βA becomes much greater than unity. This result agrees with the illustration in Fig. 8.2.

The calculation of the loop gain usually proceeds as follows. As illustrated in Fig. 8.5, we set the main input to zero, break the loop at some point, inject a test signal in the "right direction," follow the signal around the loop, and obtain the value that returns to the break point. The negative of the transfer function thus derived is the loop gain. Note that the loop gain is a dimensionless quantity. In Fig. 8.5, we have $V_t\beta(-1)A = V_F$ and hence $V_F/V_t = -\beta A$. Similarly, as depicted in Fig. 8.6, for the simple feedback circuit, we can write

$$V_t \frac{C_2}{C_1 + C_2}(-g_{m1}r_{O1}) = V_F, \tag{8.7}$$

that is,

$$\frac{V_F}{V_t} = -\frac{C_2}{C_1 + C_2} g_{m1}r_{O1}. \tag{8.8}$$

Note that the loading of C_2 on the output is neglected here. This issue will be addressed in Section 8.3.

It is also interesting to identify the four elements of feedback in the circuit of Fig. 8.3(b). Transistor M_1 and current source I_1 constitute the feedforward amplifier. Capacitor C_2

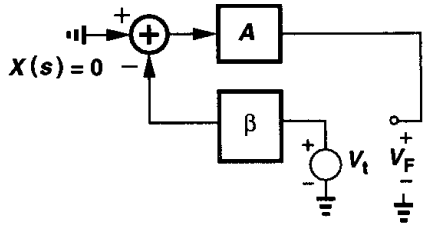


Figure 8.5 Computation of loop gain.

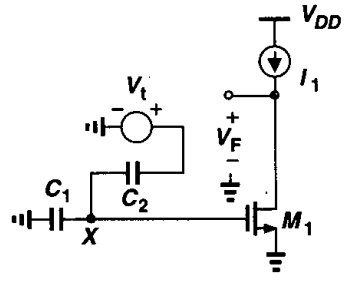


Figure 8.6 Computation of loop gain in a simple feedback circuit.

senses the output voltage and converts it to a current feedback signal, which is then added to the current produced by V_{in} through C_1. Note that the feedback is negative even though the currents through C_1 and C_2 are *added* because the feedforward amplifier itself provides a negative gain.

We should emphasize that the desensitization of gain by feedback leads to many other properties of feedback systems. Our examination of Eq. 8.6 indicates that large variations in A affect Y/X negligibly if βA is large. Such variations can arise from different sources: process, temperature, frequency, and loading. For example, if A drops at high frequencies, Y/X varies to a lesser extent, and the flat bandwidth is increased. Similarly, if A decreases because the amplifier drives a heavy load, Y/X is not affected much. These concepts become clearer below.

Terminal Impedance Modification As a second example, let us study the circuit shown in Fig. 8.7(a), where a capacitive voltage divider senses the output voltage of a common-gate stage, applying the result to the gate of current source M_2 and hence returning a current feedback signal to the input.[2] Our objective is to compute the input resistance at relatively low frequencies with and without feedback. Neglecting channel-length modulation and breaking the feedback loop [Fig. 8.7(b)], we have

$$R_{in,open} = \frac{1}{g_{m1} + g_{mb1}}. \tag{8.9}$$

For the closed-loop circuit, as depicted in Fig. 8.7(c), we write: $V_{out} = (g_{m1} + g_{mb1})V_X R_D$ and

$$V_P = V_{out}\frac{C_1}{C_1 + C_2} \tag{8.10}$$

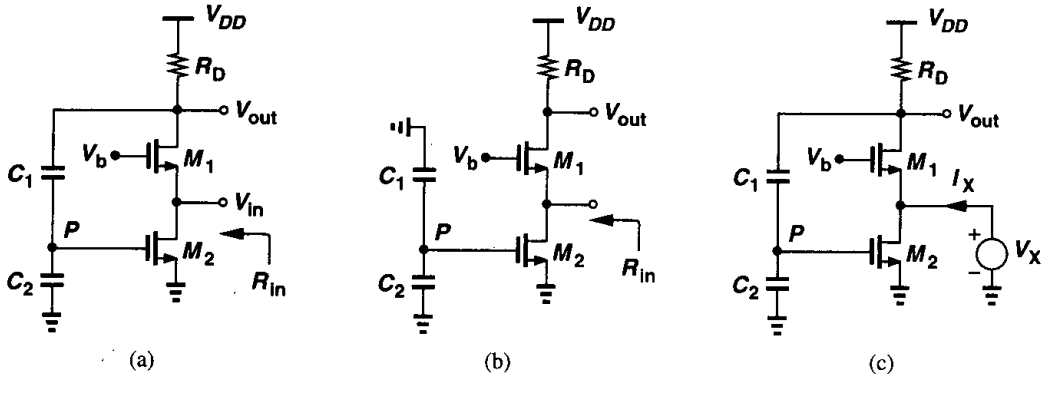


Figure 8.7 (a) Common-gate circuit with feedback, (b) open-loop circuit, (c) calculation of input resistance.

[2]The bias network for M_2 is not shown.

$$= (g_{m1} + g_{mb1})V_X R_D \frac{C_1}{C_1 + C_2}. \tag{8.11}$$

Thus, the small-signal drain current of M_2 equals $g_{m2}(g_{m1} + g_{mb1})V_X R_D C_1/(C_1 + C_2)$. Adding this current to the drain current of M_1 with proper polarity yields I_X:

$$I_X = (g_{m1} + g_{mb1})V_X + g_{m2}(g_{m1} + g_{mb1})\frac{C_1}{C_1 + C_2} R_D V_X \tag{8.12}$$

$$= (g_{m1} + g_{mb1})\left(1 + g_{m2} R_D \frac{C_1}{C_1 + C_2}\right) V_X. \tag{8.13}$$

It follows that

$$R_{in,closed} = V_X / I_X \tag{8.14}$$

$$= \frac{1}{g_{m1} + g_{mb1}} \frac{1}{1 + g_{m2} R_D \dfrac{C_1}{C_1 + C_2}}. \tag{8.15}$$

We therefore conclude that this type of feedback reduces the input resistance by a factor of $1 + g_{m2}R_D C_1/(C_1 + C_2)$. The reader can prove that the quantity $g_{m2}R_D C_1/(C_1 + C_2)$ is the loop gain.

We also identify the four elements of feedback in the circuit of Fig. 8.7(a). The feedforward amplifier consists of M_1 and R_D, the output is sensed by C_1 and C_2, the feedback network comprises C_1, C_2, and M_2, and the subtraction occurs in the current domain at the input terminal.

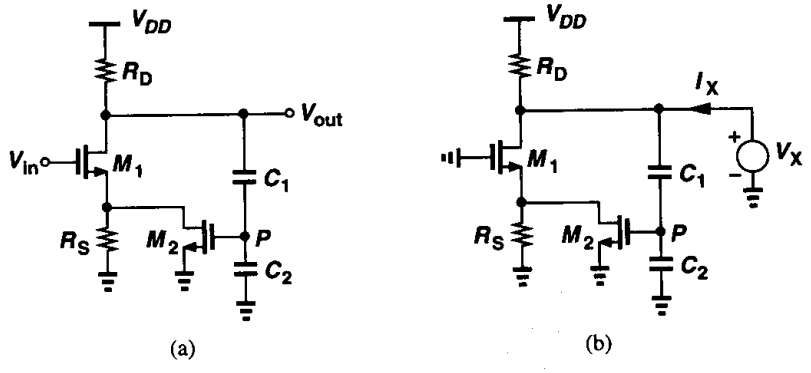


Figure 8.8 (a) CS stage with feedback, (b) calculation of output resistance.

Let us now consider the circuit of Fig. 8.8(a) as an example of output impedance modification by feedback. Here M_1, R_S, and R_D constitute a common-source stage and C_1, C_2,

and M_2 sense the output voltage,[3] returning a current equal to $[C_1/(C_1 + C_2)]V_{out}g_{m2}$ to the source of M_1. The reader can prove that the feedback is indeed negative. To compute the output resistance at relatively low frequencies, we set the input to zero [Fig. 8.8(b)] and write

$$I_{D1} = V_X \frac{C_1}{C_1 + C_2} g_{m2} \frac{R_S}{R_S + \dfrac{1}{g_{m1} + g_{mb1}}}. \tag{8.16}$$

Since $I_X = V_X/R_D + I_{D1}$, we have

$$\frac{V_X}{I_X} = \frac{R_D}{1 + \dfrac{g_{m2}R_S(g_{m1} + g_{mb1})R_D}{(g_{m1} + g_{mb1})R_S + 1} \dfrac{C_1}{C_1 + C_2}}. \tag{8.17}$$

Equation (8.17) implies that this type of feedback decreases the output resistance. The denominator of (8.17) is indeed equal to one plus the loop gain.

Bandwidth Modification. The next example illustrates the effect of negative feedback on the bandwidth. Suppose the feedforward amplifier has a one-pole transfer function:

$$A(s) = \frac{A_0}{1 + \dfrac{s}{\omega_0}}, \tag{8.18}$$

where A_0 denotes the low-frequency gain and ω_0 is the 3-dB bandwidth. What is the transfer function of the closed-loop system? From (8.5), we have

$$\frac{Y}{X}(s) = \frac{\dfrac{A_0}{1 + \dfrac{s}{\omega_0}}}{1 + \beta \dfrac{A_0}{1 + \dfrac{s}{\omega_0}}} \tag{8.19}$$

$$= \frac{A_0}{1 + \beta A_0 + \dfrac{s}{\omega_0}} \tag{8.20}$$

$$= \frac{\dfrac{A_0}{1 + \beta A_0}}{1 + \dfrac{s}{(1 + \beta A_0)\omega_0}}. \tag{8.21}$$

The numerator of (8.21) is simply the closed-loop gain at low frequencies—as predicted by (8.5)—and the denominator reveals a pole at $(1 + \beta A_0)\omega_0$. Thus, the 3-dB bandwidth has increased by a factor $1 + \beta A_0$, albeit at the cost of a proportional reduction in the gain (Fig. 8.9).

[3] Biasing of M_2 is not shown.

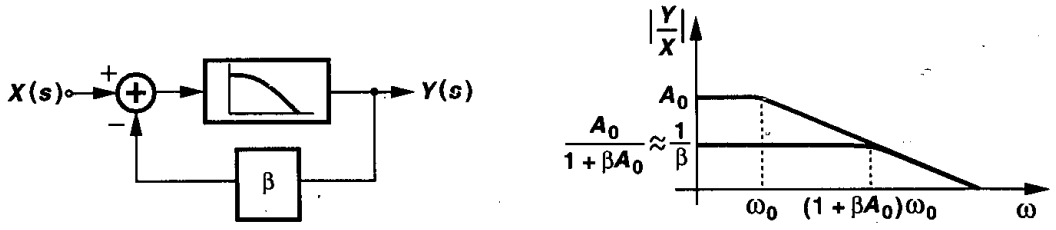


Figure 8.9 Bandwidth modification as a result of feedback.

The increase in the bandwidth fundamentally originates from the gain desensitization property of feedback. Recall from (8.6) that, if A is large enough, the closed-loop gain remains approximately equal to $1/\beta$ even if A experiences substantial variations. In the example of Fig. 8.9, A varies with frequency rather than process or temperature, but negative feedback still suppresses the effect of this variation. Of course, at high frequencies A drops to such low levels that βA becomes comparable with unity and the closed-loop gain falls below $1/\beta$.

Equation (8.21) suggests that the gain-bandwidth product of a one-pole system does not change with feedback, making the reader wonder how feedback improves the speed if a high gain is required. Suppose we need to amplify a 20-MHz square wave by a factor of 100 and maximum bandwidth but we have only a single-pole amplifier with an open-loop gain of 100 and 3-dB bandwidth of 10 MHz. If the input is applied to the open-loop amplifier, the response appears as shown in Fig. 8.10(a), exhibiting long risetime and falltime because the time constant is equal to $1/(2\pi f_{3-\text{dB}}) \approx 16$ ns.

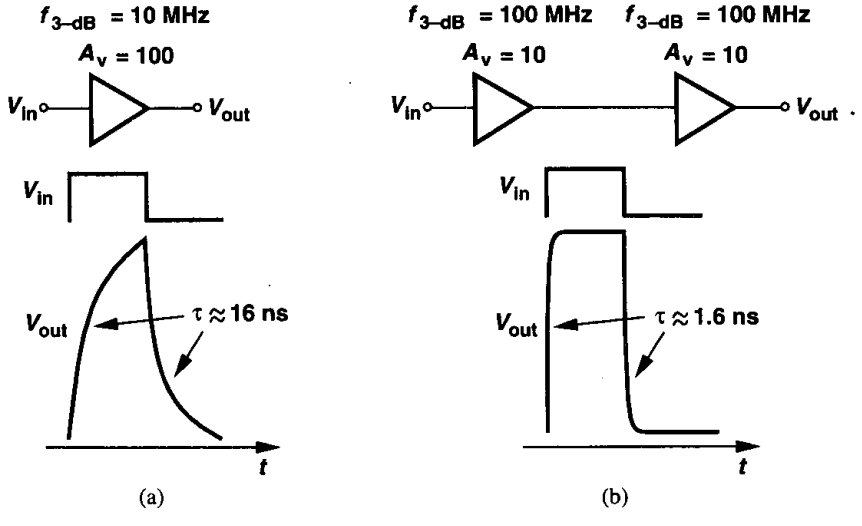


Figure 8.10 Amplification of a 20-MHz squarewave by (a) 20-MHz amplifier and (b) cascade of two 100-MHz feedback amplifiers.

Now suppose we apply feedback to the amplifier such that the gain and bandwidth are modified to 10 and 100 MHz, respectively. Placing two of these amplifiers in a cascade [Fig. 8.10(b)], we obtain a much faster response with an overall gain of 100. Of course, the cascade consumes twice as much power, but it would be quite difficult to achieve this performance by the original amplifier even if its power dissipation were doubled.

Nonlinearity Reduction A very important property of negative feedback is the suppression of nonlinearity in analog circuits. We defer the study of this effect to Chapter 13.

8.1.2 Types of Amplifiers

Most of the circuits studied thus far can be considered "voltage amplifiers" because they sense a voltage at the input and produce a voltage at the output. However, three other types of amplifiers can also be constructed such that they sense or produce currents. Shown in Fig. 8.11, the four configurations have quite different properties: (1) circuits sensing

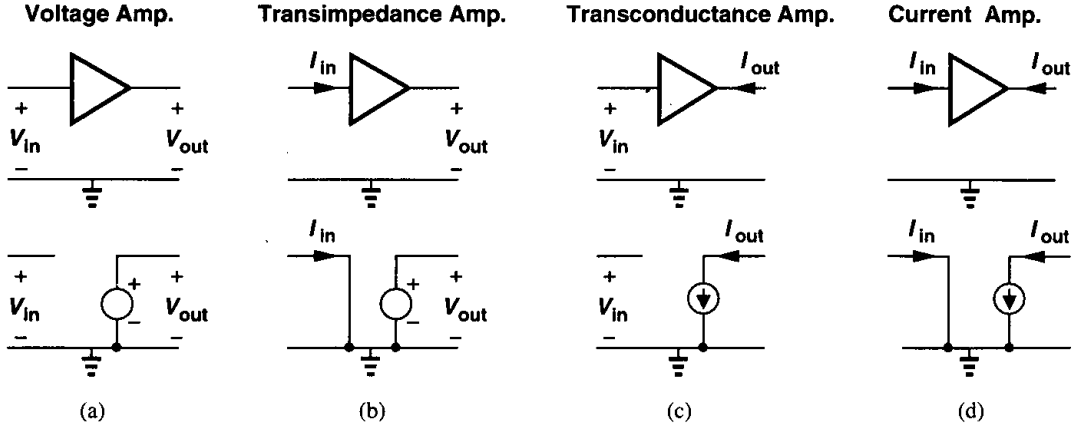


Figure 8.11 Types of amplifiers along with their idealized models.

a voltage must exhibit a high input impedance (as a voltmeter) whereas those sensing a current must provide a low input impedance (as a current meter); (2) circuits generating a voltage must exhibit a low output impedance (as a voltage source) while those generating a current must provide a high output impedance (as a current source). Note that the gains of transimpedance and transconductance[4] amplifiers have a dimension of resistance and conductance, respectively. For example, a transimpedance amplifier may have a gain of 2 kΩ, which means it produces a 2-V output in response to a 1-mA input. Also, we use the sign conventions depicted in Fig. 8.11, for example, the transimpedance $R_0 = V_{out}/I_{in}$ if I_{in} flows *into* the amplifier.

[4]This terminology is standard but not consistent. One should use either transimpedance and transadmittance or transresistance and transconductance.

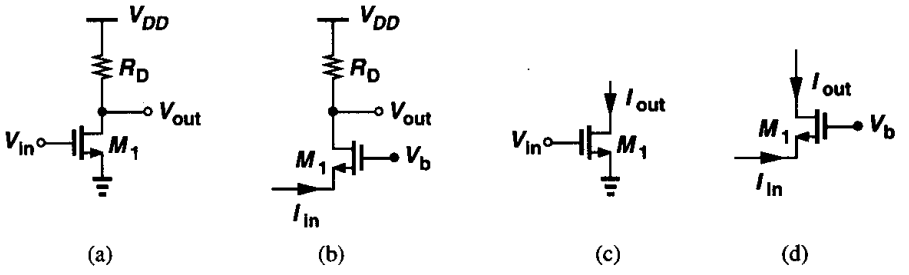


Figure 8.12 Simple implementations of four types of amplifiers.

Figure 8.12 illustrates simple implementations of each amplifier. In Fig. 8.12(a), a common-source stage senses and produces voltages and in Fig. 8.12(b), a common-gate circuit serves as a transimpedance amplifier, converting the source current to a voltage at the drain. In Fig. 8.12(c), a common-source transistor operates as a transconductance amplifier, generating an output current in response to an input voltage, and in Fig. 8.12(d), a common-gate device senses and produces currents.

The circuits of Fig. 8.12 may not provide adequate performance in many applications. For example, the circuits of Figs. 8.12(a) and (b) suffer from a relatively high output impedance. Fig. 8.13 depicts modifications that alter the output impedance or increase the gain.

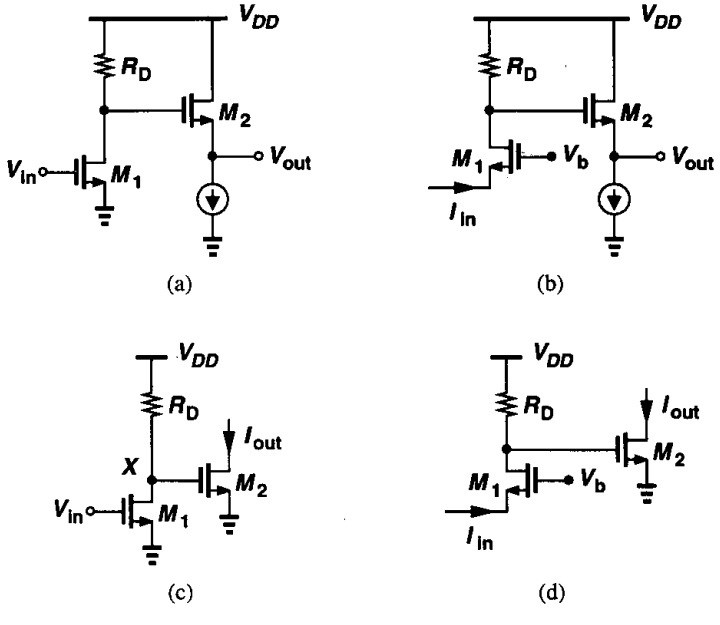


Figure 8.13 Four types of amplifiers with improved performance.

Example 8.1

Calculate the gain of the transconductance amplifier shown in Fig. 8.13(c).

Solution

The gain in this case is defined as $G_m = I_{out}/V_{in}$. That is,

$$G_m = \frac{V_X}{V_{in}} \cdot \frac{I_{out}}{V_X} \tag{8.22}$$

$$= -g_{m1}(r_{O1} \| R_D) \cdot g_{m2}. \tag{8.23}$$

While most familiar amplifiers are of voltage-voltage type, the other three configurations do find usage. For example, transimpedance amplifiers are an integral part of optical fiber receivers because they must sense the current produced by a photodiode, eventually generating a voltage that can be processed by subsequent circuits.

8.1.3 Sense and Return Mechanisms

Placing a circuit in a feedback loop requires sensing the output signal and returning (a fraction) of the result to the summing node at the input. With voltage or current quantities as input and output signals, we can identify four types of feedback: voltage-voltage, voltage-current, current-current, and current-voltage, where the first entry in each case denotes the quantity sensed at the *output* and the second the type of signal returned to the input.[5]

It is instructive to review methods of sensing and summing voltages or currents. To sense a voltage, we place a voltmeter *in parallel* with the corresponding port [Fig. 8.14(a)], ideally introducing no loading. When used in a feedback system, this type of sensing is also called "shunt feedback."

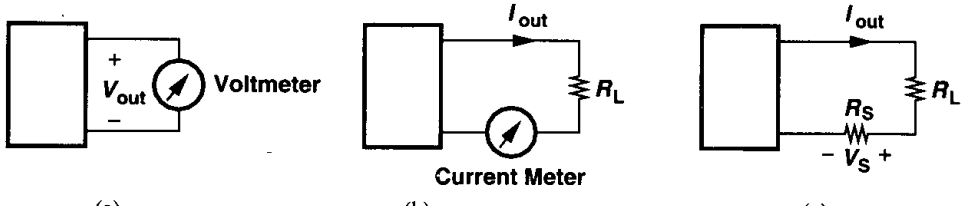


Figure 8.14 Sensing (a) a voltage by a voltmeter, (b) a current by a current meter, (c) a current by a small resistor.

To sense a current, a current meter is inserted *in series* with the signal [Fig. 8.14(b)], ideally exhibiting zero series resistance. Thus, this type of sensing is also called "series feedback." In practice, a small resistor replaces the current meter [Fig. 8.14(c)], with the voltage drop across the resistor serving as a measure of the output current.

The addition of the feedback signal and the input signal can be performed in the voltage domain or current domain. To add two quantities, we place them in series if they are

[5] Different authors use different orders or terminologies for the four types of feedback.

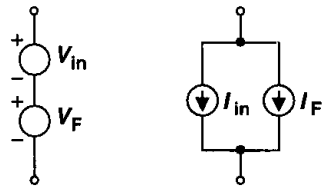


Figure 8.15 Addition of (a) voltages and (b) currents.

voltages and in parallel if they are currents (Fig. 8.15). While ideally having no influence on the operation of the open-loop amplifier itself, the feedback network in reality introduces loading effects that must be taken into account. This issue is discussed in Section 8.3.

To visualize the methods of Figs. 8.14 and 8.15, we consider a number of practical implementations. A voltage can be sensed by a resistive (or capacitive) divider in parallel with the port [Fig. 8.16(a)] and a current by placing a resistor in series with the wire and sensing

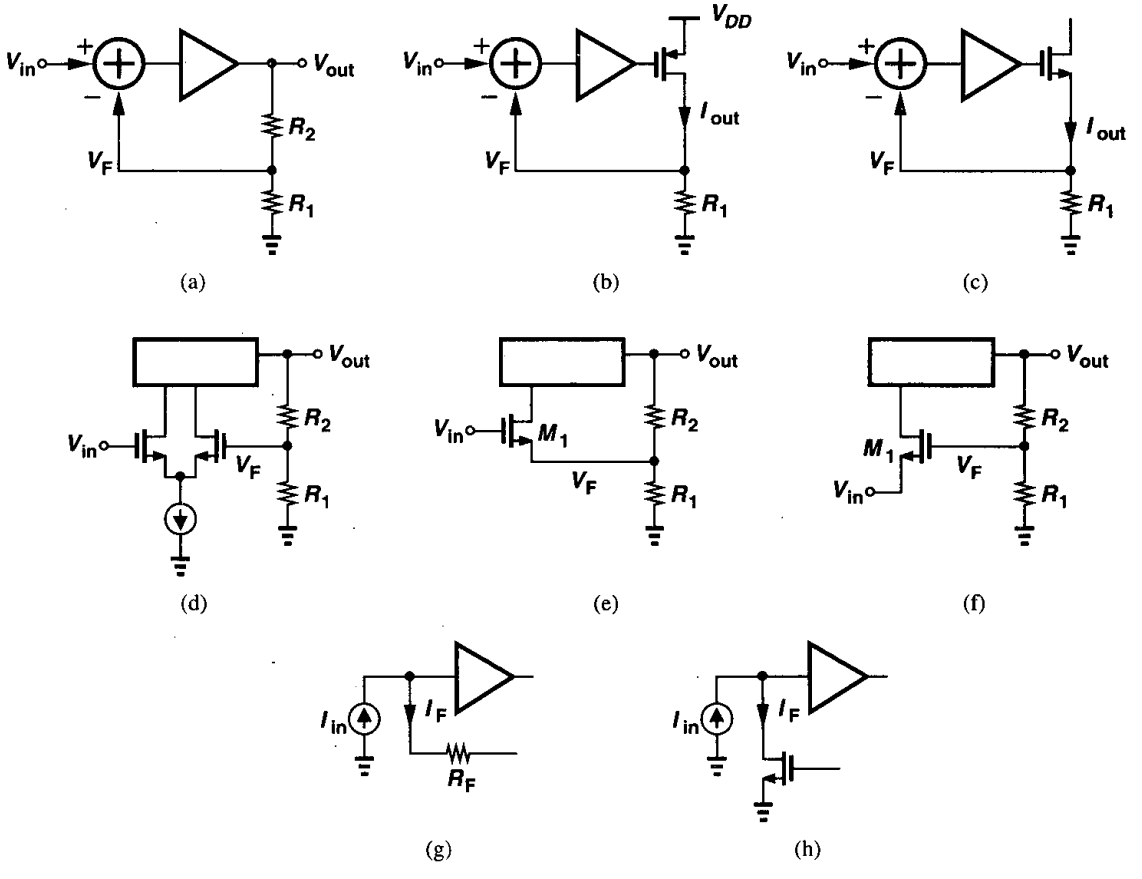


Figure 8.16 Practical means of sensing and adding voltages and currents.

the voltage across it [Figs. 8.16(b) and (c)]. To subtract two voltages, a differential pair can be used [Fig. 8.16(d)]. Alternatively, a single transistor can perform voltage subtraction as shown in Figs. 8.16(e) and (f) because I_{D1} is a function of $V_{in} - V_F$. Subtraction of currents can be accomplished as depicted in Figs. 8.16(g) or (h). Note that for voltage subtraction, the input and feedback signals are applied to *two* distinct nodes whereas for current subtraction they are applied to a single node. This observation proves helpful in identifying the type of feedback.

8.2 Feedback Topologies

8.2.1 Voltage-Voltage Feedback

This topology samples the output voltage and returns the feedback signal as a voltage.[6] Following the conceptual illustrations of Figs. 8.14 and 8.15, we note that the feedback network is connected in *parallel* with the output and in *series* with the input port (Fig. 8.17). An ideal feedback network in this case exhibits infinite input impedance and zero output

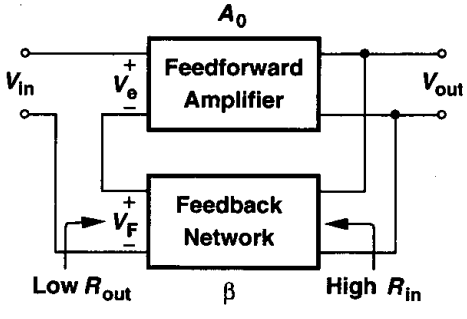


Figure 8.17 Voltage-voltage feedback.

impedance because it senses a voltage and generates a voltage. We can therefore write: $V_F = \beta V_{out}$, $V_e = V_{in} - V_F$, $V_{out} = A_0(V_{in} - \beta V_{out})$, and hence

$$\frac{V_{out}}{V_{in}} = \frac{A_0}{1 + \beta A_0}. \tag{8.24}$$

We recognize that βA_0 is the loop gain and the overall gain has dropped by $1 + \beta A_0$. Note that here both A_0 and β are dimensionless quantities.

As a simple example of voltage-voltage feedback, suppose we employ a differential voltage amplifier with single-ended output as the feedforward amplifier and a resistive divider as the feedback network [Fig. 8.18(a)]. The divider senses the output voltage, producing a fraction thereof as the feedback signal V_F. Following the block diagram of Fig. 8.17, we place V_F in series with the input of the amplifier to perform subtraction of voltages [Fig. 8.18(b)].

[6]This configuration is also called "series-shunt" feedback, where the first term refers to the *input* connection and the second to the *output* connection.

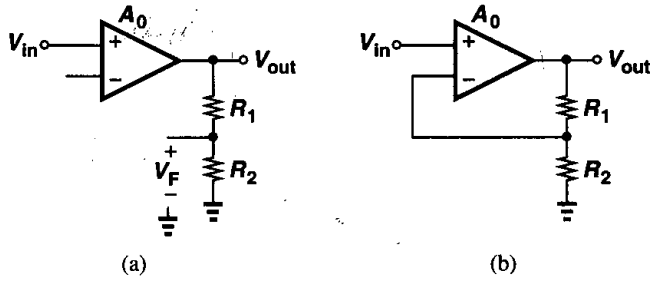


Figure 8.18 (a) Amplifier with output sensed by a resistive divider, (b) voltage-voltage feedback amplifier.

How does voltage-voltage feedback modify the input and output impedances? Let us first consider the output impedance. Recall that a negative feedback system attempts to make the output an accurate replica of the input. Now suppose, as shown in Fig. 8.19, we load the output by a resistor, gradually decreasing its value. While in the open-loop configuration the

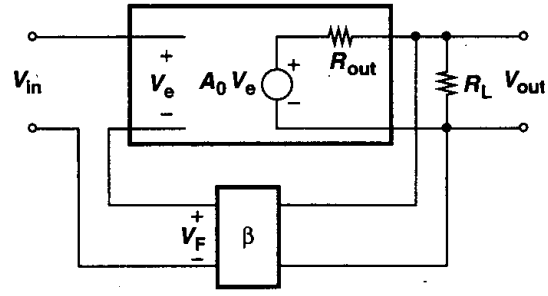


Figure 8.19 Effect of voltage-voltage feedback on output resistance.

output would simply drop in proportion to $R_L/(R_L + R_{out})$, in the feedback system, V_{out} is maintained as a reasonable replica of V_{in} even though R_L decreases. That is, so long as the loop gain remains much greater than unity, $V_{out}/V_{in} \approx 1/\beta$, regardless of the value of R_L. From another point of view, since the circuit stabilizes the output voltage amplitude despite load variations, it behaves as a *voltage* source, thus exhibiting a low output impedance. This property fundamentally originates from the gain desensitization provided by feedback.

In order to formally prove that voltage feedback lowers the output impedance, we consider the simple model in Fig. 8.20, where R_{out} represents the output impedance of the feedforward amplifier. Setting the input to zero and applying a voltage at the output, we write $V_F = \beta V_X$, $V_e = -\beta V_X$, $V_M = -\beta A_0 V_X$, and hence $I_X = [V_X - (-\beta A_0 V_X)]/R_{out}$ (if the current drawn by the feedback network is neglected). It follows that

$$\frac{V_X}{I_X} = \frac{R_{out}}{1 + \beta A_0}. \tag{8.25}$$

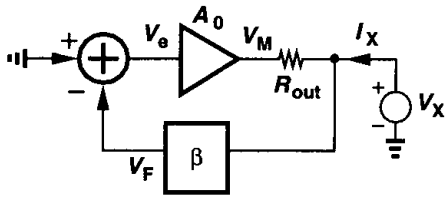
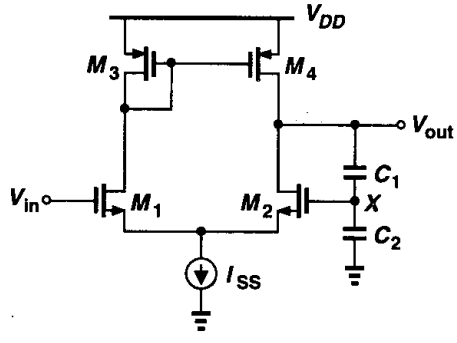


Figure 8.20 Calculation of output resistance of a voltage-voltage feedback circuit.

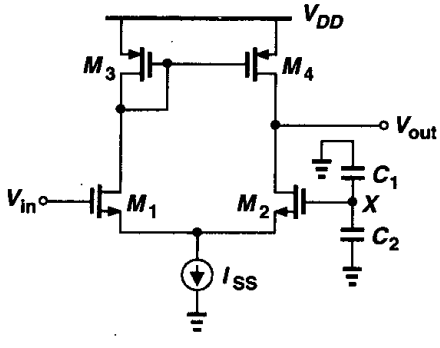
Thus, the output impedance and the gain are lowered by the same factor. In the circuit of Fig. 8.18(b), for example, the output impedance is lowered by $1 + A_0R_2/(R_1 + R_2)$.

Example 8.2

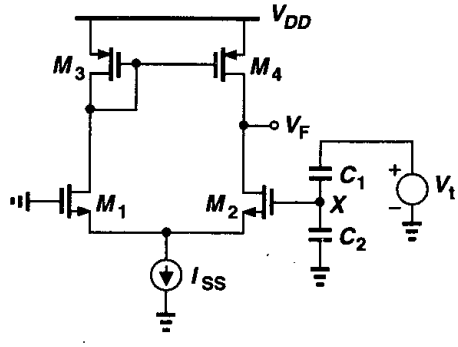
The circuit shown in Fig. 8.21(a) is an implementation of the feedback configuration depicted in Fig. 8.18(b), but with the resistors replaced by capacitors. (The bias network of M_2 is not shown.) Calculate the closed-loop gain and output resistanceof the amplifier at relatively low frequencies.



(a)



(b)



(c)

Figure 8.21

Solution

At low frequencies, C_1 and C_2 load the amplifier negligibly. To find the open-loop voltage gain, we break the feedback loop as shown in Fig. 8.21(b), grounding the top plate of C_1 to ensure zero voltage feedback. The open-loop gain is thus equal to $g_{m1}(r_{O2}\|r_{O4})$.

We must also compute the loop gain. With the aid of Fig. 8.21(c), we have

$$V_F = -V_t \frac{C_1}{C_1 + C_2} g_{m1}(r_{O2}\|r_{O4}). \tag{8.26}$$

That is,

$$\beta A_0 = \frac{C_1}{C_1 + C_2} g_{m1}(r_{O2}\|r_{O4}) \tag{8.27}$$

and hence

$$A_{closed} = \frac{g_{m1}(r_{O2}\|r_{O4})}{1 + \dfrac{C_1}{C_1 + C_2} g_{m1}(r_{O2}\|r_{O4})}. \tag{8.28}$$

As expected, if $\beta A_0 \gg 1$, then $A_{closed} \approx 1 + C_2/C_1$.

The open-loop output resistance of the circuit is equal to $r_{O2}\|r_{O4}$ (Chapter 5). It follows that

$$R_{out,closed} = \frac{r_{O2}\|r_{O4}}{1 + \dfrac{C_1}{C_1 + C_2} g_{m1}(r_{O2}\|r_{O4})}. \tag{8.29}$$

It is interesting to note that, if $\beta A_0 \gg 1$, then

$$R_{out,closed} \approx \left(1 + \frac{C_2}{C_1}\right) \frac{1}{g_{m1}}. \tag{8.30}$$

In other words, even if the open-loop amplifier suffers from a *high* output resistance, the closed-loop output resistance is independent of $r_{O2}\|r_{O4}$, simply because the open-loop *gain* scales with $r_{O2}\|r_{O4}$ as well.

Voltage-voltage feedback also modifies the input impedance. Comparing the configurations in Fig. 8.22, we note that the input impedance of the feedforward amplifier sustains the entire input voltage in Fig. 8.22(a), but only a fraction of V_{in} in Fig. 8.22(b). As a result, the current drawn by R_{in} in the feedback topology is *less* than that in the open-loop system, suggesting that returning a voltage quantity to the input *increases* the input impedance.

The foregoing observation can be confirmed analytically with the aid of Fig. 8.23. Since $V_e = I_X R_{in}$ and $V_F = \beta A_0 I_X R_{in}$, we have $V_e = V_X - V_F = V_X - \beta A_0 I_X R_{in}$. Thus, $I_X R_{in} = V_X - \beta A_0 I_X R_{in}$, and

$$\frac{V_X}{I_X} = R_{in}(1 + \beta A_0). \tag{8.31}$$

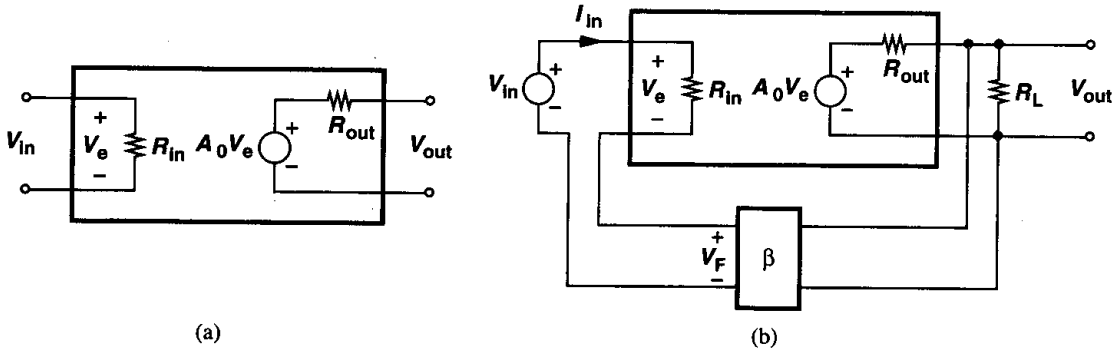


Figure 8.22 Effect of voltage-voltage feedback on input resistance.

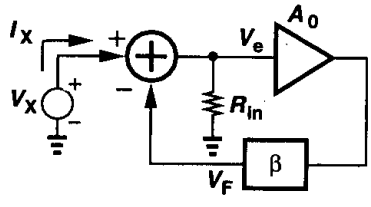


Figure 8.23 Calculation of input impedance of a voltage-voltage feedback circuit.

The input impedance therefore increases by the ubiquitous factor $1 + \beta A_0$, bringing the circuit closer to an ideal voltage amplifier.

Example 8.3

Fig. 8.24(a) shows a common-gate topology placed in a voltage-voltage feedback configuration. Note that the summation of the feedback voltage and the input voltage is accomplished by applying the former to the gate and the latter to the source.[7] Calculate the input resistance at low frequencies if channel-length modulation is negligible.

Solution

Breaking the loop as depicted in Fig. 8.24(b), we recognize that the open-loop input resistance is equal to $(g_{m1} + g_{mb1})^{-1}$. To find the loop gain, we set the input to zero and inject a test signal in the loop [Fig. 8.24(c)], obtaining $V_F/V_t = -g_{m1}R_DC_1/(C_1 + C_2)$. The closed-loop input impedance is then equal to

$$R_{in,closed} = \frac{1}{g_{m1} + g_{mb1}} \left(1 + \frac{C_1}{C_1 + C_2} g_{m1} R_D\right). \tag{8.32}$$

The increase in the input impedance can be explained as follows. Suppose the input voltage decreases by ΔV, causing the output voltage to (momentarily) fall. As a result, the gate voltage of M_1 *decreases*,

[7] This circuit is similar to the right half of the topology shown in Fig. 8.21(a).

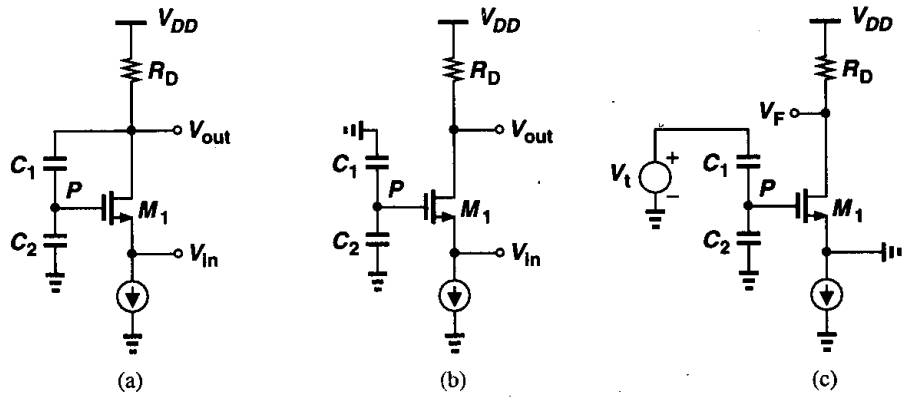


Figure 8.24

thereby lowering the gate-source voltage of M_1 and yielding a change in V_{GS1} that is *less* than ΔV. By contrast, if the gate of M_1 were connected to a constant potential, the gate-source voltage would change by ΔV, resulting in a larger current change.

In summary, voltage-voltage feedback decreases the output impedance and increases the input impedance, thereby proving useful as a "buffer" stage that can be interposed between a high-impedance source and a low-impedance load.

8.2.2 Current-Voltage Feedback

In some circuits, it is desirable or simpler to sense the output current to perform feedback. The current is actually sensed by placing a small resistor in series with the output and using the voltage drop across the resistor as the feedback information. This voltage may even serve as the return signal that is directly subtracted from the input.

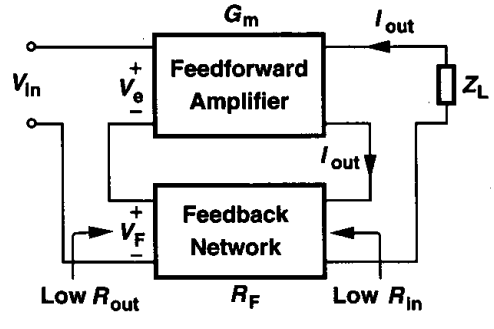


Figure 8.25 Current-voltage feedback.

Let us consider the general current-voltage feedback system illustrated in Fig. 8.25.[8] Since the feedback network senses the output current and returns a voltage, its feedback

[8]This topology is also called "series-series" feedback.

factor (β) has the dimension of resistance and is denoted by R_F. (Note that a finite load, Z_L, is connected to the output so that $I_{out} \neq 0$.) We can thus write $V_F = R_F I_{out}$, $V_e = V_{in} - R_F I_{out}$, and hence $I_{out} = G_m(V_{in} - R_F I_{out})$. It follows that

$$\frac{I_{out}}{V_{in}} = \frac{G_m}{1 + G_m R_F}. \tag{8.33}$$

An ideal feedback network in this case exhibits zero input and output impedances.

It is instructive to confirm that $G_m R_F$ is indeed the loop gain. As shown in Fig. 8.26, we set the input voltage to zero and break the loop by disconnecting the feedback network

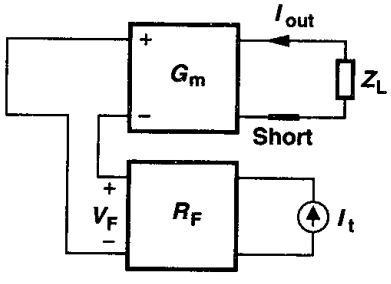


Figure 8.26 Calculation of loop gain for current-voltage feedback.

from the output and replacing it with a *short* at the output (if the feedback network is ideal). We then inject the test signal I_t, producing $V_F = R_F I_t$ and hence $I_{out} = -G_m R_F I_t$. Thus, the loop gain is equal to $G_m R_F$ and the transconductance of the amplifier is reduced by $1 + G_m R_F$ when feedback is applied.

Sensing the current at the output of a feedback system *increases* the output impedance. This is because the system attempts to make the output *current* a faithful replica of the input signal (with a proportionality factor if the input is a voltage quantity). Consequently, the system delivers the same current waveform as the load varies, in essence approaching an ideal current source and hence exhibiting a high output impedance.

To prove the above result, we consider the current-voltage feedback topology shown in Fig. 8.27, where R_{out} represents the finite output impedance of the feedforward ampli-

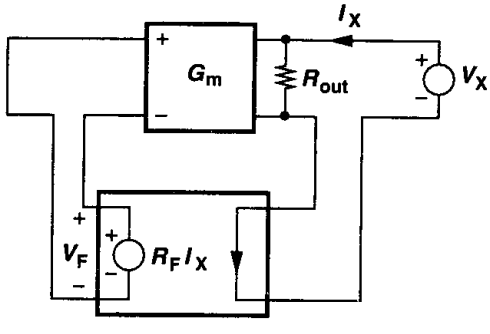


Figure 8.27 Calculation of output resistance of a current-voltage feedback amplifier.

fier.[9] The feedback network produces a voltage V_F proportional to I_X : $V_F = R_F I_X$, and the current generated by G_m equals $-R_F I_X G_m$. As a result, $-R_F I_X G_m = I_X - V_X/R_{out}$, yielding

$$\frac{V_X}{I_X} = R_{out}(1 + G_m R_F). \tag{8.34}$$

The output impedance therefore increases by a factor of $1 + G_m R_F$.

Example 8.4

Suppose we need to increase the output impedance of a common-source stage by current feedback. As shown in Fig. 8.28(a), we insert a small resistor r in the output current path, apply the voltage

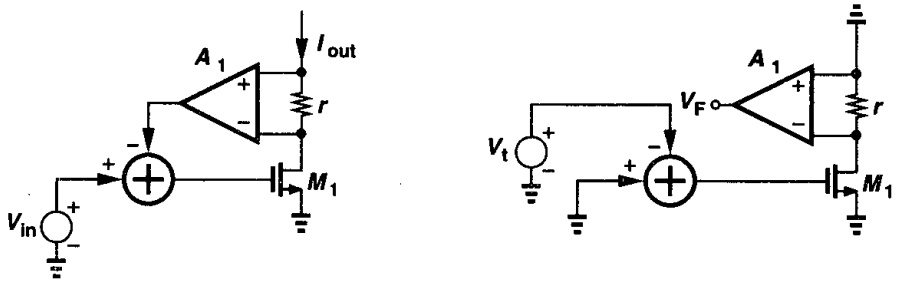


Figure 8.28

across r to an amplifier A_1, and subtract the output of A_1 from the input voltage. Calculate the output impedance of this circuit.

Solution

Using the circuit of Fig. 8.28(b) to determine the loop gain, we have

$$\frac{V_F}{V_t} = -g_m r A_1. \tag{8.35}$$

Thus, the overall output impedance is given by

$$R_{out,closed} = (1 + g_m r A_1) r_{O1}. \tag{8.36}$$

As with voltage-voltage feedback, current-voltage feedback increases the input impedance by a factor equal to one plus the loop gain. As illustrated in Fig. 8.29, we have $I_X R_{in} G_m = I_{out}$. Thus, $V_e = V_X - G_m R_F I_X R_{in}$ and

$$\frac{V_X}{I_X} = R_{in}(1 + G_m R_F). \tag{8.37}$$

[9]Note that R_{out} is placed in *parallel* with the output because the ideal transimpedance amplifier is modeled by a voltage-dependent current source.

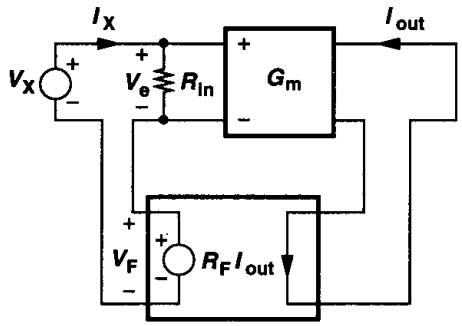


Figure 8.29 Calculation of input resistance of a current-voltage feedback amplifier.

The reader can show that the loop gain is indeed equal to $G_m R_F$.

In summary, current-voltage feedback increases both the input and the output impedances while decreasing the feedforward transconductance. As explained in Chapter 9, the high output impedance proves useful in high-gain op amps.

8.2.3 Voltage-Current Feedback

In this type of feedback, the output voltage is sensed and a proportional current is returned to the summing point at the input.[10] Note that the feedforward path incorporates a transimpedance amplifier with gain R_0 and the feedback factor has a dimension of conductance.

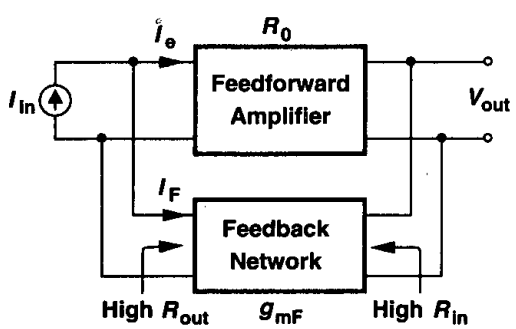


Figure 8.30 Voltage-current feedback.

A voltage-current feedback topology is shown in Fig. 8.30. Sensing a voltage and producing a current, the feedback network is characterized by a transconductance g_{mF}, ideally exhibiting infinite input and output impedances. Since $I_F = g_{mF} V_{out}$ and $I_e = I_{in} - I_F$, we have $V_{out} = R_0 I_e = R_0(I_{in} - g_{mF} V_{out})$. It follows that

$$\frac{V_{out}}{I_{in}} = \frac{R_0}{1 + g_{mF} R_0}. \tag{8.38}$$

[10]This topology is also called "shunt-shunt" feedback.

The reader can prove that $g_{mF}R_0$ is indeed the loop gain, concluding that this type of feedback lowers the transimpedance by a factor equal to one plus the loop gain.

Example 8.5

Calculate the transimpedance, V_{out}/I_{in}, of the circuit shown in Fig. 8.31(a) at relatively low frequencies.

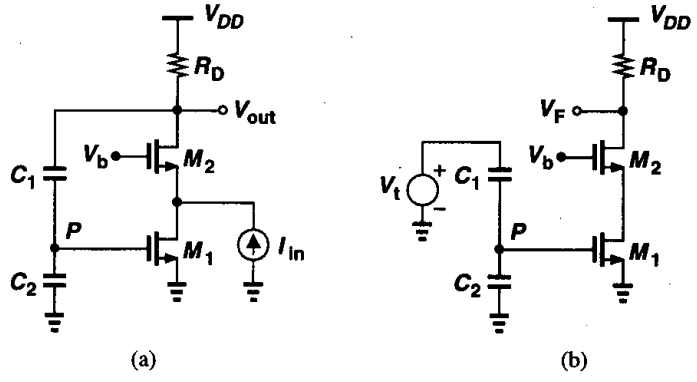


Figure 8.31

Solution

In this circuit, the capacitive divider C_1-C_2 senses the output voltage, applying the result to the gate of M_1 and producing a current that is subtracted from I_{in}. The open-loop transimpedance equals that of the core common-gate stage, R_D. The loop gain is obtained by setting I_{in} to zero and breaking the loop at the output [Fig. 8.31(b)]:

$$-V_t \frac{C_1}{C_1 + C_2} g_{m1} R_D = V_F. \tag{8.39}$$

Thus, the overall transimpedance is equal to

$$R_{tot} = \frac{R_D}{1 + \dfrac{C_1}{C_1 + C_2} g_{m1} R_D}. \tag{8.40}$$

Following our reasoning for the other two types of feedback studied above, we surmise that voltage-current feedback decreases both the input and the output impedances. As shown in Fig. 8.32(a), the input resistance of R_0 is placed in *series* because an ideal transimpedance amplifier exhibits a zero input impedance. We write $I_F = I_X - V_X/R_{in}$ and $(V_X/R_{in})R_0 g_{mF} = I_F$. Thus,

$$\frac{V_X}{I_X} = \frac{R_{in}}{1 + g_{mF} R_0}. \tag{8.41}$$

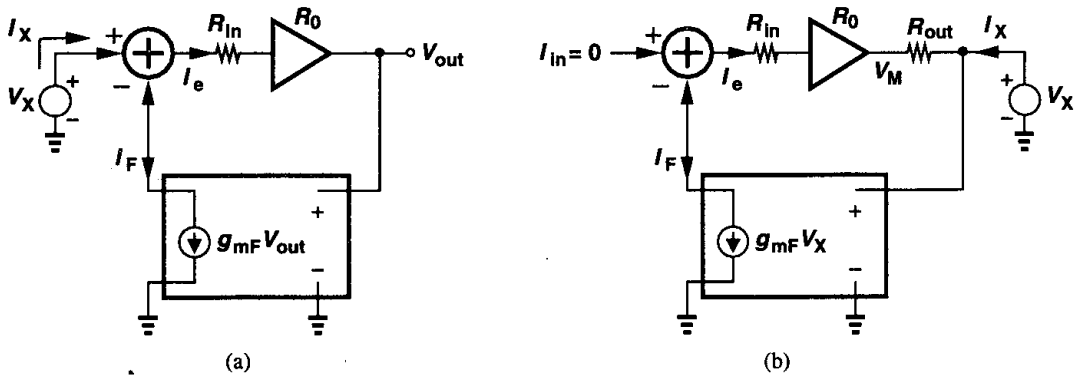


Figure 8.32 Calculation of (a) input and (b) output impedance of a voltage-current feedback amplifier.

Similarly, from Fig. 8.32(b), we have $I_F = V_X g_{mF}$, $I_e = -I_F$, and $V_M = -R_0 g_{mF} V_X$. Neglecting the input current of the feedback network, we write $I_X = (V_X - V_M)/R_{out} = (V_X + g_{mF} R_0 V_X)/R_{out}$. That is,

$$\frac{V_X}{I_X} = \frac{R_{out}}{1 + g_{mF} R_0}. \tag{8.42}$$

Example 8.6

Calculate the input and output impedances of the circuit shown in Fig. 8.33(a). For simplicity, assume $R_F \gg R_D$.

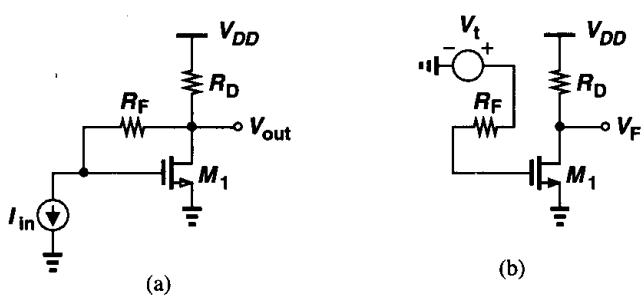


Figure 8.33

Solution

In this circuit, R_F senses the output voltage and returns a current to the input. Breaking the loop as depicted in Fig. 8.33(b), we calculate the loop gain as $g_m R_D$. Thus, the open-loop input impedance, R_F, is divided by $1 + g_m R_D$:

$$R_{in,closed} = \frac{R_F}{1 + g_m R_D}. \tag{8.43}$$

Similarly,

$$R_{out,closed} = \frac{R_D}{1 + g_m R_D}. \tag{8.44}$$

Note $R_{out,closed}$ is in fact the parallel combination of a diode-connected transistor and R_D.

An important application of amplifiers with *low* input impedance is in fiber optic receivers, where light received through a fiber is converted to a *current* by a reverse-biased photodiode. This current is typically converted to a voltage for further amplification and processing. Shown in Fig. 8.34(a), such conversion can be accomplished by a simple resistor but at the cost of bandwidth because the diode suffers from a relatively large junction capacitance. For this reason, the feedback topology of Fig. 8.34(b) is usually employed, where R_1 is placed around the voltage amplifier A to form a transimpedance circuit. The input impedance is $R_1/(1 + A)$ and the output voltage is approximately $R_1 I_{D1}$.

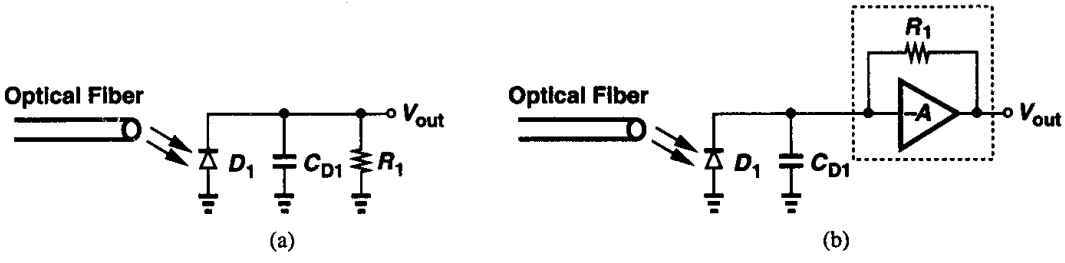


Figure 8.34 Detection of current produced by a photodiode by (a) resistor R_1 and (b) a transimpedance amplifier.

8.2.4 Current-Current Feedback

Fig. 8.35 illustrates this type of feedback.[11] Here, the feedforward amplifier is characterized by a current gain, A_I, and the feedback network by a current ratio, β. In a fashion similar to the previous derivations, the reader can easily prove that the closed-loop current gain is equal to $A_I/(1+\beta A_I)$, the input impedance is divided by $1+\beta A_I$ and the output impedance is multiplied by $1 + \beta A_I$.

Fig. 8.36 illustrates an example of current-current feedback. Here, since the source and drain currents of M_2 are equal (at low frequencies), resistor R_S is inserted in the source network to monitor the output current. Resistor R_F plays the same role as in Fig. 8.33.

[11] This topology is also called "shunt-series" feedback, where the first term refers to the input connection and the second to the output connection.

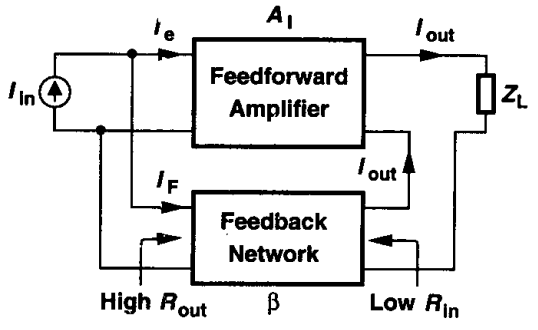


Figure 8.35 Current-current feedback.

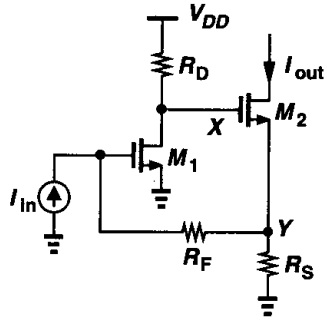


Figure 8.36

8.3 Effect of Loading

In our analysis of feedback systems thus far, we have tacitly assumed that the feedback network does not "load" the feedforward amplifier at the input or output. For example, in the voltage-voltage feedback topology of Fig. 8.21, we assumed C_1 and C_2 do not load the amplifier so that the *open-loop* gain could still be written as $g_{m1}(r_{O2} \| r_{O4})$. In reality, however, the loading due to the feedback network may not be negligible, complicating the analysis.

The problem of loading manifests itself when we need to break the feedback loop so as to identify the open-loop system, e.g., calculate the open-loop gain and the input and output impedances. To arrive at the proper procedure for including the feedback network terminal impedances, we first review models of two-port networks.

8.3.1 Two-Port Network Models

The feedback network placed around the feedforward amplifier can be considered a two-port circuit sensing and producing voltages or currents. Recall from basic circuit theory that a two-port linear (and time-invariant) network can be represented by any of the four models shown in Fig. 8.37. The "Z model" in Fig. 8.37(a) consists of input and output impedances in series with current-dependent voltage sources whereas the "Y model" in Fig. 8.37(b) comprises input and output admittances in parallel with voltage-dependent current sources. The "hybrid models" of Figs. 8.37(c) and (d) incorporate a combination of impedances

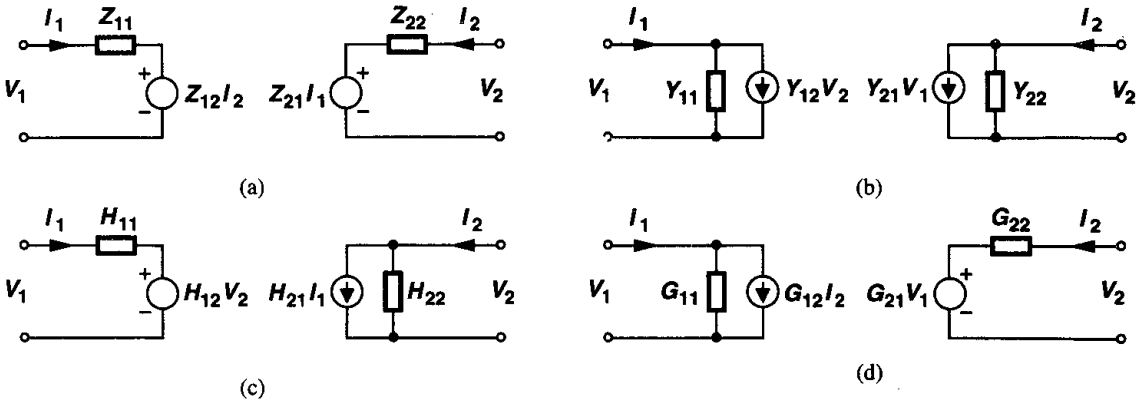


Figure 8.37 Linear two-port network models.

and admittances and voltage sources and current sources. Each model is described by two equations. For the Z model, we have

$$V_1 = Z_{11}I_1 + Z_{12}I_2 \tag{8.45}$$

$$V_2 = Z_{21}I_1 + Z_{22}I_2. \tag{8.46}$$

Each Z parameter has a dimension of impedance and is obtained by leaving one port open, e.g., $Z_{11} = V_1/I_1$ when $I_2 = 0$. Similarly, for the Y model,

$$I_1 = Y_{11}V_1 + Y_{12}V_2 \tag{8.47}$$

$$I_2 = Y_{21}V_1 + Y_{22}V_2, \tag{8.48}$$

where each Y parameter is calculated by shorting one port, e.g., $Y_{11} = I_1/V_1$ when $V_2 = 0$. For the H model,

$$V_1 = H_{11}I_1 + H_{12}V_2 \tag{8.49}$$

$$I_2 = H_{21}I_1 + H_{22}V_2, \tag{8.50}$$

and for the G model,

$$I_1 = G_{11}V_1 + G_{12}I_2 \tag{8.51}$$

$$V_2 = G_{21}V_1 + G_{22}I_2. \tag{8.52}$$

Note that, for example, Y_{11} may not be equal to the inverse of Z_{11} because the two are obtained under different conditions: the output is shorted for the former but left open for the latter.

In order to simplify the analysis of the loading due to the feedback network, we must select a suitable model from the above. We assume the *input* port of the feedback network is connected to the *output* port of the feedforward amplifier. Let us begin with voltage-voltage feedback. Which model should be used? We note that the ideal feedback network in this case must exhibit infinite input impedance and zero output impedance. The Z model is not suitable because as $Z_{11} \to \infty$, for a finite V_1, $I_1 \to 0$, and $Z_{21} I_1 \to 0$. That is, if the input impedance approaches infinity, the output voltage drops to zero. How about the Y model? In this case, if $Y_{11} \to 0$, then the output voltage remains finite, but if Y_{22} approaches ∞, the current source $Y_{21} V_1$ generates a zero output voltage. That is, if the output imepdance of the feedback network approaches zero (so that it becomes more ideal), then the output voltage of the feedback network drops to zero as well. With these observations, we surmise that the G model is the most suitable one for voltage-voltage feedback; in the ideal case $G_{11} = 0$, $G_{22} = 0$, and $G_{21} V_1 \neq 0$.

Using similar arguments, the reader can show that the other three types of feedback require the following network models: voltage-current: Y model; current-voltage: Z model; current-current: H model.

8.3.2 Loading in Voltage-Voltage Feedback

Replacing the feedback network by a G model, we arrive at the representation in Fig. 8.38(a). Unlike the simple models used in previous sections, this circuit incorporates *two* dependent sources in the feedback path: $G_{12} I_2$ and $G_{21} V_{out}$. What is the effect of $G_{12} I_2$? This current flows through the parallel combination of Z_{out} and G_{11}, contributing to the output voltage. However, if A_0 is large, the signal amplified by A_0 is much greater than the contribution of $G_{12} I_2$. In other words, the *forward* gain of the main amplifier overwhelms the *reverse* gain of the feedback network. Since this condition holds in most circuits of interest, we can neglect $G_{12} I_2$, obtaining the circuit in Fig. 8.38(b). A rigorous analysis of Fig. 8.38(a) (Problem 8.8) reveals that if $G_{12} \ll A_0 Z_{in} / Z_{out}$, then the "reverse transmission" through the feedback circuit is negligible. It is indeed expected that Z_{in} and Z_{out} play a role here. If Z_{in} is small, the voltage division between Z_{in} and G_{22} reduces the signal through the feedforward path. Similarly, if Z_{out} is large, then the voltage division between Z_{out} and G_{11} lowers the contribution of $A_0 V_e$ to the output.

Let us now compute the closed-loop gain of the circuit shown in Fig. 8.38(b). We have

$$V_e = (V_{in} - G_{21} V_{out}) \frac{Z_{in}}{Z_{in} + G_{22}}, \tag{8.53}$$

and hence

$$(V_{in} - G_{21} V_{out}) \frac{Z_{in}}{Z_{in} + G_{22}} A_0 \frac{G_{11}^{-1}}{G_{11}^{-1} + Z_{out}} = V_{out}. \tag{8.54}$$

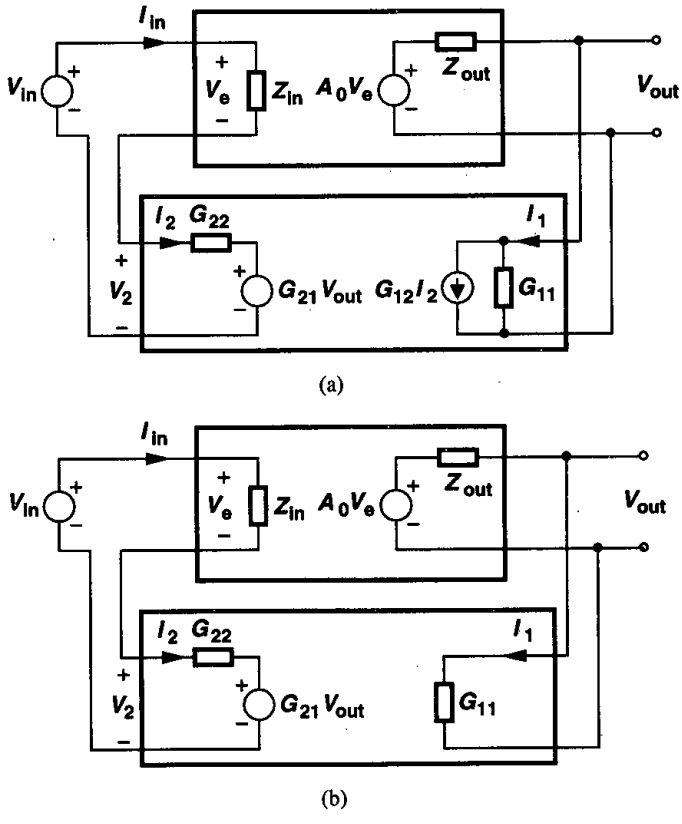


Figure 8.38 Voltage-voltage feedback circuit with (a) feedback network represented by a G model and (b) simplified G model.

It follows that

$$\frac{V_{out}}{V_{in}} = \frac{A_0 \dfrac{Z_{in}}{Z_{in}+G_{22}} \dfrac{G_{11}^{-1}}{G_{11}^{-1}+Z_{out}}}{1+\dfrac{Z_{in}}{Z_{in}+G_{22}} \dfrac{G_{11}^{-1}}{G_{11}^{-1}+Z_{out}} G_{21}A_0}. \tag{8.55}$$

Note that if the feedback network is ideal, i.e., if $G_{11}^{-1} = \infty$ and $G_{22} = 0$, then $V_{out}/V_{in} = A_0/(1 + G_{21}A_0)$, as expected.

Equation 8.55 assumes the standard form of a feedback transfer function if we define the open-loop gain in the presence of loading as

$$A_{v,open} = \frac{Z_{in}}{Z_{in}+G_{22}} \frac{G_{11}^{-1}}{G_{11}^{-1}+Z_{out}} A_0. \tag{8.56}$$

The loaded open-loop gain can be obtained from the circuit depicted in Fig. 8.39, where $G_{21}V_{out}$ is set to zero. That is, the finite input and output impedances of the feedback network reduce the output voltage and the voltage seen by the input of the main amplifier, respectively.

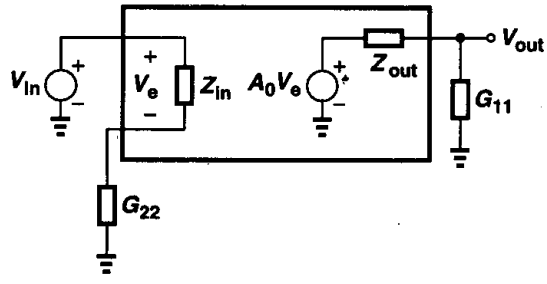


Figure 8.39 Proper method of including loading in a voltage-voltage feedback circuit.

It is important to note that G_{11} and G_{22} in Fig. 8.37 are computed as follows:

$$G_{11} = \left.\frac{I_1}{V_1}\right|_{I2=0} \tag{8.57}$$

$$G_{22} = \left.\frac{V_2}{I_2}\right|_{V1=0} \tag{8.58}$$

Thus, as illustrated in Fig. 8.40, G_{11} is obtained by leaving the output of the feedback

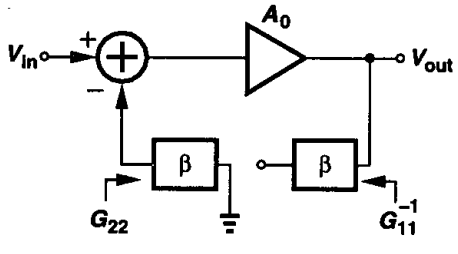


Figure 8.40 Conceptual view of opening a voltage-voltage feedback loop with proper loading.

network open whereas G_{22} is calculated by *shorting* the input of the feedback network.

Another important result of the foregoing analysis is that the loop gain, i.e., the second term in the denominator of (8.55) is simply equal to the loaded open-loop gain multiplied by G_{21}. Thus, a separate calculation of the loop gain is not necessary. Also, the open-loop input and output impedances obtained from Fig. 8.39 are scaled by $1 + G_{21}A_{v,open}$ to yield the closed-loop values.

Example 8.7

For the circuit shown in Fig. 8.41(a), calculate the open-loop and closed-loop gains.

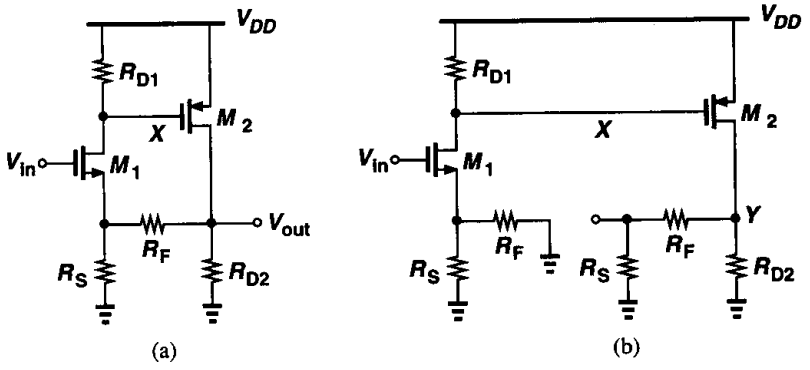


Figure 8.41

Solution

The circuit consists of two common-source stages, with R_F and R_S sensing the output voltage and returning a fraction thereof to the source of M_1. The reader can prove that the feedback is indeed negative. Following the procedure illustrated in Fig. 8.40, we identify R_F and R_S as the feedback network and construct the open-loop circuit as shown in Fig. 8.41(b). Note that the loading effect in the input network is obtained by shorting the right terminal of R_F to ground and that in the output by leaving the left terminal of R_F open. Neglecting channel-length modulation and body effect for simplicity, we have

$$A_{v,open} = \frac{V_Y}{V_{in}} = \frac{-R_{D1}}{R_F \| R_S + 1/g_{m1}} \{-g_{m2}[R_{D2} \| (R_F + R_S)]\}. \tag{8.59}$$

To compute the closed-loop gain, we first find the loop gain as $G_{21}A_{v,open}$. Recall from (8.52) that $G_{21} = V_2/V_1$ with $I_2 = 0$. For the voltage divider consisting of R_F and R_D, $G_{21} = R_S/(R_F + R_S)$. The closed-loop gain is simply equal to $A_{v,closed} = A_{v,open}/(1 + G_{21}A_{v,open})$.

8.3.3 Loading in Current-Voltage Feedback

Replacing the feedback network by a Z model, we obtain the circuit shown in Fig. 8.42(a). Using an argument similar to that for voltage-voltage feedback, we neglect the source $Z_{12}I_2$, thereby arriving at the circuit in Fig. 8.42(b). We thus have

$$(V_{in} - Z_{21}I_{out})\frac{Z_{in}}{Z_{in} + Z_{22}} G_m \frac{Z_{out}}{Z_{out} + Z_{11}} = I_{out}. \tag{8.60}$$

That is,

$$\frac{I_{out}}{V_{in}} = \frac{\dfrac{Z_{in}}{Z_{in} + Z_{22}} \dfrac{Z_{out}}{Z_{out} + Z_{11}} G_m}{1 + \dfrac{Z_{in}}{Z_{in} + Z_{22}} \dfrac{Z_{out}}{Z_{out} + Z_{11}} G_m Z_{21}}. \tag{8.61}$$

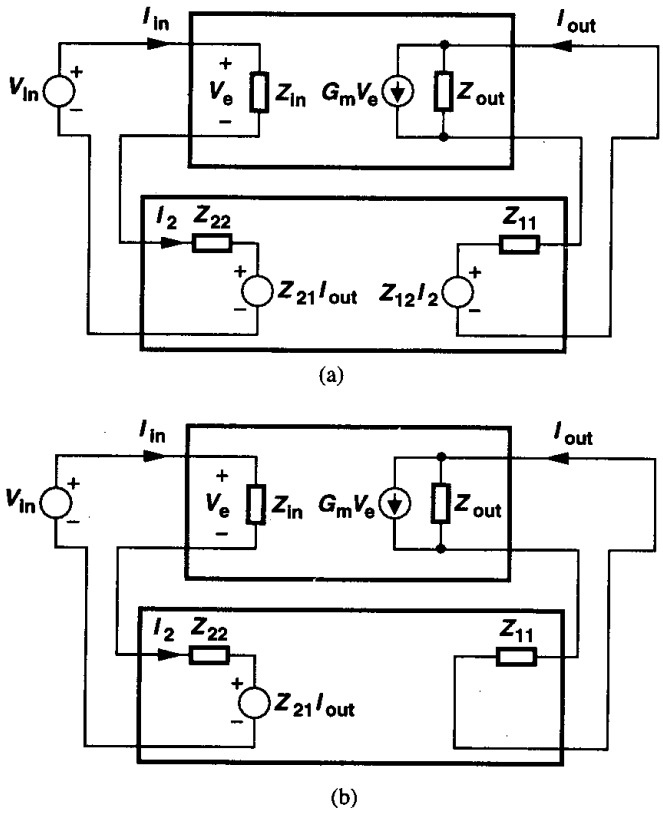


Figure 8.42 Current-voltage feedback circuit with (a) feedback network represented by a Z model and (b) simplified Z model.

Equation (8.61) suggests that the loaded open-loop gain is equal to

$$G_{m,open} = \frac{Z_{in}}{Z_{in} + Z_{22}} \frac{Z_{out}}{Z_{out} + Z_{11}} G_m, \tag{8.62}$$

revealing voltage division at the input and current division at the output (Fig. 8.43). Since Z_{22} and Z_{11} are obtained by opening the input and output ports of the feedback network, respectively, the open-loop circuit can be visualized as in Fig. 8.44. Note that the loop gain is equal to $Z_{21}G_{m,open}$.

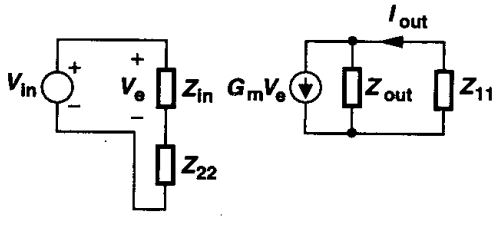


Figure 8.43 Current-voltage feedback circuit with proper loading of feedback network.

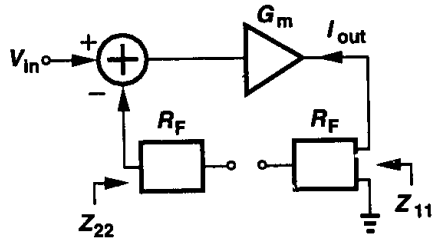


Figure 8.44 Conceptual view of opening the loop in current-voltage feedback.

Example 8.8

Calculate the open-loop and closed-loop gain of the circuit shown in Fig. 8.45(a).

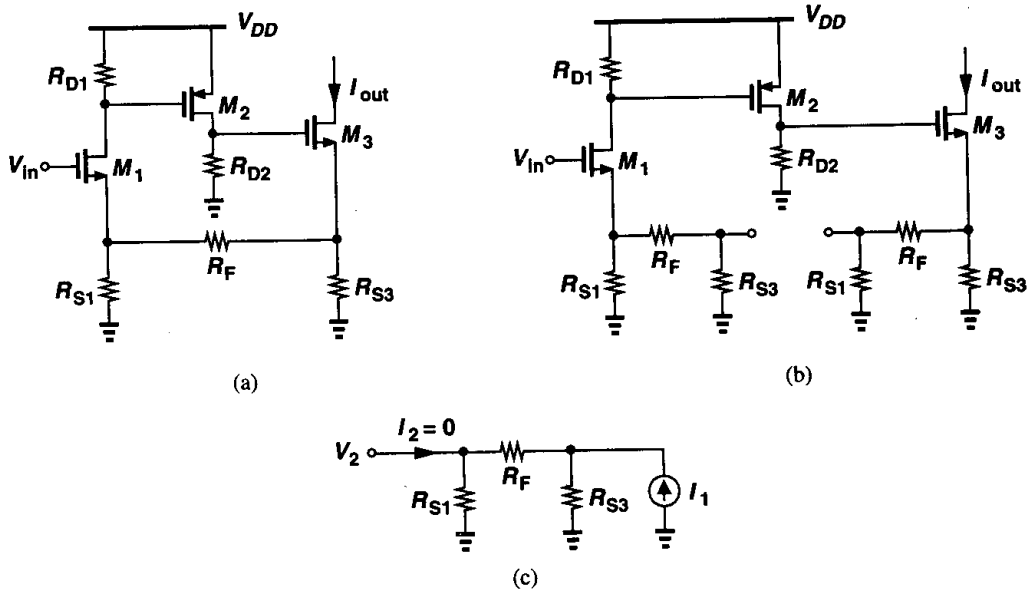


Figure 8.45

Solution

This circuit consists of two voltage gain stages, M_1 and M_2, and a voltage-to-current converter, M_3. Since the drain and source currents of M_2 are equal, the output current is monitored by R_{S3}. Thus, R_{S3}, R_F, and R_{S1} sense the output current and return a proportional voltage to the input.

If $\lambda = \gamma = 0$, the open-loop gain is equal to

$$G_{m,open} = \frac{-R_{D1}}{R_{S1} \| (R_F + R_{S3}) + 1/g_{m1}} \cdot \frac{-g_{m2} R_{D2}}{R_{S3} \| (R_F + R_{S1}) + 1/g_{m3}}. \tag{8.63}$$

The loop gain is given by $Z_{21} G_{m,open}$, where, from (8.46), $Z_{21} = V_2/I_1$ with $I_2 = 0$. For the

feedback network consisting of R_{S2}, R_F and R_{S1}, the circuit of Fig. 8.45(c) gives

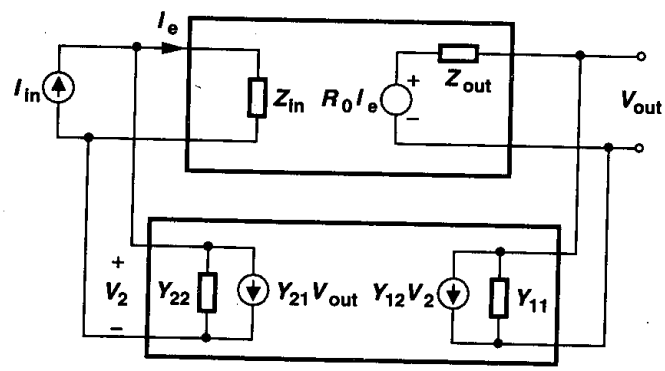
$$Z_{21} = \frac{R_{S3}}{R_{S3} + R_{S1} + R_F} R_{S1}. \tag{8.64}$$

The closed-loop gain equals $G_{m,open}/(1 + Z_{21}G_{m,open})$.

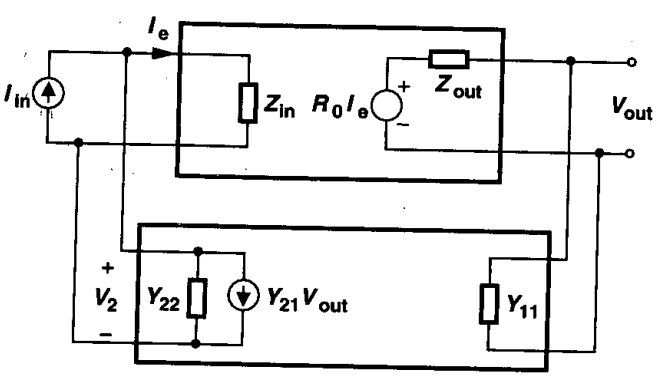
It is important to distinguish between the feedback networks in the circuits of Fig. 8.41(a) and 8.45(a). In the former, R_{D2} is part of the feedforward amplifier, rather than part of the feedback network, because it must generate a *voltage* output. In the latter, R_{S3} is part of the feedback network because it is used to *sense* the output current. If the output of interest in Fig. 8.45(a) is the voltage at the source of M_3, then R_{S3} is part of the feedforward amplifier rather than the feedback network.

8.3.4 Loading in Voltage-Current Feedback

In this type of system, we represent the feedback network by a Y model [Fig. 8.46(a)]. As with previous cases, we neglect the reverse transmission term, $Y_{12}V_2$, obtaining the circuit



(a)



(b)

Figure 8.46 Voltage-current feedback circuit with (a) feedback network represented by a Y model and (b) simplified Y model.

in Fig. 8.46(b). Writing

$$(I_{in} - Y_{21}V_{out})\frac{Y_{22}^{-1}}{Y_{22}^{-1} + Z_{in}} R_0 \frac{Y_{11}^{-1}}{Y_{11}^{-1} + Z_{out}} = V_{out}, \tag{8.65}$$

we have

$$\frac{V_{out}}{I_{in}} = \frac{\dfrac{Y_{22}^{-1}}{Y_{22}^{-1} + Z_{in}} R_0 \dfrac{Y_{11}^{-1}}{Y_{11}^{-1} + Z_{out}}}{1 + \dfrac{Y_{22}^{-1}}{Y_{22}^{-1} + Z_{in}} R_0 \dfrac{Y_{11}^{-1}}{Y_{11}^{-1} + Z_{out}} Y_{21}}. \tag{8.66}$$

It is therefore possible to define the loaded open-loop gain as

$$R_{0,open} = \frac{Y_{22}^{-1}}{Y_{22}^{-1} + Z_{in}} \frac{Y_{11}^{-1}}{Y_{11}^{-1} + Z_{out}} R_0. \tag{8.67}$$

The loading manifests itself as current division between Y_{22}^{-1} and Z_{in} and voltage division

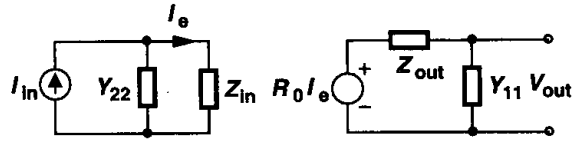


Figure 8.47 Voltage-current feedback circuit with proper loading of feedback network.

between Z_{out} and Y_{11}^{-1} (Fig. 8.47). Since Y_{22} and Y_{11} are obtained by shorting the input and output ports of the feedback network, respectively, the procedure for including the loading can be illustrated as in Fig. 8.48. The loop gain is given by $Y_{21}R_{0,open}$.

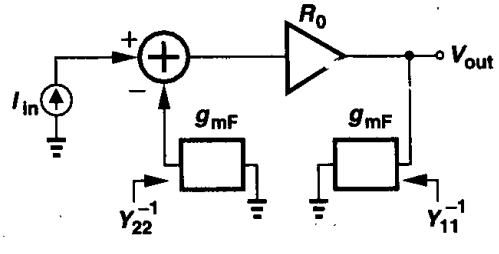


Figure 8.48 Conceptual view of opening the loop in voltage-current feedback.

Example 8.9

Calculate the voltage gain of the circuit shown in Fig. 8.49(a).

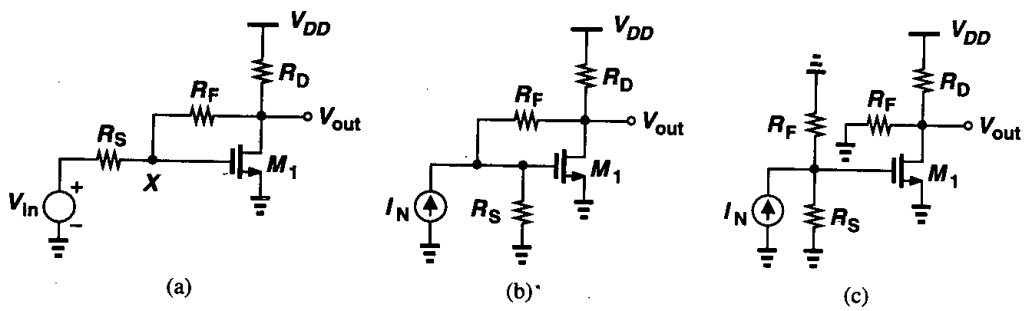


Figure 8.49

Solution

What type of feedback is used in this circuit? Resistor R_F senses the output voltage and returns a proportional current to node X. Thus, the feedback can be considered as the voltage-current type. However, in the general representation of Fig. 8.46(a), the input signal is a current quantity, whereas in this example, it is a voltage quantity. For this reason, we replace V_{in} and R_S by a Norton equivalent [Fig. 8.49(b)] and view R_S as the input resistance of the main amplifier. Opening the loop according to Fig. 8.48 and neglecting channel-length modulation, we write the open-loop gain from Fig. 8.49(c) as

$$R_{0,open} = \left.\frac{V_{out}}{I_N}\right|_{open} \tag{8.68}$$

$$= -(R_S \| R_F) g_m (R_F \| R_D), \tag{8.69}$$

where $I_N = V_{in}/R_S$. We also calculate the loop gain as $Y_{21} R_{0,open}$. From (8.48), $Y_{21} = I_2/V_1$ with $V_2 = 0$, and since the feedback network consists of only R_F, we have $Y_{21} = -1/R_F$. Thus, the circuit of Fig. 8.49(a) exhibits a voltage gain of

$$\frac{V_{out}}{V_{in}} = \frac{1}{R_S} \cdot \frac{-(R_S \| R_F) g_m (R_F \| R_D)}{1 + g_m (R_F \| R_D) R_S/(R_S + R_F)}. \tag{8.70}$$

Interestingly, if R_F is replaced by a capacitor, this analysis does not yield a zero in the transfer function because we have neglected the reverse transmission of the feedback network (from the output of the feedback network to its input.) The input and output impedances of the circuit are also interesting to calculate. This is left as an exercise for the reader. The reader is also encouraged to apply this solution to the circuit of Fig. 8.3(b).

8.3.5 Loading in Current-Current Feedback

Fig. 8.50(a) depicts a current-current feedback system with the feedback network represented by an H model. Neglecting the effect of $H_{12}V_2$ compared to the forward gain of the amplifier and drawing the circuit as in Fig. 8.50(b), we write

$$(I_{in} - H_{21}I_{out})\frac{H_{22}^{-1}}{H_{22}^{-1} + Z_{in}}A_I\frac{Z_{out}}{H_{11} + Z_{out}} = I_{out}. \tag{8.71}$$

It follows that

$$\frac{I_{out}}{I_{in}} = \frac{\dfrac{H_{22}^{-1}}{H_{22}^{-1} + Z_{in}}A_I\dfrac{Z_{out}}{H_{11} + Z_{out}}}{1 + \dfrac{H_{22}^{-1}}{H_{22}^{-1} + Z_{in}}A_I\dfrac{Z_{out}}{H_{11} + Z_{out}}H_{21}}. \tag{8.72}$$

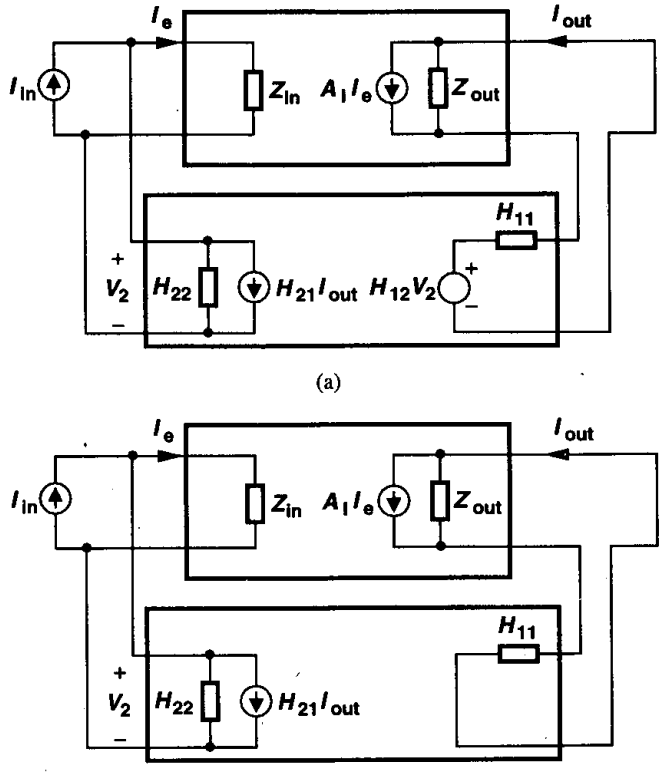


Figure 8.50 Current-current feedback circuit with (a) feedback network represented by an H model and (b) simplified H model.

We can thus define the loaded open-loop gain as

$$A_{I,open} = \frac{H_{22}^{-1}}{H_{22}^{-1} + Z_{in}} \frac{Z_{out}}{H_{11} + Z_{out}} A_I, \tag{8.73}$$

concluding that the feedback network introduces current division at both the input and the output of the system (Fig. 8.51). Note that H_{22} and H_{11} are measured with the input and the

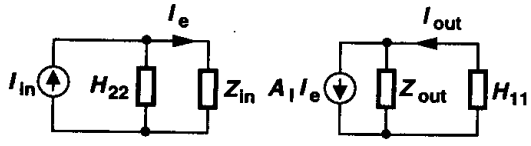


Figure 8.51 Current-current feedback circuit with proper loading of feedback network.

output ports of the feedback network open and shorted, respectively (Fig. 8.52). The loop gain is obtained as $H_{21} A_{I,open}$.

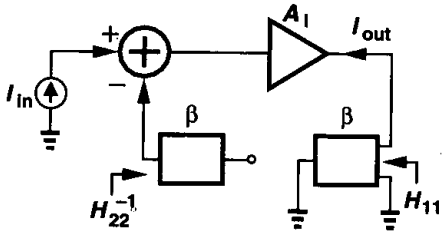


Figure 8.52 Conceptual view of including loading in current-current feedback.

Example 8.10

Calculate the open-loop and closed-loop gains of the circuit shown in Fig. 8.53(a).

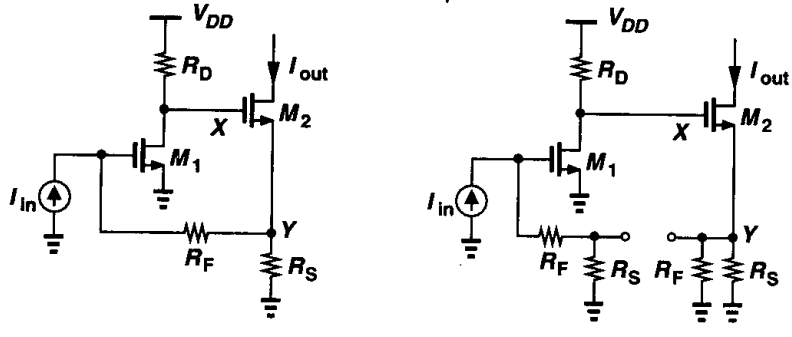


Figure 8.53

Solution

In this circuit, R_S and R_F sense the output current and return a fraction thereof to the input. Breaking the loop according to Fig. 8.52, we arrive at the circuit in Fig. 8.53(b), where we have

$$A_{I,open} = -(R_F + R_S)g_{m1}R_D \frac{1}{R_S \| R_F + 1/g_{m2}}. \tag{8.74}$$

The loop gain is given by $H_{21}A_{I,open}$, where, from (8.50), $H_{21} = I_2/I_1$ with $V_2 = 0$. For the feedback network consisting of R_S and R_F, we have $H_{21} = -R_S/(R_S + R_F)$. The closed-loop gain equals $A_{I,open}/(1 + H_{21}A_{I,open})$.

8.3.6 Summary of Loading Effects

The results of our study of loading are summarized in Fig. 8.54. The analysis is carried out in three steps: (1) open the loop with proper loading and calculate the open-loop gain, A_{OL}, and the open-loop input and output impedances; (2) determine the feedback ratio, β, and hence the loop gain, βA_{OL}; (3) calculate the closed-loop gain and input and output impedances by scaling the open-loop values by a factor of $1 + \beta A_{OL}$. Note that in the equations defining β, the subscripts 1 and 2 refer to the input and output ports of the feedback network, respectively.

In this chapter, we have described two methods of obtaining the loop gain: (1) by breaking the loop at an arbitrary point as shown in Fig. 8.5 and (2) by calculating A_{OL} and β as illustrated in Fig. 8.54. The two methods may yield slightly different results because the

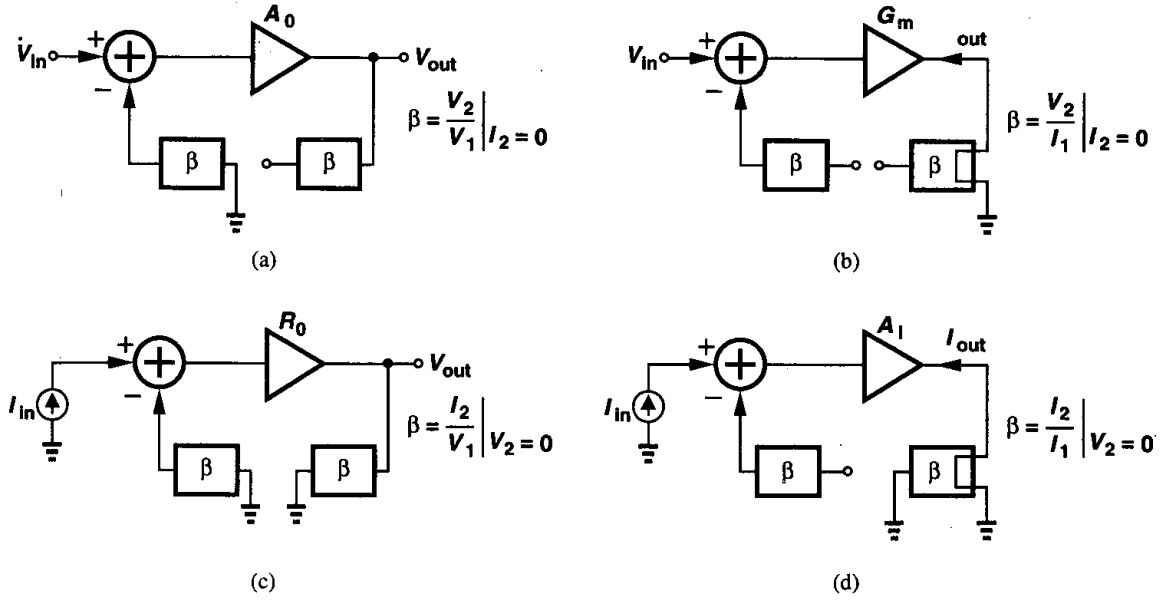


Figure 8.54 Summary of loading effects.

latter neglects the reverse transmission through the feedback network. However, the first method may be difficult to apply if loading effects must be taken into account because, if the loop can be broken at an arbitrary point, then the actual input and output ports of the overall system are unknown and the type of feedback unimportant. For example, the loop gain of the circuit of Fig. 8.53(a) does not depend on whether the output of interest is I_{out} or V_Y. In other words, since the first method does not distinguish between different types of feedback, it generally cannot utilize the loading calculations depicted in Fig. 8.54. For this reason, the second method is preferable.

We should also mention that some feedback circuits do not fall under any of the four types studied in this chapter because we have restricted our attention to cases where the output of interest is directly sensed by the feedback network. For example, if I_{out} in Fig. 8.53(a) flows through a resistor tied from the drain of M_2 to V_{DD}, then the resulting voltage is not inside the feedback loop. These cases are usually analyzed individually.

8.4 Effect of Feedback on Noise

Feedback does not improve the noise performance of circuits. Let us first consider the simple case illustrated in Fig. 8.55(a), where the open-loop voltage amplifier A_1 is characterized

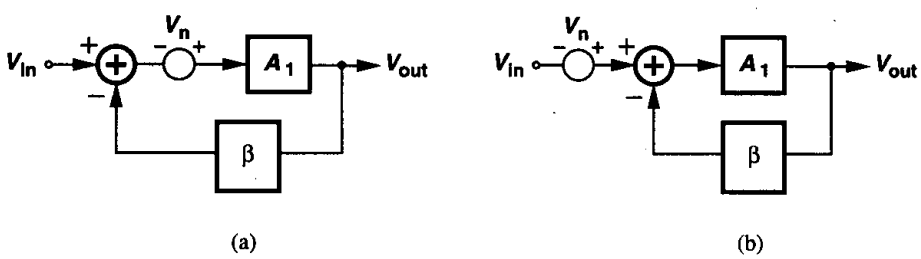


Figure 8.55 Feedback around a noisy circuit.

by only an input-referred noise voltage and the feedback network is noiseless. We have $(V_{in} - \beta V_{out} + V_n)A_1 = V_{out}$ and hence

$$V_{out} = (V_{in} + V_n)\frac{A_1}{1 + \beta A_1}. \tag{8.75}$$

Thus, the circuit can be simplified as shown in Fig. 8.55(b), revealing that the input-referred noise of the overall circuit is still equal to V_n. This analysis can be extended to all four feedback topologies to prove that the input-referred noise voltage and current remain the same if the feedback network introduces no noise. In practice, the feedback network itself may contain resistors or transistors, degrading the overall noise performance.

It is important to note that in Fig. 8.55(a) the output of interest is the same as the quantity sensed by the feedback network. This need not always be the case. For example, in the circuit of Fig. 8.56, the output is provided at the drain of M_1 whereas the feedback network senses the voltage at the source of M_1. In such cases, the input-referred noise of the closed-loop circuit may not be equal to that of the open-loop circuit even if the feedback network

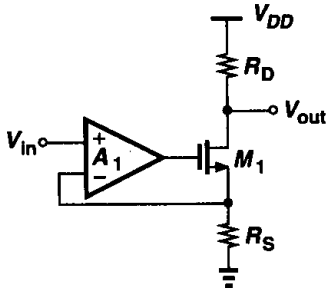


Figure 8.56 Noisy circuit with feedback sensing the source voltage.

is noiseless. As an example, let us consider the topology of Fig. 8.56 and, for simplicity, take only the noise of R_D, $V_{n,RD}$, into account. The reader can prove that the closed-loop voltage gain is equal to $-A_1 g_m R_D/[1+(1+A_1)g_m R_S]$ and hence the input-referred noise voltage due to R_D is

$$|V_{n,in,closed}| = \frac{|V_{n,RD}|}{A_1 R_D}\left[\frac{1}{g_m} + (1+A_1)R_S\right]. \tag{8.76}$$

For the open-loop circuit, on the other hand, the input-referred noise is

$$|V_{n,in,open}| = \frac{|V_{n,RD}|}{A_1 R_D}\left[\frac{1}{g_m} + R_S\right]. \tag{8.77}$$

Interestingly, as $A_1 \to \infty$, $|V_{n,in,closed}| \to |V_{n,RD}|R_S/R_D$ whereas $|V_{n,in,open}| \to 0$.

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.

8.1. Consider the circuit of Fig. 8.3(b), assuming I_1 is ideal and $g_{m1}r_{O1}$ cannot exceed 50. If a gain error of less than 5% is required, what is the maximum closed-loop voltage gain that can be achieved by this topology? What is the low-frequency closed-loop output impedance under this condition?

8.2. In the circuit of Fig. 8.7(a), assume $(W/L)_1 = 50/0.5$, $(W/L)_2 = 100/0.5$, $R_D = 2\,\text{k}\Omega$, and $C_2 = C_1$. Neglecting channel-length modulation and body effect, determine the bias current of M_1 and M_2 such that the input resistance at low frequencies is equal to 50 Ω.

8.3. Calculate the output impedance of the circuit shown in Fig. 8.8(a) at relatively low frequencies if R_D is replaced by an ideal current source.

8.4. Consider the example illustrated in Fig. 8.10. Suppose an overall voltage gain of 500 is required with maximum bandwidth. How many stages with what gain per stage must be placed in a cascade? (Hint: first find the 3-dB bandwidth of a cascade of n identical stages in terms of that of each stage.)

8.5. If in Fig. 8.18(b), amplifier A_0 exhibits an output impedance of R_0, calculate the closed-loop voltage gain and output impedance, taking into account loading effects.

8.6. Consider the circuit of Fig. 8.21(a), assuming $(W/L)_{1,2} = 50/0.5$ and $(W/L)_{3,4} = 100/0.5$. If $I_{SS} = 1$ mA, what is the maximum closed-loop voltage gain that can be achieved if the gain error is to remain below 5%?

8.7. The circuit of Fig. 8.36 can operate as a transimpedance amplifier if I_{out} flows through a resistor, R_{D2}, connected to V_{DD}, producing an output voltage. Replacing R_S with an ideal current source and assuming $\lambda = \gamma = 0$, calculate the transimpedance of the resulting circuit. Also, calculate the input-referred noise current per unit bandwidth.

8.8. For the circuit of Fig. 8.38(a), calculate the closed-loop gain without neglecting $G_{12}I_2$. Prove that this term can be neglected if $G_{12} \ll A_0 Z_{in}/Z_{out}$.

8.9. Calculate the loop gain of the circuit in Fig. 8.41 by breaking the loop at node X. Why is this result somewhat different from $G_{21}A_{v,open}$?

8.10. Using feedback techniques, calculate the input and output impedance and voltage gain of each circuit in Fig. 8.57.

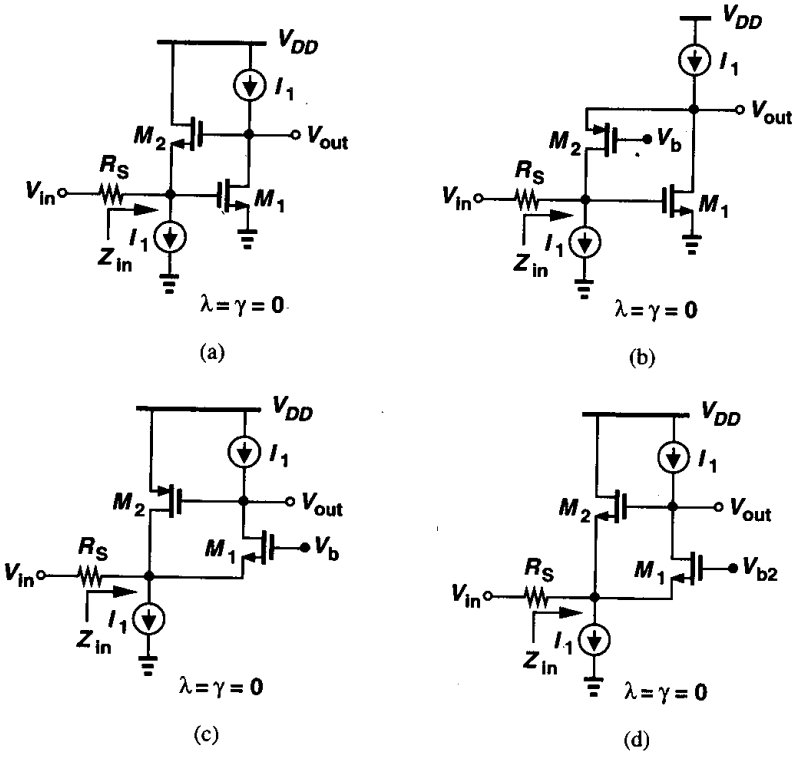


Figure 8.57

8.11. Using feedback techniques, calculate the input and output impedances of each circuit in Fig. 8.58.

8.12. Consider the circuit of Fig. 8.41(a), assuming $(W/L)_1 = (W/L)_2 = 50/0.5$, $\lambda = \gamma = 0$, and each resistor is equal to 2 kΩ. If $I_{D2} = 1$ mA, what is the bias current of M_1? What value of V_{in} gives such a current? Calculate the overall voltage gain.

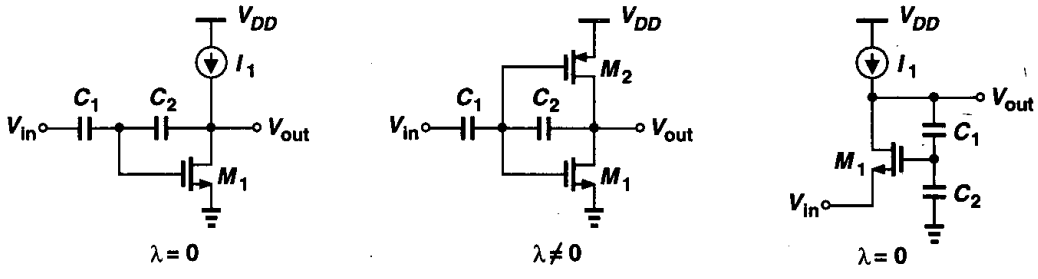


Figure 8.58

8.13. Suppose the amplifier of the circuit shown in Fig. 8.18 has an open-loop transfer function $A_0/(1+s/\omega_0)$ and an output resistance R_0. Calculate the output impedance of the closed-loop circuit and plot the magnitude as a function of frequency. Explain the behavior.

8.14. Calculate the input-referred noise voltage of the circuit shown in Fig. 8.21(a) at relatively low frequencies.

8.15. A differential pair with current-source loads can be represented as in Fig. 8.59(a), where $R_0 = r_{ON} \| r_{OP}$ and r_{ON} and r_{OP} denote the output resistance of NMOS and PMOS devices, respectively. Consider the circuit shown in Fig. 8.59(b), where G_{m1} and G_{m2} are placed in a negative feedback loop.

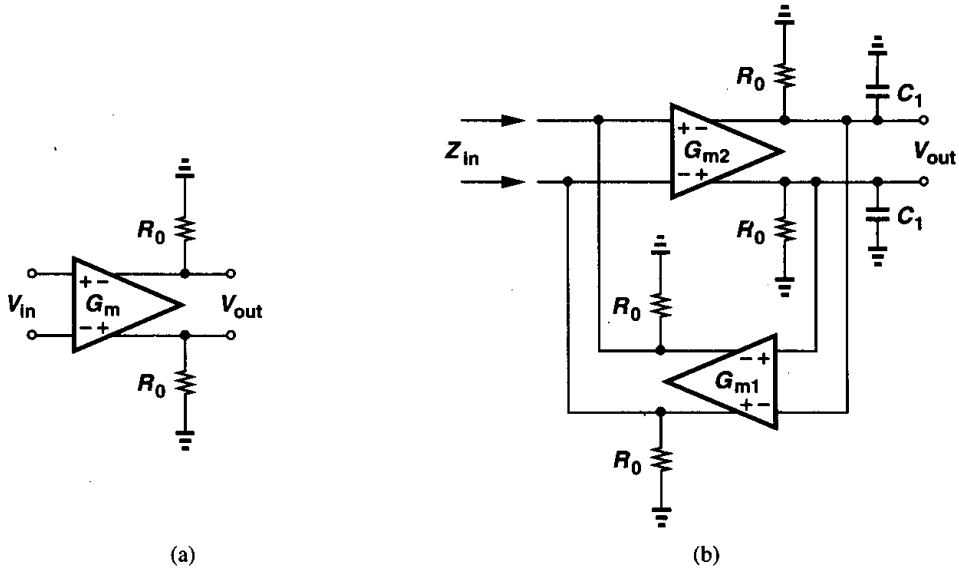


Figure 8.59

(a) Neglecting all other capacitances, derive an expression for Z_{in}. Sketch $|Z_{in}|$ versus frequency.
(b) Explain intuitively the behavior observed in part (a).
(c) Calculate the input-referred thermal noise voltage and current in terms of the input-referred noise voltage each G_m stage.

8.16. In the circuit of Fig. 8.60, $(W/L)_{1-3} = 50/0.5$, $I_{D1} = |I_{D2}| = |I_{D3}| = 0.5$ mA, and $R_{S1} = R_F = R_{D2} = 3\ \text{k}\Omega$.
(a) Determine the input bias voltage required to establish the above currents.
(b) Calculate the closed-loop voltage gain and output resistance.

8.17. The circuit of Fig. 8.60 can be modified as shown in Fig. 8.61, where a source follower, M_4, is inserted in the feedback loop. Note that M_1 and M_4 can also be viewed as a differential pair. Assume $(W/L)_{1-4} = 50/0.5$, $I_D = 0.5$ mA, for all transistors $R_{S1} = R_F = R_{D2} = 3\ \text{k}\Omega$, and $V_{b2} = 1.5$ V. Calculate the closed-loop voltage gain and output resistance and compare the results with those obtained in Problem 8.16(b).

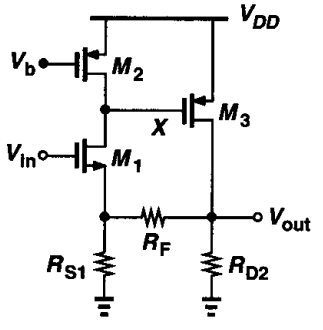


Figure 8.60

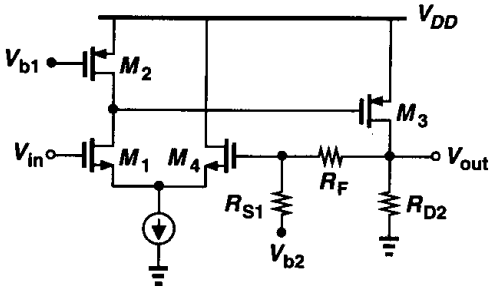


Figure 8.61

8.18. Consider the circuit of Fig. 8.62, where $(W/L)_{1-4} = 50/0.5$, $|I_{D1-4}| = 0.5$ mA and $R_2 = 3\ \text{k}\Omega$.
(a) For what range of R_1 are the above currents established while M_2 remains in saturation? What is the corresponding range of V_{in}?
(b) Calculate the closed-loop gain and output impedance for R_1 in the middle of the range obtained in part (a).

8.19. In the circuit of Fig. 8.63, suppose all resistors are equal to 2 kΩ and $g_{m1} = g_{m2} = 1/(200\ \Omega)$. Assuming $\lambda = \gamma = 0$, calculate the closed-loop gain and output impedance.

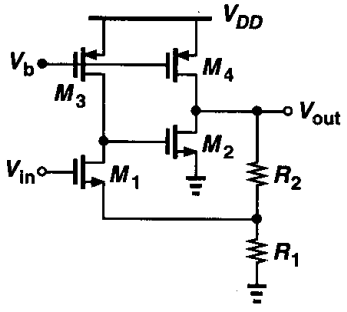


Figure 8.62

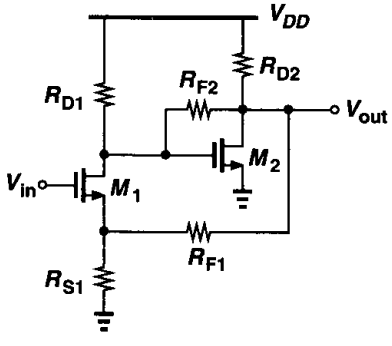


Figure 8.63

8.20. A CMOS inverter can be used as an amplifier with or without feedback (Fig. 8.64). Assume $(W/L)_{1,2} = 50/0.5$, $R_1 = 1\ \text{k}\Omega$, $R_2 = 10\ \text{k}\Omega$, and the dc levels of V_{in} and V_{out} are equal.

(a) Calculate the voltage gain and the output impedance of each circuit.

(b) Calculate the sensitivity of each circuit's output with respect to the supply voltage. That is, calculate the small-signal "gain" from V_{DD} to V_{out}. Which circuit exhibits less sensitivity?

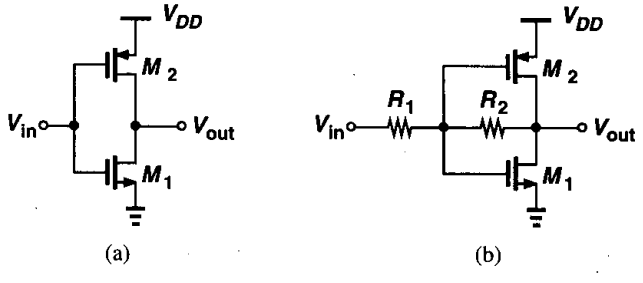


Figure 8.64

8.21. Calculate the input-referred thermal noise voltage of the circuits shown in Fig. 8.64.

8.22. The circuit shown in Fig. 8.65 employs positive feedback to produce a negative input capacitance. Using feedback analysis techniques, determine Z_{in} and identify the negative capacitance component. Assume $\lambda = \gamma = 0$.

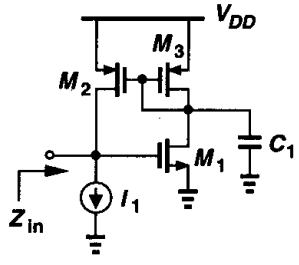


Figure 8.65

8.23. In the circuit of Fig. 8.66, assume $\lambda = 0$, $g_{m1,2} = 1/(200\ \Omega)$, $R_{1-3} = 2\ \text{k}\Omega$, and $C_1 = 100$ pF. Neglecting other capacitances, estimate the closed-loop voltage gain at very low and very high frequencies.

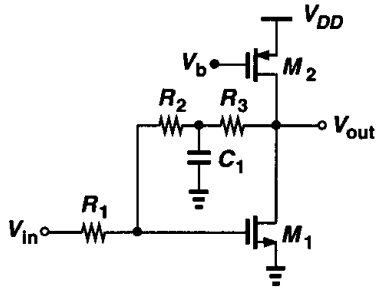


Figure 8.66

Chapter 9

Operational Amplifiers

Operational amplifiers (op amps) are an integral part of many analog and mixed-signal systems. Op amps with vastly different levels of complexity are used to realize functions ranging from dc bias generation to high-speed amplification or filtering. The design of op amps continues to pose a challenge as the supply voltage and transistor channel lengths scale down with each generation of CMOS technologies.

This chapter deals with the analysis and design of CMOS op amps. Following a review of performance parameters, we describe simple op amps such as telescopic and folded cascode topologies. Next, we study two-stage and gain-boosting configurations and the problem of common-mode feedback. Finally, we introduce the concept of slew rate and analyze the effect of supply rejection and noise in op amps.

9.1 General Considerations

We loosely define an op amp as a "high-gain differential amplifier." By "high," we mean a value that is adequate for the application, typically in the range of 10^1 to 10^5. Since op amps are usually employed to implement a feedback system, their open-loop gain is chosen according to the precision required of the closed-loop circuit.

Up to two decades ago, most op amps were designed to serve as "general-purpose" building blocks, satisfying the requirements of many different applications. Such efforts sought to create an "ideal" op amp, e.g., with very high voltage gain (several hundred thousand), high input impedance, and low output impedance, but at the cost of many other aspects of the performance, e.g., speed, output voltage swings, and power dissipation.

By contrast, today's op amp design proceeds with the recognition that the trade-offs between the parameters eventually require a multi-dimensional compromise in the overall implementation, making it necessary to know the *adequate* value that must be achieved for each parameter. For example, if the speed is critical while the gain error is not, a topology is chosen that favors the former, possibly sacrificing the latter.

9.1.1 Performance Parameters

In this section, we describe a number of op amp design parameters, providing an understanding of why and where each may become important. For this discussion, we

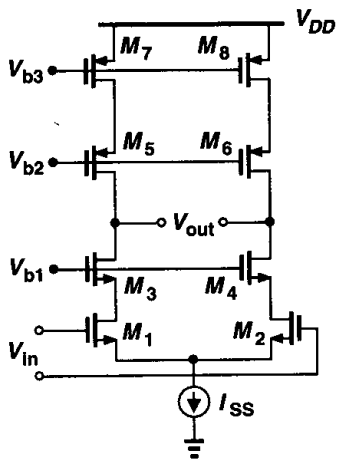


Figure 9.1 Cascode op amp.

consider the differential cascode circuit shown in Fig. 9.1 as a representative op amp design.[1] The voltages V_{b1}-V_{b3} are generated by the current mirror techniques described in Chapter 5.

Gain The open-loop gain of an op amp determines the precision of the feedback system employing the op amp. As mentioned before, the required gain may vary by four orders of magnitude according to the application. Trading with such parameters as speed and output voltage swings, the minimum required gain must therefore be known. As explained in Chapter 13, a high open-loop gain may also be necessary to suppress nonlinearity.

Example 9.1

The circuit of Fig. 9.2 is designed for a nominal gain of 10, i.e., $1 + R_1/R_2 = 10$. Determine the minimum value of A_1 for a gain error of 1%.

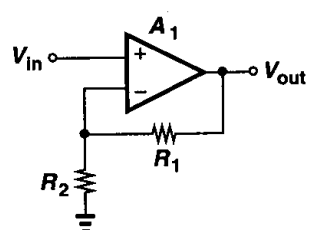


Figure 9.2

[1] Since op amps of this type have a high output resistance, they are sometimes called "operational transconductance amplifiers" (OTAs). In the limit, the circuit can be represented by a single voltage-dependent current source and called a "G_m stage."

Solution

The closed-loop gain is obtained from Chapter 8 as:

$$\frac{V_{out}}{V_{in}} = \frac{A_1}{1 + \dfrac{R_2}{R_1 + R_2} A_1} \tag{9.1}$$

$$= \frac{R_1 + R_2}{R_2} \frac{A_1}{\dfrac{R_1 + R_2}{R_2} + A_1}. \tag{9.2}$$

Predicting that $A_1 \gg 10$, we approximate (9.2) as:

$$\frac{V_{out}}{V_{in}} \approx \left(1 + \frac{R_1}{R_2}\right)\left(1 - \frac{R_1 + R_2}{R_2}\frac{1}{A_1}\right). \tag{9.3}$$

The term $(R_1 + R_2)/(R_2 A_1) = (1 + R_1/R_2)/A_1$ represents the relative gain error. To achieve a gain error less than 1%, we must have $A_1 > 1000$.

It is instructive to compare the circuit of Fig. 9.2 with an open-loop implementation such as that in Fig. 9.3. While it is possible to obtain a nominal gain of $g_m R_D = 10$ by a common-source stage, it is extremely difficult to guarantee an error less than 1%. The variations in the mobility and gate oxide thickness of the transistor and the value of the resistor typically yield an error greater than 20%.

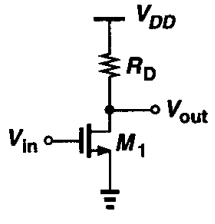


Figure 9.3 Simple common-source stage.

Small-Signal Bandwidth The high-frequency behavior of op amps plays a critical role in many applications. For example, as the frequency of operation increases, the open-loop gain begins to drop (Fig. 9.4), creating larger errors in the feedback system. The small-signal bandwidth is usually defined as the "unity-gain" frequency, f_u, which exceeds 1 GHz in

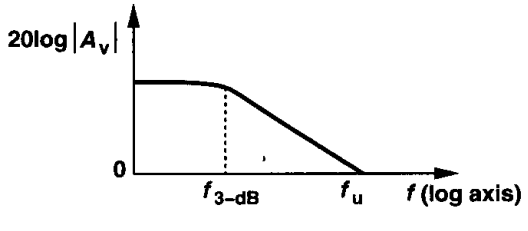


Figure 9.4 Gain roll-off with frequency.

today's CMOS op amps. The 3-dB frequency, $f_{3-\text{dB}}$, may also be specified to allow easier prediction of the closed-loop frequency response.

Example 9.2

In the circuit of Fig. 9.5, assume the op amp is a single-pole voltage amplifier. If V_{in} is a small step,

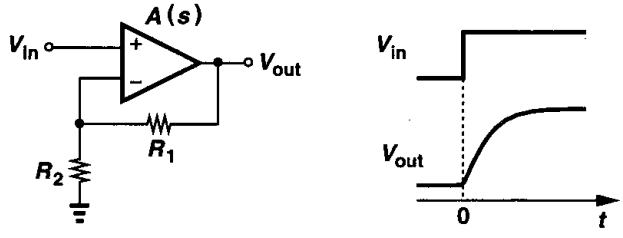


Figure 9.5

calculate the time required for the output voltage to reach within 1% of its final value. What unity-gain bandwidth must the op amp provide if $1 + R_1/R_2 \approx 10$ and the settling time is to be less than 5 ns? For simplicity, assume the low-frequency gain is much greater than unity.

Solution

Since

$$\left(V_{in} - V_{out}\frac{R_2}{R_1 + R_2}\right)A(s) = V_{out}, \tag{9.4}$$

we have

$$\frac{V_{out}}{V_{in}}(s) = \frac{A(s)}{1 + \dfrac{R_2}{R_1 + R_2}A(s)}. \tag{9.5}$$

For a one-pole system, $A(s) = A_0/(1 + s/\omega_0)$, where ω_0 is the 3-dB bandwidth and $A_0\omega_0$ the unity-gain bandwidth. Thus,

$$\frac{V_{out}}{V_{in}}(s) = \frac{A_0}{1 + \dfrac{R_2}{R_1 + R_2}A_0 + \dfrac{s}{\omega_0}} \tag{9.6}$$

$$= \frac{\dfrac{A_0}{1 + \dfrac{R_2}{R_1 + R_2}A_0}}{1 + \dfrac{s}{\left(1 + \dfrac{R_2}{R_1 + R_2}A_0\right)\omega_0}}, \tag{9.7}$$

indicating that the closed-loop amplifier is also a one-pole system with a time constant equal to

$$\tau = \frac{1}{\left(1 + \dfrac{R_2}{R_1 + R_2}A_0\right)\omega_0}. \tag{9.8}$$

Recognizing that the quantity $R_2A_0/(R_1 + R_2)$ is the low-frequency loop gain and usually much greater than unity, we have

$$\tau \approx \left(1 + \frac{R_1}{R_2}\right)\frac{1}{A_0\omega_0}. \tag{9.9}$$

The output step response for $V_{in} = au(t)$ can now be expressed as

$$V_{out}(t) \approx a\left(1 + \frac{R_1}{R_2}\right)\left(1 - \exp\frac{-t}{\tau}\right)u(t), \tag{9.10}$$

with the final value $V_F \approx a(1 + R_1/R_2)$. For 1% settling, $V_{out} = 0.99V_F$ and hence

$$1 - \exp\frac{-t_{1\%}}{\tau} = 0.99, \tag{9.11}$$

yielding $t_{1\%} = \tau \ln 100 \approx 4.6\tau$. For a 1% settling of 5 ns, $\tau \approx 1.09$ ns, and from (9.9), $A_0\omega_0 = (1 + R_1/R_2)/\tau = 9.21$ Grad/s (1.47 GHz).

The key point in the above example is that the required bandwidth depends on both the settling accuracy and the closed-loop gain that must be provided.

Large-Signal Bandwidth In many of today's applications, op amps must operate with large transient signals. Under these conditions, nonlinear phenomena make it difficult to characterize the speed by merely small-signal properties such as the open-loop response shown in Fig. 9.4. As an example, suppose the feedback circuit of Fig. 9.5 incorporates a realistic op amp (i.e., with finite output impedance) while driving a large load capacitance. How does the circuit behave if we apply a 1-V step at the input? Since the output voltage cannot change instantaneously, the voltage difference sensed by the op amp itself at $t \geq 0$ is equal to 1 V. Such a large difference momentarily drives the op amp into a nonlinear region of operation. (Otherwise, with an open-loop gain of, say, 1000, the op amp would produce 1000 V at the output.)

As explained in Section 9.8, the large-signal behavior is usually quite complex, mandating careful simulations.

Output Swing Most systems employing op amps require large voltage swings to accommodate a wide range of signal amplitudes. For example, a high-quality microphone that senses the music produced by an orchestra may generate instantaneous voltages that vary by more than four orders of magnitude, demanding that subsequent amplifiers and filters handle large swings (and/or achieve a low noise).

The need for large output swings has made fully differential op amps quite popular. Similar to the circuits described in Chapter 4, such op amps generate "complementary" outputs, roughly doubling the available swing. Nonetheless, as mentioned in Chapters 3 and 4 and explained later in this chapter, the maximum voltage swing trades with device size and bias currents and hence speed. Achieving large swings is the principal challenge in today's op amp design.

Linearity Open-loop op amps suffer from substantial nonlinearity. In the circuit of Fig. 9.1, for example, the input pair M_1-M_2 exhibits a nonlinear relationship between its differential drain current and input voltage. As explained in Chapter 13, the issue of nonlinearity is tackled by two approaches: using fully differential implementations to suppress even-order harmonics and allowing sufficient open-loop gain such that the closed-loop feedback system achieves adequate linearity. It is interesting to note that in many feedback circuits, the linearity requirement, rather than the gain error requirement, governs the choice of the open-loop gain.

Noise and Offset The input noise and offset of op amps determine the minimum signal level that can be processed with reasonable quality. In a typical op amp topology, several devices contribute noise and offset, necessitating large dimensions or bias currents. For example, in the circuit of Fig. 9.1, M_1-M_2 and M_7-M_8 contribute the most.

We should also recognize a trade-off between noise and *output swing*. For a given bias current, as the overdrive voltage of M_7 and M_8 in Fig. 9.1 is lowered to allow larger swings at the output, their transconductance increases and so does their drain noise current.

Supply Rejection Op amps are often employed in mixed-signal systems and sometimes connected to noisy digital supply lines. Thus, the performance of op amps in the presence of supply noise, especially as the noise frequency increases, is quite important. For this reason, fully differential topologies are preferred.

9.2 One-Stage Op Amps

All of the differential amplifiers studied in Chapters 4 and 5 can be considered as op amps. Fig. 9.6 shows two such topologies with single-ended and differential outputs. The small-signal, low-frequency gain of both circuits is equal to $g_{mN}(r_{ON} \| r_{OP})$, where the subscripts N and P denote NMOS and PMOS, respectively. This value hardly exceeds 20 in submicron devices with typical current levels. The bandwidth is usually determined by the load capacitance, C_L. Note that the circuit of Fig. 9.6(a) exhibits a mirror pole (Chapter 5)

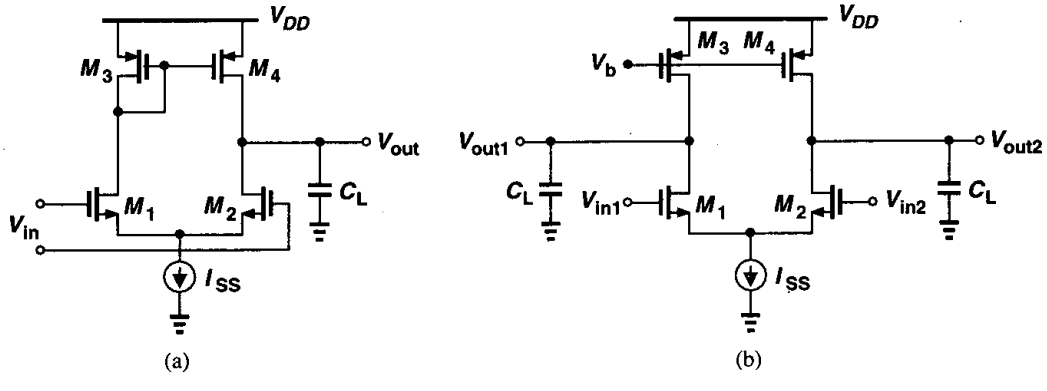


Figure 9.6 Simple op amp topologies.

whereas that of Fig. 9.6(b) does not, a critical difference in terms of the stability of feedback systems using these topologies (Chapter 10).

The circuits of Fig. 9.6 suffer from noise contributions of M_1-M_4, as calculated in Chapter 7. Interestingly, in all op amp topologies, at least four devices contribute to the input noise: two input transistors and two "load" transistors.

Example 9.3

Calculate the input common-mode voltage range and the closed-loop output impedance of the unity-gain buffer depicted in Fig. 9.7.

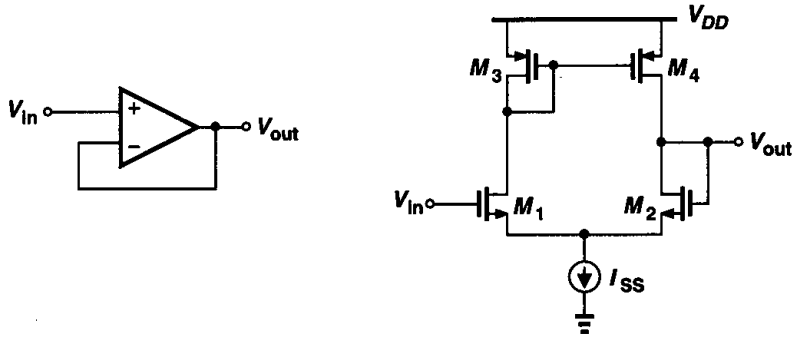


Figure 9.7

Solution

The minimum allowable input voltage is equal to $V_{CSS} + V_{GS1}$, where V_{CSS} is the voltage required across the current source. The maximum voltage is given by the level that places M_1 at the edge of the triode region: $V_{in,max} = V_{DD} - |V_{GS3}| + V_{TH1}$. For example, if each device (including the current source) has a threshold voltage of 0.7 V and an overdrive of 0.3 V, then $V_{in,min} = 0.3 + 0.3 + 0.7 = 1.3$ V and $V_{in,max} = 3 - (0.3 + 0.7) + 0.7 = 2.7$ V. Thus, the input CM range equals 1.4 V with a 3-V supply.

Since the circuit employs voltage feedback at the output, the output impedance is equal to the open-loop value, $r_{OP} \| r_{ON}$, divided by one plus the loop gain, $1 + g_{mN}(r_{OP} \| r_{ON})$. In other words, for large open-loop gain, the closed-loop output impedance is approximately equal to $(r_{OP} \| r_{ON})/[g_{mN}(r_{OP} \| r_{ON})] = 1/g_{mN}$.

It is interesting to note that the closed-loop output impedance is relatively *independent* of the open-loop output impedance. This is an important observation, allowing us to design high-gain op amps by *increasing* the open-loop output impedance while still achieving a relatively low closed-loop output impedance.

In order to achieve a high gain, the differential cascode topologies of Chapters 4 and 5 can be used. Shown in Figs. 9.8(a) and (b) for single-ended and differential output generation, respectively, such circuits display a gain on the order of $g_{mN}[(g_{mN}r_{ON}^2) \| (g_{mP}r_{OP}^2)]$, but at the cost of output swing and additional poles. These configurations are also called "telescopic" cascode op amps to distinguish them from another cascode op amp described below. The circuit providing a single-ended output suffers from a mirror pole at node X, creating stability issues (Chapter 10).

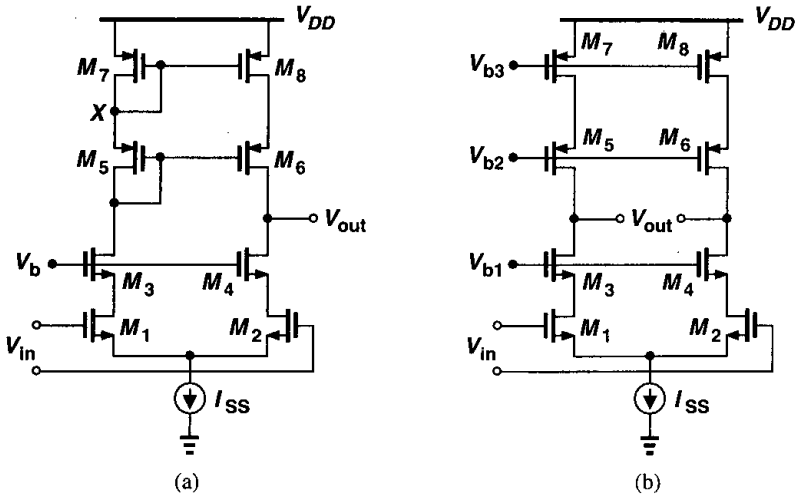


Figure 9.8 Cascode op amps.

As calculated in Chapter 4 , the output swings of telescopic op amps are relatively limited. In the fully differential version of Fig. 9.8(b), for example, the output swing is given by $2[V_{DD} - (V_{OD1} + V_{OD3} + V_{CSS} + |V_{OD5}| + |V_{OD7}|)]$, where V_{ODj} denotes the overdrive voltage of M_j.

Another drawback of telescopic cascodes is the difficulty in shorting their inputs and outputs, e.g., to implement a unity-gain buffer similar to the circuit of Fig. 9.7. To understand the issue, let us consider the unity-gain feedback topology shown in Fig. 9.9. Under what conditions are both M_2 and M_4 in saturation? We must have $V_{out} \leq V_X + V_{TH2}$ and

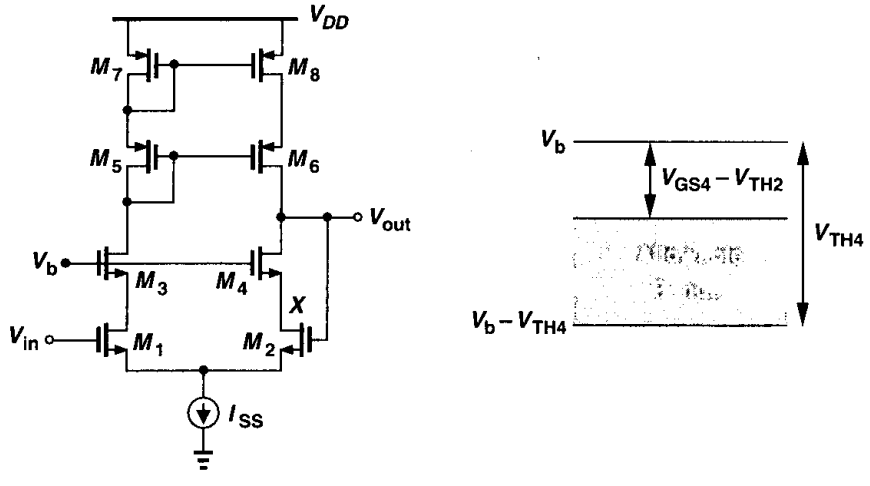


Figure 9.9 Cascode op amp with input and output shorted.

$V_{out} \geq V_b - V_{TH4}$. Since $V_X = V_b - V_{GS4}$, $V_b - V_{TH4} \leq V_{out} \leq V_b - V_{GS4} + V_{TH2}$. Depicted in Fig. 9.9, this voltage range is simply equal to $V_{max} - V_{min} = V_{TH4} - (V_{GS4} - V_{TH2})$, maximized by minimizing the overdrive of M_4 but always less than V_{TH2}.

Example 9.4

For the circuit of Fig. 9.9, explain in which region each transistor operates as V_{in} varies from below $V_b - V_{TH4}$ to above $V_b - V_{GS4} + V_{TH2}$.

Solution

Since the op amp attempts to force V_{out} to be equal to V_{in}, for $V_{in} < V_b - V_{TH4}$, we have $V_{out} \approx V_{in}$ and M_4 is in the triode region while other transistors are saturated. Under this condition, the open-loop gain of the op amp is reduced.

As V_{in} and hence V_{out} exceed $V_b - V_{TH4}$, M_4 enters saturation and the open-loop gain reaches a maximum. For $V_b - V_{TH4} < V_{in} < V_b - (V_{GS4} - V_{TH2})$, both M_2 and M_4 are saturated and for $V_{in} > V_b - (V_{GS4} - V_{TH2})$, M_2 and M_1 enter the triode region, degrading the gain.

While a cascode op amp is rarely used as a unity-gain buffer, some other topologies such as the switched-capacitor circuits of Chapter 12 require that the input and output of the op amp be shorted for part of the operation period.

At this point, the reader may wonder how exactly we design an op amp. With so many devices and performance parameters, it may not be clear where the starting point is and how the numbers are chosen. Indeed, the actual design methodology of an op amp somewhat depends on the specifications that the circuit must meet. For example, a high-gain op amp may be designed quite differently from a low-noise op amp. Nevertheless, in most cases, some aspects of the performance, e.g., output voltage swings and open-loop gain, are of primary concern, pointing to a specific design procedure. The following example illustrates these ideas.

Example 9.5

Design a fully differential telescopic op amp with the following specifications: $V_{DD} = 3$ V, differential output swing = 3 V, power dissipation = 10 mW, voltage gain = 2000. Assume $\mu_n C_{ox} = 60\ \mu\text{A/V}^2$, $\mu_p C_{ox} = 30\ \mu\text{A/V}^2$, $\lambda_n = 0.1\ \text{V}^{-1}$, $\lambda_p = 0.2\ \text{V}^{-1}$ (for an effective channel length of 0.5 μm), $\gamma = 0$, $V_{THN} = |V_{THP}| = 0.7$ V.

Solution

Fig. 9.10 shows the op amp topology along with two current mirrors defining the drain currents of M_7-M_9. We begin with the power budget, allocating 3 mA to M_9 and the remaining 330 μA to M_{b1} and M_{b2}. Thus, each cascode branch of the op amp carries a current of 1.5 mA. Next, we consider the required output swings. Each of nodes X and Y must be able to swing by 1.5 V without driving M_3-M_6 into the triode region. With a 3-V supply, therefore, the total voltage available for M_9 and each cascode branch is equal to 1.5 V, i.e., $|V_{OD7}| + |V_{OD5}| + V_{OD3} + V_{OD1} + V_{OD9} = 1.5$ V. Since M_9 carries the largest current, we choose $V_{OD9} \approx 0.5$ V, leaving 1 V for the four transistors in the cascode. Moreover, since M_5-M_8 suffer from low mobility, we allocate an overdrive of approximately 300 mV to each, obtaining 400 mV for $V_{OD1} + V_{OD3}$. As an initial guess, $V_{OD1} = V_{OD3} = 200$ mV.

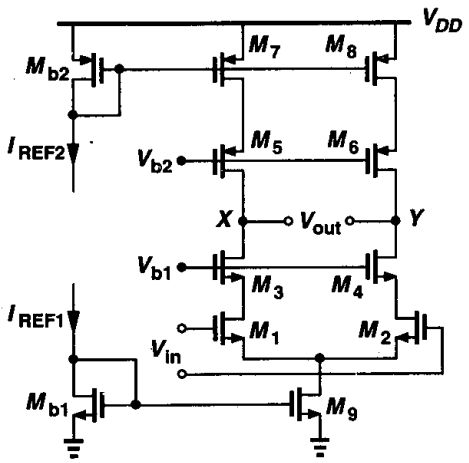


Figure 9.10

With the bias current and overdrive voltage of each transistor known, we can easily determine the aspect ratios from $I_D = (1/2)\mu C_{ox}(W/L)(V_{GS} - V_{TH})^2$. To minimize the device capacitances, we choose the minimum length for each transistor, obtaining a corresponding width. We then have $(W/L)_{1-4} = 1250$, $(W/L)_{5-8} = 1111$, $(W/L)_9 = 400$.

The design has thus far satisfied the swing, power dissipation, and supply voltage specifications. But, how about the gain? Using $A_v \approx g_{m1}[(g_{m3}r_{O3}r_{O1})\|(g_{m5}r_{O5}r_{O7})]$ and assuming minimum channel length for all of the transistors, we have $A_v = 1416$, quite lower than the required value.

In order to increase the gain, we recognize that $g_m r_O = \sqrt{2\mu C_{ox}(W/L)I_D}/(\lambda I_D)$. Now, recall that $\lambda \propto 1/L$, and hence $g_m r_O \propto \sqrt{WL/I_D}$. We can therefore increase the width or length or *decrease* the bias current of the transistors. In practice, speed or noise requirements may dictate the bias current, leaving only the dimensions as the variables. Of course, the width of each transistor must at least scale with its length so as to maintain a constant overdrive voltage.

Which transistors in the circuit of Fig. 9.10 should be made longer? Since M_1-M_4 appear in the signal path, it is desirable to keep their capacitances to a minimum. The PMOS devices, M_5-M_8, on the other hand, affect the signal to a much lesser extent and can therefore have larger dimensions.[2] Doubling the (effective) length and width of each of these transistors in fact *doubles* their $g_m r_O$ because g_m remains constant while r_O increases by a factor of 2. Choosing $(W/L)_{5-8} = 1111\ \mu\text{m}/1.0\ \mu\text{m}$ and hence $\lambda_p = 0.1\ \text{V}^{-1}$, we obtain $A_v \approx 4000$. Thus, the PMOS dimensions can be somewhat smaller. Note that with such large dimensions for PMOS transistors, we may revisit our earlier distribution of the overdrive voltages, possibly reducing that of M_9 by 100 to 200 mV and allocating more to the PMOS devices.

In the op amp of Fig. 9.10, the input CM level and the bias voltages V_{b1} and V_{b2} must be chosen so as to allow maximum output swings. The minimum allowable input CM level equals $V_{GS1} + V_{OD9} = V_{TH1} + V_{OD1} + V_{OD9} = 1.4$ V. The minimum value of V_{b1} is given by $V_{GS3} + V_{OD1} + V_{OD9} = 1.6$ V, placing M_1-M_2 at the edge of the triode region. Similarly, $V_{b2,max} = V_{DD} - (|V_{GS5}| + |V_{OD7}|) = 1.7$ V. In practice, some margin must be included in the value of V_{b1} and V_{b2} to allow for process variations. Also, the increase in the threshold voltages due to body effect must be taken into account.

[2]This point is studied in Chapter 10.

In order to alleviate the drawbacks of telescopic cascode op amps, namely, limited output swings and difficulty in shorting the input and output, a "folded cascode" op amp can be used. As described in Chapter 3 and illustrated in Fig. 9.11, in an NMOS or PMOS cascode amplifier, the input device is replaced by the opposite type while still converting the

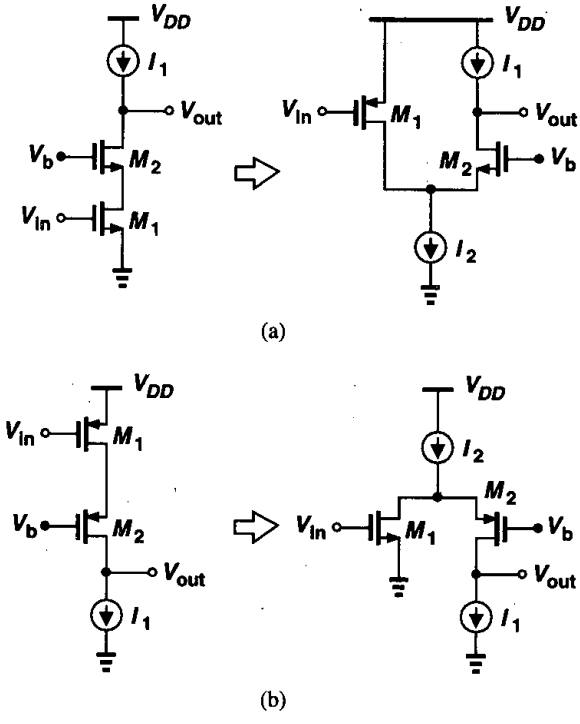


Figure 9.11 Folded cascode circuits.

input voltage to a current. In the four circuits shown in Fig. 9.11, the small-signal current generated by M_1 flows through M_2 and subsequently the load, producing an output voltage approximately equal to $g_{m1}R_{out}V_{in}$. The primary advantage of the folded structure lies in the choice of the voltage levels because it does not "stack" the cascode transistor on top of the input device. We will return to this point later.

The folding idea depicted in Fig. 9.11 can easily be applied to differential pairs and hence operational amplifiers as well. Shown in Fig. 9.12, the resulting circuit replaces the input NMOS pair with a PMOS counterpart. Note two important differences between the two circuits. (1) In Fig. 9.12(a), one bias current, I_{SS}, provides the drain current of both the input transistors and the cascode devices, whereas in Fig. 9.12(b) the input pair requires an additional bias current. In other words, $I_{SS1} = I_{SS}/2 + I_{D3}$. Thus, the folded-cascode configuration generally consumes higher power. (2) In Fig. 9.12(a), the input CM level cannot exceed $V_{b1} - V_{GS3} + V_{TH1}$, whereas in Fig. 9.12(b), it cannot be *less* than $V_{b1} - V_{GS3} + |V_{THP}|$. It is therefore possible to design the latter to allow shorting its input and output terminals with negligible swing limitation. This is in contrast to the behavior

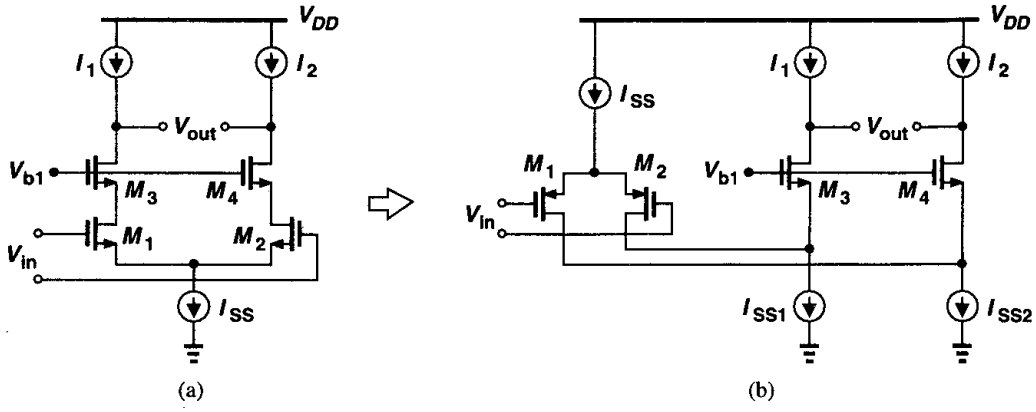


Figure 9.12 Folded cascode op amp topology.

depicted in Fig. 9.9. In Fig. 9.12(b), it is possible to tie the n-well of M_1 and M_2 to their common source point. We return to this idea in Chapters 13 and 18.

Let us now calculate the maximum output voltage swing of the folded-cascode op amp shown in Fig. 9.13, where M_5-M_{10} replace the ideal current sources of Fig. 9.12(b). With proper choice of V_{b1} and V_{b2}, the lower end of the swing is given by $V_{OD3} + V_{OD5}$ and the upper end by $V_{DD} - (|V_{OD7}| + |V_{OD9}|)$. Thus, the peak-to-peak swing on each side is equal to $V_{DD} - (V_{OD3} + V_{OD5} + |V_{OD7}| + |V_{OD9}|)$. In the telescopic cascode of Fig. 9.12(a), on the other hand, the swing is less by the overdrive of the tail current source. We should nonetheless note that, carrying a large current, M_5 and M_6 in Fig. 9.13 may require a high overdrive voltage if their capacitance contribution to nodes X and Y is to be minimized.

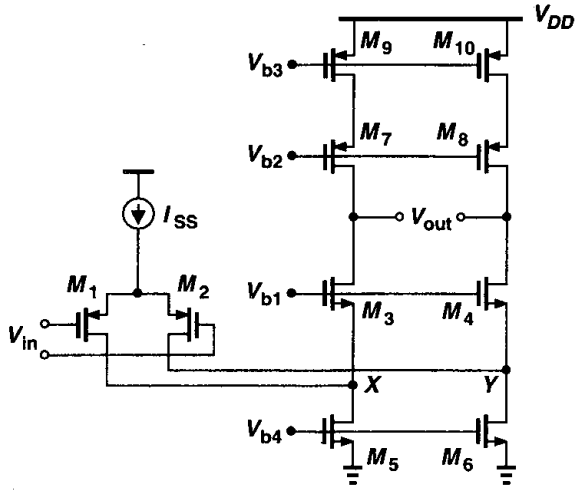


Figure 9.13 Folded cascode op amp with cascode PMOS loads.

We now determine the small-signal voltage gain of the folded-cascode op amp of Fig. 9.13. Using the half circuit depicted in Fig. 9.14(a) and writing $|A_v| = G_m R_{out}$, we must calculate G_m and R_{out}. As shown in Fig. 9.14(b), the output short-circuit current is approximately equal to the drain current of M_1 because the impedance seen looking into the source of M_3, that is, $(g_{m3} + g_{mb3})^{-1} \| r_{O3}$, is typically much lower than $r_{O1} \| r_{O5}$. Thus, $G_m \approx g_{m1}$. To calculate R_{out}, we use Fig. 9.14(c), with $R_{OP} \approx (g_{m7} + g_{mb7}) r_{O7} r_{O9}$, to write $R_{out} \approx R_{OP} \| [(g_{m3} + g_{mb3}) r_{O3} (r_{O1} \| r_{O5})]$. It follows that

$$|A_v| \approx g_{m1} \{ [(g_{m3} + g_{mb3}) r_{O3} (r_{O1} \| r_{O5})] \| [(g_{m7} + g_{mb7}) r_{O7} r_{O9}] \}. \tag{9.12}$$

How does this value compare with the gain of a telescopic op amp? For comparable device dimensions and bias currents, the PMOS input differential pair exhibits a lower transconductance than does an NMOS pair. Furthermore, r_{O1} and r_{O5} appear in parallel,

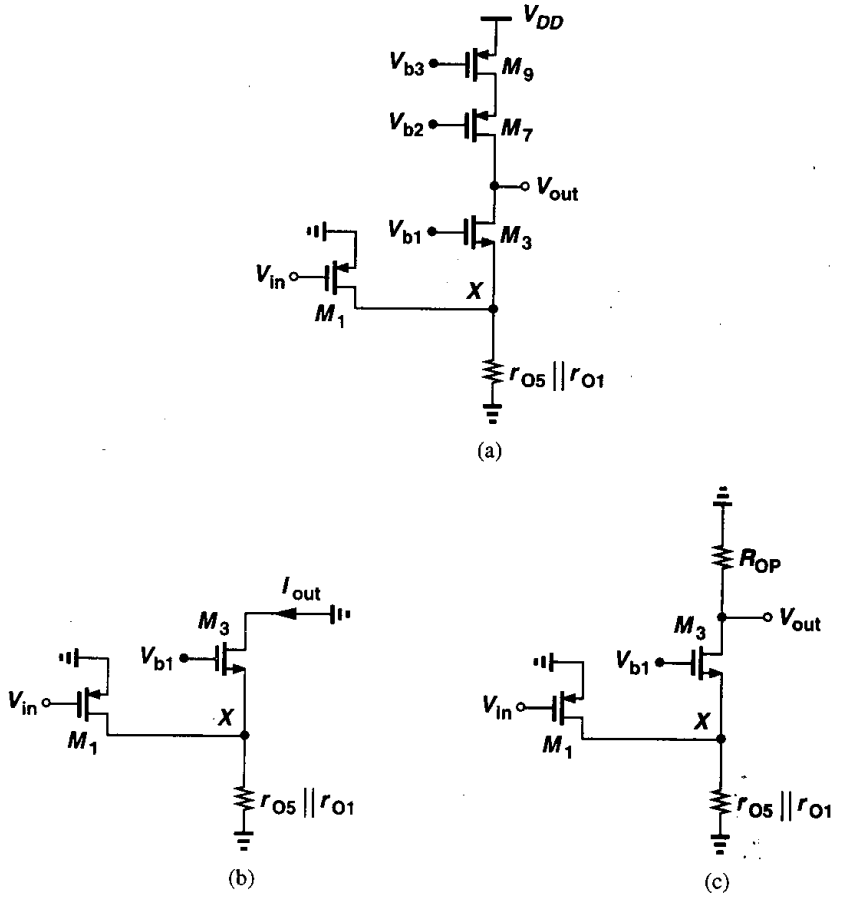


Figure 9.14 (a) Half circuit of folded cascode op amp, (b) equivalent circuit with output shorted to ground, (c) equivalent circuit with output open.

reducing the output impedance, especially because M_5 carries the currents of both the input device and the cascode branch. As a consequence, the gain in (9.12) is usually two to three times lower than that of a comparable telescopic cascode.

It is also worth noting that the pole at the "folding point," i.e., the sources of M_3 and M_4, is quite closer to the origin than that associated with the source of cascode devices in a telescopic topology. In Fig. 9.15(a), C_{tot} arises from C_{GS3}, C_{SB3}, C_{DB1}, and C_{GD1}.

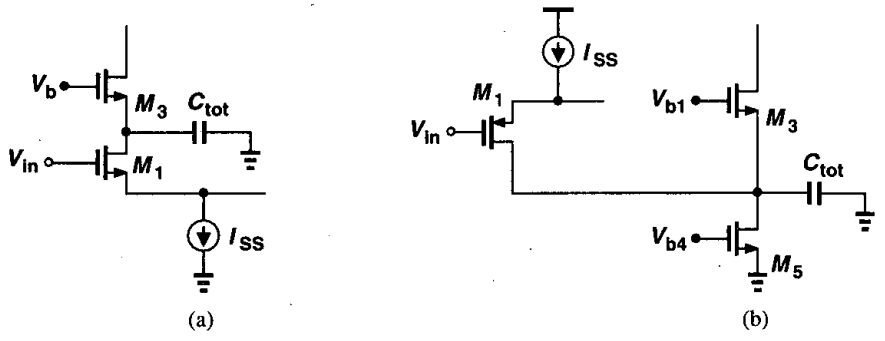


Figure 9.15 Effect of device capacitance on the nondominant pole in telescopic and folded-cascode op amps.

By contrast, in Fig. 9.15(b), C_{tot} contains additional contributions due to C_{GD5} and C_{DB5}, typically significant components because M_5 must be wide enough to carry a large current with a small overdrive.

A folded-cascode op amp may incorporate NMOS input devices and PMOS cascode transistors. Illustrated in Fig. 9.16, such a circuit potentially provides a higher gain than the op amp of Fig. 9.13 because of the greater mobility of NMOS devices, but at the cost of lowering the pole at the folding point. To understand why, note that the pole at node X is given by the product of $1/(g_{m3}+g_{mb3})$ and the total capacitance at this node. The magnitude of both of these components is relatively high: M_3 suffers from a low transconductance and

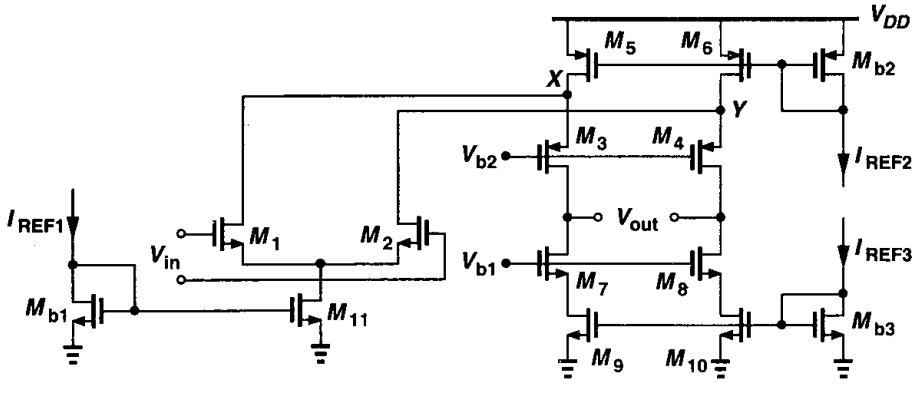


Figure 9.16 Realization of a folded-cascode op amp.

M_5 contributes substantial capacitance because it must be wide enough to carry the drain currents of both M_1 and M_3. In fact, for comparable bias currents, M_5-M_6 in Fig. 9.16 may be several times wider than M_5-M_6 in Fig. 9.13.

Our study thus far suggests that the overall voltage swing of a folded-cascode op amp is only slightly higher than that of a telescopic configuration. This advantage comes at the cost of higher power dissipation, lower voltage gain, lower pole frequencies, and, as explained in Section 9.10, higher noise. Nonetheless, folded-cascode op amps are used quite widely, even more than telescopic topologies, because the inputs and outputs can be shorted together and the choice of the input common-mode level is easier. In a telescopic op amp, *three* voltages must be defined carefully: the input CM level and the gate bias voltages of the PMOS and NMOS cascode transistors, whereas in folded-cascode configurations only the latter two are critical.

We now carry out the design of a folded-cascode op amp to reinforce the foregoing concepts.

Example 9.6

Design a folded-cascode op amp with an NMOS input pair (Fig. 9.16) to satisfy the following specifications: $V_{DD} = 3$ V, differential output swing = 3 V, power dissipation = 10 mW, voltage gain = 2000. Use the same device parameters as in Example 9.5.

Solution

As with the telescopic cascode of the previous example, we begin with the power and swing specifications. Allocating 1.5 mA to the input pair, 1.5 mA to the two cascode branches, and the remaining 330 μA to the three current mirrors, we first consider the devices in each cascode branch. Since M_5 and M_6 must each carry 1.5 mA, we allow an overdrive of 500 mV for these transistors so as to keep their width to a reasonable value. To M_3-M_4, we allocate 400 mV and to M_7-M_{10}, 300 mV. Thus, $(W/L)_{5,6} = 400$, $(W/L)_{3,4} = 313$, $(W/L)_{7-10} = 278$. Since the minimum and maximum output levels are equal to 0.6 V and 2.1 V, respectively, the optimum output common-mode level is 1.35 V.

The minimum dimensions of M_1-M_2 are dictated by the minimum input common-mode level, $V_{GS1} + V_{OD11}$. For example, if the input and the output are shorted for part of the operation period (Fig. 9.17), then $V_{GS2} + V_{OD11} = 1.35$ V. With $V_{OD11} = 0.4$ V as an initial guess, we have $V_{GS1} =$

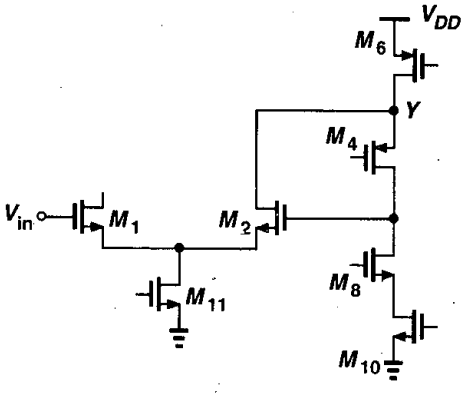


Figure 9.17 Folded-cascode op amp with input and output shorted.

0.95 V, obtaining $V_{OD1,2} = 0.95 - 0.7 = 0.25$ V and hence $(W/L)_{1,2} = 400$. The maximum dimensions of M_1 and M_2 are determined by the tolerable input capacitance and the capacitance at nodes X and Y in Fig. 9.16.

We now calculate the small-signal gain. Using $g_m = 2I_D/(V_{GS} - V_{TH})$, we have $g_{m1,2} = 0.006$ A/V, $g_{m3,4} = 0.0038$ A/V, and $g_{m7,8} = 0.05$ A/V. For $L = 0.5$ μm, $r_{O1,2} = r_{O7-10} = 13.3$ kΩ, and $r_{O3,4} = 2r_{O5,6} = 6.67$ kΩ. It follows that the impedance seen looking into the drain of M_7 (or M_8) is equal to 8.8 MΩ whereas, owing to the limited intrinsic gain of M_3 (or M_4), that seen looking into the drain of M_3 is equal to 66.5 kΩ. The overall gain is therefore limited to about 400.

In order to increase the gain, we first observe that $r_{O5,6}$ is quite lower than $r_{O1,2}$. Thus, the length of M_5-M_6 must be increased. Also, the transconductance of M_1-M_2 is relatively low and can be increased by widening these transistors. Finally, we may decide to double the intrinsic gain of M_3 and M_4 by doubling both their length and width, but at the cost of increasing the capacitance at nodes X and Y. We leave the exact choice of the device dimensions as an exercise for the reader.

An important property of folded-cascode op amps is the capability of handling input common-mode levels close to one of the supply rails. In Fig. 9.16, for example, the CM voltage at the gates of M_1 and M_2 can be equal to V_{DD} because $V_X = V_Y = V_{DD} - 500$ mV. By the same token, a similar topology using a PMOS input pair can accommodate input CM levels as low as zero.

Telescopic and folded-cascode op amps can also be designed to provide a single-ended output. Shown in Fig. 9.18(a) is an example, where a PMOS cascode current mirror converts the differential currents of M_3 and M_4 to a single-ended output voltage. In this implementation, however, $V_X = V_{DD} - |V_{GS5}| - |V_{GS7}|$, limiting the maximum value of V_{out} to $V_{DD} - |V_{GS5}| - |V_{GS7}| + |V_{TH6}|$ and "wasting" one PMOS threshold voltage in the swing (Chapter 5). To resolve this issue, the PMOS load can be modified as shown in Fig. 9.18(b)

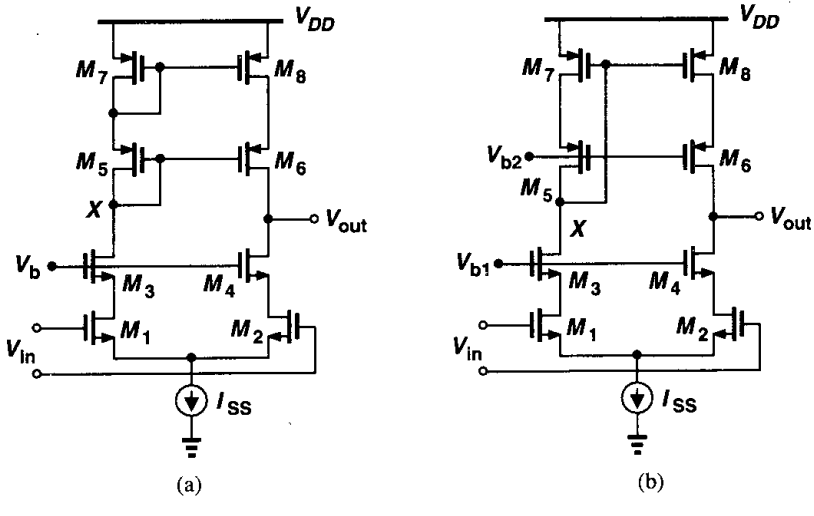


Figure 9.18 Cascode op amps with single-ended output.

so that M_7 and M_8 are biased at the edge of the triode region. Similar ideas apply to folded-cascode op amps as well.

The circuit of Fig. 9.18(a) suffers from two disadvantages with respect to its differential counterpart in Fig. 9.8(b). First, it provides only half the output voltage swing. Second, it contains a mirror pole at node X (Chapter 5), thus limiting the speed of feedback systems employing such an amplifier. It is therefore preferable to use the differential topology, although it requires a feedback loop to define the output common-mode level (Section 9.6).

As a final note, we recognize that to achieve a higher gain, additional cascode devices can be inserted in each branch. Shown in Fig. 9.19 is a "triple cascode," providing a gain on

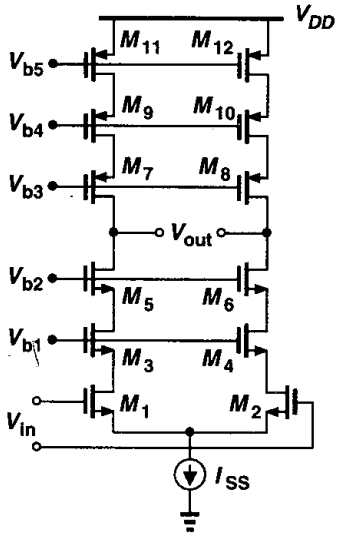


Figure 9.19 Triple-cascode op amp.

the order of $(g_m r_O)^3/2$ but further limiting the output swings. With six overdrive voltages subtracted from V_{DD} in this circuit, it is difficult to operate the amplifier from a supply voltage of 3 V or lower while obtaining reasonable output swings.

9.3 Two-Stage Op Amps

The op amps studied thus far exhibit a "one-stage" nature in that they allow the small-signal current produced by the input pair to flow directly through the output impedance. The gain of these topologies is therefore limited to the product of the input pair transconductance and the output impedance. We have also observed that cascoding in such circuits increases the gain while limiting the output swings.

In some applications, the gain and/or the output swings provided by cascode op amps are not adequate. For example, an op amp used in a hearing aid must operate with supply voltages as low as 0.9 V while delivering single-ended output swings as large as 0.5 V.

In such cases, we resort to "two-stage" op amps, with the first stage providing a high gain and the second, large swings (Fig. 9.20). In contrast to cascode op amps, a two-stage configuration isolates the gain and swing requirements.

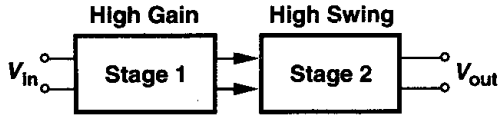


Figure 9.20 Two-stage op amp.

Each stage in Fig. 9.20 can incorporate various amplifier topologies studied in previous sections, but the second stage is typically configured as a simple common-source stage so as to allow maximum output swings. Fig. 9.21 shows an example, where the first and second stages exhibit gains equal to $g_{m1,2}(r_{O1,2} \| r_{O3,4})$ and $g_{m5,6}(r_{O5,6} \| r_{O7,8})$, respectively. The overall gain is therefore comparable with that of a cascode op amp, but the swing at V_{out1} and V_{out2} is equal to $V_{DD} - |V_{OD5,6}| - V_{OD7,8}$.

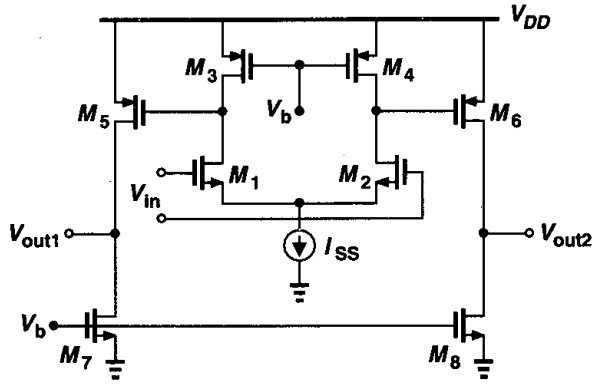


Figure 9.21 Simple implementation of a two-stage op amp.

To obtain a higher gain, the first stage can incorporate cascode devices, as depicted in Fig. 9.22. With a gain of, say, 10 in the output stage, the voltage swings at X and Y are quite small, allowing optimization of M_1-M_8 for higher gain. The overall voltage gain can be expressed as

$$A_v \approx \{g_{m1,2}[(g_{m3,4} + g_{mb3,4})r_{O3,4}r_{O1,2}] \| [(g_{m5,6} + g_{mb5,6})r_{O5,6}r_{O7,8}]\} \times [g_{m9,10}(r_{O9,10} \| r_{O11,12})]. \quad (9.13)$$

A two-stage op amp may provide a single-ended output. One method is to convert the differential currents of the two output stages to a single-ended voltage. Illustrated in Fig. 9.23, this approach maintains the differential nature of the first stage, using only the current mirror M_7-M_8 to generate a single-ended output. Note, however, that if the gate of M_1 is shorted to V_{out2} to form a unity-gain buffer, then the minimum allowable output level is equal to $V_{GS1} + V_{ISS}$, severely limiting the output swing.

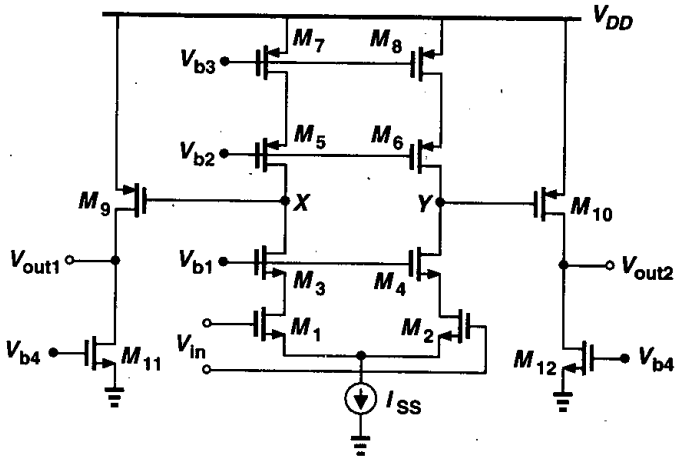


Figure 9.22 Two-stage op amp employing cascoding.

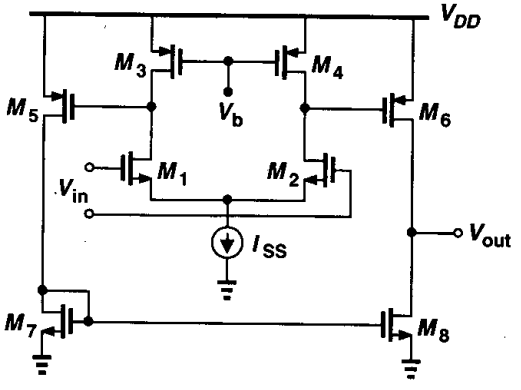


Figure 9.23 Two-stage op amp with single-ended output.

Can we cascade more than two stages to achieve a higher gain? As explained in Chapter 10, each gain stage introduces at least one pole in the open-loop transfer function, making it difficult to guarantee stability in a feedback system using such an op amp. For this reason, op amps having more than two stages are rarely used. Exceptions are described in [1, 2, 3].

9.4 Gain Boosting

The limited gain of one-stage op amps studied in Section 9.2 and the difficulties in using two-stage op amps at high speeds have motivated extensive work on new topologies. Recall that in one-stage op amps such as telescopic and folded-cascode topologies the objective is to maximize the output impedance so as to attain a high voltage gain. The idea behind gain boosting is to further increase the output impedance without adding more cascode devices.

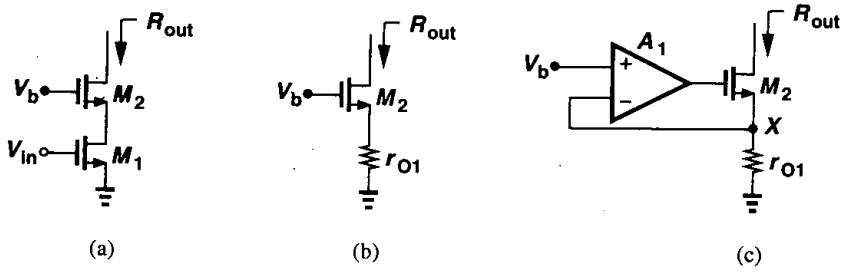


Figure 9.24 Increasing the output impedance by feedback.

Consider the simple cascode in Fig. 9.24(a), whose output impedance is given by $R_{out} = g_{m2}r_{O2}r_{O1}$. As far as R_{out} is concerned, M_1 operates as a degeneration resistor [Fig. 9.24(b)], sensing the output current and converting it to a voltage. The observation that the small-signal voltage produced across r_{O1} is proportional to the output current suggests that this voltage can be subtracted from V_b so as to place M_2 in current-voltage feedback, thereby increasing the output impedance. Illustrated in Fig. 9.24(c), the idea is to drive the gate of M_2 by an amplifier that forces V_X to be equal to V_b. Thus, voltage variations at the drain of M_2 now affect V_X to a lesser extent because A_1 "regulates" this voltage. With smaller variations at X, the current through r_{O1} and hence the output current remain more constant than those in Fig. 9.24(b), yielding a higher output impedance. The reader can prove that

$$R_{out} \approx A_1 g_{m2} r_{O2} r_{O1}, \tag{9.14}$$

concluding that R_{out} can be "boosted" substantially without stacking more cascode devices on top of M_2.

Since for small-signal operation, V_b is set to zero, the circuit can be simplified as shown in Fig. 9.25(a), with the amplifier possibly implemented as in Fig. 9.25(b). Called a "regulated cascode," the overall stage is illustrated in Fig. 9.25(c), exhibiting a gain equal to $|A_v| \approx g_{m1}(g_{m2}r_{O2}r_{O1})(g_{m3}r_{O3})$, similar to the gain of a *triple* cascode. This topology was first invented in 1976 [4] and applied to boost the gain of op amps in 1989 [5, 6].

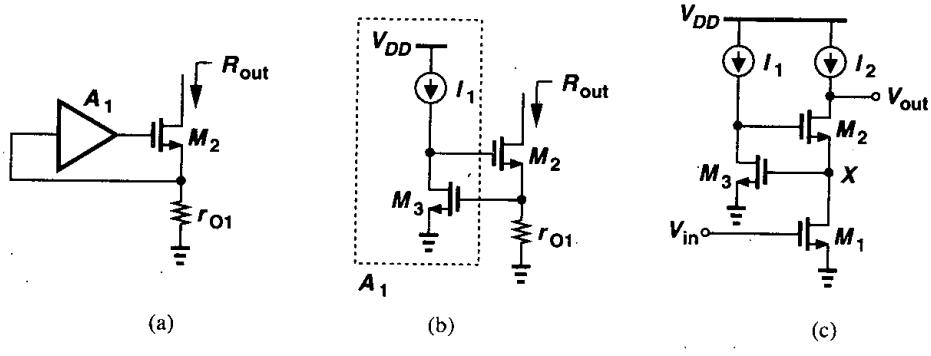


Figure 9.25 Gain boosting in cascode stage.

Before incorporating the technique of Fig. 9.25(c) in an op amp, let us examine the output voltage swings, in particular, the minimum allowable level. Since $V_X = V_{GS3}$, the minimum value of V_{out} is $V_{OD2} + V_{GS3}$, whereas, in a simple cascode with proper choice of V_{G2}, it would be $V_{OD2} + V_{OD1}$. Thus, the auxiliary amplifier in this case limits the output swing.

We now apply gain boosting to a differential cascode stage, as shown in Fig. 9.26(a). Since the signals at nodes X and Y are differential, we surmise that the two single-ended gain boosting amplifiers A_1 and A_2 can be replaced by one differential amplifier [Fig. 9.26(b)]. Following the topology of Fig. 9.25(c), we implement the differential auxiliary amplifier as shown in Fig. 9.26(c), but noting that the minimum level at the drain of M_3 is equal to $V_{OD3} + V_{GS5} + V_{ISS2}$, where V_{ISS2} denotes the voltage required across I_{SS2}. In a simple differential cascode, on the other hand, the minimum would be approximately one threshold voltage lower.

The voltage swing limitation in Fig. 9.26(c) results from the fact that the gain-boosting amplifier incorporates an NMOS differential pair. If nodes X and Y are sensed by a PMOS pair, the minimum value of V_X and V_Y is not dictated by the gain-boosting amplifier. Now

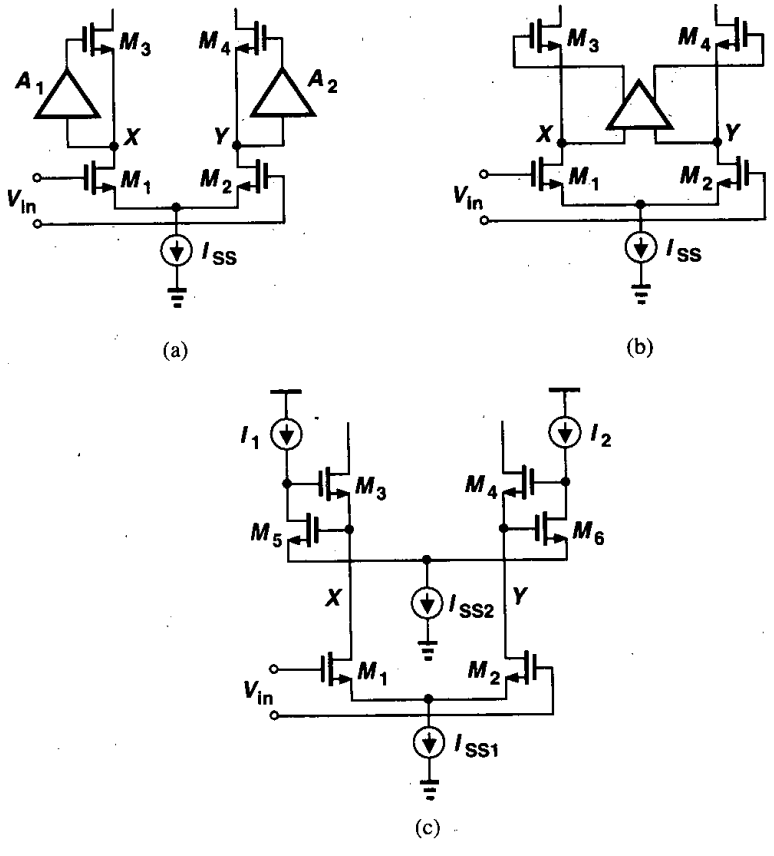


Figure 9.26 Boosting the output impedance of a differential cascode stage.

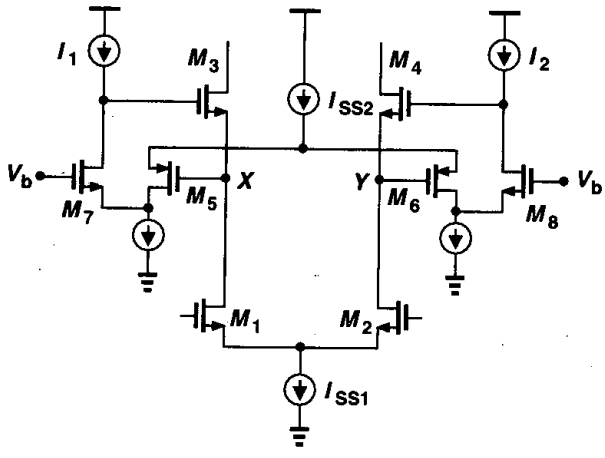


Figure 9.27 Folded-cascode circuit used as auxiliary amplifier.

recall from Section 9.2 that the minimum input CM level of a folded-cascode stage using a PMOS input pair can be zero. Thus, we employ such a topology for the gain-boosting amplifier, arriving at the circuit shown in Fig. 9.27. Here, the minimum allowable level of V_X and V_Y is given by $V_{OD1,2} + V_{ISS1}$.

Example 9.7

Calculate the output impedance of the circuit shown in Fig. 9.27.

Solution

Using the half-circuit concept and replacing the ideal current sources with transistors, we obtain the equivalent depicted in Fig. 9.28. The voltage gain from X to P is approximately equal to

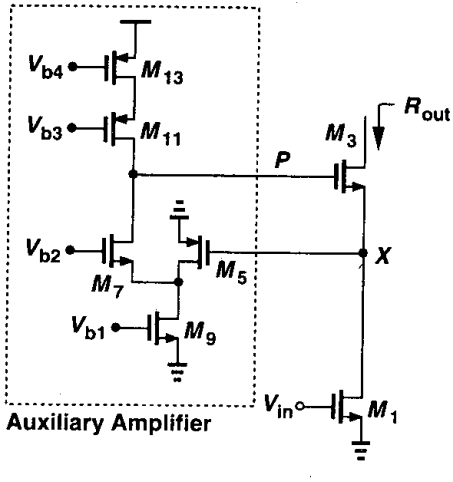


Figure 9.28

$g_{m5}R_{out1}$, where $R_{out1} \approx [g_{m7}r_{O7}(r_{O9}\|r_{O5})]\|(g_{m11}r_{O11}r_{O13})$. Thus, $R_{out} \approx g_{m3}r_{O3}r_{O1}g_{m5}R_{out1}$. In essence, since the output impedance of a cascode is boosted by a folded-cascode stage, the overall output impedance is similar to that of a "quadruple" cascode.

Regulated cascodes can also be utilized in the load current sources of a cascode op amp. Shown in Fig. 9.29(a), such a topology boosts the output impedance of the PMOS current sources as well, thereby achieving a very high voltage gain. To allow maximum swings at the output, amplifier A_2 must employ an NMOS input differential pair. Similar ideas apply to folded-cascode op amps [Fig. 9.29(b)].

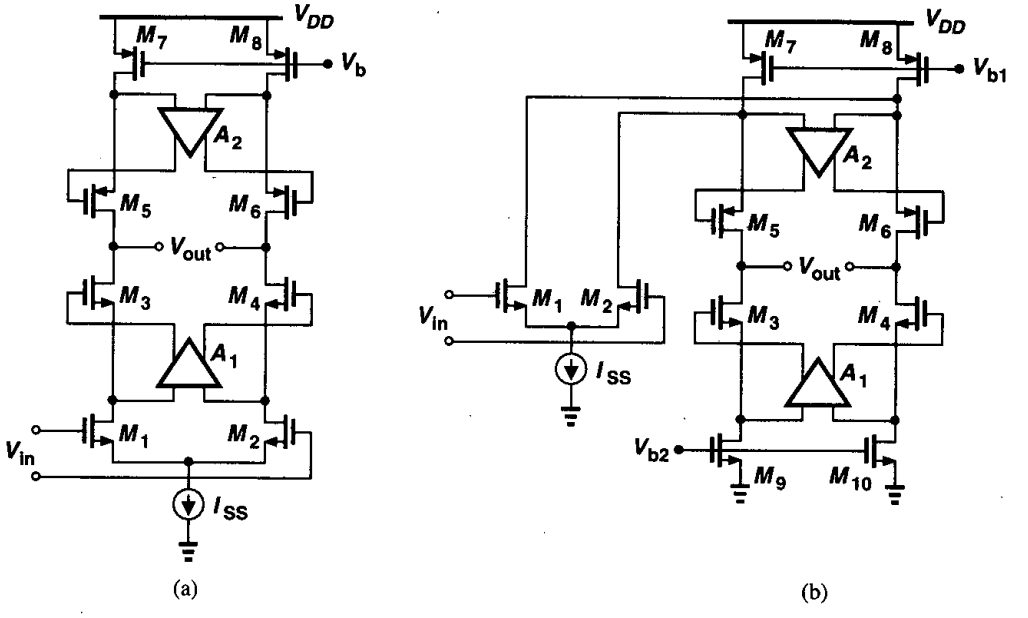


Figure 9.29 Gain boosting applied to both signal path and load devices.

Now recall that the premise behind gain boosting is to increase the gain without adding a second stage or more cascode devices. Does this mean that the op amps of Fig. 9.29 have a one-stage nature? After all, the gain-boosting amplifier introduces its own poles. In contrast to two-stage op amps, where the entire signal experiences the poles associated with each stage, in a gain-boosted op amp, most of the signal directly flows through the cascode devices to the output. Only a small "error" component is processed by the gain-boosting amplifier and "slowed down."

9.5 Comparison

Our study of op amps in this chapter has introduced four principal topologies: telescopic cascode, folded cascode, two-stage op amp, and gain boosting. It is instructive to compare

Table 9.1 Comparison of performance of various op amp topologies.

	Gain	Output Swing	Speed	Power Dissipation	Noise
Telescopic	Medium	Medium	Highest	Low	Low
Folded-Cascode	Medium	Medium	High	Medium	Medium
Two-Stage	High	Highest	Low	Medium	Low
Gain-Boosted	High	Medium	Medium	High	Medium

various performance aspects of these circuits to gain a better view of their applicability. Table 9.1 comparatively presents important attributes of each op amp topology. We study the speed differences in Chapter 10.

9.6 Common-Mode Feedback

In this and previous chapters, we have described many advantages of fully differential circuits over their single-ended counterparts. In addition to greater output swings, differential op amps avoid mirror poles, thus achieving a higher closed-loop speed. However, high-gain differential circuits require "common-mode feedback" (CMFB).

To understand the need for CMFB, let us begin with a simple realization of a differential amplifier [Fig. 9.30(a)]. In some applications, we short the inputs and outputs for part of the operation [Fig. 9.30(b)], providing *differential* negative feedback. The input and output common-mode levels in this case are quite well-defined, equal to $V_{DD} - I_{SS}R_D/2$.

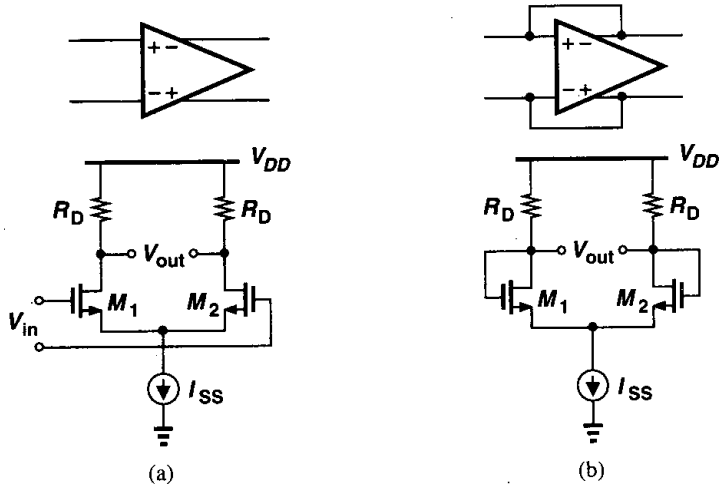


Figure 9.30 (a) Simple differential pair, (b) circuit with inputs shorted to outputs.

Now suppose the load resistors are replaced by PMOS current sources so as to increase the differential voltage gain [Fig. 9.31(a)]. What is the common-mode level at nodes X and Y? Since each of the input transistors carries a current of $I_{SS}/2$, the CM level depends on how close I_{D3} and I_{D4} are to this value. In practice, as exemplified by Fig. 9.31(b), mismatches in the PMOS and NMOS current mirrors defining I_{SS} and $I_{D3,4}$ create a finite error between $I_{D3,4}$ and $I_{SS}/2$. Suppose, for example, that the drain currents of M_3 and M_4 in the saturation region are slightly greater than $I_{SS}/2$. As a result, to satisfy Kirchoff current law at nodes X and Y, both M_3 and M_4 must enter the triode region so that their drain currents fall to $I_{SS}/2$. Conversely, if $I_{D3,4} < I_{SS}/2$, then both V_X and V_Y must drop so that M_5 enters the triode region, thereby producing only $2I_{D3,4}$.

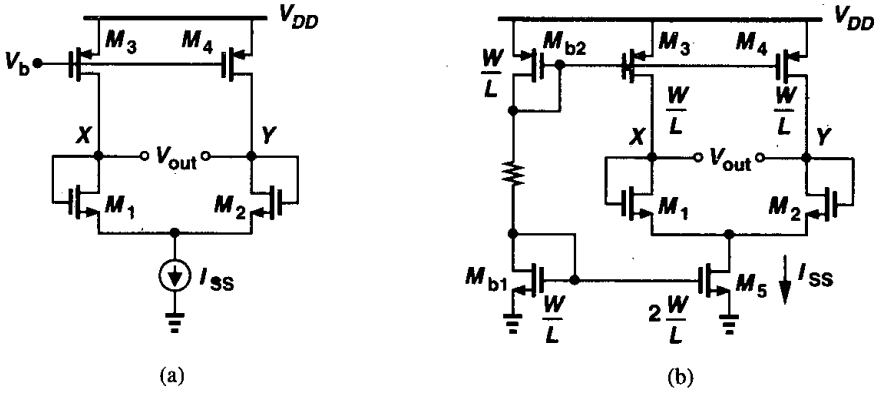


Figure 9.31 (a) High-gain differential pair with inputs shorted to outputs, (b) effect of current mismatches.

The above difficulties fundamentally arise because in high-gain amplifiers, we wish a p-type current source to balance an n-type current source. As illustrated in Fig. 9.32, the difference between I_P and I_N must flow through the intrinsic output impedance of the amplifier, creating an output voltage change of $(I_P - I_N)(R_P \| R_N)$. Since the current error depends on mismatches and $R_P \| R_N$ is quite high, the voltage error may be large, thus driving the p-type or n-type current source into the triode region. As a general rule, if the output CM level cannot be determined by "visual inspection" and requires calculations based on device properties, then it is poorly defined. This is the case in Fig. 9.31 but not in Fig. 9.30. We emphasize that differential feedback cannot define the CM level.

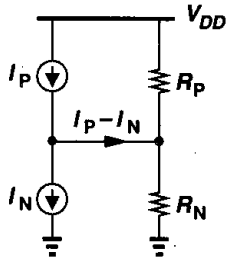


Figure 9.32 Simplified model of high-gain amplifier.

Example 9.8

Consider the telescopic op amp designed in Example 9.5 and repeated in Fig. 9.33 with bias current mirrors. Suppose M_9 suffers from a 1% current mismatch with respect to M_{10}, producing $I_{SS} =$

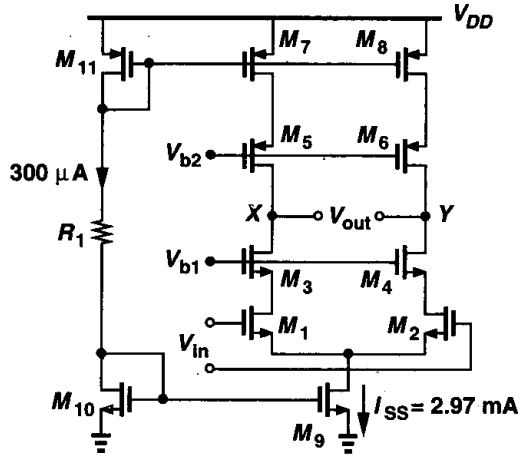


Figure 9.33

2.97 mA rather than 3 mA. Assuming perfect matching for other transistors, explain what happens in the circuit.

Solution

From Example 9.5, the single-ended output impedance of the circuit equals 266 kΩ. Since the difference between the drain currents of M_3 and M_5 (and M_4 and M_6) is 30 μA/2 = 15 μA, the output voltage error would be 266 kΩ × 15 μA= 3.99 V. Since this large error cannot be produced, V_X and V_Y must rise so much that M_5-M_6 and M_7-M_8 enter the triode region, yielding $I_{D7,8} = 1.485$ mA. We should also mention that another important source of CM error in the simple biasing scheme of Fig. 9.33 is the *deterministic* error between $I_{D7,8}$ and I_{11} (and also between I_{D9} and I_{D10}) due to their different drain-source voltages. This error can nonetheless be reduced by means of the current mirror techniques of Chapter 5.

The foregoing study implies that in high-gain amplifiers, the output CM level is quite sensitive to device properties and mismatches and it cannot be stabilized by means of *differential* feedback. Thus, a common-mode feedback network must be added to sense the CM level of the two outputs and accordingly adjust one of the bias currents in the amplifier. Following our view of feedback systems in Chapter 8, we divide the task of CMFB into three operations: sensing the output CM level, comparison with a reference, and returning the error to the amplifier's bias network. Fig. 9.34 conceptually illustrates the idea.

In order to sense the output CM level, we recall that $V_{out,CM} = (V_{out1} + V_{out2})/2$, where V_{out1} and V_{out2} are the single-ended outputs. It therefore seems plausible to employ a resistive divider as shown in Fig. 9.35, generating $V_{out,CM} = (R_1 V_{out2} + R_2 V_{out1})/(R_1 + R_2)$, which reduces to $(V_{out1} + V_{out2})/2$ if $R_1 = R_2$. The difficulty, however, is that R_1 and R_2 must be much greater than the output impedance of the op amp so as to avoid lowering the open-loop gain. For example, in the design of Fig. 9.33, the output impedance equals 266 kΩ,

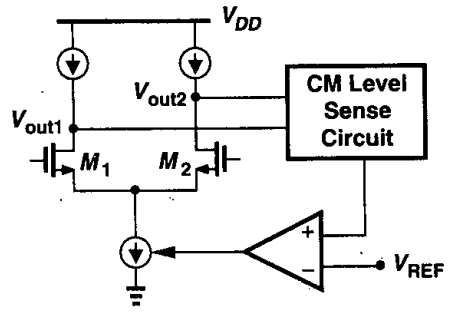


Figure 9.34 Conceptual topology for common-mode feedback.

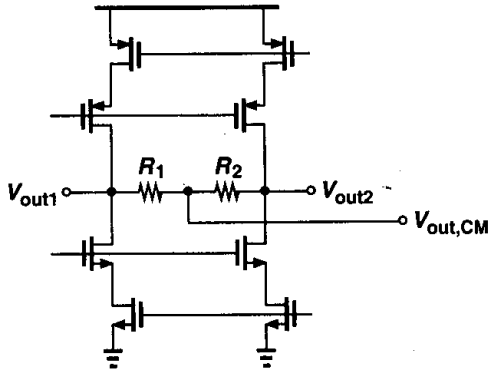


Figure 9.35 Common-mode feedback with resistive sensing.

necessitating a value of several megaohms for R_1 and R_2. As explained in Chapter 17, such large resistors occupy a very large area and, more importantly, suffer from substantial parasitic capacitance to the substrate.

To eliminate the resistive loading, we can interpose source followers between each output and its corresponding resistor. Illustrated in Fig. 9.36, this technique produces a CM level that is in fact lower than the output CM level by $V_{GS7,8}$, but this shift can be taken into account in the comparison operation. Note that R_1 and R_2 or I_1 and I_2 must be large enough to ensure that M_7 or M_8 is not "starved" when a large differential swing appears at the output. As conceptually depicted in Fig. 9.37, if, say, V_{out2} is quite higher than V_{out1}, then I_1 must sink both $I_X \approx (V_{out2} - V_{out1})/(R_1 + R_2)$ and I_{D7}. Consequently, if $R_1 + R_2$ or I_1 is not sufficiently large, I_{D7} drops to zero and $V_{out,CM}$ no longer represents the true output CM level.

The sensing method of Fig. 9.36 nevertheless suffers from an important drawback: it limits the differential output swings (even if $R_{1,2}$ and $I_{1,2}$ are large enough.) To understand why, let us determine the minimum allowable level of V_{out1} (and V_{out2}), noting that without CMFB it would be equal to $V_{OD3} + V_{OD5}$. With the source followers in place, $V_{out1,min} = V_{GS7} + V_{I1}$, where V_{I1} denotes the minimum voltage required across I_1. This is roughly equal to two overdrive voltages plus one threshold voltage. Thus, the swing at each output is reduced by approximately V_{TH}, a significant value in low-voltage design.

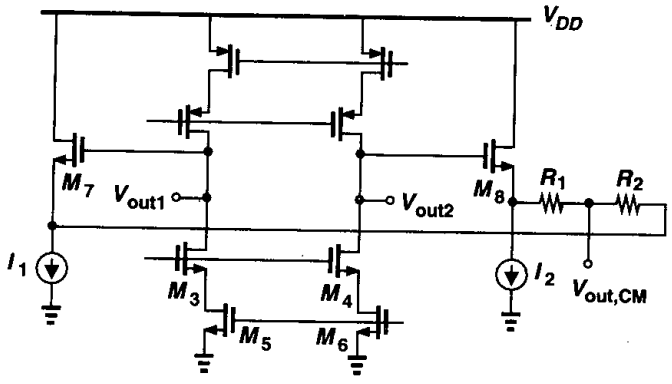


Figure 9.36 Common-mode feedback using source followers.

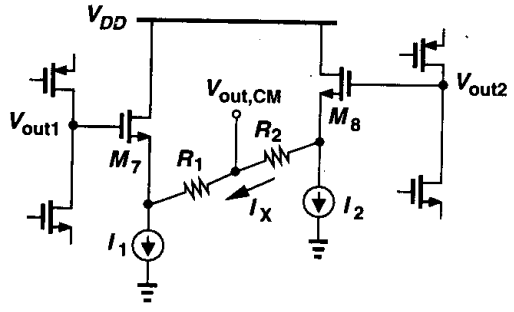


Figure 9.37 Current starvation of source followers for large swings.

Looking at Fig. 9.35, the reader may wonder if the output CM level can be sensed by means of *capacitors*, rather than resistors, so as to avoid degrading the low-frequency open-loop gain of the op amp. This is indeed possible in some cases and will be studied in Chapter 12.

Another type of CM sensing is depicted in Fig. 9.38. Here, identical transistors M_7 and M_8 operate in deep triode region, introducing a total resistance between P and ground equal to

$$R_{tot} = R_{on7} \| R_{on8} \tag{9.15}$$

$$= \frac{1}{\mu_n C_{ox} \frac{W}{L}(V_{out1} - V_{TH})} \Bigg\| \frac{1}{\mu_n C_{ox} \frac{W}{L}(V_{out2} - V_{TH})} \tag{9.16}$$

$$= \frac{1}{\mu_n C_{ox} \frac{W}{L}(V_{out2} + V_{out1} - 2V_{TH})}, \tag{9.17}$$

where W/L denotes the aspect ratio of M_7 and M_8. Equation (9.17) indicates that R_{tot} is a function of $V_{out2} + V_{out1}$ but independent of $V_{out2} - V_{out1}$. From Fig. 9.38, we observe that if the outputs rise together, then R_{tot} drops, whereas if they change differentially, one R_{on} increases and the other decreases.

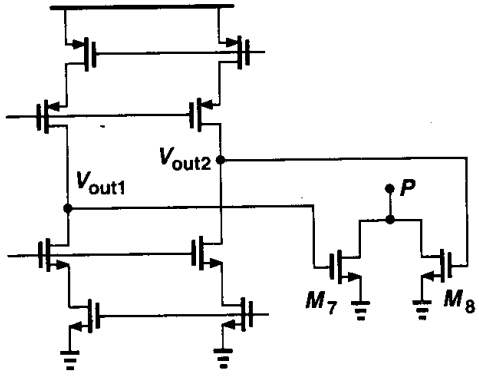


Figure 9.38 Common-mode sensing using MOSFETs operating in deep triode region.

In the circuit of Fig. 9.38, the use of M_7 and M_8 limits the output voltage swings. Here, it may seem that $V_{out,min} = V_{TH7,8}$, which is relatively close to two overdrive voltages, but the difficulty arises from the assumption above that both M_7 and M_8 operate in deep triode region. In fact, if, say, V_{out1} drops from the equilibrium CM level to one threshold voltage above ground and V_{out2} rises by the same amount, then M_7 enters the saturation region, thus exhibiting a variation in its on-resistance that is not counterbalanced by that of M_8.

We now study techniques of comparing the measured CM level with a reference and returning the error to the op amp's bias network. In the circuit of Fig. 9.39, we employ a simple amplifier to detect the difference between $V_{out,CM}$ and a reference voltage, V_{REF}, applying the result to the NMOS current sources with negative feedback. If both V_{out1} and V_{out2} rise, so does V_E, thereby increasing the drain currents of M_3-M_4 and lowering the output CM level. In other words, if the loop gain is large, the feedback network forces the CM level of V_{out1} and V_{out2} to approach V_{REF}. Note that the feedback can be applied to the PMOS current sources as well. Also, the feedback may control only a fraction of the current to allow optimization of the settling behavior. For example, each of M_3 and M_4

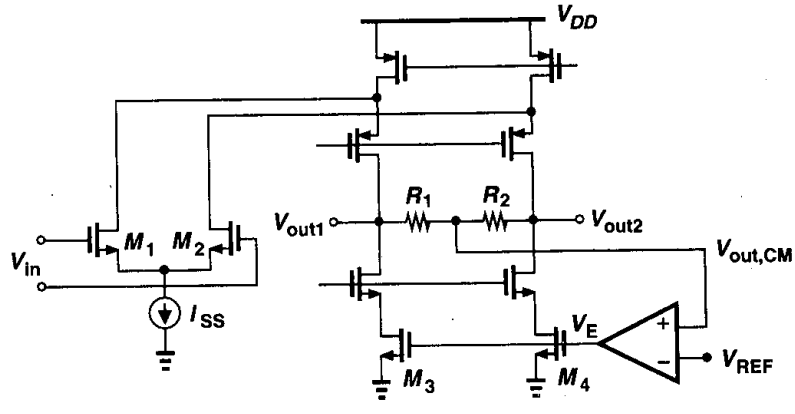


Figure 9.39 Sensing and controlling output CM level.

can be decomposed into two parallel devices, one biased at a constant current and the other driven by the error amplifier.

In a folded-cascode op amp, the CM feedback may control the tail current of the input differential pair. Illustrated in Fig. 9.40, this method increases the tail current if V_{out1} and V_{out2} rise, lowering the drain currents of M_5-M_6 and restoring the output CM level.

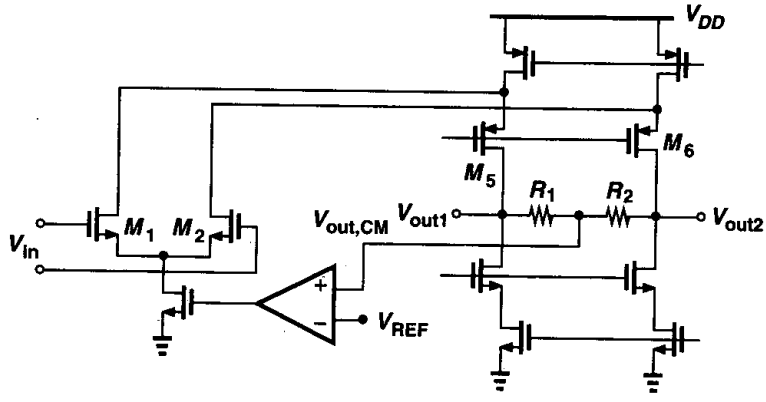


Figure 9.40 Alternative method of controlling output CM level.

How do we perform comparison and feedback with the sensing scheme of Fig. 9.38? Here, the output CM voltage is directly converted to a resistance or a current, prohibiting comparison with a reference voltage. A simple feedback topology utilizing this technique is depicted in Fig. 9.41, where $R_{on7} \| R_{on8}$ adjusts the bias current of M_5 and M_6. The output CM level sets $R_{on7} \| R_{on8}$ such that I_{D5} and I_{D6} exactly balance I_{D9} and I_{D10}, respectively. Assuming $I_{D9} = I_{D10} = I_D$, we must have $V_b - V_{GS5} = 2I_D(R_{on7} \| R_{on8})$ and hence

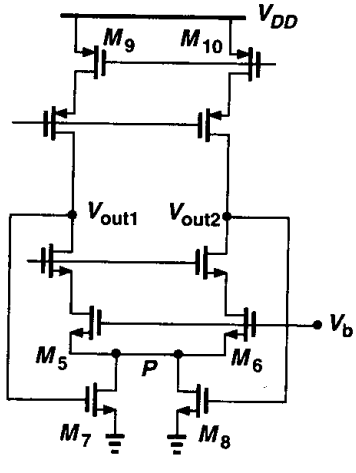


Figure 9.41 CMFB using triode devices.

$R_{on7} \| R_{on8} = (V_b - V_{GS5})/(2I_D)$. From (9.17),

$$\frac{1}{\mu_n C_{ox}\left(\frac{W}{L}\right)_{7,8}(V_{out2} + V_{out1} - 2V_{TH})} = \frac{V_b - V_{GS5}}{2I_D}, \tag{9.18}$$

that is,

$$V_{out1} + V_{out2} = \frac{2I_D}{\mu_n C_{ox}\left(\frac{W}{L}\right)_{7,8}} \frac{1}{V_b - V_{GS5}} + 2V_{TH}. \tag{9.19}$$

The CM level can thus be obtained by noting that $V_{GS5} = \sqrt{2I_D/[\mu_n C_{ox}(W/L)_5]} + V_{TH5}$.

The CMFB network of Fig. 9.41 suffers from several drawbacks. First, the value of the output CM level is a function of device parameters. Second, the voltage drop across $R_{on7} \| R_{on8}$ limits the output voltage swings. Third, to minimize this drop, M_7 and M_8 are usually quite wide devices, introducing substantial capacitance at the output. The second issue can be alleviated by applying the feedback to the tail current of the input differential pair (Fig. 9.42), but the other two remain.

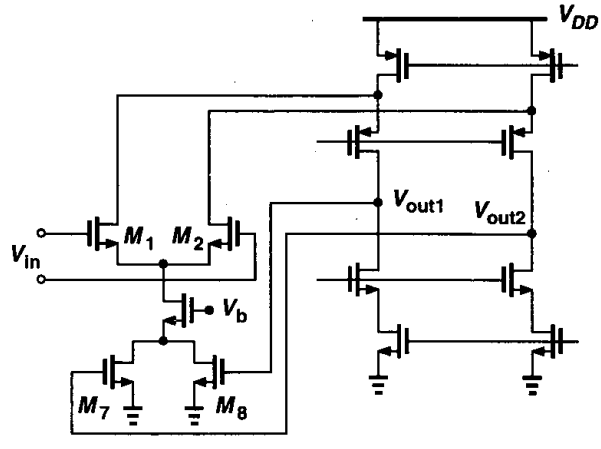


Figure 9.42 Alternative method of controlling output CM level.

How is V_b generated in Fig. 9.42? We note that $V_{out,CM}$ is somewhat sensitive to the value of V_b: if V_b is higher than expected, the tail current of M_1 and M_2 increases and the output CM level falls. Since the feedback through M_7 and M_8 attempts to correct this error, the overall change in $V_{out,CM}$ depends on the loop gain in the CMFB network. This is studied in the following example.

Example 9.9

For the circuit of Fig. 9.42, determine the sensitivity of $V_{out,CM}$ to V_b, i.e., $dV_{out,CM}/dV_b$.

Solution

Setting V_{in} to zero and following the procedure depicted in Fig. 4.25, we simplify the circuit as shown in Fig. 9.43. Note that g_{m7} and g_{m8} must be calculated in the triode region: $g_{m7} = g_{m8} = \mu_n C_{ox}(W/L)_{7,8} V_{DS7,8}$, where $V_{DS7,8}$ denotes the bias value of the drain-source voltage of M_7 and M_8. Since M_7 and M_8 operate in deep triode region, $V_{DS7,8}$ typically does not exceed a few hundred millivolts.

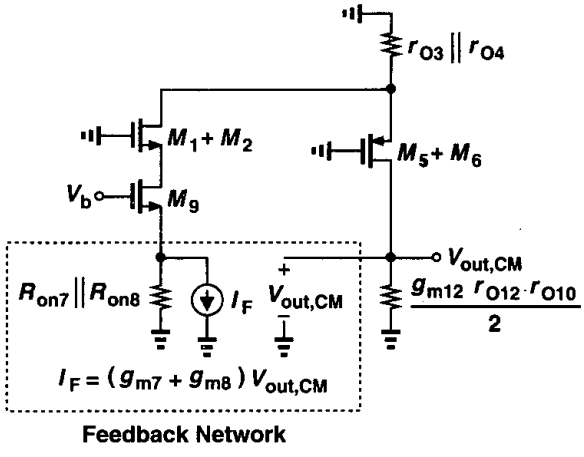


Figure 9.43

In a well-designed circuit, the loop gain must be relatively high. We therefore surmise that the closed-loop gain is approximately equal to $1/\beta$, where β represents the feedback factor. We write from Chapter 8:

$$\beta = \left.\frac{V_2}{V_1}\right|_{I2=0} \tag{9.20}$$

$$= -(g_{m7} + g_{m8})(R_{on7} \| R_{on8}) \tag{9.21}$$

$$= -2\mu_n C_{ox} \left(\frac{W}{L}\right)_{7,8} V_{DS7,8} \cdot \frac{1}{2\mu_n C_{ox}(W/L)_{7,8}(V_{GS7,8} - V_{TH7,8})} \tag{9.22}$$

$$= -\frac{V_{DS7,8}}{V_{GS7,8} - V_{TH7,8}}, \tag{9.23}$$

where $V_{GS7,8} - V_{TH7,8}$ denotes the overdrive voltage of M_7 and M_8. Thus,

$$\left|\frac{dV_{out,CM}}{dV_b}\right|_{closed} \approx \frac{V_{GS7,8} - V_{TH7,8}}{V_{DS7,8}}. \tag{9.24}$$

This is an important result. Since $V_{GS7,8}$ (i.e., the output CM level) is typically in the vicinity of $V_{DD}/2$, the above equation suggests that $V_{DS7,8}$ must be maximized.

We now introduce a modification to the circuit of 9.42 that both makes the output level relatively independent of device parameters and lowers the sensitivity to the value of V_b. Illustrated in Fig. 9.44(a), the idea is to define V_b by a current mirror arrangement such that I_{D9} "tracks" I_1 and V_{REF}. For simplicity, suppose $(W/L)_{15} = (W/L)_9$ and

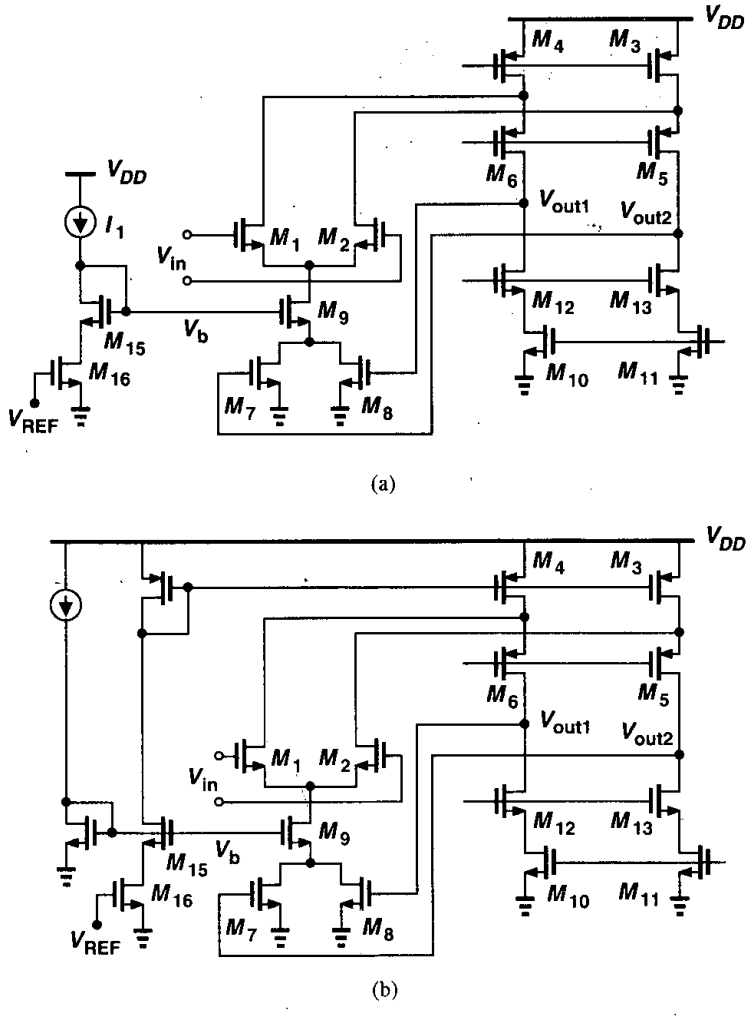


Figure 9.44 Modification of CMFB for more accurate definition of output MC level.

$(W/L)_{16} = (W/L)_7 + (W/L)_8$. Thus, $I_{D9} = I_1$ only if $V_{out,CM} = V_{REF}$. In other words, as with Fig. 9.40, the circuit produces an output CM level equal to a reference but it requires no resistors in sensing $V_{out,CM}$. The overall design can be simplified as shown in Fig. 9.44(b).

In practice, since $V_{DS15} \neq V_{DS9}$, channel-length modulation results in a finite error. Figure 9.45 depicts a modification that suppresses this error. Here, transistors M_{17} and M_{18}

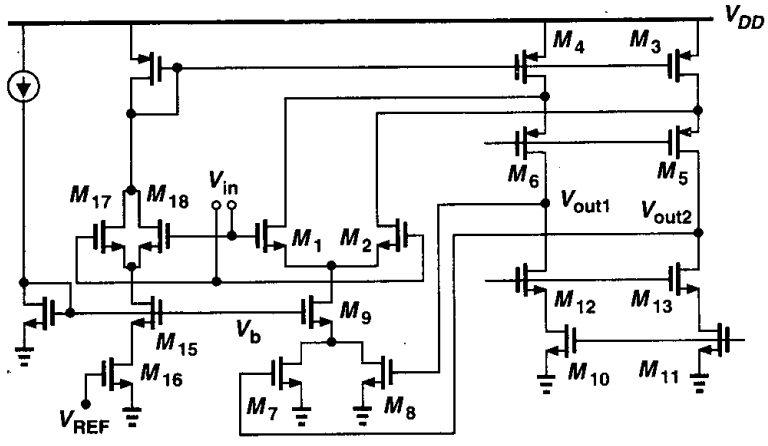


Figure 9.45 Modification to suppress error due to channel-length modulation.

reproduce at the drain of M_{15} a voltage equal to the source voltage of M_1 and M_2, ensuring that $V_{DS15} = V_{DS9}$.

To arrive at another CM feedback topology, let us consider the simple differential pair shown in Fig. 9.46(a). Here, the output CM level, $V_{DD} - V_{GS3,4}$, is relatively well-defined, but the voltage gain is quite low. To increase the differential gain, the PMOS devices must operate as current sources for *differential* signals. We therefore modify the circuit as depicted in Fig. 9.46(b), where for differential changes at V_{out1} and V_{out2}, node P is a virtual ground and the gain can be expressed as $g_{m1,2}(r_{O1,2} \| r_{O3,4} \| R_F)$. For common-mode levels, on the other hand, M_3 and M_4 operate as diode-connected devices. The circuit proves useful in low-gain applications.

It is important to note that fully-differential two-stage op amps such as that in Fig. 9.22 require *two* CMFB networks, one for the output of each stage. An example is described in [10].

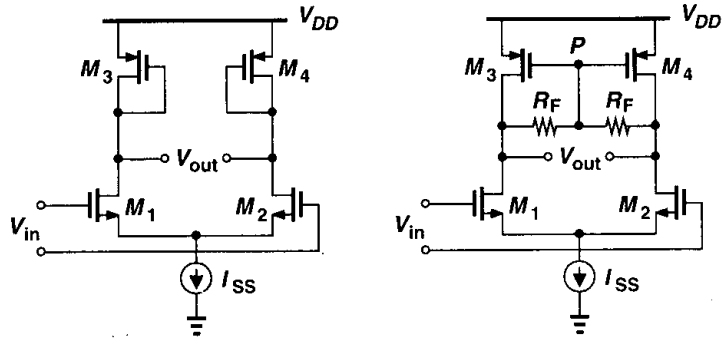


Figure 9.46 (a) Differential pair using diode-connected loads, (b) resistive CMFB.

9.7 Input Range Limitations

The op amp circuits studied thus far have evolved to achieve large differential output swings. While the differential input swings are usually much smaller (by a factor equal to the open-loop gain), the input *common-mode* level may need to vary over a wide range in some applications. For example, consider the simple unity-gain buffer shown in Fig. 9.47, where the input swing is nearly equal to the output swing. Interestingly, in this case the voltage swings are limited by the input differential pair rather than the output cascode branch. Specifically, $V_{in,min} \approx V_{out,min} = V_{GS1,2} + V_{ISS}$, approximately one threshold voltage higher than the allowable minimum provided by M_5-M_8.

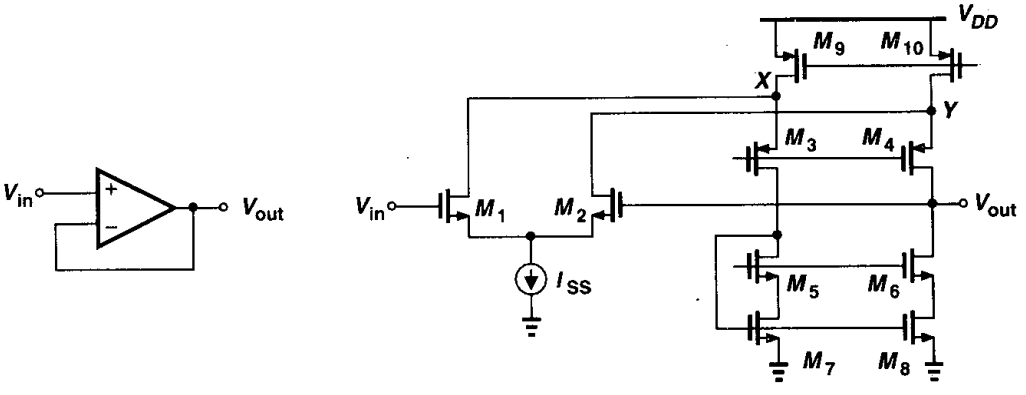


Figure 9.47 Unity-gain buffer.

What happens if V_{in} falls below the minimum given above? The MOS transistor operating as I_{SS} enters the triode region, decreasing the bias current of the differential pair and hence lowering the transconductance. We then postulate that the limitation is overcome if the transconductance can somehow be restored.

A simple approach to extending the input CM range is to incorporate both NMOS and PMOS differential pairs such that when one is "dead," the other is "alive." Illustrated in Fig. 9.48, the idea is to combine two folded-cascode op amps with NMOS and PMOS input differential pairs. Here, as the input CM level approaches the ground potential, the NMOS pair's transconductance drops, eventually falling to zero. Nonetheless, the PMOS pair remains active, allowing normal operation. Conversely, if the input CM level approaches V_{DD}, M_{1P} and M_{2P} begin to turn off but M_1 and M_2 function properly.

An important concern in the circuit of Fig. 9.48 is the *variation* of the overall transconductance of the two pairs as the input CM level changes. Considering the operation of each pair, we anticipate the behavior depicted in Fig. 9.49. Thus, many properties of the circuit, including gain, speed, and noise, vary. More sophisticated techniques of minimizing this variation are described in [7].

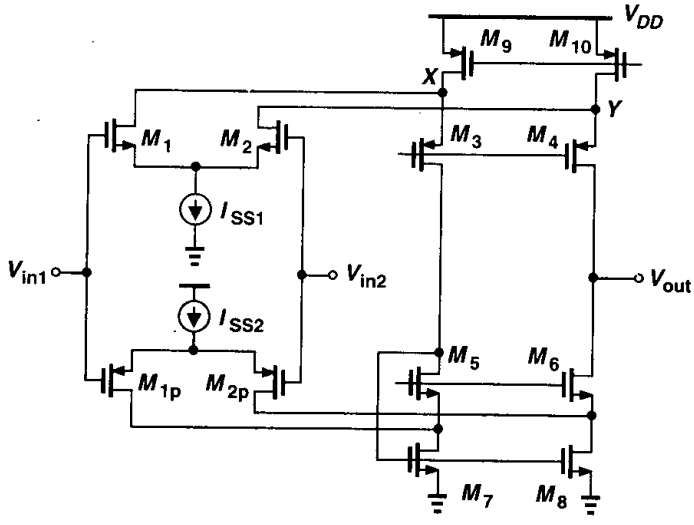


Figure 9.48 Extension of input CM range.

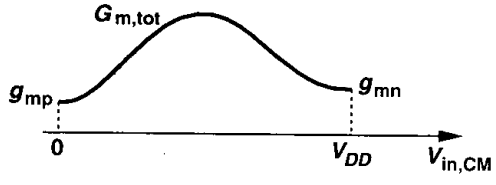


Figure 9.49 Variation of equivalent transconductance with the input CM level.

9.8 Slew Rate

Op amps used in feedback circuits exhibit a large-signal behavior called "slewing." We first describe an interesting property of *linear* systems that vanishes during slewing. Consider the simple RC network shown in Fig. 9.50, where the input is an ideal voltage step of height V_0. Since $V_{out} = V_0[1 - \exp(-t/\tau)]$, where $\tau = RC$, we have

$$\frac{dV_{out}}{dt} = \frac{V_0}{\tau} \exp \frac{-t}{\tau}. \tag{9.25}$$

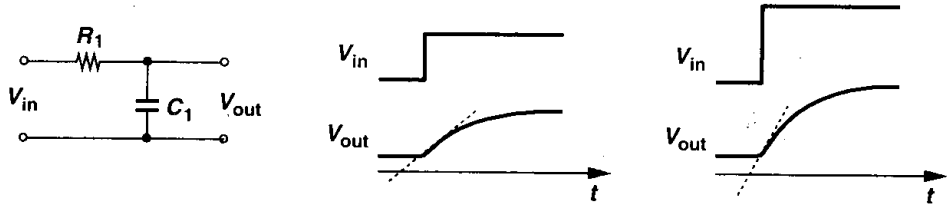


Figure 9.50 Response of a linear circuit to input step.

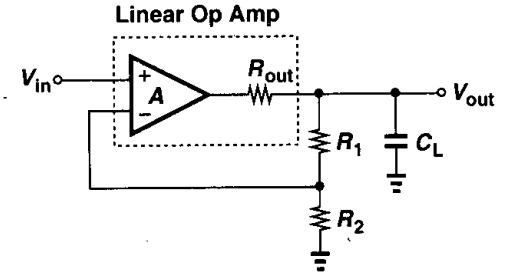


Figure 9.51 Response of linear op amp to step response.

That is, the slope of the step response is proportional to the final value of the output; if we apply a larger input step, the output rises more rapidly. This is a fundamental property of linear systems: if the input amplitude is, say, doubled while other parameters remain constant, the output signal level must double at *every* point, leading to a twofold increase in the slope.

The foregoing observation applies to linear feedback systems as well. Shown in Fig. 9.51 is an example, where the op amp is assumed linear. Here, we can write

$$\left[\left(V_{in} - V_{out}\frac{R_2}{R_1+R_2}\right)A - V_{out}\right]\frac{1}{R_{out}} = \frac{V_{out}}{R_1+R_2} + V_{out}C_L s. \tag{9.26}$$

Assuming $R_1 + R_2 \gg R_{out}$, we have

$$\frac{V_{out}}{V_{in}}(s) \approx \frac{A}{\left(1+A\dfrac{R_2}{R_1+R_2}\right)\left[1+\dfrac{R_{out}C_L}{1+AR_2/(R_1+R_2)}s\right]}. \tag{9.27}$$

As expected, both the low-frequency gain and the time constant are divided by $1 + AR_2/(R_1 + R_2)$. The step response is therefore given by

$$V_{out} = V_0\frac{A}{1+A\dfrac{R_2}{R_1+R_2}}\left(1 - \exp\frac{-t}{\dfrac{C_L R_{out}}{1+AR_2/(R_1+R_2)}}\right)u(t), \tag{9.28}$$

indicating that the slope is proportional to the final value. This type of response is called "linear settling."

With a realistic op amp, on the other hand, the step response of the circuit begins to deviate from (9.28) as the input amplitude increases. Illustrated in Fig. 9.52, the response to sufficiently small inputs follows the exponential of Eq. (9.28), but with large input steps, the output displays a linear *ramp* having a *constant slope*. Under this condition, we say the op amp experiences slewing and call the slope of the ramp the "slew rate."

To understand the origin of slewing, let us replace the op amp of Fig. 9.52 by a simple CMOS implementation (Fig. 9.53), assuming for simplicity that $R_1 + R_2$ is quite large. We first examine the circuit with a small input step. If V_{in} experiences a change of ΔV, I_{D1} increases by $g_m \Delta V/2$ and I_{D2} decreases by $g_m \Delta V/2$. Since the mirror action of M_3

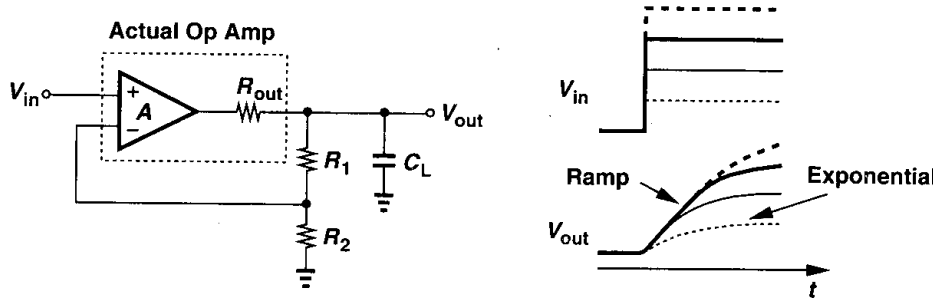


Figure 9.52 Slewing in an op amp circuit.

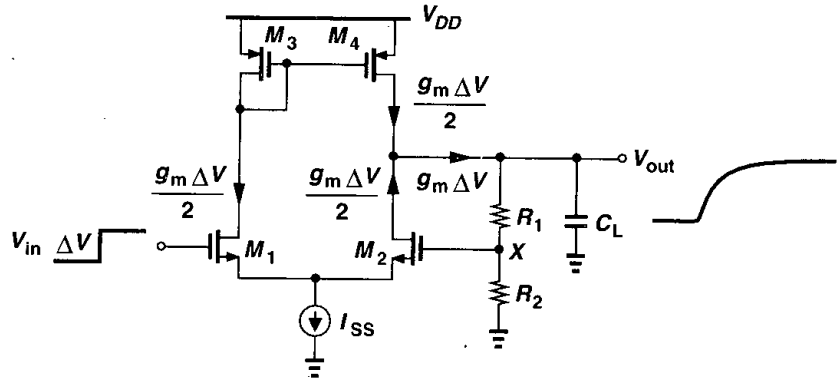


Figure 9.53 Small-signal operation of a simple op amp.

and M_4 raises $|I_{D4}|$ by $g_m\Delta V/2$, the total small-signal current provided by the op amp equals $g_m\Delta V$. This current begins to charge C_L, but as V_{out} rises, so does V_X, reducing the difference between V_{G1} and V_{G2} and hence the output current of the op amp. As a result, V_{out} varies according to (9.28).

Now suppose ΔV is so large that M_1 absorbs all of I_{SS}, turning off M_2. The circuit then reduces to that shown in Fig. 9.54, generating a ramp output with a slope equal to I_{SS}/C_L (if the channel-length modulation of M_4 and the current drawn by $R_1 + R_2$ are neglected). Note that so long as M_2 remains off, the feedback loop is broken and the current charging C_L is constant and independent of the input level. As V_{out} rises, V_X eventually approaches V_{in}, M_2 turns on, and the circuit returns to linear operation.

In Fig. 9.53, slewing occurs for falling edges at the input as well. If the input drops so much that M_1 turns off, then the circuit is simplified as in Fig. 9.55, discharging C_L by a current approximately equal to I_{SS}. After V_{out} decreases sufficiently, the difference between V_X and V_{in} is small enough to allow M_1 to turn on, leading to linear behavior thereafter.

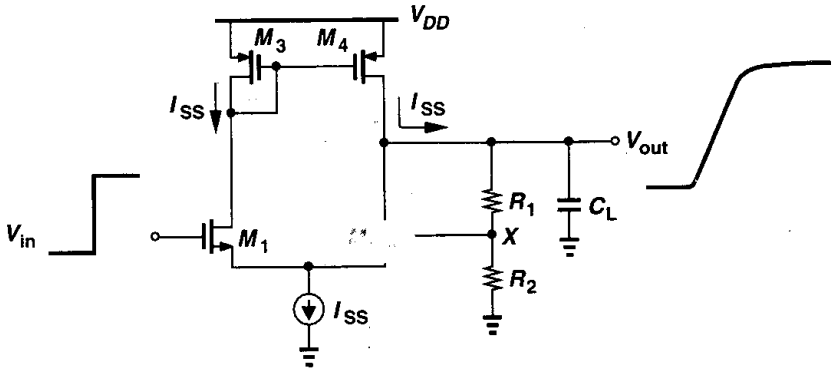


Figure 9.54 Slewing during low-to-high transition.

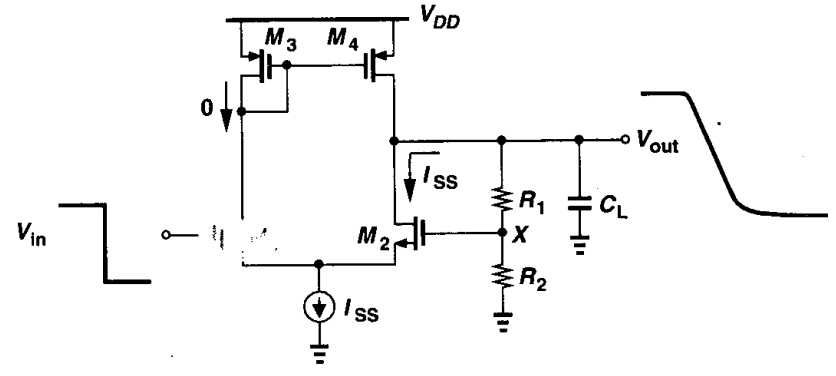


Figure 9.55 Slewing during high-to-low transition.

The foregoing observations explain why slewing is a nonlinear phenomenon. If the input amplitude, say, doubles, the output level does not double at *all* points because the ramp exhibits a slope independent of the input.

Slewing is an undesirable effect in high-speed circuits that process large signals. While the small-signal bandwidth of a circuit may suggest a fast time-domain response, the large-signal speed may be limited by the slew rate simply because the current available to charge and discharge the dominant capacitor in the circuit is small. Moreover, since the input-output relationship during slewing is nonlinear, the output of a slewing amplifier exhibits substantial distortion. For example, if a circuit is to amplify a sinusoid $V_0 \sin \omega_0 t$ (in the steady state), then its slew rate must exceed $V_0\omega_0$.

Example 9.10

Consider the feedback amplifier depicted in Fig. 9.56(a), where C_1 and C_2 set the closed-loop gain. (The bias network for the gate of M_2 is not shown.) (a) Determine the small-signal step response of the circuit. (b) Calculate the positive and negative slew rates.

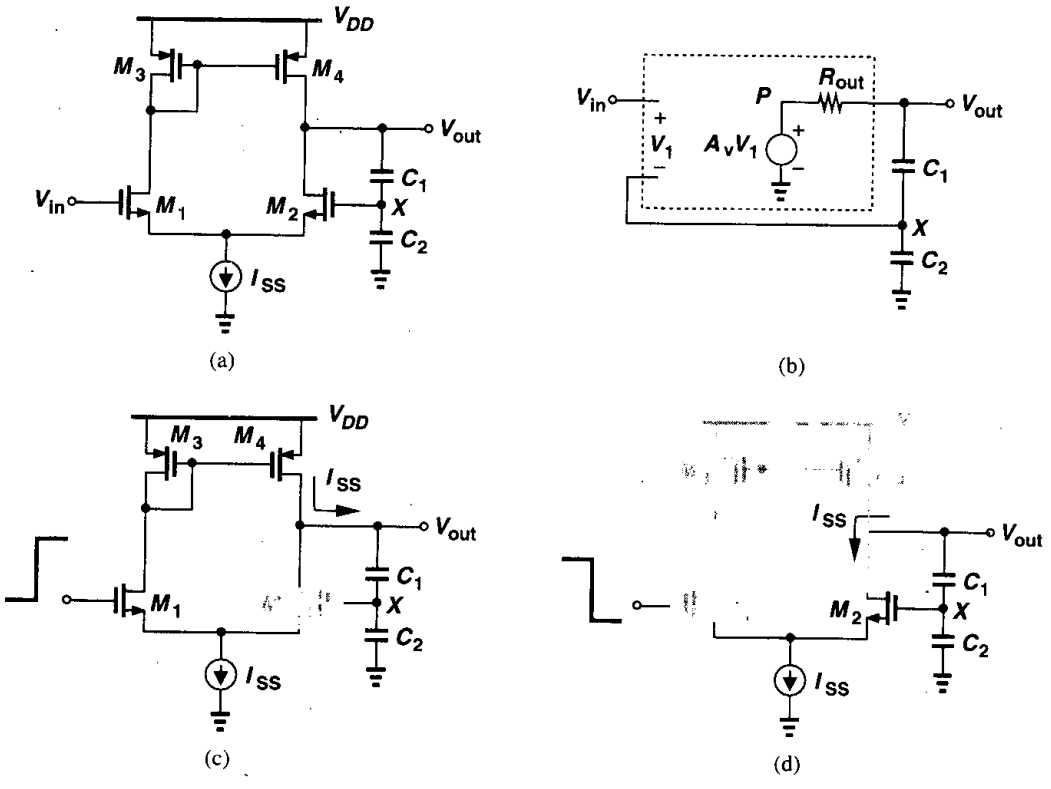


Figure 9.56

Solution

(a) Modeling the op amp as in Fig. 9.56(b), where $A_v = g_{m1,2}(r_{O2} \| r_{O4})$ and $R_{out} = r_{O2} \| r_{O4}$, we have $V_X = C_1 V_{out}/(C_1 + C_2)$ and hence

$$V_P = \left(V_{in} - \frac{C_1}{C_1 + C_2} V_{out} \right) A_v, \tag{9.29}$$

obtaining

$$\left[\left(V_{in} - \frac{C_1}{C_1 + C_2} V_{out} \right) A_v - V_{out} \right] \frac{1}{R_{out}} = V_{out} \frac{C_1 C_2}{C_1 + C_2} s. \tag{9.30}$$

It follows that

$$\frac{V_{out}}{V_{in}}(s) = \frac{A_v}{1 + A_v \dfrac{C_1}{C_1 + C_2} + \dfrac{C_1 C_2}{C_1 + C_2} R_{out} s} \tag{9.31}$$

$$= \frac{A_v/\left(1 + A_v \dfrac{C_1}{C_1 + C_2}\right)}{1 + \dfrac{C_1 C_2}{C_1 + C_2} R_{out} s/\left(1 + A_v \dfrac{C_1}{C_1 + C_2}\right)}, \tag{9.32}$$

revealing that both the low-frequency gain and the time constant of the circuit have decreased by a factor of $1 + A_v C_1/(C_1 + C_2)$. The response to a unity step is thus given by

$$V_{out}(t) = \frac{A_v}{1 + A_v \dfrac{C_1}{C_1 + C_2}} V_0 \left(1 - \exp \frac{-t}{\tau}\right) u(t), \tag{9.33}$$

where

$$\tau = \frac{C_1 C_2}{C_1 + C_2} R_{out}/\left(1 + A_v \frac{C_1}{C_1 + C_2}\right). \tag{9.34}$$

(b) Suppose a large positive step is applied to the gate of M_1 in Fig. 9.56(a) while the initial voltage across C_1 is zero. Then, M_2 turns off and, as shown in Fig. 9.56(c), V_{out} rises according to $V_{out}(t) = I_{SS}/[C_1 C_2/(C_1 + C_2)]t$. Similarly, for a large negative step at the input, Fig. 9.56(d) yields $V_{out} = -I_{SS}/[C_1 C_2/(C_1 + C_2)]t$.

As another example, let us find the slew rate of the telescopic op amp shown in Fig. 9.57(a). When a large differential input is applied, M_1 or M_2 turns off, reducing the circuit to that shown in Fig. 9.57(b). Thus, V_{out1} and V_{out2} appear as ramps with slopes equal to $\pm I_{SS}/(2C_L)$, and consequently $V_{out1} - V_{out2}$ exhibits a slew rate equal to I_{SS}/C_L. (Of course, the circuit is usually used in closed-loop form.)

It is also instructive to study the slewing behavior of a folded-cascode op amp with single-ended output [Fig. 9.58(a)]. Figs. 9.58(a) and (b) depict the equivalent circuit for

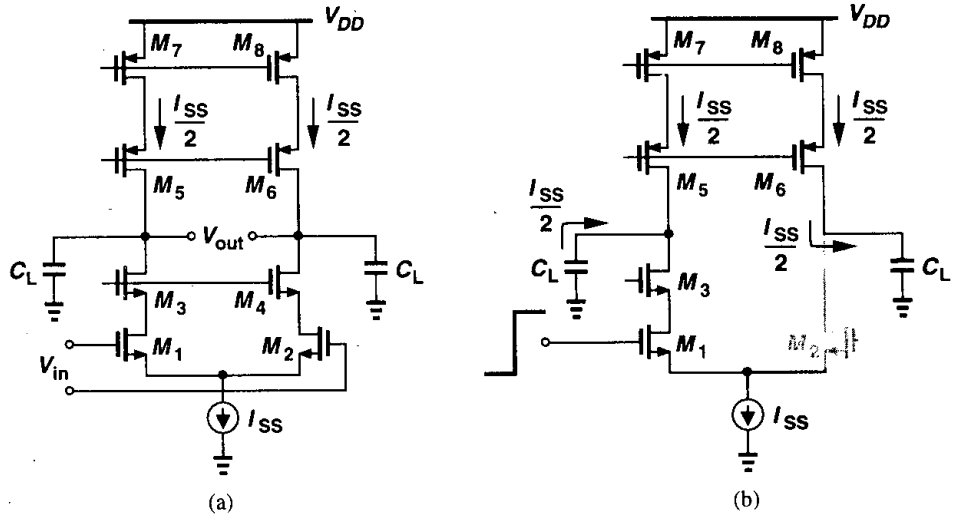


Figure 9.57 Slewing in telescopic op amp.

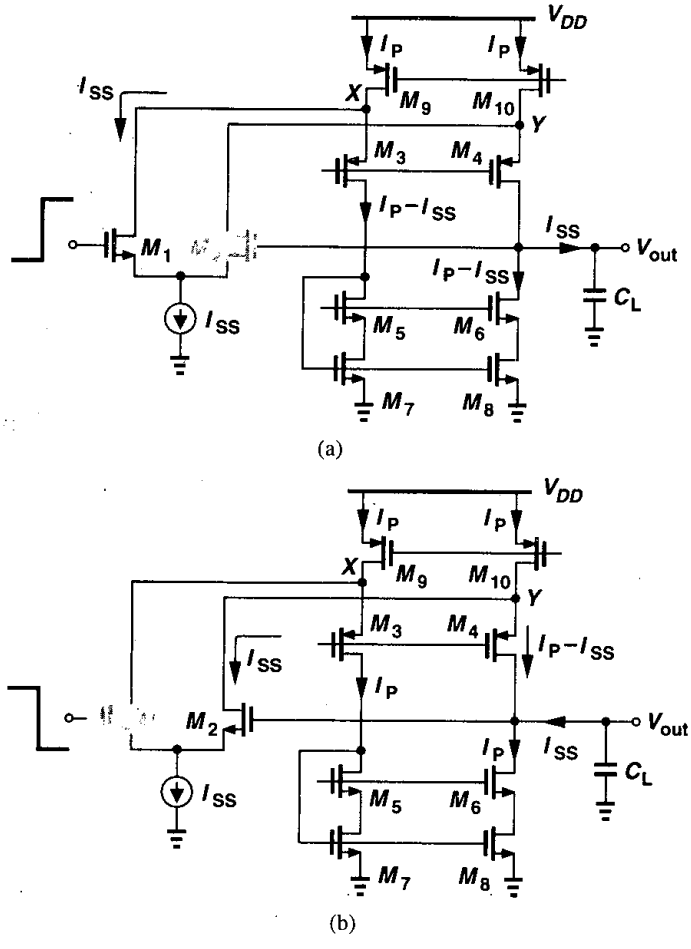


Figure 9.58 Slewing in folded-cascode op amp.

positive and negative input steps, respectively. Here, the PMOS current sources provide a current of I_P, and the current that charges or discharges C_L is equal to I_{SS}, yielding a slew rate of I_{SS}/C_L. Note that the slew rate is independent of I_P if $I_P \geq I_{SS}$. In practice, we choose $I_P \approx I_{SS}$.

In Fig. 9.58(a), if $I_{SS} > I_P$, then during slewing M_3 turns off and V_X falls to a low level such that M_1 and the tail current source enter the triode region. Thus, for the circuit to return to equilibrium after M_2 turns on, V_X must experience a large swing, slowing down the settling. This phenomenon is illustrated in Fig. 9.59.

To alleviate this issue, two "clamp" transistors can be added as shown in Fig. 9.60(a) [8]. The idea is that the difference between I_{SS} and I_P now flows through M_{11} or M_{12}, requiring only enough drop in V_X or V_Y to turn on one of these transistors. Fig. 9.60(b) illustrates a more aggressive approach, where M_{11} and M_{12} clamp the two nodes directly to V_{DD}. Since

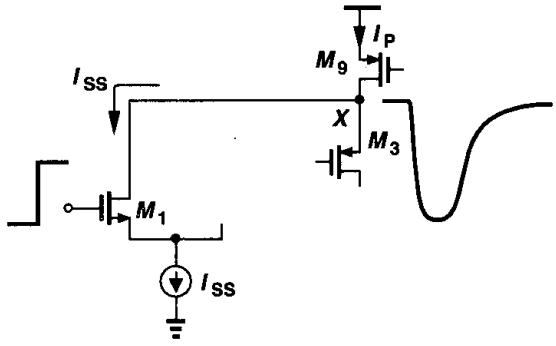


Figure 9.59 Long settling due to overdrive recovery after slewing.

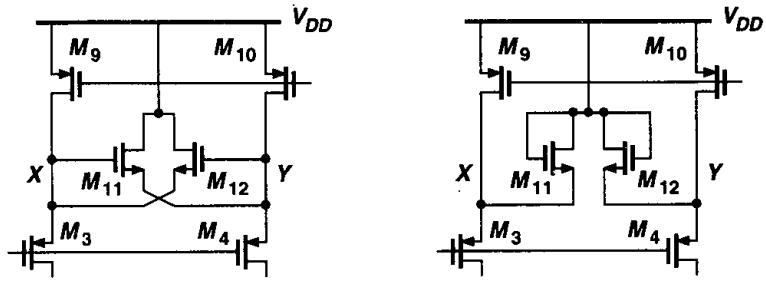


Figure 9.60 Clamp circuit to limit swings at X and Y.

the equilibrium value of V_X and V_Y is usually higher than $V_{DD} - V_{THN}$, M_{11} and M_{12} are off during small-signal operation.

What trade-offs are encountered in increasing the slew rate? In the examples of Figs. 9.57 and 9.58, for a given load capacitance, I_{SS} must be increased and to maintain the same maximum output swing, all of the transistors must be made proportionally wider. As a result, the power dissipation and the input capacitance are increased. Note that if the device currents and widths scale together, $g_m r_O$ of each transistor and hence the open-loop gain of the op amp remain constant.

How does an op amp leave the slewing regime and enter the linear-settling regime? Since the point at which one of the input transistors "turns on" is ambiguous, the distinction between slewing and linear settling is somewhat arbitrary. The following example illustrates the point.

Example 9.11

Consider the circuit of Fig. 9.56(a) in the slewing regime [Fig. 9.56(c)]. As V_{out} rises, so does V_X, eventually turning M_2 on. As I_{D2} increases from zero, the differential pair becomes more linear. Considering M_1 and M_2 to operate linearly if the difference between their drain currents is less than αI_{SS} (e.g., $\alpha = 0.1$), determine how long the circuit takes to enter linear settling. Assume the input step has an amplitude of V_0.

Solution

The circuit displays a slew rate of $I_{SS}/[C_1C_2/(C_1 + C_2)]$ until $|V_{in1} - V_{in2}|$ is sufficiently small. From Chapter 4, we can write

$$\alpha I_{SS} = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{in1} - V_{in2})\sqrt{\frac{4I_{SS}}{\mu_n C_{ox}\frac{W}{L}} - (V_{in1} - V_{in2})^2}, \tag{9.35}$$

obtaining

$$\Delta V_G^4 - \Delta V_G^2\frac{4I_{SS}}{\mu_n C_{ox}\frac{W}{L}} + \left(\frac{2\alpha I_{SS}}{\mu_n C_{ox}\frac{W}{L}}\right)^2 = 0, \tag{9.36}$$

where $\Delta V_G = V_{in1} - V_{in2}$. Thus,

$$\Delta V_G \approx \alpha\sqrt{\frac{I_{SS}}{\mu_n C_{ox}\frac{W}{L}}}. \tag{9.37}$$

(Recall that $\sqrt{I_{SS}/[\mu_n C_{ox}(W/L)]}$ is the equilibrium overdrive voltage of each transistor in the differential pair.) Alternatively, we recognize that for a small difference, αI_{SS}, between I_{D1} and I_{D2}, a small-signal approximation is valid: $\alpha I_{SS} = g_m \Delta V_G$. Thus, $\Delta V_G = \alpha I_{SS}/g_m \approx \alpha I_{SS}/\sqrt{\mu_n C_{ox}(W/L)I_{SS}}$. Note that this calculation is quite rough because as M_2 turns on, the current charging the load capacitance is no longer constant.

Since V_X must rise to $V_0 - \Delta V_G$ for M_2 to carry the required current, V_{out} increases by $(V_0 - \Delta V_G)(1 + C_2/C_1)$, requiring a time given by

$$t = \frac{C_2}{I_{SS}}\left(V_0 - \alpha\sqrt{\frac{I_{SS}}{\mu_n C_{ox}\frac{W}{L}}}\right). \tag{9.38}$$

In the above example, the value of α that determines the onset of linear settling depends, among other things, on the actual required linearity. In other words, for a nonlinearity of 1%, α can be quite larger than for a nonlinearity of 0.1%.

The slewing behavior of two-stage op amps is somewhat different from that of the circuits studied above. This case is studied in Chapter 10.

9.9 Power Supply Rejection

As other analog circuits, op amps are often supplied from noisy lines and must therefore "reject" the noise adequately. For this reason, it is important to understand how noise on the supply manifests itself at the output of an op amp.

Let us consider the simple op amp shown in Fig. 9.61, assuming the supply voltage varies slowly. If the circuit is perfectly symmetric, $V_{out} = V_X$. Since the diode-connected device "clamps" node X to V_{DD}, V_X and hence V_{out} experience approximately the same change as does V_{DD}. In other words, the gain from V_{DD} to V_{out} is close to unity. The power supply

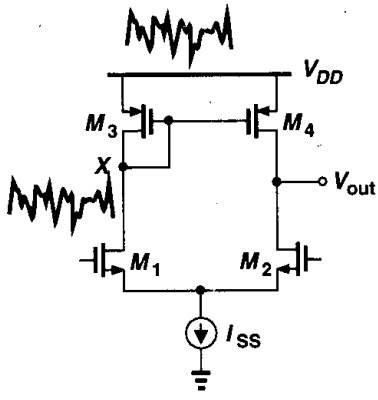


Figure 9.61 Supply rejection of differential pair with active current mirror.

rejection ratio (PSRR) is defined as the gain from the input to the output divided by the gain from the supply to the output. At low frequencies:

$$PSRR \approx g_{mN}(r_{OP} \| r_{ON}). \tag{9.39}$$

Example 9.12

Calculate the low-frequency PSRR of the feedback circuit shown in Fig. 9.62(a).

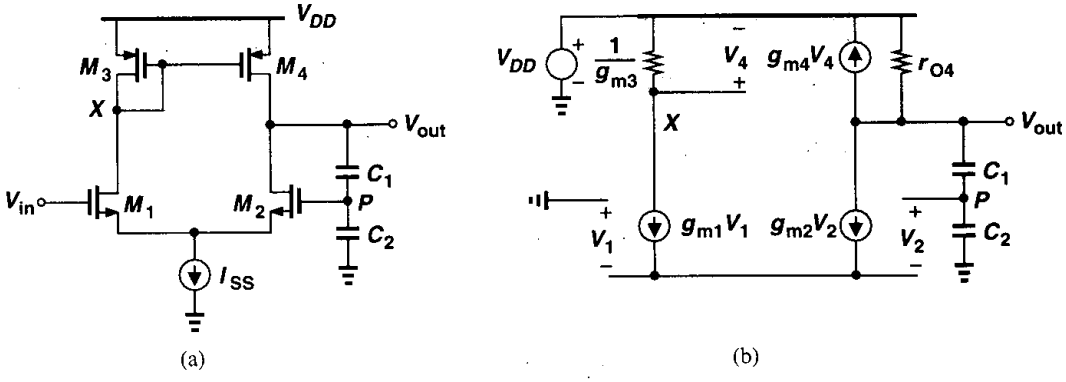


Figure 9.62

Solution

From the foregoing analysis, we may surmise that a change ΔV in V_{DD} appears unattenuated at the output. But, we should note that if V_{out} changes, so do V_P and I_{D2}, thereby opposing the change. Using Fig. 9.62(b) and neglecting channel-length modulation in M_1-M_3 for simplicity, we can write:

$$V_{out}\frac{C_1}{C_1 + C_2} - V_2 = -V_1, \tag{9.40}$$

and $g_{m1}V_1 + g_{m2}V_2 = 0$. Thus, if the circuit is symmetric,

$$V_2 = \frac{V_{out}}{2}\frac{C_1}{C_1 + C_2}. \tag{9.41}$$

We also have

$$-\frac{g_{m1}V_1}{g_{m3}}g_{m4} - \frac{V_{DD} - V_{out}}{r_{O4}} + g_{m2}V_2 = 0. \tag{9.42}$$

It follows that

$$\frac{V_{out}}{V_{DD}} = \frac{1}{g_{m2}r_{O4}\dfrac{C_1}{C_1 + C_2} + 1}. \tag{9.43}$$

Thus,

$$PSRR \approx \frac{1 + \dfrac{C_2}{C_1}}{g_{m2}r_{O4}\dfrac{C_1}{C_1 + C_2} + 1}. \tag{9.44}$$

9.10 Noise in Op Amps

In low-noise applications, the input-referred noise of op amps becomes critical. We now extend the noise analysis of differential amplifiers in Chapter 7 to more sophisticated topologies. With many transistors in an op amp, it may seem difficult to intuitively identify the dominant sources of noise. A simple rule for inspection is to (mentally) change the gate voltage of each transistor by a small amount and predict the effect at the output.

Let us first consider the telescopic op amp shown in Fig. 9.63. At relatively low frequencies, the cascode devices contribute negligible noise, leaving M_1-M_2 and M_7-M_8 as

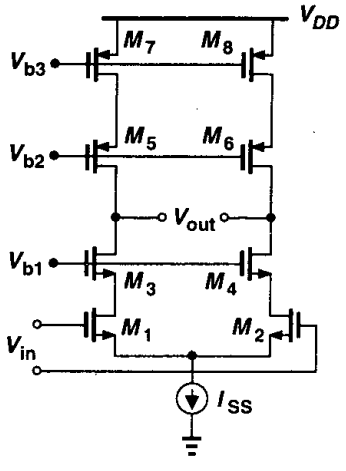


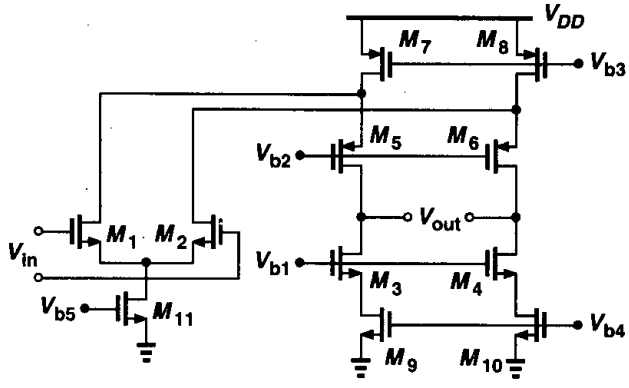
Figure 9.63 Noise in a telescopic op amp.

the primary noise sources. The input-referred noise voltage per unit bandwidth is therefore similar to that in Fig. 7.47(a) and given by:

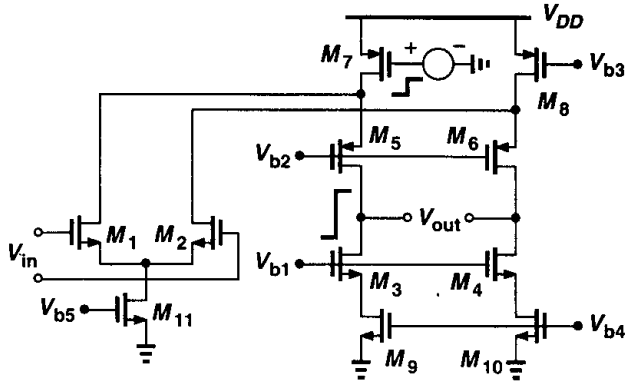
$$\overline{V_n^2} = 4kT\left(2\frac{2}{3g_{m1,2}} + 2\frac{2g_{m7,8}}{3g_{m1,2}^2}\right) + 2\frac{K_N}{(WL)_{1,2}C_{ox}f} + 2\frac{K_P}{(WL)_{7,8}C_{ox}f}\frac{g_{m7,8}^2}{g_{m1,2}^2}, \tag{9.45}$$

where K_N and K_P denote the $1/f$ noise coefficients of NMOS and PMOS devices, respectively.

Next, we study the noise behavior of the folded-cascode op amp of Fig. 9.64(a), considering only thermal noise at this point. Again, the noise of the cascode devices is negligible at low frequencies, leaving M_1-M_2, M_7-M_8, and M_9-M_{10} as potentially significant sources. Do both pairs M_7-M_8 and M_9-M_{10} contribute noise? Using our simple rule, we change the gate voltage of M_7 by a small amount [Fig. 9.64(b)], noting that the output indeed



(a)



(b)

Figure 9.64 Noise in a folded-cascode op amp.

changes considerably. The same observation applies to M_8-M_{10} as well. To determine the input-referred thermal noise, we first refer the noise of M_7-M_8 and M_9-M_{10} to the output:

$$\overline{V_{n,out}^2}|_{M7,8} = 2\left(4kT\frac{2}{3g_{m7,8}}g_{m7,8}^2R_{out}^2\right), \tag{9.46}$$

where the factor 2 accounts for (uncorrelated) noise of M_7 and M_8 and R_{out} denotes the open-loop output resistance of the op amp. Similarly,

$$\overline{V_{n,out}^2}|_{M9,10} = 2\left(4kT\frac{2}{3g_{m9,10}}g_{m9,10}^2R_{out}^2\right). \tag{9.47}$$

Dividing these quantities by $g_{m1,2}^2R_{out}^2$ and adding the contribution of M_1-M_2, we obtain the overall noise:

$$\overline{V_{n,int}^2} = 8kT\left(\frac{2}{3g_{m1,2}} + \frac{2}{3}\frac{g_{m7,8}}{g_{m1,2}^2} + \frac{2}{3}\frac{g_{m9,10}}{g_{m1,2}^2}\right). \tag{9.48}$$

The effect of flicker noise can be included in a similar manner (Problem 9.15). Note that the folded-cascode topology potentially suffers from greater noise than the telescopic counterpart.

As observed for the differential amplifiers in Chapter 7, the noise contribution of the PMOS and NMOS current sources *increases* in proportion to their transconductance. This trend results in a trade-off between output voltage swings and input-referred noise: for a given current, as implied by $g_m = 2I_D/(V_{GS} - V_{TH})$, if the overdrive voltage of the current sources is minimized to allow large swings, then their transconductance is maximized.

As another case, we calculate the input-referred thermal noise of the two-stage op amp shown in Fig. 9.65. Beginning with the second stage, we note that the noise current of M_5

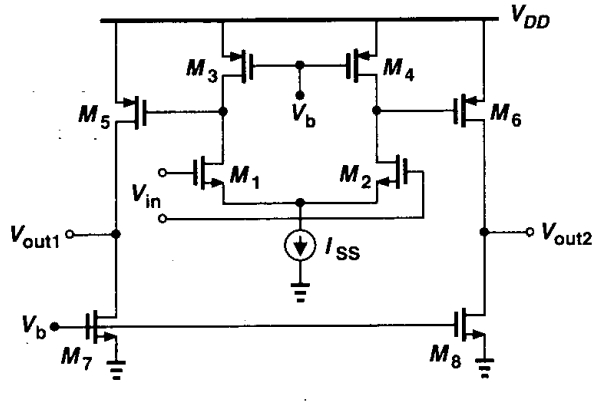


Figure 9.65 Noise in a two-stage op amp.

and M_7 flows through $r_{O5}\|r_{O7}$. Dividing the resulting output noise voltage by the total gain, $g_{m1}(r_{O1}\|r_{O3}) \times g_{m5}(r_{O5}\|r_{O7})$, and doubling the power, we obtain the input-referred contribution of M_5-M_8:

$$\overline{V_n^2}|_{M5-8} = 2 \times 4kT\frac{2}{3}(g_{m5}+g_{m7})(r_{O5}\|r_{O7})^2\frac{1}{g_{m1}^2(r_{O1}\|r_{O3})^2 g_{m5}^2(r_{O5}\|r_{O7})^2} \tag{9.49}$$

$$= \frac{16kT}{3}\frac{g_{m5}+g_{m7}}{g_{m1}^2 g_{m5}^2(r_{O1}\|r_{O3})^2}. \tag{9.50}$$

The noise due to M_1-M_4 is simply equal to

$$\overline{V_n^2}|_{M1-4} = 2 \times 4kT\frac{2}{3}\frac{g_{m1}+g_{m3}}{g_{m1}^2}. \tag{9.51}$$

It follows that

$$\overline{V_{n,tot}^2} = \frac{16kT}{3}\frac{1}{g_{m1}^2}\left[g_{m1}+g_{m3}+\frac{g_{m5}+g_{m7}}{g_{m5}^2(r_{O1}\|r_{O3})^2}\right]. \tag{9.52}$$

Note the noise resulting from the second stage is usually negligible because it is divided by the gain of the first stage when referred to the main input.

Example 9.13

A simple amplifier is constructed as shown in Fig. 9.66. Note that the first stage incorporates diode-connected—rather than current-source—loads. Assuming all of the transistors are in saturation and $(W/L)_{1,2} = 50/0.6$, $(W/L)_{3,4} = 10/0.6$, $(W/L)_{5,6} = 20/0.6$, and $(W/L)_{7,8} = 56/0.6$, calculate the input-referred noise voltage if $\mu_n C_{ox} = 75\ \mu\text{A/V}^2$ and $\mu_p C_{ox} = 30\ \mu\text{A/V}^2$.

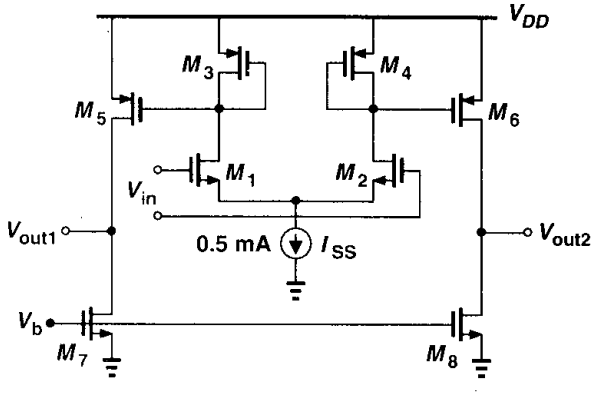


Figure 9.66

Solution

We first calculate the small-signal gain of the first stage.:

$$A_{v1} \approx \frac{g_{m1}}{g_{m3}} \tag{9.53}$$

$$= \sqrt{\frac{50 \times 75}{10 \times 30}} \tag{9.54}$$

$$\approx 3.54. \tag{9.55}$$

The noise of M_5 and M_7 referred to the gate of M_5 is equal to $4kT(2/3)(g_{m5} + g_{m7})/g_{m5}^2 = 2.87 \times 10^{-17}$ V^2/Hz, which is divided by A_{v1}^2 when referred to the main input: $\overline{V_n^2}|_{M5,7} = 2.29 \times 10^{-18}$ V^2/Hz. Transistors M_1 and M_3 produce an input-referred noise of $\overline{V_n^2}|_{M1,3} = (8kT/3)(g_{m3} + g_{m1})/g_{m1}^2 = 1.10 \times 10^{-17}$ V^2/Hz. Thus, the total input-referred noise equals

$$\overline{V_{n,in}^2} = 2(2.29 \times 10^{-18} + 1.10 \times 10^{-17}) \tag{9.56}$$

$$= 2.66 \times 10^{-17} \text{ V}^2/\text{Hz}, \tag{9.57}$$

where the factor 2 accounts for the noise produced by both odd-numbered and even-numbered transistors in the circuit. This value corresponds to an input noise voltage of 5.16 nV/$\sqrt{\text{Hz}}$.

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.

9.1. **(a)** Derive expressions for the transconductance and output resistance of a MOSFET in the triode region. Plot these quantities and $g_m r_O$ as a function of V_{DS}, covering both triode and saturation regions.

(b) Consider the amplifier of Fig. 9.6(b), with $(W/L)_{1-4} = 50/0.5$, $I_{SS} = 1$ mA, and input CM level of 1.3 V. Calculate the small-signal gain and the maximum output swing if all transistors remain in saturation.

(c) For the circuit of part (b), suppose we allow each PMOS device to enter the triode region by 50 mV so as to increase the allowable differential swing by 100 mV. What is the small-signal gain at the peaks of the output swing?

9.2. In the circuit of Fig. 9.9, assume $(W/L)_{1-4} = 100/0.5$, $I_{SS} = 1$ mA, $V_b = 1.4$ V, and $\gamma = 0$.

(a) If M_5-M_8 are identical and have a length of 0.5 μm, calculate their minimum width such that M_3 operates in saturation.

(b) Calculate the maximum output voltage swing.

(c) What is the open-loop voltage gain?

(d) Calculate the input-referred thermal noise voltage.

9.3. Design the folded-cascode op amp of Fig. 9.13 for the following requirements: maximum differential swing = 2.4 V, total power dissipation = 6 mW. If all of the transistors have a channel length of 0.5 μm, what is the overall voltage gain? Can the input common-mode level be as low as zero?

9.4. In the op amp of Fig. 9.18(b), $(W/L)_{1-8} = 100/0.5$, $I_{SS} = 1$ mA, and $V_{b1} = 1.7$ V. Assume $\gamma = 0$.
(a) What is the maximum allowable input CM level?
(b) What is V_X?
(c) What is the maximum allowable output swing if the gate of M_2 is connected to the output?
(d) What is the acceptable range of V_{b2}?
(e) What is the input-referred thermal noise voltage?

9.5. Design the op amp of Fig. 9.18(b) for the following requirements: maximum differential swing $= 2.4$ V, total power dissipation $= 6$ mW. (Assume the gate of M_2 is never shorted to the output.)

9.6. If in Fig. 9.21, $(W/L)_{1-8} = 100/0.5$ and $I_{SS} = 1$ mA,
(a) What CM level must be established at the drains of M_3 and M_4 so that $I_{D5} = I_{D6} = 1$ mA? How does this constrain the maximum input CM level?
(b) With the choice made in part (a), calculate the overall voltage gain and the maximum output swing.

9.7. Design the op amp of Fig. 9.21 for the following requirements: maximum differential swing $= 4$ V, total power dissipation $= 6$ mW, $I_{SS} = 0.5$ mA.

9.8. Suppose the circuit of Fig. 9.22 is designed with I_{SS} equal to 1 mA, I_{D9}-I_{D12} equal to 0.5 mA, and $(W/L)_{9-12} = 100/0.5$.
(a) What CM level is required at X and Y?
(b) If I_{SS} requires a minimum voltage of 400 mV, choose the minimum dimensions of M_1-M_8 to allow a peak-to-peak swing of 200 mV at X and at Y.
(c) Calculate the overall voltage gain.

9.9. In Fig. 9.25(c), calculate the input-referred thermal noise if I_1 and I_2 are implemented by PMOS devices.

9.10. Suppose in Fig. 9.25(c), $I_1 = 100\ \mu$A, $I_2 = 0.5$ mA, and $(W/L)_{1-3} = 100/0.5$. Assuming I_1 and I_2 are implemented with PMOS devices having $(W/L)_P = 50/0.5$,
(a) Calculate the gate bias voltages of M_2 and M_3.
(b) Determine the maximum allowable output voltage swing.
(c) Calculate the overall voltage gain and the input-referred thermal noise voltage.

9.11. In the circuit of Fig. 9.41, each branch is biased at a current of 0.5 mA. Choose the dimensions of M_7 and M_8 such that the output CM level is equal to 1.5 V and $V_P = 100$ mV.

9.12. Consider the CMFB network in Fig. 9.39. The amplifier sensing $V_{out,CM}$ is to be implemented as a different pair with active current mirror load.
(a) Should the input pair of the amplifier use PMOS devices or NMOS devices?
(b) Calculate the loop gain for the CMFB network.

9.13. Repeat Problem 9.12(b) for the circuit of Fig. 9.40.

9.14. In the circuit of Fig. 9.56(a), assume $(W/L)_{1-4} = 100/0.5$, $C_1 = C_2 = 0.5$ pF, and $I_{SS} = 1$ mA.
(a) Calculate the small-signal time constant of the circuit.
(b) With a 1-V step at the input [Fig. 9.56(c)], how long does it take for I_{D2} to reach $0.1I_{SS}$?

9.15. It is possible to argue that the auxiliary amplifier in the circuit of Fig. 9.24(c) *reduces* the output impedance. Consider the circuit as drawn in Fig. 9.67, where the drain voltage of M_2 is changed by ΔV to measure the output impedance. It seems that, since the feedback provided by A_1 attempts to hold V_X constant, the change in the current through r_{O2} is much *greater* than in the circuit of Fig. 9.24(b), suggesting that $R_{out} \approx r_{O2}$. Explain the flaw in this argument.

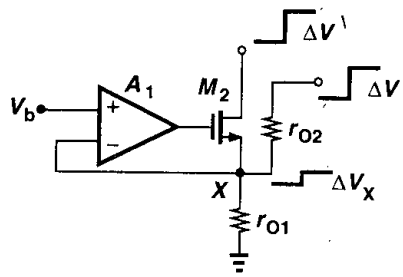


Figure 9.67

9.16. Calculate the CMRR of the circuit shown in Fig. 9.56(a).

9.17. Calculate the input-referred flicker noise of the op amp shown in Fig. 9.64(a).

9.18. In this problem, we design a two-stage op amp based on the topology shown in Fig. 9.68. Assume a power budget of 6 mW, a required output swing of 2.5 V, and $L_{eff} = 0.5\ \mu$m for all devices.

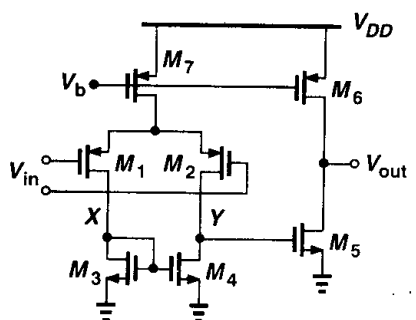


Figure 9.68

(a) Allocating a current of 1 mA to the output stage and roughly equal overdrive voltages to M_5 and M_6, determine $(W/L)_5$ and $(W/L)_6$. Note that the gate-source capacitance of M_5 is in the signal path whereas that of M_6 is not. Thus, M_6 can be quite larger than M_5.

(b) Calculate the small-signal gain of the output stage.

(c) With the remaining 1 mA flowing through M_7, determine the aspect ratio of M_3 (and M_4) such that $V_{GS3} = V_{GS5}$. This is to guarantee that if $V_{in} = 0$ and hence $V_X = V_Y$, then M_5 carries the expected current.

(d) Calculate the aspect ratios of M_1 and M_2 such that the overall voltage gain of the op amp is equal to 500.

9.19. Consider the op amp of Fig. 9.68, assuming that the second stage is to provide a voltage gain of 20 with a bias current of 1 mA.

(a) Determine $(W/L)_5$ and $(W/L)_6$ such that M_5 and M_6 have equal overdrive voltages.

(b) What is the small-signal gain of this stage if M_6 is driven into the triode region by 50 mV?

9.20. The op amp designed in Problem 9.18(d) is placed in unity-gain feedback. Assume $|V_{GS7} - V_{TH7}| = 0.4$ V.

(a) What is the allowable input voltage range?

(b) At what input voltage are the input and output voltages *exactly* equal?

9.21. Calculate the input-referred noise of the op amp designed in Problem 9.18(d).

9.22. It is possible to use the bulk terminal of PMOS devices as an input [9]. Consider the amplifier shown in Fig. 9.69 as an example.

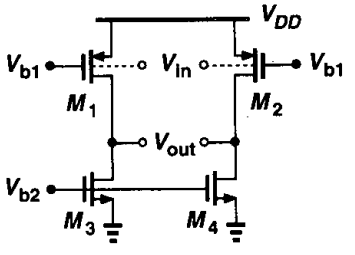


Figure 9.69

(a) Calculate the voltage gain.
(b) What is the acceptable input common-mode range?
(c) How does the small-signal gain vary with the input common-mode level?
(d) Calculate the input-referred thermal noise voltage and compare the result with that of a regular PMOS differential pair having NMOS current-source loads.

9.23. The idea of the active current mirror can be applied to the output stage of a two-stage op amp as well. That is, the load current source can become a function of the signal. Figure 9.70 shows an example [10]. Here, the first stage consists of M_1-M_4 and the output is produced by M_5-M_8. Transistors M_7 and M_8 operate as active current sources because their current varies with the signal voltage at nodes Y and X, respectively.
(a) Calculate the differential voltage gain of the op amp.
(b) Estimate the magnitude of the three major poles of the circuit.

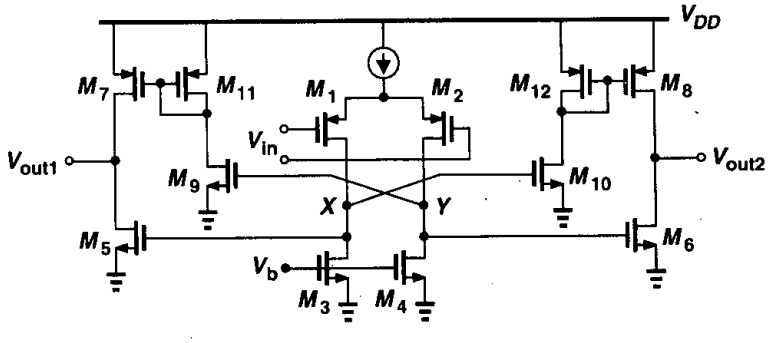


Figure 9.70

9.24. The circuit of Fig. 9.71 employs a fast path (M_1' and M_2') in parallel with the slow path. Calculate the differential voltage gain of the circuit. Which transistors typically limit the output swing?

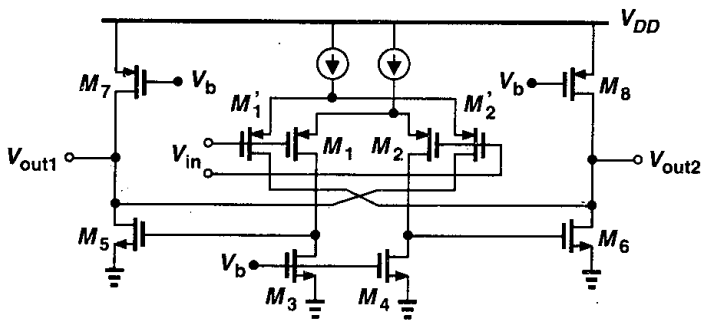


Figure 9.71

9.25. Calculate the input-referred thermal noise of the op amp in Fig. 9.71.

References

1. R. G. Eschauzier, L. P. T. Kerklaan, and J. H. Huising, "A 100-MHz 100-dB Operational Amplifier with Multipath Nested Miller Compensation Structure," *IEEE J. of Solid-State Circuits,* vol. 27, pp. 1709–1717, Dec. 1992.
2. R. M. Ziazadeh, H.-T. Ng, and D. J. Allstot, "A Multistage Amplifier Topology with Embeded Tracking Compensation," *CICC Proc.*, pp. 361–364, May 1998.
3. F. You, S. H. Embabi, and E. Sanchez-Sincencio, "A Multistage Amplifier Topology with Nested G_m-C Compensation for Low-Voltage Application," *ISSCC Dig. of Tech. Papers,* pp. 348–349, Feb. 1997.
4. B. J. Hosticka, "Improvement of the Gain of CMOS Amplifiers," *IEEE J. of Solid-State Circuits,* vol. 14, pp. 1111–1114, Dec. 1979.
5. K. Bult and G. J. G. H. Geelen, "A Fast-Settling CMOS Operational Amplifier for SC Circuits with 90-dB DC Gain," *IEEE J. of Solid-State Circuits,* vol. 25, pp. 1379–1384, Dec. 1990.
6. E. Sackinger and W. Guggenbuhl, "A High-Swing High-Impedance MOS Cascode Circuit," *IEEE J. of Solid-State Circuits,* vol. 25, pp. 289–298, Feb. 1990.
7. R. Hogervost et al., "A Compact Power-Efficient 3-V CMOS Rail-to-Rail Input/Output Operational Amplifier for VLSI Cell Libraries," *IEEE J. of Solid-State Circuits,* vol. 29, pp. 1505–1513, Dec. 1994.
8. D. A. Johns and K. Martin, *Analog Integrated Circuit Design,* New York: Wiley, 1997.
9. P. E. Allen, B. J. Blalock, and G. A. Rincon, "A 1-V CMOS Op Amp Using Bulk-Driven MOSFETs," ISSCC Dig. of Tech. Papers, pp. 192–193, Feb. 1995.
10. S. Rabii and B. A. Wooley, "A 1.8-V Digital-Audio Sigma-Delta Modulator in 0.8-μm CMOS," *IEEE J. of Solid-State Circuits,* vol. 32, pp. 783–796, June 1997.

Chapter 10

Stability and Frequency Compensation

Negative feedback finds wide application in the processing of analog signals. The properties of feedback described in Chapter 8 allow precise operations by suppressing variations of the open-loop characteristics. Feedback systems, however, suffer from potential instability, that is, they may oscillate.

In this chapter, we deal with the stability and frequency compensation of linear feedback systems to the extent necessary to understand design issues of analog feedback circuits. Beginning with a review of stability criteria and the concept of phase margin, we study frequency compensation, introducing various techniques suited to different op amp topologies. We also analyze the impact of frequency compensation on the slew rate of two-stage op amps.

10.1 General Considerations

Let us consider the negative feedback system shown in Fig. 10.1, where β is assumed constant. Writing the closed-loop transfer function as

$$\frac{Y}{X}(s) = \frac{H(s)}{1 + \beta H(s)}, \tag{10.1}$$

we note that if $\beta H(s = j\omega_1) = -1$, the "gain" goes to infinity, and the circuit can amplify its own noise until it eventually begins to oscillate. In other words, if $\beta H(j\omega_1) = -1$, then

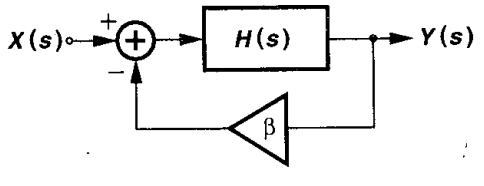


Figure 10.1 Basic negative-feedback system.

the circuit may oscillate at frequency ω_1. This condition can be expressed as

$$|\beta H(j\omega_1)| = 1 \tag{10.2}$$

$$\angle\beta H(j\omega_1) = -180°, \tag{10.3}$$

which are called "Barkhausen's Criteria." Note that the total phase shift around the loop at ω_1 is 360° because *negative* feedback itself introduces 180° of phase shift. The 360° phase shift is necessary for oscillation since the feedback signal must add *in phase* to the original noise to allow oscillation buildup. By the same token, a loop gain of unity (or greater) is also required to enable growth of the oscillation amplitude.

In summary, a negative feedback system may oscillate at ω_1 if (1) the phase shift around the loop at this frequency is so much that the feedback becomes *positive* and (2) the loop gain is still enough to allow signal buildup. Illustrated in Fig. 10.2, the situation can be viewed as excessive loop gain at the frequency for which the phase shift reaches −180° or, equivalently, excessive phase at the frequency for which the loop gain drops to unity. Thus, to avoid instability, we must minimize the total phase shift so that for $|\beta H| = 1$, $\angle\beta H$ is still more positive than −180°. In this chapter, we assume β is less than or equal to unity and does not depend on the frequency.

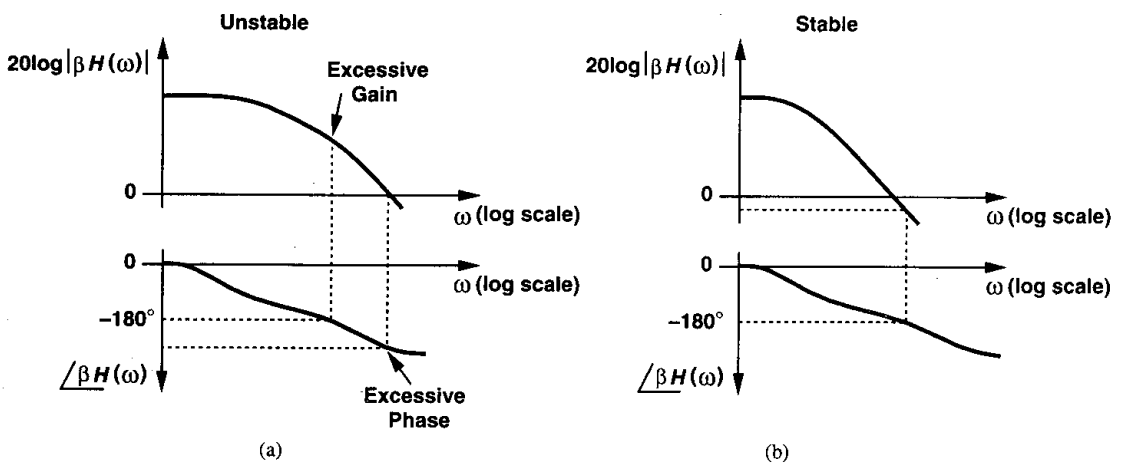


Figure 10.2 Bode plots of loop gain for unstable and stable systems.

The frequencies at which the magnitude and phase of the loop gain are equal to unity and −180°, respectively, play a critical role in the stability and are called the "gain crossover point" and the "phase crossover point," respectively. In a stable system, the gain crossover must occur well before the phase crossover. For the sake of brevity, we denote the gain crossover by GX and the phase crossover by PX. Note that if β is reduced (i.e., less feedback is applied), then the magnitude plots of Fig. 10.2 are shifted down, thereby moving the gain crossover closer to the origin and making the feedback system more stable. Thus, the

worst-case stability corresponds to $\beta = 1$, i.e, unity-gain feedback. For this reason, we often analyze the magnitude and phase plots for $\beta H = H$.

Before studying more specific cases, let us review a few basic rules of constructing Bode plots. A Bode plot illustrates the asymptotic behavior of the magnitude and phase of a complex function according to the magnitude of the poles and zeros. The following two rules are used. (1) The slope of the magnitude plot changes by +20 dB/dec at every zero frequency and by −20 dB/dec at every pole frequency. (2) For a pole (zero) frequency of ω_m, the phase begins to fall (rise) at approximately $0.1\omega_m$, experiences a change of −45° (+45°) at ω_m, and approaches a change of −90° (+90°) at approximately $10\omega_m$. The key point here is that the phase may be much more significantly affected by high-frequency poles and zeros than the magnitude is.

It is also instructive to plot the location of the poles of a closed-loop system on a complex plane. Expressing each pole frequency as $s_p = j\omega_p + \sigma_p$ and noting that the impulse response of the system includes a term $\exp(j\omega_p + \sigma_p)t$, we observe that if s_p falls in the right half plane, i.e., if $\sigma_p > 0$, then the system is likely to oscillate because its time-domain response exhibits a growing exponential [Fig. 10.3(a)]. Even if $\sigma_p = 0$, the system may sustain oscillations [Fig. 10.3(b)]. Conversely, if the poles lie in the left half plane, all time-domain exponential terms decay to zero [Fig. 10.3(c)].[1] In practice, we plot

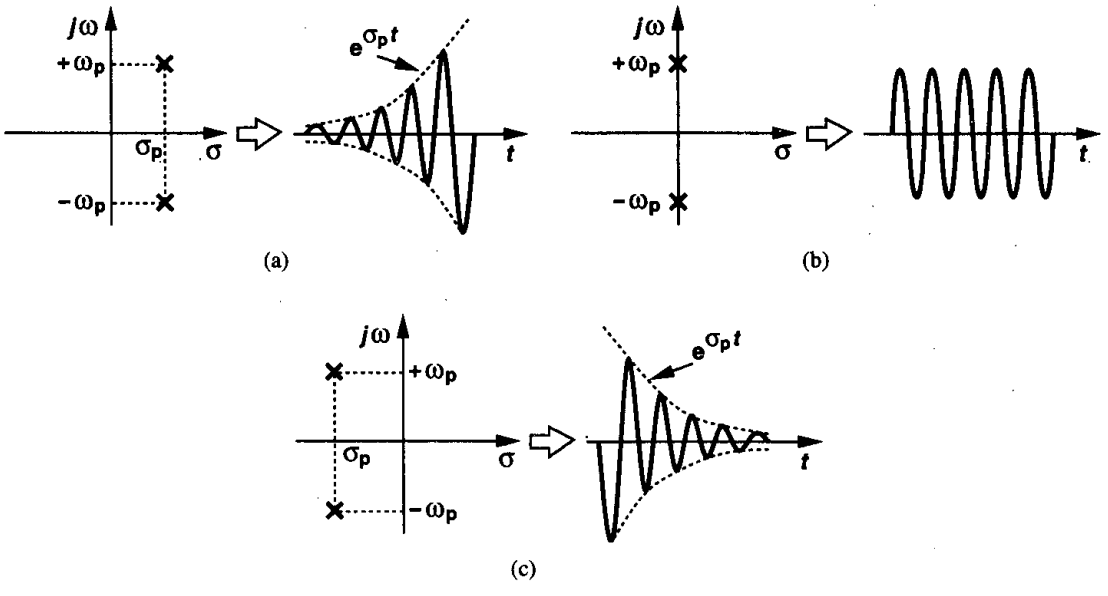


Figure 10.3 Time-domain response of a system versus the position of poles, (a) unstable with growing amplitude, (b) unstable with constant-amplitude oscillation, (c) stable.

[1] We ignore the effect of zeros for now.

the location of the poles as the loop gain varies, thereby revealing how close to oscillation the system may come. Such a plot is called a "root locus."

We now study a feedback system incorporating a one-pole feedforward amplifier. Assuming $H(s) = A_0/(1 + s/\omega_0)$, we have from (10.1),

$$\frac{Y}{X}(s) = \frac{\dfrac{A_0}{1+\beta A_0}}{1 + \dfrac{s}{\omega_0(1+\beta A_0)}}. \tag{10.4}$$

In order to analyze the stability behavior, we plot $|\beta H(s = j\omega)|$ and $\angle \beta H(s = j\omega)$ (Fig. 10.4), observing that a single pole cannot contribute a phase shift greater than 90° and the system is unconditionally stable for all non-negative values of β. Note that $\angle \beta H$ is independent of β.

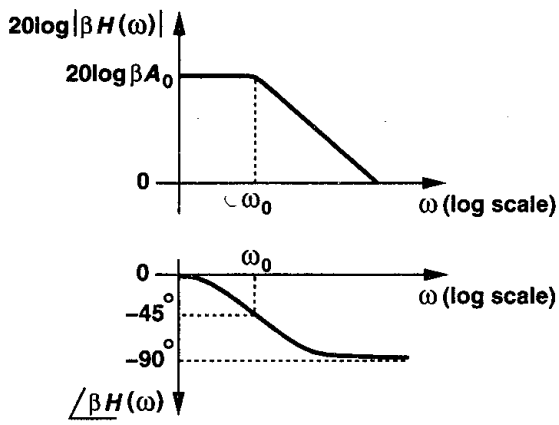


Figure 10.4 Bode plots of loop gain for a one-pole system.

Example 10.1

Construct the root locus for a one-pole system.

Solution

Equation (10.4) implies that the closed-loop system has a pole $s_p = -\omega_0(1 + \beta A_0)$, i.e., a real-valued pole in the left half plane that moves away from the origin as the loop gain increases (Fig. 10.5).

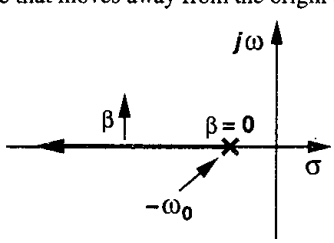


Figure 10.5

10.2 Multipole Systems

Our study of op amps in Chapter 9 indicates that such circuits generally contain multiple poles. In two-stage op amps, for example, each gain stage introduces a "dominant" pole. It is therefore important to study a feedback system whose core amplifier exhibits more than one pole.

Let us consider a two-pole system first. For stability considerations, we plot $|\beta H|$ and $\angle \beta H$ as a function of the frequency. Shown in Fig. 10.6, the magnitude begins to drop at 20 dB/dec at $\omega = \omega_{p1}$ and at 40 dB/dec at $\omega = \omega_{p2}$. Also, the phase begins to change at $\omega = 0.1\omega_{p1}$, reaches $-45°$ at $\omega = \omega_{p1}$ and $-90°$ at $\omega = 10\omega_{p1}$, begins to change again at $\omega = 0.1\omega_{p2}$ (if $0.1\omega_{p2} > 10\omega_{p1}$), reaches $-135°$ at $\omega = \omega_{p2}$, and asymptotically approaches $-180°$. The system is therefore stable because $|\beta H|$ drops to below unity at a frequency for which $\angle \beta H < -180°$.

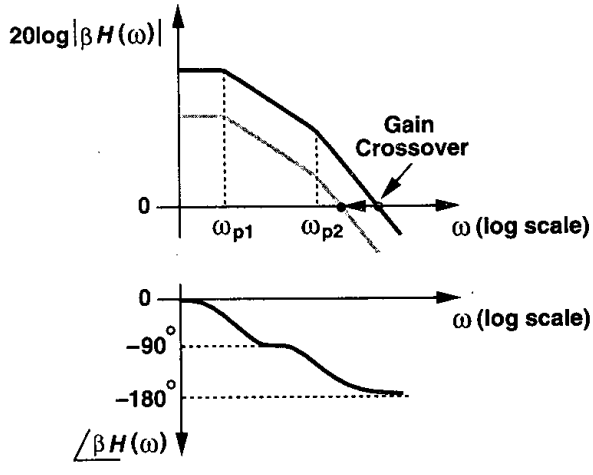


Figure 10.6 Bode plots of loop gain for a two-pole system.

What happens if the feedback is made "weaker?" To reduce the amount of feedback, we decrease β, obtaining the gray magnitude plot in Fig. 10.6. For a logarithmic vertical axis, a change in β translates the magnitude plot vertically. Note that the phase plot does not change. The key point is that as the feedback becomes weaker, the gain crossover point moves toward the origin while the phase crossover point remains constant, resulting in a more stable system. The stability is obtained at the cost of weaker feedback.

Example 10.2

Construct the root locus for a two-pole system.

Solution

Writing the open-loop transfer function as:

$$H(s) = \frac{A_0}{\left(1 + \frac{s}{\omega_{p1}}\right)\left(1 + \frac{s}{\omega_{p2}}\right)}, \tag{10.5}$$

we have

$$\frac{Y}{X}(s) = \frac{A_0}{\left(1+\frac{s}{\omega_{p1}}\right)\left(1+\frac{s}{\omega_{p2}}\right)+\beta A_0} \tag{10.6}$$

$$= \frac{A_0\omega_{p1}\omega_{p2}}{s^2+(\omega_{p1}+\omega_{p2})s+(1+\beta A_0)\omega_{p1}\omega_{p2}}. \tag{10.7}$$

Thus, the closed-loop poles are given by

$$s_{1,2} = \frac{-(\omega_{p1}+\omega_{p2}) \pm \sqrt{(\omega_{p1}+\omega_{p2})^2 - 4(1+\beta A_0)\omega_{p1}\omega_{p2}}}{2}. \tag{10.8}$$

As expected, for $\beta = 0$, $s_{1,2} = -\omega_{p1}, -\omega_{p2}$. As β increases, the term under the square root drops, taking on a value of zero for

$$\beta_1 = \frac{1}{A_0}\frac{(\omega_{p1}-\omega_{p2})^2}{4\omega_{p1}\omega_{p2}}. \tag{10.9}$$

As shown in Fig. 10.7, the poles begin at $-\omega_{p1}$ and $-\omega_{p2}$, move toward each other, coincide for $\beta = \beta_1$, and become complex for $\beta > \beta_1$.

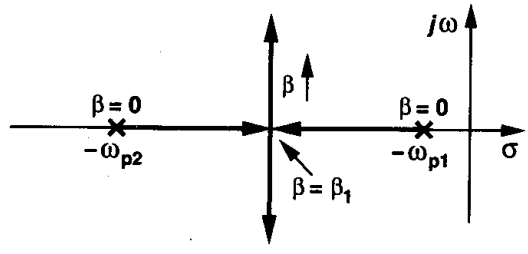


Figure 10.7

The foregoing calculations point to the complexity of the algebra required to construct a root locus for higher-order systems. For this reason, many root locus techniques have been devised so as to minimize such computations [1].

We now study a three-pole system. Shown in Fig. 10.8 are the Bode plots of the magnitude and phase of the loop gain. The third pole gives rise to additional phase shift, possibly moving the phase crossover to frequencies lower than the gain crossover and leading to oscillation.

Since the third pole also decreases the *magnitude* of the loop gain at a greater rate, the reader may wonder why the gain crossover does not move as much as the phase crossover does. As mentioned before, the phase begins to change at approximately one-tenth of the pole frequency whereas the magnitude begins to drop only near the pole frequency. For this reason, additional poles (and zeros) impact the phase to a much greater extent than they do the magnitude.

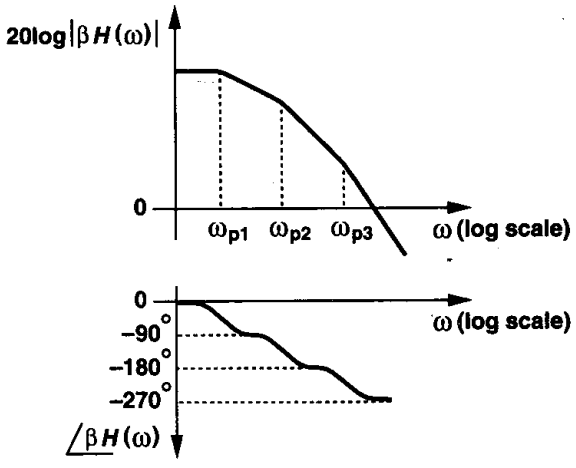


Figure 10.8 Bode plots of loop gain for a three-pole system.

As with a two-pole system, if the feedback factor in Fig. 10.8 decreases, the circuit becomes more stable because the gain crossover moves toward the origin while the phase crossover remains constant.

10.3 Phase Margin

Our foregoing study indicates that to ensure stability, $|\beta H|$ must drop to unity before $\angle \beta H$ crosses $-180°$. We may naturally ask: how far should PX be from GX? Let us first consider a "marginal" case where, as depicted in Fig. 10.9(a), GX is only slightly below PX; sharp peak for example, at GX the phase equals $-175°$. How does the closed-loop system respond in this case? Noting that at GX, $\beta H(j\omega_1) = 1 \times \exp(-j175°)$, we have

$$\frac{Y}{X}(j\omega_1) = \frac{H(j\omega_1)}{1 + \beta H(j\omega_1)} \tag{10.10}$$

$$= \frac{\frac{1}{\beta}\exp(-j175°)}{1 + \exp(-j175°)} \tag{10.11}$$

$$= \frac{1}{\beta} \cdot \frac{-0.9962 - j0.0872}{0.0038 - j0.0872}, \tag{10.12}$$

and hence

$$\left|\frac{Y}{X}(j\omega_1)\right| = \frac{1}{\beta} \cdot \frac{1}{0.0872} \tag{10.13}$$

$$\approx \frac{11.5}{\beta}. \tag{10.14}$$

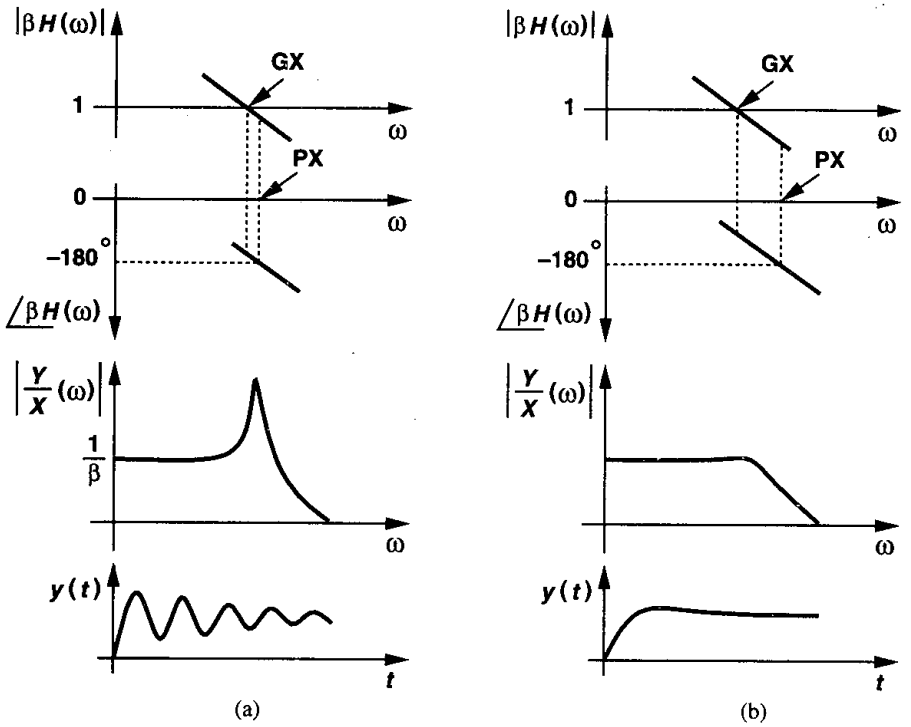


Figure 10.9 Closed-loop frequency and time response for (a) small and (b) large margin between gain and phase crossover points.

Since at low frequencies, $|Y/X| \approx 1/\beta$, the closed-loop frequency response exhibits a sharp peak in the vicinity of $\omega = \omega_1$. In other words, the closed-loop system is near oscillation and its step response exhibits a very underdamped behavior. This point also reveals that a second-order system may suffer from ringing although it is stable.

Now suppose, as shown in Fig. 10.9(b), GX precedes PX by a greater margin. Then, we expect a relatively "well-behaved" closed-loop response in both the frequency domain and the time domain. It is therefore plausible to conclude that the greater the spacing between GX and PX (while GX remains below PX), the more stable the feedback system. Alternatively, the phase of βH at the gain crossover frequency can serve as a measure of stability: the smaller $|\angle \beta H|$ at this point, the more stable the system.

This observation leads us to the concept of "phase margin" (PM), defined as $PM = 180° + \angle \beta H(\omega = \omega_1)$, where ω_1 is the gain crossover frequency.

Example 10.3

A two-pole feedback system is designed such that $|\beta H(\omega_{p2})| = 1$ and $|\omega_{p1}| \ll |\omega_{p2}|$ (Fig. 10.10). What is the phase margin?

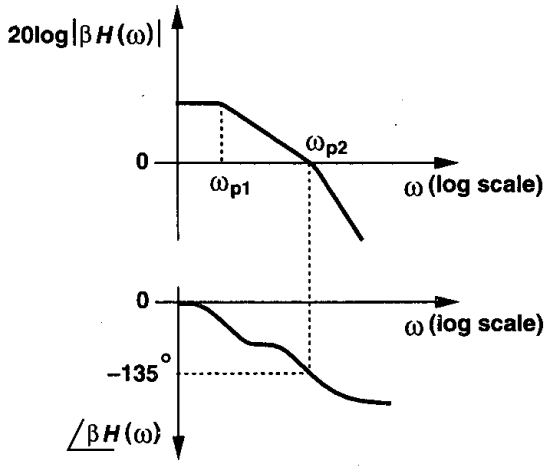


Figure 10.10

Solution

Since $\angle \beta H$ reaches $-135°$ at $\omega = \omega_{p2}$, the phase margin is equal to 45°.

How much phase margin is adequate? It is instructive to examine the closed-loop frequency response for different phase margins [1]. For $PM = 45°$, at the gain crossover frequency $\angle \beta H(\omega_1) = -135°$ and $|\beta H(\omega_1)| = 1$ (Fig. 10.11), yielding

$$\frac{Y}{X} = \frac{H(j\omega_1)}{1 + 1 \times \exp(-j135°)} \tag{10.15}$$

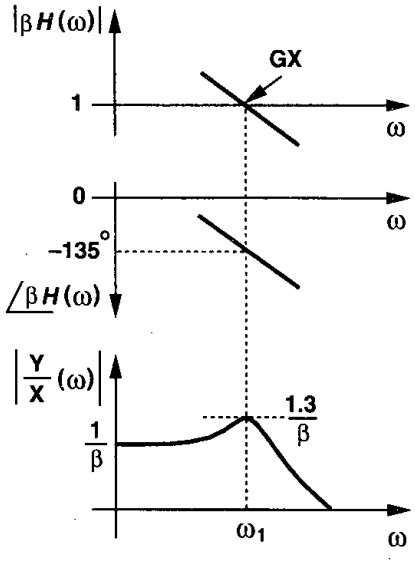


Figure 10.11 Closed-loop frequency response for 45° phase margin.

$$= \frac{H(j\omega_1)}{0.29 - 0.71j}. \tag{10.16}$$

It follows that

$$\left|\frac{Y}{X}\right| = \frac{1}{\beta} \cdot \frac{1}{|0.29 - 0.71j|} \tag{10.17}$$

$$\approx \frac{1.3}{\beta}. \tag{10.18}$$

Consequently, the frequency response of the feedback system suffers from a 30% peak at $\omega = \omega_1$.

It can be shown that for $PM = 60°$, $Y(j\omega_1)/X(j\omega_1) = 1/\beta$, suggesting a negligible frequency peaking. This typically means that the step response of the feedback system exhibits little ringing, providing a fast settling. For greater phase margins, the system is more stable but the time response slows down (Fig. 10.12). Thus, $PM = 60°$ is typically considered the optimum value.

The concept of phase margin is well-suited to the design of circuits that process *small* signals. In practice, the large-signal step response of feedback amplifiers does not follow the illustration of Fig. 10.12. This is not only due to slewing but also because of the nonlinear behavior resulting from large excursions in the bias voltages and currents of the amplifier. Such excursions in fact cause the pole and zero frequencies to *vary* during the transient, leading to a complicated time response. Thus, for large-signal applications, time-domain simulations of the closed-loop system prove more relevant and useful than small-signal ac computations of the open-loop amplifier.

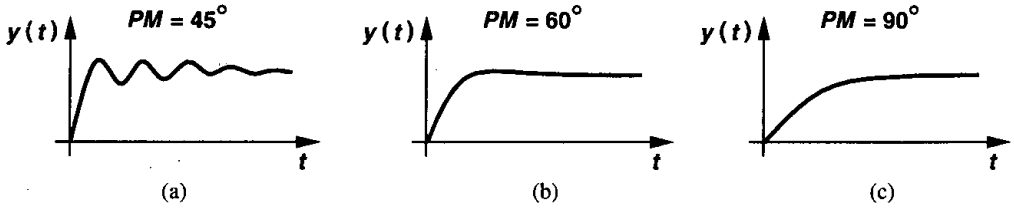


Figure 10.12 Closed-loop time response for 45°, 60°, and 90° phase margins.

As an example of a feedback circuit exhibiting a reasonable phase margin but poor settling behavior, consider the unity-gain amplifier of Fig. 10.13, where the aspect ratio of all transistors is equal to 50 μm / 0.6 μm. With the choice of the device dimensions, bias currents, and capacitor values shown here, SPICE yields a phase margin of approximately 65° and a unity-gain frequency of 150 MHz. The large-signal step response, however, suffers from significant ringing.

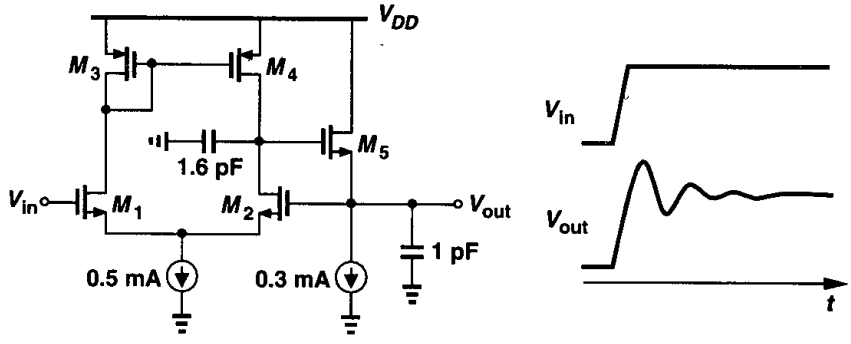


Figure 10.13 Unity-gain buffer.

10.4 Frequency Compensation

Typical op amp circuits contain many poles. In a folded-cascode topology, for example, both the folding node and the output node contribute poles. For this reason, op amps must usually be "compensated," that is, their open-loop transfer function must be modified such that the closed-loop circuit is stable and the time response is well-behaved.

The need for compensation arises because $|\beta H|$ does not drop to unity well before $\angle \beta H$ reaches $-180°$. We then postulate that stability can be achieved by (1) minimizing the overall phase shift, thus pushing the phase crossover *out* [Fig. 10.14(a)]; or (2) dropping the gain, thereby pushing the gain crossover *in* [Fig. 10.14(b)]. The first approach requires that we attempt to minimize the number of poles in the signal path by proper design. Since each additional stage contributes at least one pole, this means the number of stages must be

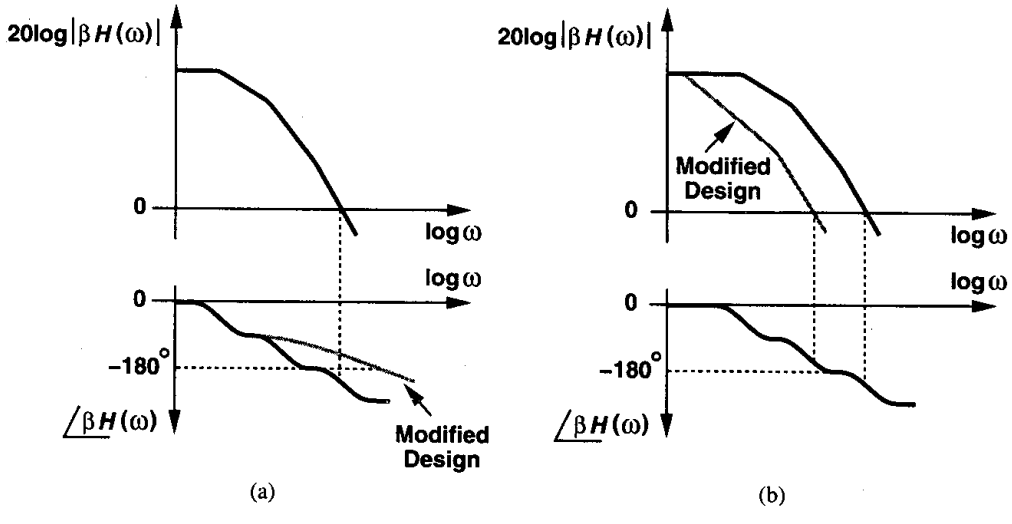


Figure 10.14 Frequency compensation by (a) moving PX out, (b) pushing GX in.

minimized, a remedy that yields low voltage gain and/or limited output swings (Chapter 9). The second approach, on the other hand, retains the low-frequency gain and the output swings but it reduces the bandwidth by forcing the gain to fall at lower frequencies.

In practice, we first try to design an op amp so as to minimize the number of poles while meeting other requirements. Since the resulting circuit may still suffer from insufficient phase margin, we then compensate the op amp, i.e., modify the design so as to move the gain crossover toward the origin.

Let us apply the above procedures to various op amp topologies. We begin with the telescopic cascode op amp shown in Fig. 10.15, where a PMOS current mirror performs differential to single-ended conversion. We identify a number of poles in the signal paths: path 1 contains a high-frequency pole at the source of M_3, a mirror pole at node A, and another high-frequency pole at the source of M_7, whereas path 2 contains a high-frequency pole at the source of M_4. The two paths share a pole at the output.

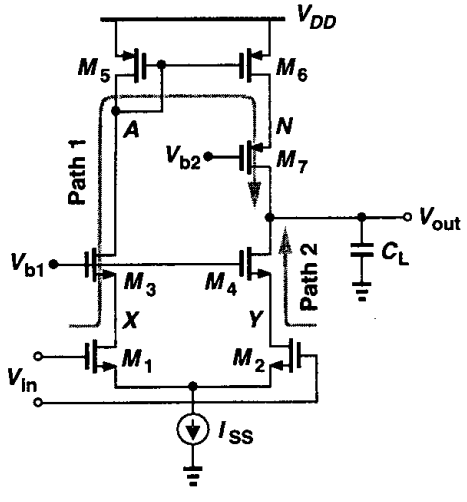


Figure 10.15 Telescopic op amp with single-ended output.

It is instructive to estimate the relative position of these poles. Since the output resistance of the op amp is much higher than the small-signal resistances seen at the other nodes in the circuit, we expect that, even with a moderate load capacitance, the output pole, $\omega_{p,out}$, is the closest to the origin. Called the "dominant pole," $\omega_{p,out}$ usually sets the open-loop 3-dB bandwidth.

We also surmise that the first "nondominant pole," i.e., the closest pole to the origin after the dominant pole, arises at node A. This is because the total capacitance at this node, roughly equal to $C_{GS5} + C_{GS6} + C_{DB5} + 2C_{GD6} + C_{DB3} + C_{GD3}$, is typically quite larger than that at nodes X, Y, and N and the small-signal resistance of M_5, approximately $1/g_{m5}$, is relatively large.

Which node yields the next nondominant pole: N or X (and Y)? Recall from Chapter 9 that, to obtain a low overdrive and consume a reasonable voltage headroom, the PMOS devices in the op amp are typically quite wider than the NMOS transistors. Comparing M_4 and M_7 and neglecting body effect, we note that since $g_m = 2I_D/|V_{GS} - V_{TH}|$, if

the two transistors are designed to have the same overdrive, they also exhibit the same transconductance. However, from square-law characteristics, we have $W_4/W_7 = \mu_p/\mu_n$, which is about 1/3 in today's technologies. Thus, nodes N and X (or Y) see roughly equal small-signal resistances to ground but node N suffers from much more capacitance. It is therefore plausible to assume that node N contributes the next nondominant pole. Figure 10.16 illustrates the results, denoting the capacitance at nodes A, N, and X by C_A, C_N, and C_X, respectively. The poles at nodes X and Y are nearly equal and their

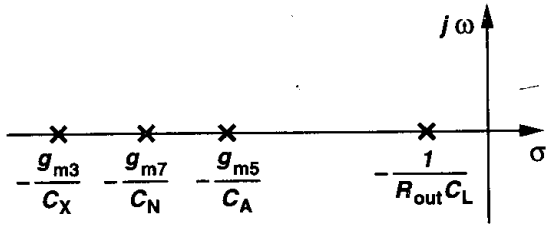


Figure 10.16 Pole locations for the op amp of Fig. 10.15.

corresponding terms in the transfer functions of path 1 and path 2 can be factored out. Thus, they count as one pole rather than two.

With the position of the poles roughly determined, we can construct the magnitude and phase plots for βH, using $\beta = 1$ for the worst case. Shown in Fig. 10.17, such characteristics indicate that the mirror pole typically limits the phase margin because its phase contribution occurs at lower frequencies than that of other nondominant poles.

Recall from Chapter 6 that differential pairs using active current mirrors exhibit a zero located at twice the mirror pole frequency. The circuit of Fig. 10.15 contains such a zero as well. Located at $2\omega_{p,A}$, the zero has some effect on the magnitude and phase characteristics. The analysis is left to the reader.

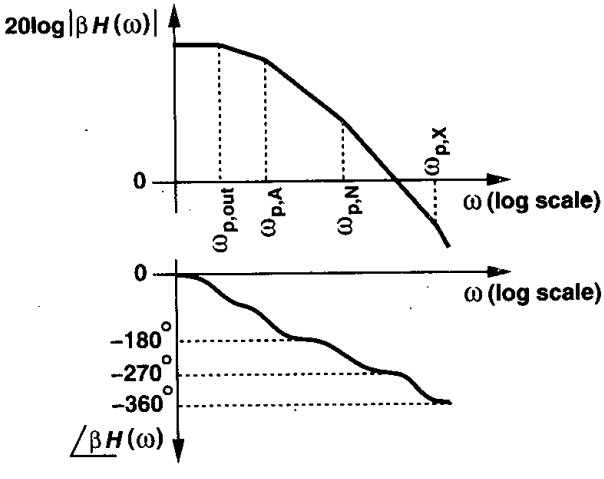


Figure 10.17 Bode plots of loop gain for op amp of Fig. 10.15.

How should we compensate the op amp? Let us assume that the number and location of the nondominant poles and hence the phase plot at frequencies higher than roughly $10\omega_{p,out}$ remain constant. Thus, we must force the loop gain to drop such that the gain crossover point moves toward the origin. To accomplish this, we simply lower the frequency of the dominant pole by increasing the load capacitance. The key point is that the phase contribution of the dominant pole in the vicinity of the gain or phase crossover points is close to 90° and relatively independent of the location of the pole. That is, as illustrated in Fig. 10.18, translating the dominant pole toward the origin affects the magnitude plot but not the critical part of the phase plot.

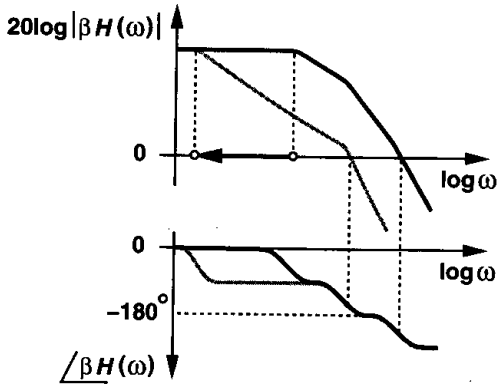


Figure 10.18 Translating the dominant pole toward origin.

In order to understand how much the dominant pole must be shifted down as well as arrive at an important conclusion, let us assume (1) the second nondominant pole ($\omega_{p,N}$) in Fig. 10.15 is quite higher than the mirror pole so that the phase shift at $\omega = \omega_{p,A}$ is equal to $-135°$ and (2) a phase margin of 45° (which is usually inadequate) is necessary. To compensate the circuit, we first identify the frequency at which the phase plot yields the required phase margin, in this case, $\omega_{p,A}$. Since the dominant pole must drop the gain to unity at $\omega_{p,A}$ with a slope of 20 dB/dec, we draw a straight line from $\omega_{p,A}$ toward the origin with such a slope (Fig. 10.19), thus obtaining the new magnitude of the dominant pole, $\omega'_{p,out}$. Therefore, the load capacitance must be increased by a factor of $\omega_{p,out}/\omega'_{p,out}$.

From the new magnitude plot, we note that the unity-gain bandwidth of the compensated op amp is equal to the *frequency of the first nondominant pole* (of course with a phase margin of 45°). This is a fundamental result, indicating that to achieve a wideband in a feedback system employing an op amp, the first nondominant pole must be as far as possible. For this reason, the mirror pole proves undesirable.

We should also mention that although $\omega_{p,out} = (R_{out}C_L)^{-1}$, increasing R_{out} does *not* compensate the op amp. As shown in Fig. 10.20, a higher R_{out} results in a greater gain, only affecting the low-frequency portion of the characteristics. Also, moving one of the nondominant poles toward the origin does not improve the phase margin. (Why?)

Now consider the fully differential telescopic cascode depicted in Fig. 10.21. In addition to achieving various useful properties of differential operation, this topology avoids the mirror pole, thereby exhibiting stable behavior for a greater bandwidth. In fact, we can identify one dominant pole at each output node and only *one* nondominant pole arising from node X (or Y). This suggests that fully differential telescopic cascode circuits are quite stable.

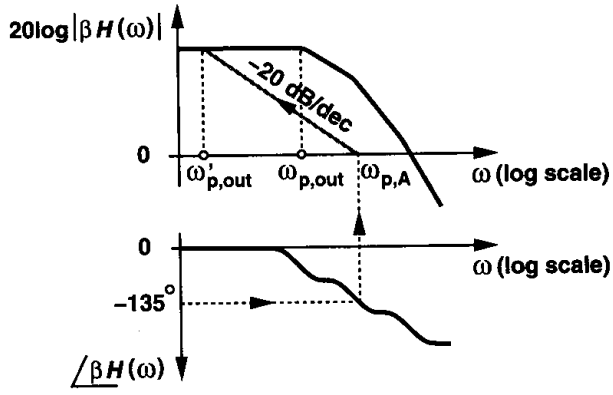


Figure 10.19 Translating the dominant pole toward the origin for 45° phase margin.

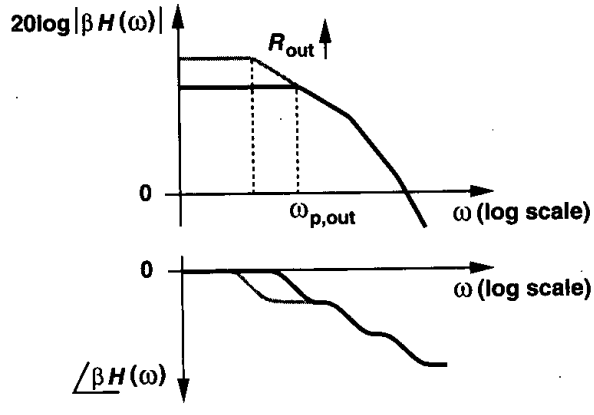


Figure 10.20 Bode plots of loop gain for higher output resistance.

But how about the pole at node N (or K) in Fig. 10.21? Considering one of the PMOS cascodes as shown in Fig. 10.22(a), we may think that the capacitance at node N, $C_N = C_{GS5} + C_{SB5} + C_{GD7} + C_{DB7}$, shunts the output resistance of M_7 at high frequencies, thereby dropping the output impedance of the cascode. To quantify this effect, we first determine Z_{out} in Fig. 10.22(a):

$$Z_{out} = (1 + g_{m5}r_{O5})Z_N + r_{O5}, \tag{10.19}$$

where body effect is neglected and $Z_N = r_{O7}||(C_N s)^{-1}$. Assuming the first term is much greater than the second, we have

$$Z_{out} \approx (1 + g_{m5}r_{O5})\frac{r_{O7}}{r_{O7}C_N s + 1}. \tag{10.20}$$

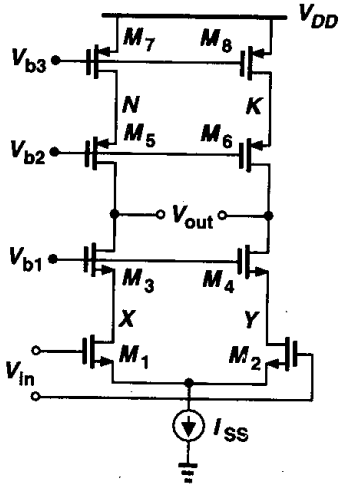


Figure 10.21 Fully differential telescopic op amp.

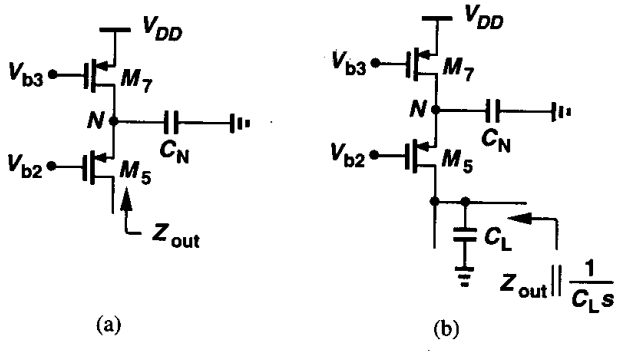


Figure 10.22 Effect of device capacitance at internal node of a cascode current source.

Now, as illustrated in Fig. 10.22(b), we take the output load capacitance into account:

$$Z_{out} \| \frac{1}{C_L s} = \frac{(1 + g_{m5} r_{O5}) \dfrac{r_{O7}}{r_{O7} C_N s + 1} \cdot \dfrac{1}{C_L s}}{(1 + g_{m5} r_{O5}) \dfrac{r_{O7}}{r_{O7} C_N s + 1} + \dfrac{1}{C_L s}} \tag{10.21}$$

$$= \frac{(1 + g_{m5} r_{O5}) r_{O7}}{[(1 + g_{m5} r_{O5}) r_{O7} C_L + r_{O7} C_N] s + 1}. \tag{10.22}$$

Thus, the parallel combination of Z_{out} and the load capacitance still contains a single pole corresponding to a time constant $(1 + g_{m5} r_{O5}) r_{O7} C_L + r_{O7} C_N$. Note that $(1 + g_{m5} r_{O5}) r_{O7} C_L$

is simply due to the low-frequency output resistance of the cascode. In other words, the overall time constant equals the "output" time constant plus $r_{O7}C_N$. The key point in this calculation is that the pole in the PMOS cascode is *merged* with the output pole, thus creating no *additional* pole. It merely lowers the dominant pole by a slight amount. For this reason, we loosely say that the signal does not "see" the pole in the cascode current sources.[2]

Comparison of the circuits shown in Figs. 10.15 and 10.21 now reveals that the fully differential configuration avoids both the mirror pole *and* the pole at node N. With the approximation made in (10.22), the circuit of Fig. 10.21 contains only one nondominant pole located at relatively high frequencies owing to the high transconductance of the NMOS transistors. This is a remarkable advantage of fully differential cascode op amps.

We have thus far observed that nondominant poles give rise to instability, requiring frequency compensation. It is possible to cancel one or more of these poles by introducing *zeros* in the transfer function? For example, following the analysis of Fig. 6.31, we surmise that if a low-gain but fast path is placed in parallel with the main amplifier, a zero is created that can be positioned atop the first nondominant pole. However, cancellation of a pole by a zero in the presence of mismatches leads to long settling components in the step response of the closed-loop circuit. This effect is studied in Problem 10.19.

10.5 Compensation of Two-Stage Op Amps

Our study of op amps in Chapter 9 indicates that two-stage topologies may prove inevitable if the output voltage swing must be maximized. Thus, the stability and compensation of such op amps is of interest.

Consider the circuit shown in Fig. 10.23. We identify three poles: a pole at X (or Y), another at E (or F), and a third at A (or B). From our foregoing discussions, we know that the pole at X lies at relatively high frequencies. But how about the other two? Since the small-signal resistance seen at E is quite high, even the capacitances of M_3, M_5, and M_9 can create a pole relatively close to the origin. At node A, the small-signal resistance is lower but the value of C_L may be quite high. Consequently, we say the circuit exhibits *two* dominant poles.

From these observations, we can construct the magnitude and phase plots shown in Fig. 10.24. Here, $\omega_{p,E}$ is assumed more dominant but the relative position of $\omega_{p,E}$ and $\omega_{p,A}$ depends on the design and the load capacitance. Note that, since the poles at E and A are relatively close to the origin, the phase approaches $-180°$ well below the third pole. In other words, the phase margin may be quite close to zero even before the third pole contributes significant phase shift.

Let us now investigate the frequency compensation of two-stage op amps. In Fig. 10.24, one of the dominant poles must be moved toward the origin so as to place the gain crossover well below the phase crossover. However, recall from Section 10.4 that the unity-gain bandwidth after compensation cannot exceed the frequency of the second pole of the open-loop system. Thus, if in Fig. 10.24 the magnitude of $\omega_{p,E}$ is to be reduced, the available

[2] If the second term in Eq. (10.19) is included in subsequent derivations, a pole and a zero that are nearly equal appear in the overall output impedance. Nonetheless, for $g_m r_O \gg 1$ and $C_L > C_N$, their effect is negligible.

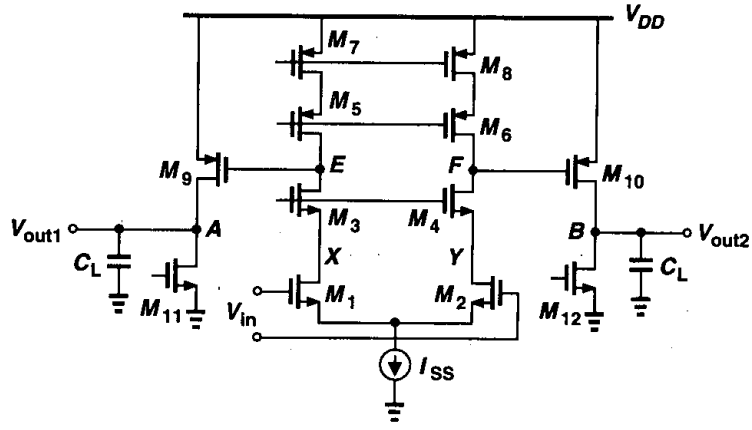


Figure 10.23 Two-stage op amp.

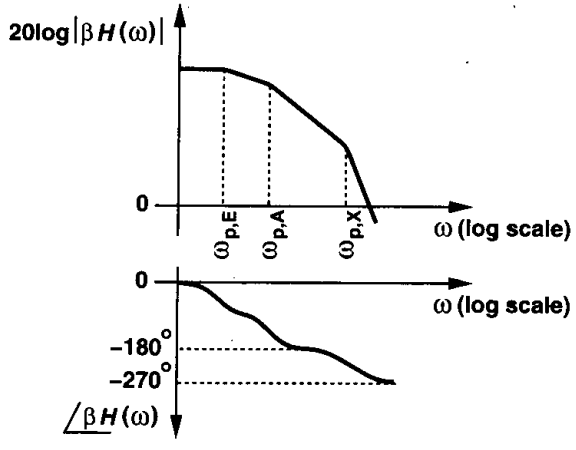


Figure 10.24 Bode plots of loop gain of two-stage op amp.

bandwidth is limited to approximately $\omega_{p,A}$, a low value. Furthermore, the very small magnitude of the required dominant pole translates to a very large compensation capacitor.

Fortunately, a more efficient method of compensation can be applied to the circuit of Fig. 10.23. To arrive at this method, we note that, as illustrated in Fig. 10.25(a), the first stage exhibits a high output impedance and the second stage provides a moderate gain, thereby providing a suitable environment for Miller multiplication of capacitors. Shown in Fig. 10.25(b), the idea is to create a large capacitance at node E, equal to $(1+A_{v2})C_C$, moving the corresponding pole to $R_{out1}^{-1}[C_E + (1 + A_{v2})C_C]^{-1}$, where C_E denotes the capacitance at node E before C_C is added. As a result, a low-frequency pole can be established with a moderate capacitor value, saving considerable chip area. This technique is called "Miller compensation."

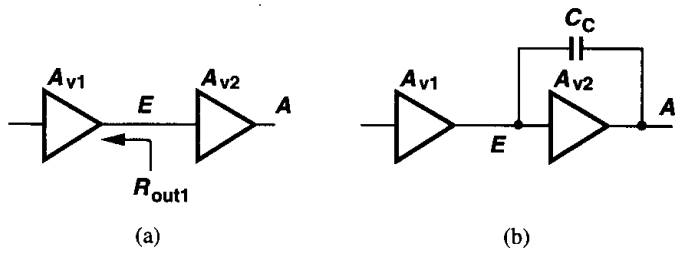


Figure 10.25 Miller compensation of a two-stage op amp.

In addition to lowering the required capacitor value, Miller compensation entails a very important property: it moves the *output* pole *away* from the origin. Illustrated in Fig. 10.26, this effect is called "pole splitting." To understand the underlying principle, we simplify the output stage of Fig. 10.23 as in Fig. 10.27, where R_S denotes the output resistance of the first stage and $R_L = r_{O9} || r_{O11}$. From our analysis in Chapter 6, we note that this circuit contains two poles:

$$\omega_{p1} \approx \frac{1}{R_S[(1 + g_{m9}R_L)(C_C + C_{GD9}) + C_E] + R_L(C_C + C_{GD9} + C_L)} \tag{10.23}$$

$$\omega_{p2} \approx \frac{R_S[(1 + g_{m9}R_L)(C_C + C_{GD9}) + C_E] + R_L(C_C + C_{GD9} + C_L)}{R_S R_L[(C_C + C_{GD9})C_E + (C_C + C_{GD9})C_L + C_E C_L)]}. \tag{10.24}$$

These expressions are based on the assumption $|\omega_{p1}| \ll |\omega_{p2}|$. Before compensation, however, ω_{p1} and ω_{p2} are of the same order of magnitude. For $C_C = 0$ and relatively large C_L, we may approximate the magnitude of the output pole as $\omega_{p2} \approx 1/(R_L C_L)$.

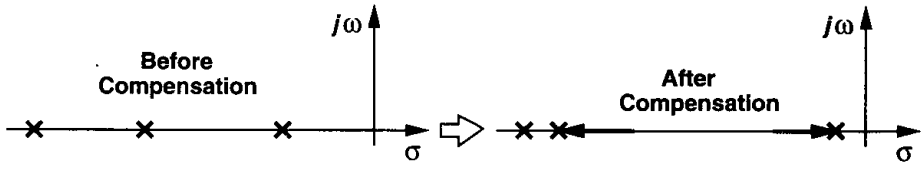


Figure 10.26 Pole splitting as a result of Miller compensation.

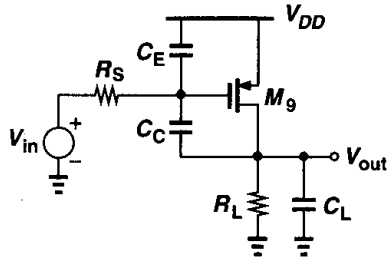


Figure 10.27 Simplified circuit of a two-stage op amp.

To compare the magnitudes of ω_{p2} before and after compensation, we consider a typical case: $C_C + C_{GD9} \gg C_E$, reducing (10.24) to $\omega_{p2} \approx g_{m9}/(C_E + C_L)$. Noting that typically $C_E \ll C_L$, we conclude that Miller compensation increases the magnitude of the output pole by roughly a factor of $g_{m9}R_L$, a relatively large value. Intuitively, this is because at high frequencies, C_C provides a low impedance between the gate and drain of M_9, reducing the resistance seen by C_L from R_L to roughly $R_S||g_{m9}^{-1}||R_L \approx g_{m9}^{-1}$.

In summary, Miller compensation moves the interstage pole toward the origin and the output pole away from the origin, allowing a much greater bandwidth than that obtained by merely connecting the compensation capacitor from one node to ground. In practice, the choice of the compensation capacitor for proper phase margin requires some iteration.

Our study of stability and compensation has thus far neglected the effect of *zeros* of the transfer function. While in cascode topologies, the zeros are quite far from the origin, in two-stage op amps incorporating Miller compensation, a nearby zero appears in the circuit. Recall from Chapter 6 that the circuit of Fig. 10.27 contains a right-half-plane zero at $\omega_z = g_{m9}/(C_C + C_{GD9})$. This is because $C_C + C_{GD9}$ forms a "parasitic" signal path from the input to the output. What is the effect of such a zero? The numerator of the transfer function reads $(1 - s/\omega_z)$, yielding a phase of $-\tan^{-1}(\omega/\omega_z)$, a negative value because ω_z is positive. In other words, as with poles in the left half plane, a zero in the right half plane contributes more phase shift, thus moving the phase crossover toward the origin. Furthermore, from Bode approximations, the zero slows down the drop of the magnitude, thereby pushing the gain crossover away from the origin. As a result, the stability degrades considerably.

To better understand the foregoing discussion, let us construct the Bode plots for a third-order system containing a dominant pole ω_{p1}, two nondominant poles ω_{p2} and ω_{p3}, and a zero in the right half plane ω_z. For two-stage op amps, typically $|\omega_{p1}| < |\omega_z| < |\omega_{p2}|$. As shown in Fig. 10.28, the zero introduces significant phase shift while preventing the gain from falling sufficiently.

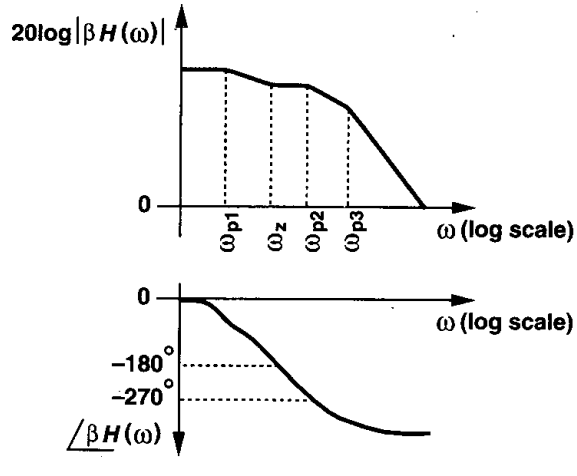


Figure 10.28 Effect of right half plane zero.

The right half plane zero in two-stage CMOS op amps, given by $g_m/(C_C + C_{GD})$, is a serious issue because g_m is relatively small and C_C is chosen large enough to position the dominant pole properly. Various techniques of eliminating or moving the zero have been invented. Illustrated in Fig. 10.29, one approach places a resistor in series with the com-

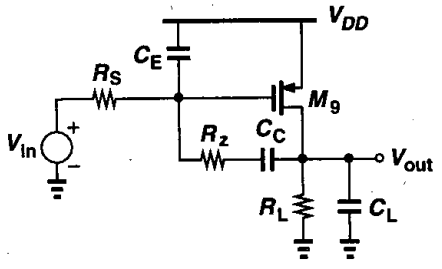


Figure 10.29 Addition of R_z to move the right half plane zero.

pensation capacitor, thereby modifying the zero frequency. The output stage now exhibits *three* poles, but for moderate values of R_z the third pole is located at high frequencies and the first two poles are close to the values calculated with $R_z = 0$. Moreover, it can be shown (Problem 10.8) that the zero frequency is given by

$$\omega_z \approx \frac{1}{C_C\left(g_{m9}^{-1} - R_z\right)}. \tag{10.25}$$

Thus, if $R_z \geq g_{m9}^{-1}$, then $\omega_z \leq 0$. While $R_z = g_{m9}^{-1}$ seems a natural choice, in practice we may even move the zero well into the left half plane so as to cancel the first nondominant pole. This occurs if

$$\frac{1}{C_C\left(g_{m9}^{-1} - R_z\right)} = \frac{-g_{m9}}{C_L + C_E}, \tag{10.26}$$

that is,

$$R_z = \frac{C_L + C_E + C_C}{g_{m9}C_C} \tag{10.27}$$

$$\approx \frac{C_L + C_C}{g_{m9}C_C}, \tag{10.28}$$

because C_E is typically much less than $C_L + C_C$.

The possibility of canceling the nondominant pole makes this technique quite attractive, but in reality two important drawbacks must be considered. First, it is difficult to guarantee the relationship given by (10.28), especially if C_L is unknown or variable. For example, as explained in Chapter 12, the load capacitance seen by an op amp may vary from one part of the period to another in switched-capacitor circuits, necessitating a corresponding change in R_z and complicating the design. The second drawback relates to the actual implementation of R_z. Typically realized by a MOS transistor in the triode region (Fig. 10.30), R_z changes

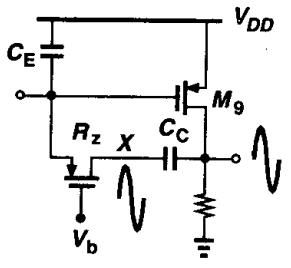


Figure 10.30 Effect of large output swings on R_z.

substantially as output voltage excursions are coupled through C_C to node X, thereby degrading the large-signal settling response.

Generating V_b in Fig. 10.30 is not straightforward because R_Z must remain equal to $(1 + C_L/C_C)g_{m9}^{-1}$ despite process and temperature variations. A common approach is illustrated in Fig. 10.31 [2], where diode-connected devices M_{13} and M_{14} are placed in series. If I_1 is chosen with respect to I_{D9} such that $V_{GS13} = V_{GS9}$, then $V_{GS15} = V_{GS14}$. Since $g_{m14} = \mu_p C_{ox}(W/L)_{14}(V_{GS14} - V_{TH14})$ and $R_{on15} = [\mu_p C_{ox}(W/L)_{15}(V_{GS15} - V_{TH15})]^{-1}$, we have $R_{on15} = g_{m14}^{-1}(W/L)_{14}/(W/L)_{15}$. For pole-zero cancellation to occur,

$$g_{m14}^{-1}\frac{(W/L)_{14}}{(W/L)_{15}} = g_{m9}^{-1}\left(1 + \frac{C_L}{C_C}\right), \tag{10.29}$$

and hence

$$(W/L)_{15} = \sqrt{(W/L)_{14}(W/L)_9}\sqrt{\frac{I_{D9}}{I_{D14}}}\frac{C_C}{C_C + C_L}. \tag{10.30}$$

If C_L is constant, (10.30) can be established with reasonable accuracy because it contains only the *ratio* of quantities.

Another method of guaranteeing Eq. (10.28) is to use a simple resistor for R_Z and define g_{m9} with respect to a resistor that closely matches R_Z [3]. Depicted in Fig. 10.32, this technique incorporates M_{b1}-M_{b4} along with R_S to generate $I_b \propto R_S^{-2}$. (This circuit is studied in detail in Chapter 11.) Thus, $g_{m9} \propto \sqrt{I_{D9}} \propto \sqrt{I_{D11}} \propto R_S^{-1}$. Proper ratioing of R_Z and R_S therefore ensures (10.28) is valid even with temperature and process variations.

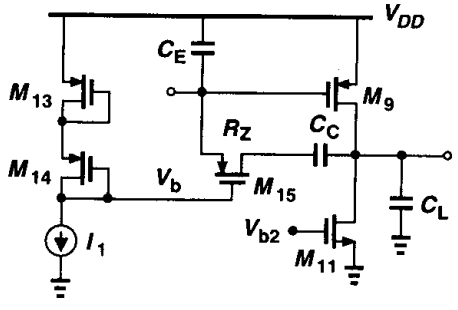


Figure 10.31 Generation of V_b for proper temperature and process tracking.

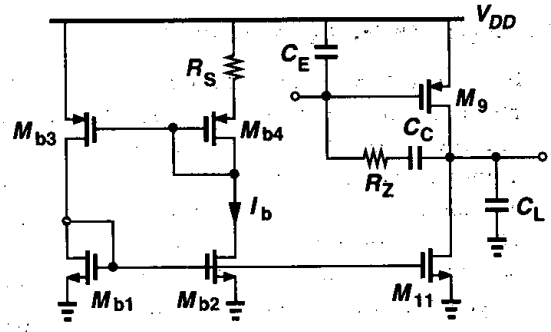


Figure 10.32 Method of defining g_{m9} with respect to R_S.

The principal drawback of the two methods described above is that they assume square-law characteristics for all of the transistors. As described in Chapter 16, short-channel MOSFETs may substantially deviate from the square-law regime, creating errors in the foregoing calculations. In particular, transistor M_9 is typically a short-channel device because it appears in the signal path and its raw speed is critical.

An attribute of two-stage op amps that makes them inferior to "one-stage" op amps is the susceptibility to the load capacitance. Since Miller compensation establishes the dominant pole at the output of the first stage, a higher load capacitance presented to the second stage moves the second pole toward the origin, degrading the phase margin. By contrast, in one-stage op amps, a higher load capacitance brings the *dominant* pole closer to the origin, *improving* the phase margin (albeit making the feedback system more overdamped). Illustrated in Fig. 10.33 is the step response of a unity-gain feedback amplifier employing a one-stage or a two-stage op amp, suggesting that the response approaches an oscillatory behavior if the load capacitance seen by the two-stage op amp increases.

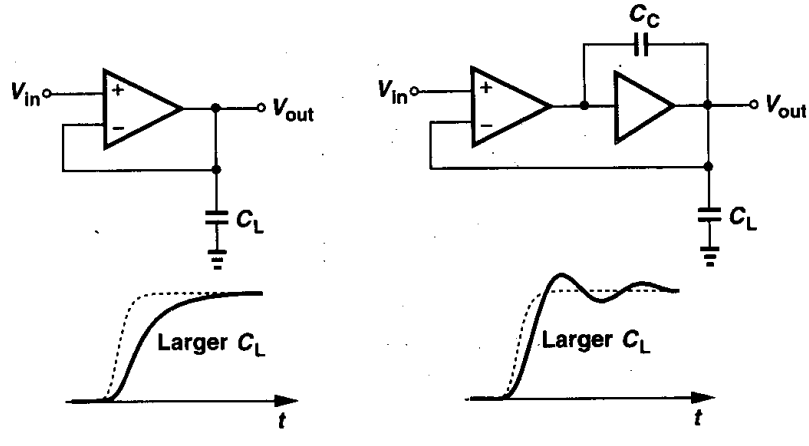


Figure 10.33 Effect of increased load capacitance on step response of one- and two-stage op amps.

10.5.1 Slewing in Two-Stage Op Amps

It is instructive to study the slewing characteristics of two-stage op amps. Suppose in Fig. 10.34(a) V_{in} experiences a large positive step at $t = 0$, turning off M_2, M_4, and M_3. The circuit can then be simplified to that in Fig. 10.34(b), revealing that C_C is charged by a constant current I_{SS} if parasitic capacitances at node X are negligible. Recognizing that the gain of the output stage makes node X a virtual ground, we write: $V_{out} \approx I_{SS}t/C_C$. Thus, the positive slew rate[3] equals I_{SS}/C_C. Note that during slewing, M_5 must provide *two* currents: I_{SS} and I_1. If M_5 is not wide enough to sustain $I_{SS} + I_1$ in saturation, then V_X drops significantly, possibly driving M_1 into the triode region.

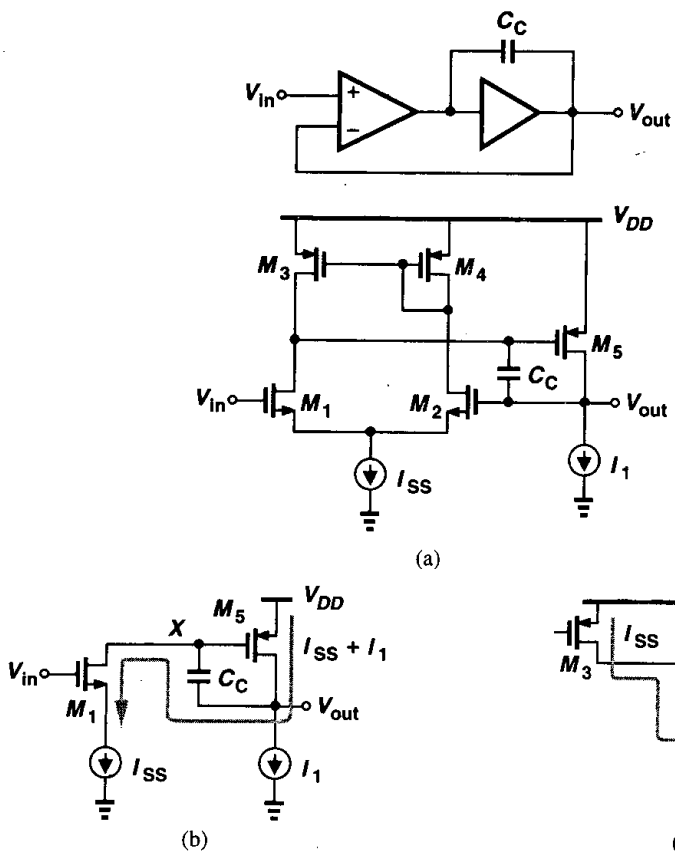


Figure 10.34 (a) Simple two-stage op amp, (b) simplified circuit during positive slewing, (c) simplified circuit during negative slewing.

[3]The term positive refers to the slope of the waveform at the output of the op amp.

For the negative slew rate, we simplify the circuit as shown in Fig. 10.34(c). Here I_1 must support both I_{SS} and I_{D5}. For example, if $I_1 = I_{SS}$, then V_X rises so as to turn off M_5. If $I_1 < I_{SS}$, then M_3 enters the triode region and the slew rate is given by I_{D3}/C_C.

10.6 Other Compensation Techniques

The difficulty in compensating two-stage CMOS op amps arises from the feedforward path formed by the compensation capacitor [Fig. 10.35(a)]. If C_C could conduct current from the output node to node X but not vice versa, then the zero would move to a very high

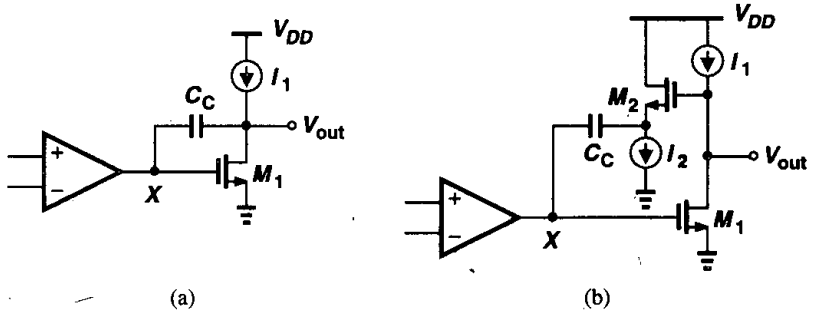


Figure 10.35 (a) Two-stage op amp with right half plane zero due to C_C, (b) addition of a source follower to remove the zero.

frequency. As shown in Fig. 10.35(b), this can be accomplished by inserting a source follower in series with the capacitor. Since the gate-source capacitance of M_2 is typically much less than C_C, we expect the right half plane zero to occur at high frequencies. Assuming $\gamma = \lambda = 0$ for the source follower, neglecting some of the device capacitances, and simplifying the circuit as shown in Fig. 10.36, we can write $-g_{m1}V_1 = V_{out}(R_L^{-1} + C_L s)$ and hence

$$V_1 = \frac{-V_{out}}{g_{m1}R_L}(1 + R_L C_L s). \tag{10.31}$$

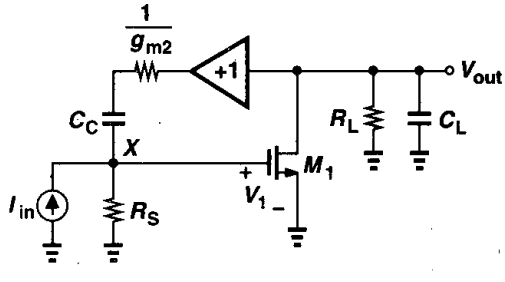


Figure 10.36 Simplified equivalent circuit of Fig. 10.35(b).

We also have

$$\frac{V_{out} - V_1}{\frac{1}{g_{m2}} + \frac{1}{C_C s}} + I_{in} = \frac{V_1}{R_S}. \tag{10.32}$$

Substituting for V_1 from (10.31) yields:

$$\frac{V_{out}}{I_{in}} = \frac{-g_{m1} R_L R_S (g_{m2} + C_C s)}{R_L C_L C_C (1 + g_{m2} R_S) s^2 + [(1 + g_{m1} g_{m2} R_L R_S) C_C + g_{m2} R_L C_L] s + g_{m2}}. \tag{10.33}$$

Thus, the circuit contains a zero in the *left* half plane, which can be chosen to cancel one of the poles. The zero can also be derived as illustrated in Fig. 6.15.

We can also compute the magnitudes of the two poles assuming that they are widely separated. Since typically $1 + g_{m2} R_S \gg 1$ and $(1 + g_{m1} g_{m2} R_L R_S) C_C \gg g_{m2} R_L C_L$, we have

$$\omega_{p1} \approx \frac{g_{m2}}{g_{m1} g_{m2} R_L R_S C_C} \tag{10.34}$$

$$\approx \frac{1}{g_{m1} R_L R_S C_C}, \tag{10.35}$$

and

$$\omega_{p2} \approx \frac{g_{m1} g_{m2} R_L R_S C_C}{R_L C_L C_C g_{m2} R_S} \tag{10.36}$$

$$\approx \frac{g_{m1}}{C_L}. \tag{10.37}$$

Thus, the new values of ω_{p1} and ω_{p2} are similar to those obtained by simple Miller approximation. For example, the output pole has moved from $(R_L C_L)^{-1}$ to g_{m1}/C_L.

The primary issue in the circuit of Fig. 10.35(b) is that the source follower limits the lower end of the output voltage to $V_{GS2} + V_{I2}$, where V_{I2} is the voltage required across I_2. For this reason, it is desirable to utilize the compensation capacitor to isolate the dc levels in the active feedback stage from that at the output. Such a topology is depicted in Fig. 10.37, where C_C and the common-gate stage M_2 convert the output voltage swing to a current, returning the result to the gate of M_1 [4]. If V_1 changes by ΔV and V_{out} by $A_v \Delta V$, then the current through the capacitor is nearly equal to $A_v \Delta V C_C s$ because $1/g_{m2}$ can be relatively small. Thus, a change ΔV at the gate of M_1 creates a current change of $A_v \Delta V C_C s$, providing a capacitor multiplication factor equal to A_v.

Assuming $\lambda = \gamma = 0$ for the common-gate stage, we redraw the circuit of Fig. 10.37 in Fig. 10.38, where we have:

$$V_{out} + \frac{g_{m2} V_2}{C_C s} = -V_2 \tag{10.38}$$

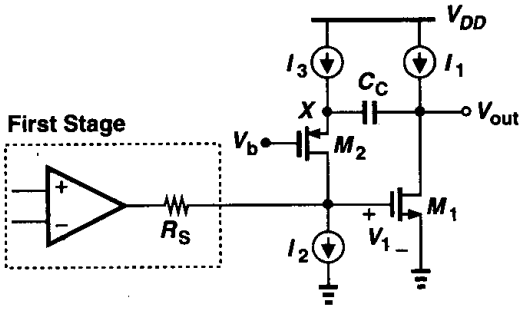


Figure 10.37 Compensation technique using a common-gate stage.

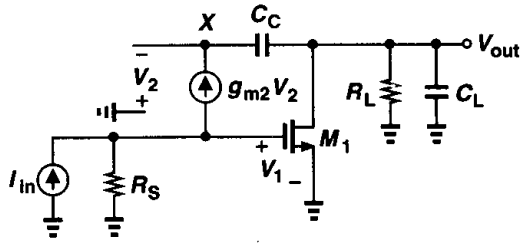


Figure 10.38 Simplified equivalent circuit of Fig. 10.37.

and hence

$$V_2 = -V_{out}\frac{C_C s}{C_C s + g_{m2}}. \tag{10.39}$$

Also,

$$g_{m1}V_1 + V_{out}\left(\frac{1}{R_L} + C_L s\right) = g_{m2}V_2 \tag{10.40}$$

and $I_{in} = V_1/R_S + g_{m2}V_2$. Solving these equations, we obtain

$$\frac{V_{out}}{I_{in}} = \frac{-g_{m1}R_S R_L(g_{m2} + C_C s)}{R_L C_L C_C s^2 + [(1 + g_{m1}R_S)g_{m2}R_L C_C + C_C + g_{m2}R_L C_L]s + g_{m2}}. \tag{10.41}$$

As with the circuit of Fig. 10.35(b), this topology contains a zero in the left half plane. Using similar approximations, we compute the poles as

$$\omega_{p1} \approx \frac{1}{g_{m1}R_L R_S C_C} \tag{10.42}$$

$$\omega_{p2} \approx \frac{g_{m2}R_s g_{m1}}{C_L}. \tag{10.43}$$

Interestingly, the second pole has considerably risen in magnitude — by a factor of $g_{m2}R_S$ with respect to that of the circuit of Fig. 10.35. This is because at very high frequencies,

the feedback loop consisting of M_2 and R_S in Fig. 10.37 lowers the output resistance by the same factor. Of course, if the capacitance at the gate of M_1 is taken into account, pole splitting is less pronounced. Nevertheless, this technique can potentially provide a high bandwidth in two-stage op amps.

The op amp of Fig. 10.37 entails important slewing issues. For positive slewing at the output, the simplified circuit of Fig. 10.39(a) suggests that M_2 and hence I_1 must support

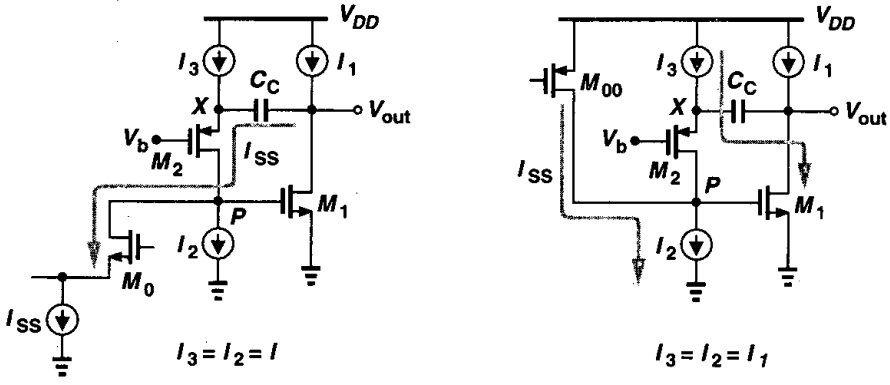


Figure 10.39 Circuit of Fig. 10.37 during (a) positive and (b) negative slewing.

I_{SS}, requiring that $I_1 \geq I_{SS} + I_{D1}$. If I_1 is less, then V_P drops, turning M_1 off, and if $I_1 < I_{SS}$, M_0 and its tail current source must enter the triode region, yielding a slew rate equal to I_1/C_C.

For negative slewing, I_2 must support both I_{SS} and I_{D2} [Fig. 10.39(b)]. As I_{SS} flows into node P, V_P tends to rise, increasing I_{D1}. Thus, M_1 absorbs the current produced by I_3 through C_C, turning off M_2 and opposing the increase in V_P. We can therefore consider P a virtual ground node. This means that, for equal positive and negative slew rates, I_3 (and hence I_2) must be as large as I_{SS}, raising the power dissipation.

Op amps using a cascode topology as their first stage can incorporate a variant of the technique illustrated in Fig. 10.37. Shown in Fig. 10.40(a), this approach places the compensation capacitor between the *source* of the cascode devices and the output nodes. Using the simplified model of Fig. 10.40(b) and the method of Fig. 6.15, the reader can prove that the zero appears at $(g_{m4}R_{eq})(g_{m9}/C_C)$, a much greater magnitude than g_{m9}/C_C. If other capacitances are neglected, it can also be proved that the dominant pole is located at approximately $(R_{eq}g_{m9}R_LC_C)^{-1}$, as if C_C were connected to the gate of M_9 rather than the source of M_4. Also, the first nondominant pole is given by $g_{m4}g_{m9}R_{eq}/C_L$, an effect similar to that described by Eq. (10.43). In reality, the capacitance at X may not be negligible because the resistance seen at this node is quite large. The analysis of the slew rate is left as an exercise for the reader.

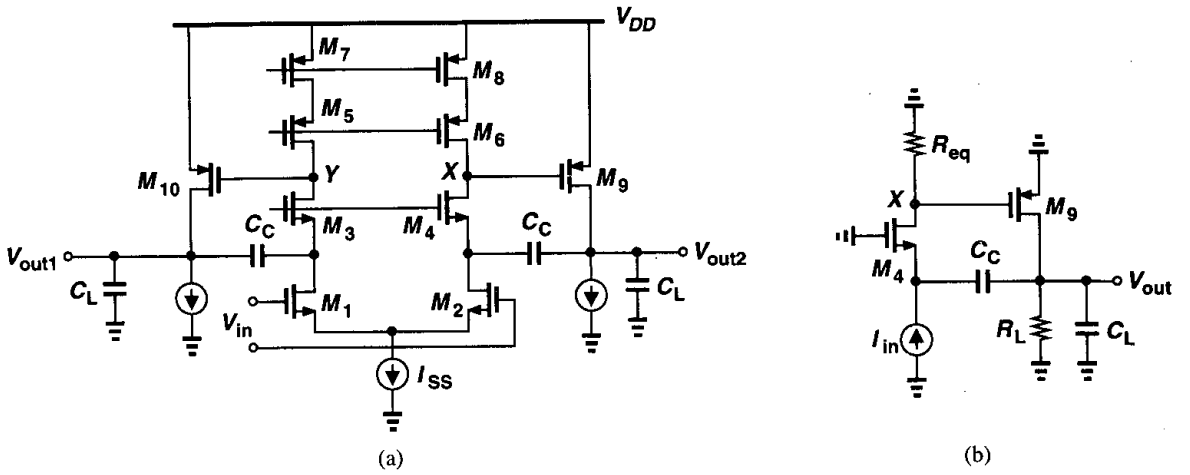


Figure 10.40 (a) Alternative method of compensating two-stage op amps, (b) simplified equivalent circuit of (a).

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.

10.1. An amplifier with a forward gain of A_0 and two poles at 10 MHz and 500 MHz is placed in a unity-gain feedback loop. Calculate A_0 for a phase margin of 60°.

10.2. An amplifier with a forward gain of A_0 has two coincident poles at ω_p. Calculate the maximum value of A_0 for a 60° phase margin with a closed-loop gain of **(a)** unity, **(b)** 4.

10.3. An amplifier has a forward gain of $A_0 = 1000$ and two poles at ω_{p1} and ω_{p2}. For $\omega_{p1} =$ 1 MHz, calculate the phase margin of a unity-gain feedback loop if **(a)** $\omega_{p2} = 2\omega_{p1}$, **(b)** $\omega_{p2} = 4\omega_{p1}$.

10.4. A unity-gain closed-loop amplifier exhibits a frequency peaking of 50% in the vicinity of the gain crossover. What is the phase margin?

10.5. Consider the transimpedance amplifier shown in Fig. 10.41, where $R_D = 1$ kΩ, $R_F =$ 10 kΩ, $g_{m1} = g_{m2} = 1/(100\ \Omega)$, and $C_A = C_X = C_Y = 100$ fF. Neglecting all other

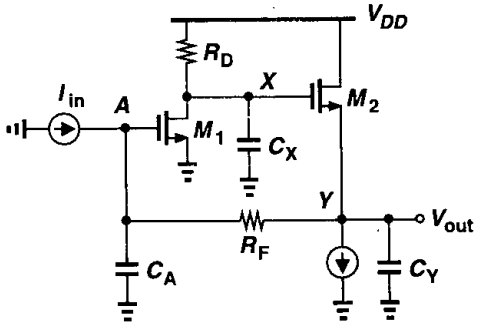


Figure 10.41

capacitances and assuming $\lambda = \gamma = 0$, compute the phase margin of the circuit. (Hint: break the loop at node X.)

10.6. In Problem 10.5, what is the phase margin if R_D is increased to 2 kΩ?

10.7. If the phase margin required of the amplifier of Problem 10.5 is 45°, what is the maximum value of (a) C_Y, (b) C_A, (c) C_X while the other two capacitances remain constant?

10.8. Prove that the zero of the circuit shown in Fig. 10.29 is given by Eq. (10.25). Apply the technique illustrated in Fig. 6.15.

10.9. Consider the amplifier of Fig. 10.42, where $(W/L)_{1-4} = 50/0.5$ and $I_{SS} = I_1 = 0.5$ mA.

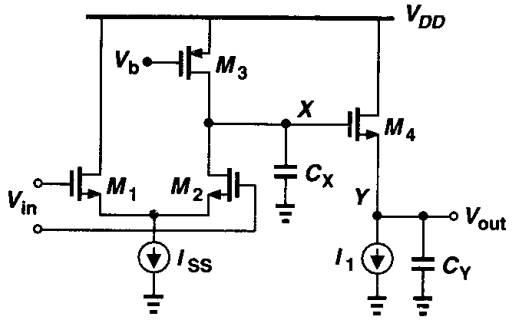


Figure 10.42

(a) Estimate the poles at nodes X and Y by multiplying the small-signal resistance and capacitance to ground. Assume $C_X = C_Y = 0.5$ pF. What is the phase margin for unity-gain feedback?

(b) If $C_X = 0.5$ pF, what is the maximum tolerable value of C_Y that yields a phase margin of 60° for unity-gain feedback?

10.10. Estimate the slew rate of the op amp of Problem 10.9(b) for both parts **(a)** and **(b)**.

10.11. In the two-stage op amp of Fig. 10.43, $W/L = 50/0.5$ for all transistors except for $M_{5,6}$, for which $W/L = 60/0.5$. Also, $I_{SS} = 0.25$ mA and each output branch is biased at 1 mA.

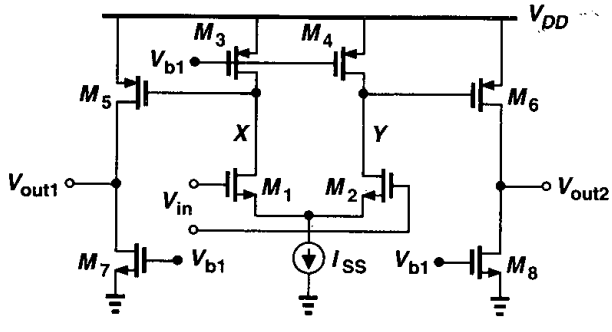


Figure 10.43

(a) Determine the CM level at nodes X and Y.

(b) Calculate the maximum output voltage swing.

(c) If each output is loaded by a 1-pF capacitor, compensate the op amp by Miller multiplication for a phase margin of 60° in unity-gain feedback. Calculate the pole and zero positions after compensation.

(d) Calculate the resistance that must be placed in series with the compensation capacitors to position the zero atop the nondominant pole.
(e) Determine the slew rate.

10.12. In Problem 10.11(e), the pole-zero cancellation resistor is implemented with a PMOS device as in Fig. 10.31. Calculate the dimensions of M_{13}-M_{15} if $I_1 = 100\ \mu$A.

10.13. Calculate the input-referred thermal noise voltage of the op amp shown in Fig. 10.43.

10.14. Figure 10.44 depicts a transimpedance amplifier employing voltage-current feedback. Note that the feedback factor may exceed unity because of M_3. Assume I_1-I_3 are ideal, $I_1 = I_2 = 1$ mA, $I_3 = 10\ \mu$A, $(W/L)_{1,2} = 50/0.5$, and $(W/L)_3 = 5/0.5$.

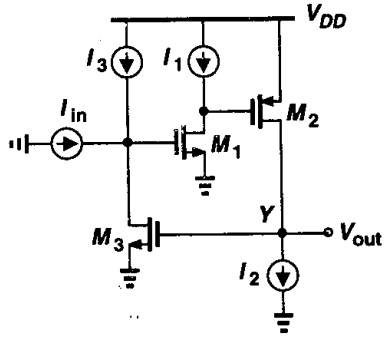


Figure 10.44

(a) Breaking the loop at the gate of M_3, estimate the poles of the open-loop transfer function.
(b) If the circuit is compensated by adding a capacitor C_C between the gate and the drain of M_1, what value of C_C achieves a phase margin of 60°? Determine the poles after compensation.
(c) What resistance must be placed in series with C_C to position the zero of the output stage atop the first nondominant pole?

10.15. Repeat Problem 10.14(c) if the output node is loaded by a 0.5-pF capacitor.

10.16. Suppose in the circuit of Fig. 10.44 a large negative input current is applied such that M_1 turns off momentarily. What is the slew rate at the output?

10.17. Explain why in the circuit of Fig. 10.44, the compensation capacitor should not be placed between the gate and the drain of M_2 or M_3.

10.18. Determine the input-referred noise current of the circuit shown in Fig. 10.44 and described in Problem 10.14(c).

10.19. The cancellation of a pole by a zero, e.g., in a two-stage op amp, entails an issue called the "doublet" problem [5, 6]. If the pole and the zero do not exactly coincide, we say they constitute a doublet. The step response of feedback circuits in the presence of doublets is of great interest. Suppose the open-loop transfer function of a two-stage op amp is expressed as

$$H_{open}(s) = \frac{A_0\left(1+\dfrac{s}{\omega_z}\right)}{\left(1+\dfrac{s}{\omega_{p1}}\right)\left(1+\dfrac{s}{\omega_{p2}}\right)}. \tag{10.44}$$

Ideally, $\omega_z = \omega_{p2}$ and the feedback circuit exhibits a first-order behavior, i.e., its step response contains a single time constant and no overshoot.

(a) Prove that the transfer function of the amplifier in a unity-gain feedback loop is given by

$$H_{closed}(s) = \frac{A_0\left(1+\frac{s}{\omega_z}\right)}{\frac{s^2}{\omega_{p1}\omega_{p2}} + \left(\frac{1}{\omega_{p1}} + \frac{1}{\omega_{p2}} + \frac{A_0}{\omega_z}\right)s + A_0 + 1}. \tag{10.45}$$

(b) Determine the two poles of $H_{closed}(s)$, assuming they are widely spaced.

(c) Assuming $\omega_z \approx \omega_{p2}$ and $\omega_{p2} \ll (1 + A_0)\omega_{p1}$, write $H_{closed}(s)$ in the form

$$H_{closed}(s) = \frac{A\left(1+\frac{s}{\omega_z}\right)}{\left(1+\frac{s}{\omega_{pA}}\right)\left(1+\frac{s}{\omega_{pB}}\right)}, \tag{10.46}$$

and determine the small-signal step response of the closed-loop amplifier.

(d) Prove that the step response contains an exponential term of the form $(1 - \omega_z/\omega_{p2})$ $\exp(-\omega_{p2}t)$. This is an important result, indicating that if the zero does not exactly cancel the pole, the step response exhibits an exponential with an amplitude proportional to $1 - \omega_z/\omega_{p2}$ (which depends on the mismatch between ω_z and ω_{p2}) and a time constant of $1/\omega_z$.

10.20. Using the results of Problem 10.19(d), determine the step response of the amplifier described in Problem 10.11(e) with **(a)** perfect pole-zero cancellation, **(b)** 10% mismatch between the pole and the zero magnitudes.

References

1. P. R. Gray and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*, Third Ed., New York: Wiley, 1993.
2. W. C. Black, D. J. Allstot, and R. A. Reed, "A High Performance Low Power CMOS Channel Filter," *IEEE J. of Solid-State Circuits,* vol. 15, pp. 929–938, Dec. 1983.
3. R. M. Ziazadeh, H.-T. Ng, and D. J. Allstot, "A Multistage Amplifier Topology with Embeded Tracking Compensation," *CICC Proc.*, pp. 361–364, May 1998.
4. B. K. Ahuja, "An Improved Frequency Compensation Technique for CMOS Operational Amplifiers," *IEEE J. of Solid-State Circuits,* vol. 18, pp. 629–633, Dec. 1983.
5. P. R. Gray and R. G. Meyer, "MOS Operational Amplifier Design—A Tutorial Overview," *IEEE J. of Solid-State Circuits,* vol. 17, pp. 969–982, Dec. 1982.
6. B. Y. Kamath, R. G. Meyer, and P. R. Gray, "Relationship between Frequency Response and Settling Time of Operational Amplifiers," *IEEE J. of Solid-State Circuits,* vol. 9, pp. 347–352, Dec. 1974.

Chapter 11

Bandgap References

Analog circuits incorporate voltage and current references extensively. Such references are dc quantities that exhibit little dependence on supply and process parameters and a *well-defined* dependence on the temperature. For example, the bias current of a differential pair must be generated according to a reference, for it affects the voltage gain and noise of the circuit. Also, in systems such as A/D and D/A converters, a reference is required to define the input or output full-scale range.

In this chapter, we deal with the design of reference generators in CMOS technology, focusing on well-established "bandgap" techniques. First, we study supply-independent biasing and the problem of start-up. Next, we describe temperature-independent references and examine issues such as the effect of offset voltages. Finally, we present constant-G_m biasing and study an example of state-of-the-art bandgap references.

11.1 General Considerations

As mentioned above, the objective of reference generation is to establish a dc voltage or current that is independent of the supply and process and has a well-defined behavior with temperature. In most applications, the required temperature dependence assumes one of three forms: (a) proportional to absolute temperature (PTAT); (2) constant-G_m behavior, i.e., such that the transconductance of certain transistors remains constant; (3) temperature independent. We can therefore divide the task into two design problems: supply-independent biasing and definition of the temperature variation.

In addition to supply, process, and temperature variability, several other parameters of reference generators may be critical as well. These include output impedance, output noise, and power dissipation. We return to these issues later in this chapter.

11.2 Supply-Independent Biasing

Our use of bias currents and current mirrors in previous chapters has implicitly assumed that a "golden" reference current is available. As shown in Fig. 11.1(a), if I_{REF} does not

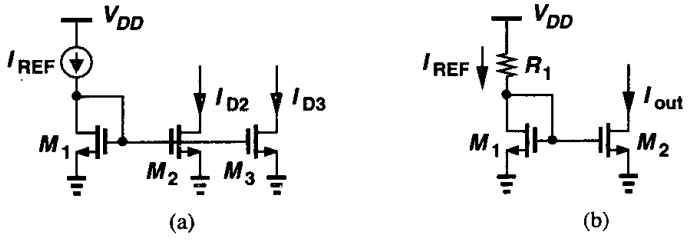


Figure 11.1 Current-mirror biasing using (a) an ideal current source, (b) a resistor.

vary with V_{DD} and channel-length modulation of M_2 and M_3 is neglected, then I_{D2} and I_{D3} remain independent of the supply voltage. The question then is: how do we generate I_{REF}?

As an approximation of a current source, we tie a resistor from V_{DD} to the gate of M_1 [Fig. 11.1(b)]. However, the output current of this circuit is quite sensitive to V_{DD}:

$$\Delta I_{out} = \frac{\Delta V_{DD}}{R_1 + 1/g_{m1}} \cdot \frac{(W/L)_2}{(W/L)_1}. \tag{11.1}$$

In order to arrive at a less sensitive solution, we postulate that the circuit must bias *itself*, i.e., I_{REF} must be somehow derived from I_{out}. The idea is that if I_{out} is to be ultimately independent of V_{DD}, then I_{REF} can be a replica of I_{out}. Fig. 11.2 illustrates an implementation where M_3 and M_4 copy I_{out}, thereby defining I_{REF}. In essence, I_{REF} is "bootstrapped" to I_{out}. With the sizes chosen here, we have $I_{out} = KI_{REF}$ if channel-length modulation is neglected. Note that, since each diode-connected device feeds from a current source, I_{out} and I_{REF} are relatively independent of V_{DD}.

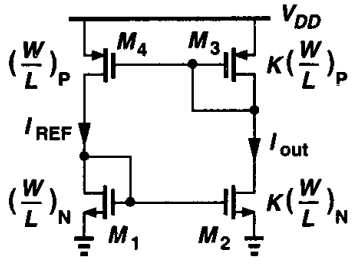


Figure 11.2 Simple circuit to establish supply-independent currents.

Since I_{out} and I_{REF} in Fig. 11.2 display little dependence on V_{DD}, their magnitude is set by other parameters. How do we calculate these currents? Interestingly, if M_1-M_4 operate in saturation and $\lambda \approx 0$, then the circuit is governed by only one equation, $I_{out} = KI_{REF}$, and hence can support *any* current level! For example, if we initially force I_{REF} to be 10 μA, the resulting I_{out} of $K \times 10$ μA "circulates" around the loop, sustaining these current levels in the left and right branches indefinitely.

To uniquely define the currents, we add another constraint to the circuit, e.g., as shown in Fig. 11.3(a). Here, resistor R_S decreases the current of M_2 while the PMOS devices

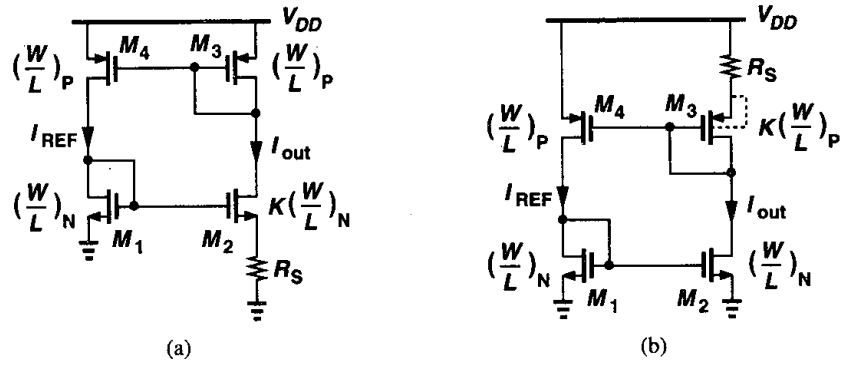


Figure 11.3 (a) Addition of R_S to define the currents, (b) alternative implementation eliminating body effect.

require that $I_{out} = I_{REF}$ because they have identical dimensions. We can write $V_{GS1} = V_{GS2} + I_{D2}R_S$, or

$$\sqrt{\frac{2I_{out}}{\mu_n C_{ox}(W/L)_N}} + V_{TH1} = \sqrt{\frac{2I_{out}}{\mu_n C_{ox} K(W/L)_N}} + V_{TH2} + I_{out}R_S. \tag{11.2}$$

Neglecting body effect, we have

$$\sqrt{\frac{2I_{out}}{\mu_n C_{ox}(W/L)_N}}\left(1 - \frac{1}{\sqrt{K}}\right) = I_{out}R_S, \tag{11.3}$$

and hence

$$I_{out} = \frac{2}{\mu_n C_{ox}(W/L)_N} \cdot \frac{1}{R_S^2}\left(1 - \frac{1}{\sqrt{K}}\right)^2. \tag{11.4}$$

As expected, the current is independent of the supply voltage (but still a function of process and temperature).

The assumption $V_{TH1} = V_{TH2}$ introduces some error in the foregoing calculations because the sources of M_1 and M_2 are at different voltages. Shown in Fig. 11.3(b), a simple remedy is to place the resistor in the source of M_3 while eliminating body effect by tying the source and bulk of each PMOS transistor. Another solution is described in Problem 11.1.

The circuits of Figs. 11.3(a) and (b) exhibit little supply dependence if channel-length modulation is negligible. For this reason, relatively long channels are used for all of the transistors in the circuit.

Example 11.1

Assuming $\lambda \neq 0$ in Fig. 11.3(a), estimate the change in I_{out} for a small change ΔV_{DD} in the supply voltage.

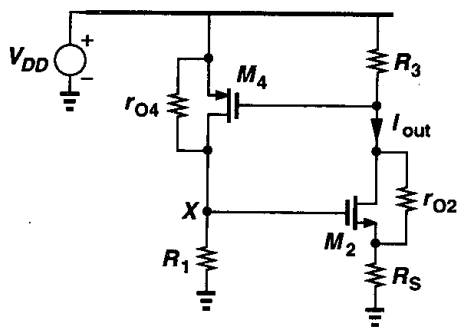


Figure 11.4

Solution

Simplifying the circuit as depicted in Fig. 11.4, where $R_1 = r_{O1} \| (1/g_{m1})$ and $R_3 = r_{O3} \| (1/g_{m3})$, we calculate the "gain" from V_{DD} to I_{out}. The small-signal gate-source voltage of M_4 equals $-I_{out} R_3$ and the current through r_{O4} is $(V_{DD} - V_X)/r_{O4}$. Thus,

$$\frac{V_{DD} - V_X}{r_{O4}} + I_{out} R_3 g_{m4} = \frac{V_X}{R_1}. \tag{11.5}$$

If we denote the equivalent transconductance of M_2 and R_S by $G_{m2} = I_{out}/V_X$, then

$$\frac{I_{out}}{V_{DD}} = \frac{1}{r_{O4}} \left[\frac{1}{G_{m2}(r_{O4} \| R_1)} - g_{m4} R_3 \right]^{-1}. \tag{11.6}$$

Note from Chapter 3 that

$$G_{m2} = \frac{g_{m2} r_{O2}}{R_S + r_{O2} + (g_{m2} + g_{mb2}) R_S r_{O2}}. \tag{11.7}$$

Interestingly, the sensitivity vanishes if $r_{O4} = \infty$.

In some applications, the sensitivity given by (11.6) is prohibitively large. Also, owing to various capacitive paths, the supply sensitivity of the circuit typically rises at high frequencies. For these reasons, the supply voltage of the core is often derived from a locally-generated, less sensitive voltage. We return to this point in Section 11.7.

An important issue in supply-independent biasing is the existence of "degenerate" bias points. In the circuit of Fig. 11.3(a), for example, if all of the transistors carry zero current when the supply is turned on, they may remain off indefinitely because the loop can support a zero current in both branches. This condition is not predicted by (11.4) because in manipulating (11.3) we divided both sides by $\sqrt{I_{out}}$, tacitly assuming $I_{out} \neq 0$. In other words, the circuit can settle in one of *two* different operating conditions.

Called the "start-up" problem, the above issue is resolved by adding a mechanism that drives the circuit out of the degenerate bias point when the supply is turned on. Shown in Fig. 11.5 is a simple example, where the diode-connected device M_5 provides a current path from V_{DD} through M_3 and M_1 to ground upon start-up. Thus, M_3 and M_1, and hence M_2

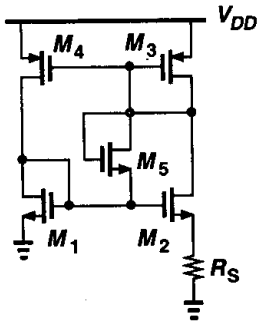


Figure 11.5 Addition of start-up device to the circuit of Fig. 11.3(a).

and M_4, cannot remain off. Of course, this technique is practical only if $V_{TH1} + V_{TH5} + |V_{TH3}| < V_{DD}$ and $V_{GS1} + V_{TH5} + |V_{GS3}| > V_{DD}$, the latter to ensure M_5 remains off after start-up. Another start-up circuit is analyzed in Problem 11.2.

The problem of start-up generally requires careful analysis and simulation. The supply voltage must be ramped from zero in a dc sweep simulation (such that parasitic capacitances do not cause false start-up) as well as in a transient simulation and the behavior of the circuit examined for each supply voltage. In complex implementations, more than one degenerate point may exist.

11.3 Temperature-Independent References

Reference voltages or currents that exhibit little dependence on temperature prove essential in many analog circuits. It is interesting to note that, since most process parameters vary with temperature, if a reference is temperature-independent, then it is usually process-independent as well.

How do we generate a quantity that remains constant with temperature? We postulate that if two quantities having opposite temperature coefficients (TCs) are added with proper weighting, the result displays a zero TC. For example, for two voltages V_1 and V_2 that vary in opposite directions with temperature, we choose α_1 and α_2 such that $\alpha_1 \partial V_1/\partial T + \alpha_2 \partial V_2/\partial T = 0$, obtaining a reference voltage, $V_{REF} = \alpha_1 V_1 + \alpha_2 V_2$, with zero TC.

We must now identify two voltages that have positive and negative TCs. Among various device parameters in semiconductor technologies, the characteristics of bipolar transistors have proven the most reproducible and well-defined quantities that can provide positive and negative TCs. Even though many parameters of MOS devices have been considered for the task of reference generation [1, 2], bipolar operation still forms the core of such circuits.

11.3.1 Negative-TC Voltage

The base-emitter voltage of bipolar transistors or, more generally, the forward voltage of a *pn*-junction diode exhibits a negative TC. We first obtain the expression for the TC in terms of readily-available quantities.

For a bipolar device we can write $I_C = I_S \exp(V_{BE}/V_T)$, where $V_T = kT/q$. The saturation current I_S is proportional to $\mu k T n_i^2$, where μ denotes the mobility of minority

carriers and n_i is the intrinsic minority carrier concentration of silicon. The temperature dependence of these quantities is represented as $\mu \propto \mu_0 T^m$, where $m \approx -3/2$, and $n_i^2 \propto T^3 \exp[-E_g/(kT)]$, where $E_g \approx 1.12$ eV is the bandgap energy of silicon. Thus,

$$I_S = bT^{4+m} \exp \frac{-E_g}{kT}, \tag{11.8}$$

where b is a proportionality factor. Writing $V_{BE} = V_T \ln(I_C/I_S)$, we can now compute the TC of the base-emitter voltage. In taking the derivative of V_{BE} with respect to T, we must know the behavior of I_C as a function of the temperature. To simplify the analysis, we assume for now that I_C is held constant. Thus,

$$\frac{\partial V_{BE}}{\partial T} = \frac{\partial V_T}{\partial T} \ln \frac{I_C}{I_S} - \frac{V_T}{I_S} \frac{\partial I_S}{\partial T}. \tag{11.9}$$

From (11.8), we have

$$\frac{\partial I_S}{\partial T} = b(4+m)T^{3+m} \exp \frac{-E_g}{kT} + bT^{4+m} \left(\exp \frac{-E_g}{kT} \right) \left(\frac{E_g}{kT^2} \right). \tag{11.10}$$

Therefore,

$$\frac{V_T}{I_S} \frac{\partial I_S}{\partial T} = (4+m) \frac{V_T}{T} + \frac{E_g}{kT^2} V_T. \tag{11.11}$$

With the aid of (11.9) and (11.11), we can write

$$\frac{\partial V_{BE}}{\partial T} = \frac{V_T}{T} \ln \frac{I_C}{I_S} - (4+m) \frac{V_T}{T} - \frac{E_g}{kT^2} V_T \tag{11.12}$$

$$= \frac{V_{BE} - (4+m)V_T - E_g/q}{T}. \tag{11.13}$$

Equation (11.13) gives the temperature coefficient of the base-emitter voltage at a given temperature T, revealing dependence on the magnitude of V_{BE} itself. With $V_{BE} \approx 750$ mV and $T = 300°$K, $\partial V_{BE}/\partial T \approx -1.5$ mV/°K.

From (11.13), we note that the temperature coefficient of V_{BE} itself depends on the temperature, creating error in constant reference generation if the positive-TC quantity exhibits a *constant* temperature coefficient.

11.3.2 Positive-TC Voltage

It was recognized in 1964 [3] that if two bipolar transistors operate at unequal current densities, then the *difference* between their base-emitter voltages is directly proportional to the absolute temperature. For example, as shown in Fig. 11.6, if two identical transistors ($I_{S1} = I_{S2}$) are biased at collector currents of nI_0 and I_0 and their base currents are neg-

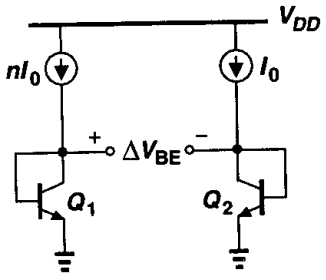


Figure 11.6 Generation of PTAT voltage.

ligible, then

$$\Delta V_{BE} = V_{BE1} - V_{BE2} \tag{11.14}$$

$$= V_T \ln \frac{nI_0}{I_{S1}} - V_T \ln \frac{I_0}{I_{S2}} \tag{11.15}$$

$$= V_T \ln n. \tag{11.16}$$

Thus, the V_{BE} difference exhibits a positive temperature coefficient:

$$\frac{\partial \Delta V_{BE}}{\partial T} = \frac{k}{q} \ln n. \tag{11.17}$$

Interestingly, this TC is independent of the temperature or behavior of the collector currents.[1]

Example 11.2

Calculate ΔV_{BE} in the circuit of Fig. 11.7.

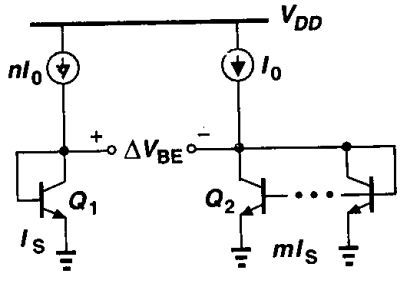


Figure 11.7

[1] Nonidealities in the characteristics of bipolar transistors introduce a small temperature dependence in this TC.

Solution

Neglecting base currents, we can write

$$= V_T \ln \frac{nI_0}{I_S} - V_T \ln \frac{I_0}{mI_S} \tag{11.18}$$

$$= V_T \ln(nm). \tag{11.19}$$

The temperature coefficient is therefore equal to $(k/q)\ln(nm)$.

11.3.3 Bandgap Reference

With the negative- and positive-TC voltages obtained above, we can now develop a reference having a nominally zero temperature coefficient. We write $V_{REF} = \alpha_1 V_{BE} + \alpha_2(V_T \ln n)$, where $V_T \ln n$ is the difference between the base-emitter voltages of the two bipolar transistors operating at different current densities. How do we choose α_1 and α_2? Since at room temperature $\partial V_{BE}/\partial T \approx -1.5$ mV/°K whereas $\partial V_T/\partial T \approx +0.087$ mV/°K, we may set $\alpha_1 = 1$ and choose $\alpha_2 \ln n$ such that $(\alpha_2 \ln n)(0.087 \text{ mV/°K}) = 1.5$ mV/°K. That is, $\alpha_2 \ln n \approx 17.2$, indicating that for zero TC:

$$V_{REF} \approx V_{BE} + 17.2V_T \tag{11.20}$$

$$\approx 1.25 \text{ V}. \tag{11.21}$$

Let us now devise a circuit that adds V_{BE} to $17.2V_T$. First, consider the circuit shown in Fig. 11.8, where base currents are assumed negligible, transistor Q_2 consists of n unit transistors in parallel, and Q_1 is a unit transistor. Suppose we somehow force V_{O1} and V_{O2} to be equal. Then, $V_{BE1} = RI + V_{BE2}$ and $RI = V_{BE1} - V_{BE2} = V_T \ln n$. Thus, $V_{O2} = V_{BE2} + V_T \ln n$, suggesting that V_{O2} can serve as a temperature-independent reference if $\ln n \approx 17.2$ (while V_{O1} and V_{O2} remain equal).

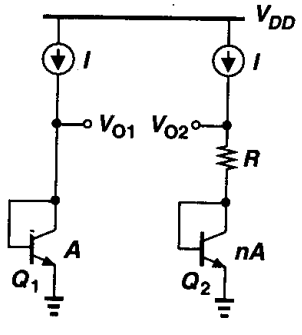


Figure 11.8 Conceptual generation of temperature-independent voltage.

The circuit of Fig. 11.8 requires two modifications to become practical. First, a mechanism must be added to guarantee $V_{O1} = V_{O2}$. Second, since $\ln n = 17.2$ translates to a prohibitively large n, the term $RI = V_T \ln n$ must be scaled up by a reasonable factor. Shown in Fig. 11.9 is an implementation accomplishing both tasks [4]. Here, amplifier

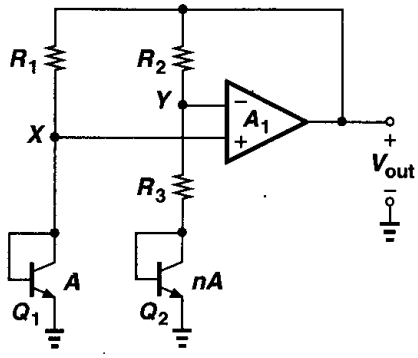


Figure 11.9 Actual implementation of the concept shown in Fig. 11.8.

A_1 senses V_X and V_Y, driving the top terminals of R_1 and R_2 ($R_1 = R_2$) such that X and Y settle to approximately equal voltages. The reference voltage is obtained at the output of the amplifier (rather than node Y). Following the analysis of Fig. 11.8, we have $V_{BE1} - V_{BE2} = V_T \ln n$, arriving at a current equal to $V_T \ln n / R_3$ through the right branch and hence an output voltage of

$$V_{out} = V_{BE2} + \frac{V_T \ln n}{R_3}(R_3 + R_2) \tag{11.22}$$

$$= V_{BE2} + (V_T \ln n)\left(1 + \frac{R_2}{R_3}\right). \tag{11.23}$$

For a zero TC, we must have $(1 + R_2/R_3) \ln n \approx 17.2$. For example, we may choose $n = 31$ and $R_2/R_3 = 4$. Note these results do not depend on the TC of the resistors.

The circuit of Fig. 11.9 entails a number of design issues. We consider each one below.

Collector Current Variation The circuit of Fig. 11.9 violates one of our earlier assumptions: the collector currents of Q_1 and Q_2, given by $(V_T \ln n)/R_3$, are proportional to T, whereas $\partial V_{BE}/\partial T \approx -1.5$ mV/°K was derived for a constant current. What happens to the temperature coefficient of V_{BE} if the collector currents are PTAT? As a first-order iterative solution, let us assume $I_{C1} = I_{C2} \approx (V_T \ln n)/R_3$. Returning to Eq. (11.9) and including $\partial I_C/\partial T$, we have

$$\frac{\partial V_{BE}}{\partial T} = \frac{\partial V_T}{\partial T} \ln \frac{I_C}{I_S} + V_T \left(\frac{1}{I_C} \frac{\partial I_C}{\partial T} - \frac{1}{I_S} \frac{\partial I_S}{\partial T} \right). \tag{11.24}$$

Since $\partial I_C/\partial T \approx (V_T \ln n)/(R_3 T) = I_C/T$, we can write

$$\frac{\partial V_{BE}}{\partial T} = \frac{\partial V_T}{\partial T} \ln \frac{I_C}{I_S} + \frac{V_T}{T} - \frac{V_T}{I_S} \frac{\partial I_S}{\partial T}. \tag{11.25}$$

Equation (11.13) is therefore modified as

$$\frac{\partial V_{BE}}{\partial T} = \frac{V_{BE} - (3 + m)V_T - E_g/q}{T}, \tag{11.26}$$

indicating that the TC is slightly less negative than -1.5 mV/°K. In practice, accurate simulations are necessary to predict the temperature coefficient.

Compatibility with CMOS Technology Our derivation of a temperature-independent voltage relies on the exponential characteristics of bipolar devices for both negative- and positive-TC quantities. We must therefore seek structures in a standard CMOS technology that exhibit such characteristics.

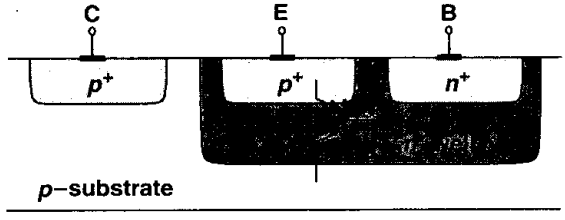


Figure 11.10 Realization of a *pnp* bipolar transistor in CMOS technology.

In n-well processes, a *pnp* transistor can be formed as depicted in Fig. 11.10. A p^+ region (the same as the S/D region of PFETs) inside an n-well serves as the emitter and the n-well itself as the base. The p-type substrate acts as the collector and it is inevitably connected to the most negative supply (usually ground). The circuit of Fig. 11.9 can therefore be redrawn as shown in Fig. 11.11.

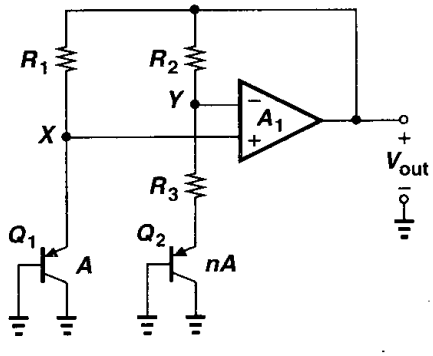


Figure 11.11 Circuit of Fig. 11.9 implemented with *pnp* transistors.

Op Amp Offset and Output Impedance As explained in Chapter 13, owing to asymmetries, op amps suffer from input "offsets," i.e., the output voltage of the op amp is not zero if the input is set to zero. The input offset voltage of the op amp in Fig. 11.9 introduces error in the output voltage. Included in Fig. 11.12, the effect is quantified as $V_{BE1} - V_{OS} \approx V_{BE2} + R_3 I_{C2}$ (if A_1 is large) and $V_{out} = V_{BE2} + (R_3 + R_2) I_{C2}$. Thus,

$$V_{out} = V_{BE2} + (R_3 + R_2)\frac{V_{BE1} - V_{BE2} - V_{OS}}{R_3} \tag{11.27}$$

$$= V_{BE2} + \left(1 + \frac{R_2}{R_3}\right)(V_T \ln n - V_{OS}), \tag{11.28}$$

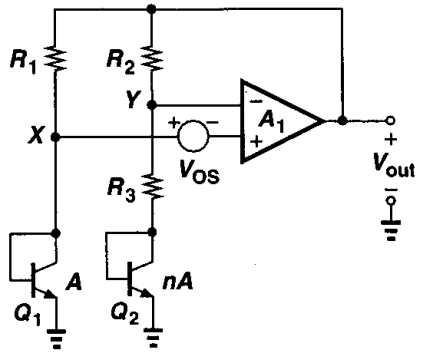


Figure 11.12 Effect of op amp offset on the reference voltage.

where we have assumed $I_{C2} \approx I_{C1}$ despite the offset voltage. The key point here is that V_{OS} is amplified by $1 + R_2/R_3$, introducing error in V_{out}. More importantly, as explained in Chapter 13, V_{OS} itself varies with temperature, raising the temperature coefficient of the output voltage.

Several methods are employed to lower the effect of V_{OS}. First, the op amp incorporates large devices in a carefully chosen topology so as to minimize the offset (Chapter 18). Second, as illustrated in Fig. 11.7, the collector currents of Q_1 and Q_2 can be ratioed by a factor of m such that $\Delta V_{BE} = V_T \ln(mn)$. Third, each branch may use two pn junctions in series to double ΔV_{BE}. Fig. 11.13 depicts a realization using the last two techniques. Here,

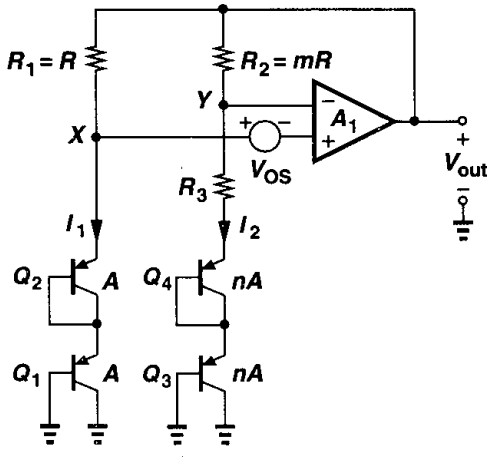


Figure 11.13 Reduction of the effect of op amp offset.

R_1 and R_2 are ratioed by a factor of m, producing $I_1 \approx mI_2$. Neglecting base currents and assuming A_1 is large, we can now write $V_{BE1} + V_{BE2} - V_{OS} = V_{BE3} + V_{BE4} + R_3 I_2$ and $V_{out} = V_{BE3} + V_{BE4} + (R_3 + R_2)I_2$. It follows that

$$V_{out} = V_{BE3} + V_{BE4} + (R_3 + R_2)\frac{2V_T \ln(mn) - V_{OS}}{R_3} \tag{11.29}$$

$$= 2V_{BE} + \left(1 + \frac{R_2}{R_3}\right)[2V_T \ln(mn) - V_{OS}]. \tag{11.30}$$

Thus, the effect of the offset voltage is reduced by increasing the first term in the square brackets. The issue, however, is that $V_{out} \approx 2 \times 1.25\ \text{V} = 2.5\ \text{V}$, a value difficult to generate by the op amp at low supply voltages.

The implementation of Fig. 11.13 is not feasible in a standard CMOS technology because the collectors of Q_2 and Q_4 are not grounded. In order to utilize the bipolar structure shown in Fig. 11.10, we modify the series combination of the diodes as illustrated in Fig. 11.14(a), converting one of the diodes to an emitter follower. However, we must ensure that the bias currents of both transistors have the same behavior with temperature. Thus, we bias each transistor by a PMOS current source rather than a resistor [Fig. 11.14(b)]. The overall circuit then assumes the topology shown in Fig. 11.15, where the op amp adjusts the gate voltage of the PMOS devices so as to equalize V_X and V_Y. Interestingly, in this circuit the op amp experiences no resistive loading, but the mismatch and channel-length modulation of the PMOS devices introduce error at the output [Problem 11.3(d)].

An important concern in the circuit of Fig. 11.15 is the relatively low current gain of the "native" *pnp* transistors. Since the base currents of Q_2 and Q_4 generate an error in

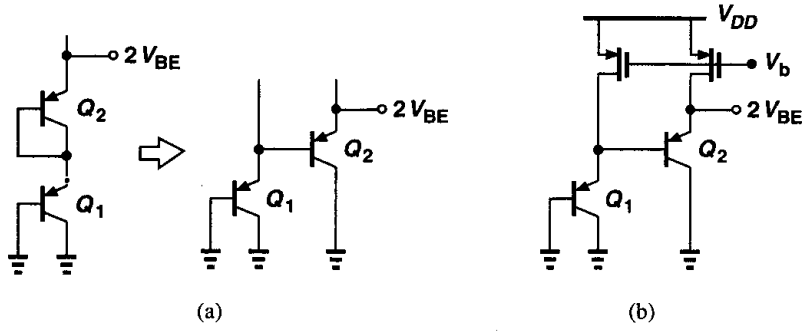


Figure 11.14 (a) Conversion of series diodes to a topology with grounded collectors, (b) circuit of part (a) biased by PMOS current sources.

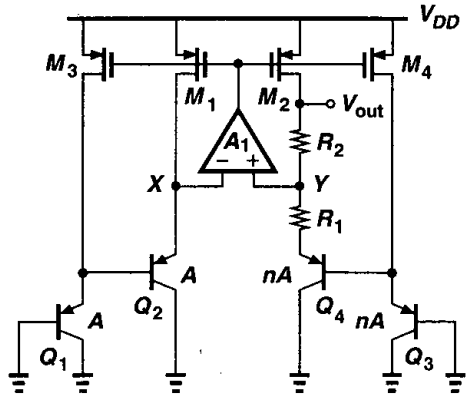


Figure 11.15 Reference generator incorporating two series base-emitter voltages.

the emitter currents of Q_1 and Q_3, a means of base current cancellation may be necessary (Problem 11.5).

Feedback Polarity In the circuit of Fig. 11.9, the feedback signal produced by the op amp returns to both of its inputs. The negative feedback factor is given by

$$\beta_N = \frac{1/g_{m2} + R_3}{1/g_{m2} + R_3 + R_2}, \tag{11.31}$$

and the positive feedback factor by

$$\beta_P = \frac{1/g_{m1}}{1/g_{m1} + R_1}. \tag{11.32}$$

To ensure an overall negative feedback, β_P must be less than β_N, preferably by roughly a factor of two so that the circuit's transient response remains well-behaved with large capacitive loads.

Bandgap Reference The voltage generated according to (11.20) is called a "bandgap reference." To understand the origin of this terminology, let us write the output voltage as

$$V_{REF} = V_{BE} + V_T \ln n \tag{11.33}$$

and hence:

$$\frac{\partial V_{REF}}{\partial T} = \frac{\partial V_{BE}}{\partial T} + \frac{V_T}{T} \ln n. \tag{11.34}$$

Setting this to zero and substituting for $\partial V_{BE}/\partial T$ from (11.13), we have

$$\frac{V_{BE} - (4+m)V_T - E_g/q}{T} = -\frac{V_T}{T} \ln n. \tag{11.35}$$

If $V_T \ln n$ is found from this equation and inserted in (11.33), we obtain:

$$V_{REF} = \frac{E_g}{q} + (4+m)V_T. \tag{11.36}$$

Thus, the reference voltage exhibiting a nominally-zero TC is given by a few *fundamental* numbers: the bandgap voltage of silicon, E_g/q, the temperature exponent of mobility, m, and the thermal voltage, V_T. The term "bandgap" is used here because as $T \to 0$, $V_{REF} \to E_g/q$.

Supply Dependence and Start-Up In the circuit of Fig. 11.9, the output voltage is relatively independent of the supply voltage so long as the open-loop gain of the op amp is sufficiently high. The circuit may require a start-up mechanism because if V_X and V_Y are equal to zero, the input differential pair of the op amp may turn off. Start-up techniques similar to those of Fig. 11.5 can be added to ensure the op amp turns on when the supply is applied.

The supply rejection of the circuit typically degrades at high frequencies owing to the op amp's rejection properties, often mandating "supply regulation." An example is described in Section 11.7.

Curvature Correction If plotted as a function of temperature, bandgap voltages exhibit a finite "curvature," i.e., their TC is typically zero at one temperature and positive or negative at other temperatures (Fig. 11.16). The curvature arises from temperature variation of base-emitter voltages, collector currents, and offset voltages.

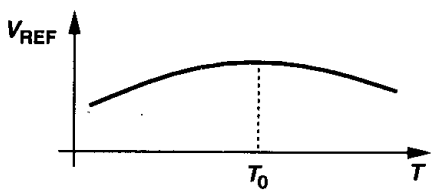


Figure 11.16 Curvature in temperature dependence of a bandgap voltage.

Many curvature correction techniques have been devised to suppress the variation of V_{REF} [5, 6] in bipolar bandgap circuits but they are seldom used in CMOS counterparts. This is because, due to large offsets and process variations, samples of a bandgap reference display substantially different zero-TC temperatures (Fig. 11.17), making it difficult to correct the curvature reliably.

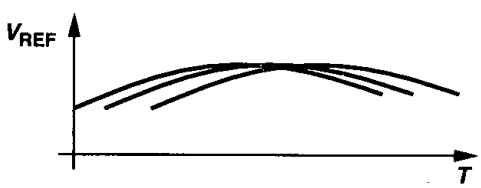


Figure 11.17 Variation of the zero-TC temperature for difference samples.

11.4 PTAT Current Generation

In the analysis of bandgap circuits, we noted that the bias currents of the bipolar transistors are in fact proportional to absolute temperature. Useful in many applications, PTAT currents can be generated by a topology such as that shown in Fig. 11.18. Alternatively, we can combine the supply-independent biasing scheme of Fig. 11.2 with a bipolar core, arriving at Fig. 11.19.[2] Assuming for simplicity that M_1-M_2 and M_3-M_4 are identical pairs, we note that for $I_{D1} = I_{D2}$, the circuit must ensure that $V_X = V_Y$. Thus, $I_{D1} = I_{D2} = (V_T \ln n)/R_1$, yielding the same behavior for I_{D5}. In practice, due to mismatches between the transistors and, more importantly, the temperature coefficient of R_1 the variation of I_{D5} deviates from the ideal equation.

The circuit of Fig. 11.19 can be readily modified to provide a bandgap reference voltage as well. Illustrated in Fig. 11.20, the idea is to add a PTAT voltage $I_{D5}R_2$ to a base-emitter

[2]The the two circuits in Figs. 11.18 and 11.19 exhibit difference supply rejections. With a carefully designed op amp, the former achieves a higher rejection.

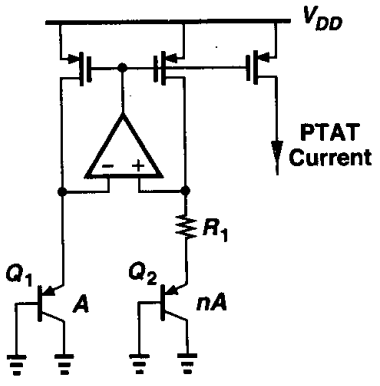


Figure 11.18 Generation of a PTAT current.

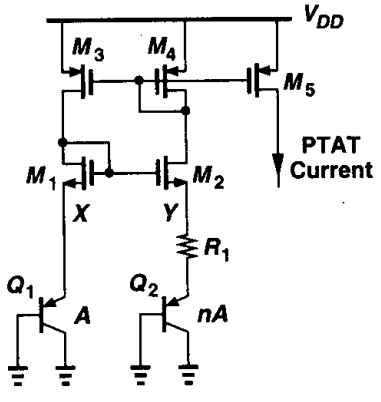


Figure 11.19 Generation of a PTAT current using a simple amplifier.

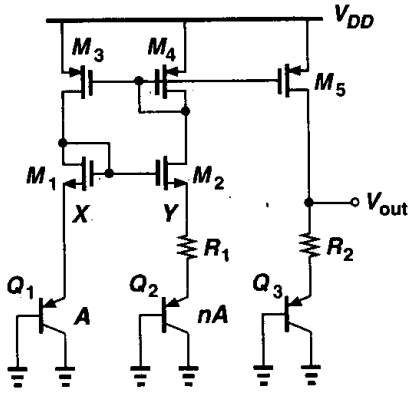


Figure 11.20 Generation of a temperature-independent voltage.

voltage. The output therefore equals

$$V_{REF} = V_{BE3} + \frac{R_2}{R_1} V_T \ln n, \tag{11.37}$$

where all PMOS transistors are assumed identical. Note that the value of V_{BE3} and hence the size of Q_3 are somewhat arbitrary so long as the sum of the two terms in (11.37) gives a zero TC. In reality, mismatches of the PMOS devices introduce error in V_{out}.

11.5 Constant-G_m Biasing

The transconductance of MOSFETs plays a critical role in analog circuits, determining such performance parameters as noise, small-signal gain, and speed. For this reason, it is often desirable to bias the transistors such that their transconductance does not depend on the temperature, process, or supply voltage.

A simple circuit used to define the transconductance is the supply-independent bias topology of Fig. 11.3. Recall that the bias current is given by

$$I_{out} = \frac{2}{\mu_n C_{ox} (W/L)_N} \frac{1}{R_S^2} \left(1 - \frac{1}{\sqrt{K}}\right)^2. \tag{11.38}$$

Thus, the transconductance of M_1 equals

$$g_{m1} = \sqrt{2\mu_n C_{ox} \left(\frac{W}{L}\right)_N I_{D1}} \tag{11.39}$$

$$= \frac{2}{R_S} \left(1 - \frac{1}{\sqrt{K}}\right), \tag{11.40}$$

a value independent of the supply voltage and MOS device parameters.

In reality, the value of R_S in (11.40) does vary with temperature and process. If the temperature coefficient of the resistor is known, bandgap and PTAT reference generation techniques can be utilized to cancel the temperature dependence. Process variations, however, limit the accuracy with which g_{m1} is defined.

In systems where a precise clock frequency is available, the resistor R_S in Fig. 11.3 can be replaced by a switched-capacitor equivalent (Chapter 12) to achieve a somewhat higher accuracy. Depicted in Fig. 11.21, the idea is to establish an average resistance equal to $(C_S f_{CK})^{-1}$ between the source of M_2 and ground, where f_{CK} denotes the clock frequency. Capacitor C_B is added to shunt the high-frequency components resulting from switching to ground. Since the absolute value of capacitors is typically more tightly controlled and since the TC of capacitors is much smaller than that of resistors, this technique provides a higher reproducibility in the bias current and transconductance.

The switched-capacitor approach of Fig. 11.21 can be applied to other circuits as well. For example, as shown in Fig. 11.22, a voltage-to-current converter with a relatively high accuracy can be constructed.

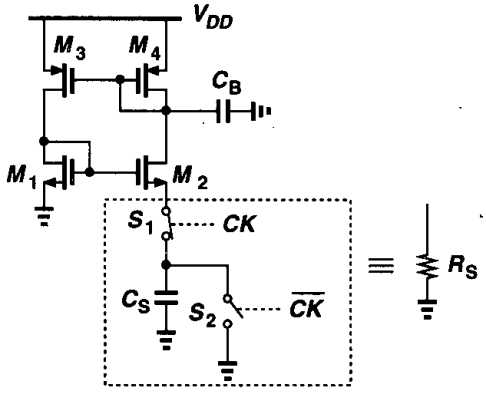


Figure 11.21 Constant-G_m biasing by means of a switched-capacitor "resistor."

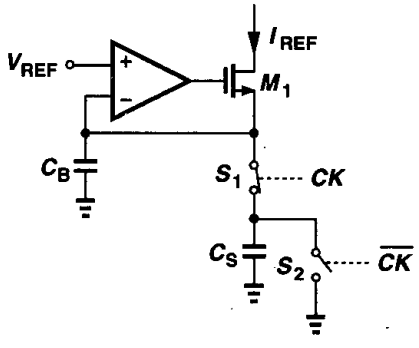


Figure 11.22 Voltage-to-current conversion by means of a switched-capacitor resistor.

11.6 Speed and Noise Issues

Even though reference generators are low-frequency circuits, they may impact the speed of the circuits that they feed. Furthermore, various building blocks may experience "crosstalk" through reference lines. These difficulties arise because of the finite output impedance of reference voltage generators, especially if they incorporate op amps. As an example, let us consider the configuration shown in Fig. 11.23, assuming the voltage at node N is heavily disturbed by the circuit fed by M_5. For fast changes in V_N, the op amp cannot maintain V_P constant and the bias currents of M_5 and M_6 experience large transient changes. Also, the duration of the transient at node P may be quite long if the op amp suffers from a slow response. For this reason, many applications may require a high-speed op amp in the reference generator.

In systems where the power consumed by the reference circuit must be small, the use of a high-speed op amp may not be feasible. Alternatively, the critical node, e.g., node P in Fig. 11.23, can be bypassed to ground by means of a large capacitor (C_B) so as to suppress the effect of external disturbances. This approach involves two issues. First, the stability of the op amp must not degrade with the addition of the capacitor, requiring the op amp to be of one-stage nature (Chapter 10). Second, since C_B generally slows down the transient response of the op amp, its value must be much greater than the capacitance

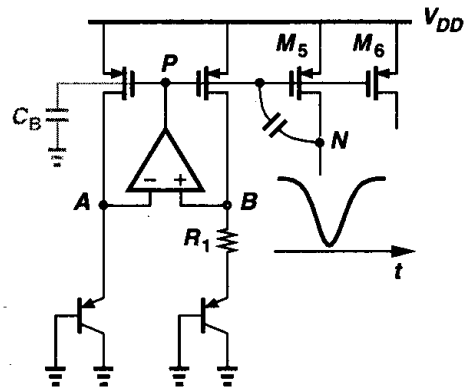


Figure 11.23 Effect of circuit transients on reference voltages and currents.

that couples the disturbance to node P. As illustrated in Fig. 11.24, if C_B is not sufficiently large, then V_P experiences a change and takes a long time to return to its original value, possibly degrading the settling speed of the circuits biased by the reference generator. In other words, depending on the environment, it may be preferable to leave node P agile so that it can quickly recover from transients. In general, as depicted in Fig. 11.25, the response of the circuit must be analyzed by applying a disturbance at the output and observing the settling behavior.

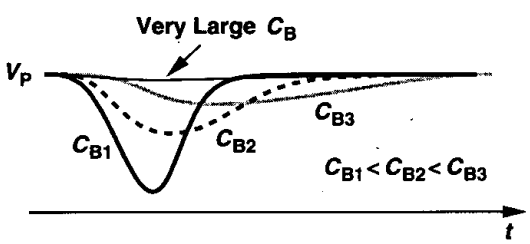


Figure 11.24 Effect of increasing bypass capacitor on the response of reference generator.

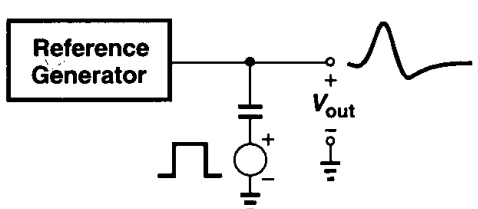


Figure 11.25 Setup for testing the transient response of a reference generator.

Example 11.3

Determine the small-signal output impedance of the bandgap reference shown in Fig. 11.23 and examine its behavior with frequency.

Solution

Fig. 11.26 depicts the equivalent circuit, modeling the open-loop op amp by a one-pole transfer function $A(s) = A_0/(1 + s/\omega_0)$ and an output resistance R_{out} and each bipolar transistor by a

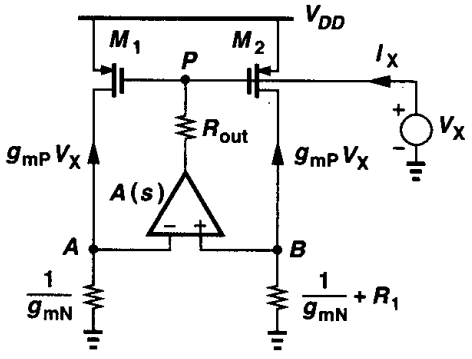


Figure 11.26 Circuit for calculation of the output impedance of a reference generator.

resistance $1/g_{mN}$. If M_1 and M_2 are identical, each having a transconductance of g_{mP}, then their drain currents are equal to $g_{mP}V_X$, producing a differential voltage at the input of the op amp equal to

$$V_{AB} = -g_{mP}V_X\frac{1}{g_{mN}} + g_{mP}V_X\left(\frac{1}{g_{mN}} + R_1\right) \tag{11.41}$$

$$= g_{mP}V_X R_1. \tag{11.42}$$

The current flowing through R_{out} is therefore given by

$$I_X = \frac{V_X + g_{mP}V_X R_1 A(s)}{R_{out}}, \tag{11.43}$$

yielding

$$\frac{V_X}{I_X} = \frac{R_{out}}{1 + g_{mP}R_1 A(s)} \tag{11.44}$$

$$= \frac{R_{out}}{1 + g_{mP}R_1\dfrac{A_0}{1 + s/\omega_0}} \tag{11.45}$$

$$= \frac{R_{out}}{1 + g_{mP}R_1 A_0}\,\frac{1 + \dfrac{s}{\omega_0}}{1 + \dfrac{s}{(1 + g_{mP}R_1 A_0)\omega_0}}. \tag{11.46}$$

Thus, the output impedance exhibits a zero at ω_0 and a pole at $(1 + g_{mP}R_1A_0)\omega_0$, with the magnitude behavior plotted in Fig. 11.27. Note that $|Z_{out}|$ is quite low for $\omega < \omega_0$, but it rises to a high value as the frequency approaches the pole. In fact, setting $\omega = (1 + g_{mP}R_1A_0)\omega_0$ and assuming $g_{mP}R_1A_0 \gg 1$, we have

$$|Z_{out}| = \frac{R_{out}}{1 + g_{mP}R_1A_0}\left|\frac{1 + j(1 + g_{mP}R_1A_0)}{1 + j}\right| \tag{11.47}$$

$$= \frac{R_{out}}{\sqrt{2}}, \tag{11.48}$$

which is only 30% lower than the open-loop value.

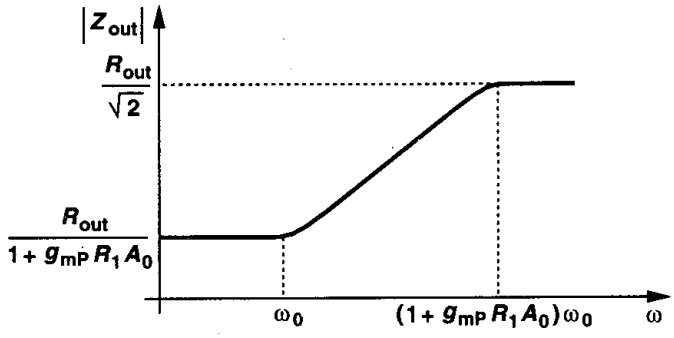


Figure 11.27 Variation of the reference generator output impedance with frequency.

The output noise of reference generators may impact the performance of low-noise circuits considerably. For example, if a high-precision A/D converter employs a bandgap voltage as the reference with which the analog input signal is compared (Fig. 11.28), then the noise in the reference is directly added to the input.

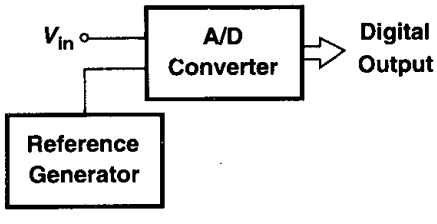


Figure 11.28 A/D converter using a reference generator.

As a simple example, let us calculate the output noise voltage of the circuit shown in Fig. 11.29, taking into account only the input-referred noise voltage of the op amp, $V_{n,op}$. Since the small-signal drain currents of M_1 and M_2 are equal to $V_{n,out}/(R_1 + g_{mN}^{-1})$, we have $V_P = -g_{mP}^{-1} V_{n,out}/(R_1 + g_{mN}^{-1})$, obtaining the differential voltage at the input of the op amp as $-g_{mP}^{-1} A_0^{-1} V_{n,out}/(R_1 + g_{mN}^{-1})$. Beginning from node A, we can then write

$$\frac{V_{n,out}}{R_1 + g_{mN}^{-1}} \cdot \frac{1}{g_{mN}} - \frac{V_{n,out}}{g_{mP} A_0 (R_1 + g_{mN}^{-1})} = V_{n,op} + V_{n,out} \tag{11.49}$$

and hence

$$V_{n,out} \left[\frac{1}{R_1 + g_{mN}^{-1}} \left(\frac{1}{g_{mN}} - \frac{1}{g_{mP} A_0} \right) - 1 \right] = V_{n,op}. \tag{11.50}$$

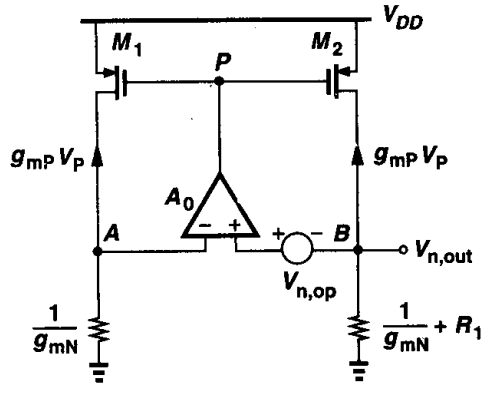


Figure 11.29 Circuit for calculation of noise in a reference generator.

Since typically $g_{mP}A_0 \gg g_{mN} \gg R_1^{-1}$,

$$|V_{n,out}| \approx V_{n,op}, \tag{11.51}$$

suggesting that the noise of the op amp directly appears at the output. Note that even the addition of a large capacitor from the output to ground may not suppress low-frequency $1/f$ noise components, a serious difficulty in low-noise applications. The noise contributed by other devices in the circuit is studied in Problem 11.6.

11.7 Case Study

In this section, we study a bandgap reference circuit designed for high-precision analog systems [7]. The reference generator incorporates the topology of Fig. 11.19 but with two series base-emitter voltages in each branch so as to reduce the effect of MOSFET mismatches. A simplified version of the core is depicted in Fig. 11.30, where the PMOS current mirror arrangement ensures equal collector currents for Q_1-Q_4.

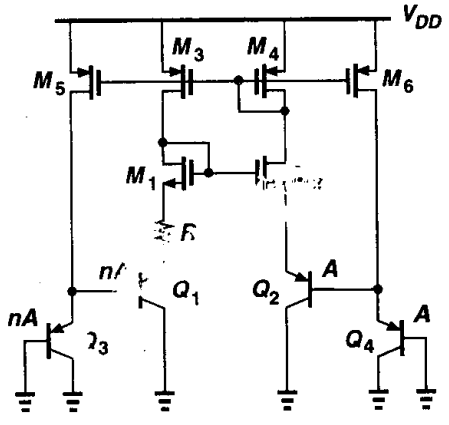


Figure 11.30 Simplified core of the bandgap circuit reported in [7].

Channel-length modulation of the MOS devices in Fig. 11.30 still results in significant supply dependence. To resolve this issue, each branch can employ both NMOS and PMOS cascode topologies. Fig. 11.31(a) shows an example where the low-voltage cascode current mirror described in Chapter 5 is utilized. To obviate the need for V_{b1} and V_{b2}, this design actually introduces a "self-biased" cascode, shown in Fig. 11.31(b), where R_2 and R_3 sustain proper voltages to allow all MOSFETs to remain in saturation. This cascode topology is analyzed in Problem 11.7.

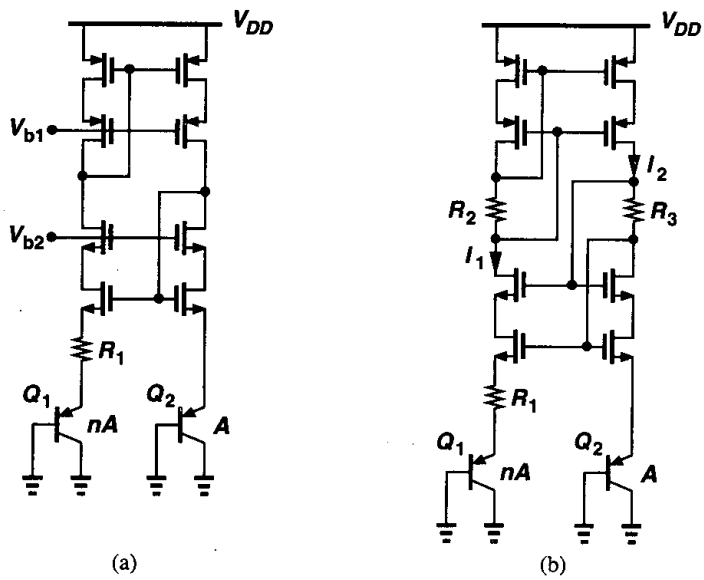


Figure 11.31 (a) Addition of cascode devices to improve supply rejection, (b) use of self-biased cascode to eliminate V_{b1} and V_{b2}.

The bandgap circuit reported in [7] is designed to generate a *floating* reference. This is accomplished by the modification shown in Fig. 11.32, where the drain currents of M_9 and M_{10} flow through R_4 and R_5, respectively. Note that M_{11} sets the gate voltage of M_9 at $V_{BE4} + V_{GS11}$, establishing a voltage equal to V_{BE4} across R_6 if M_9 and M_{11} are identical. Thus, $I_{D9} = V_{BE4}/R_6$, yielding $V_{R4} = V_{BE4}(R_4/R_6)$. Also, if M_{10} is identical to M_2, then $|I_{D10}| = 2(V_T \ln n)/R_1$ and hence $V_{R5} = 2(V_T \ln n)(R_5/R_1)$. Since the op amp ensures that $V_E \approx V_F$, we have

$$V_{out} = \frac{R_4}{R_6} V_{BE4} + 2\frac{R_5}{R_1} V_T \ln n. \tag{11.52}$$

Proper choice of the resistor ratios and n therefore provides a zero temperature coefficient.

In order to further enhance the supply rejection, this design regulates the supply voltage of the core and the op amp. Illustrated in Fig. 11.33, the idea is to generate a local supply, V_{DDL}, that is defined by a reference V_{R1} and the ratio of R_{r1} and R_{r2} and hence

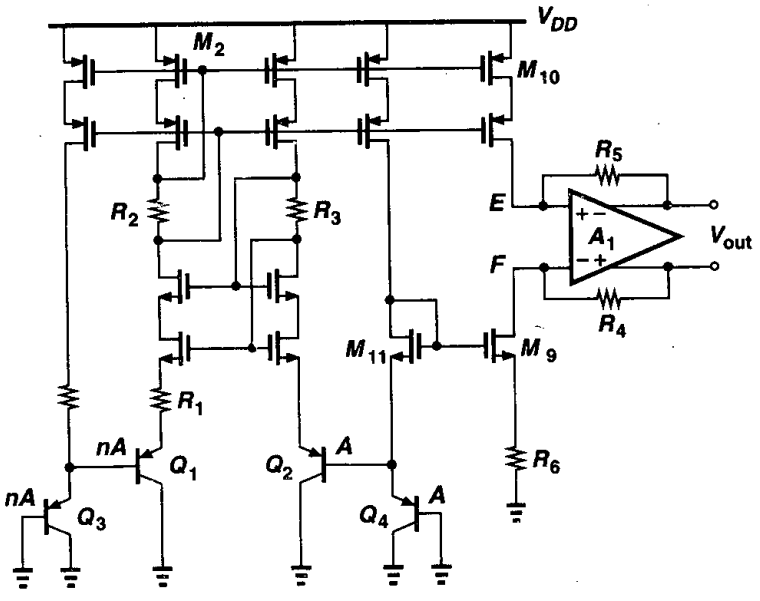


Figure 11.32 Generation of a floating reference voltage.

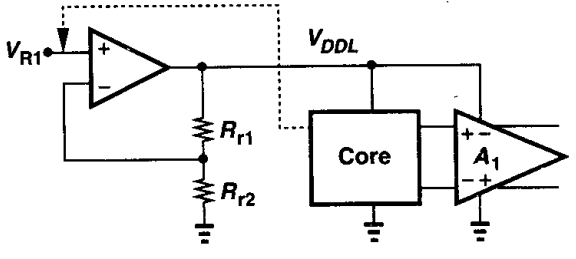


Figure 11.33 Regulation of the supply voltage of the core and op amp to improve supply rejection.

remains relatively independent of the global supply voltage. But how is V_{R1} itself generated? To minimize the dependence of V_{R1} upon the supply, this voltage is established *inside* the core, as depicted in Fig. 11.34. In fact, R_M is chosen such that V_{R1} is a bandgap reference.

Fig. 11.35 shows the overall implementation, omitting a few details for simplicity. A start-up circuit is also used. Operating from a 5-V supply, the reference generator produces a 2.00-V output while consuming 2.2 mW. The supply rejection is 94 dB at low frequencies, dropping to 58 dB at 100 kHz [7].

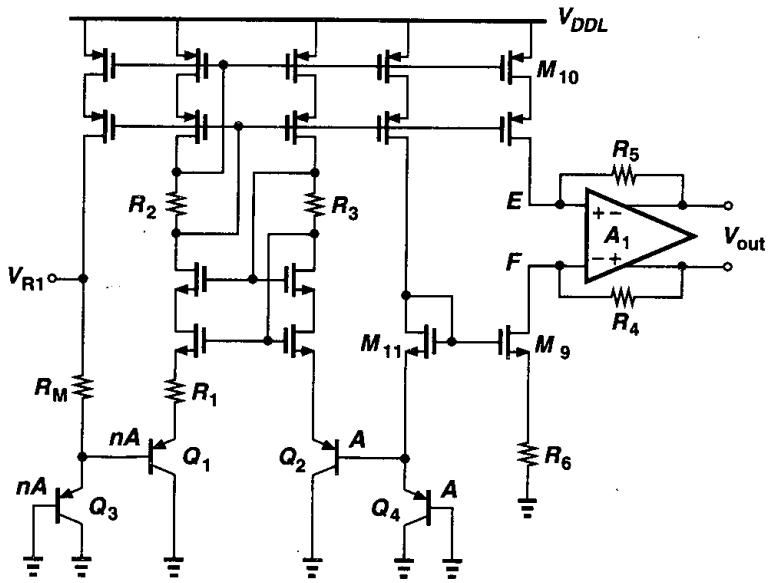


Figure 11.34 Generation of V_{R1}, used in Fig. 11.33.

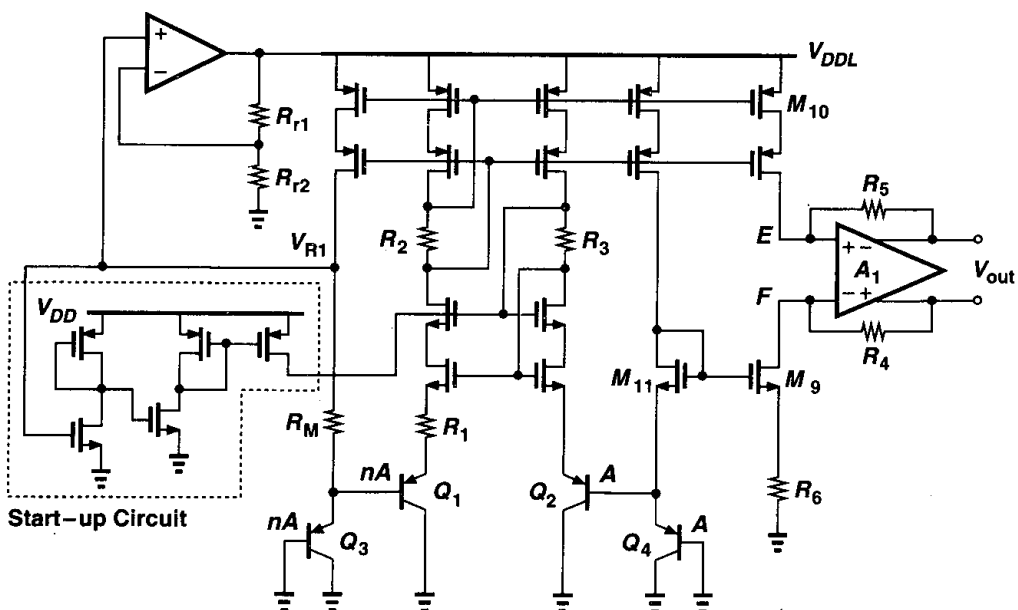


Figure 11.35 Overall circuit of the bandgap generator reported in [7].

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.

11.1. Derive an expression for I_{out} in Fig. 11.36.

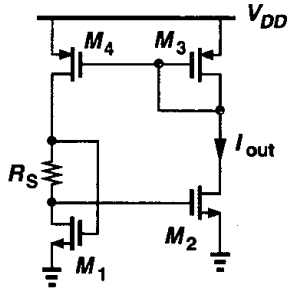


Figure 11.36

11.2. Explain how the start-up circuit shown in Fig. 11.37 operates. Derive a relationship that guarantees $V_X < V_{TH}$ after the circuit turns on.

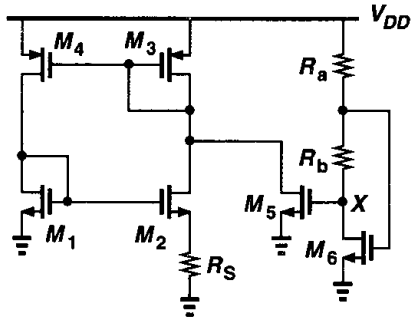


Figure 11.37

11.3. Consider the circuit of Fig. 11.15.

(a) If M_1 and M_2 suffer from channel-length modulation, what is the error in the output voltage?

(b) Repeat part (a) for M_3 and M_4.

(c) If M_1 and M_2 have a threshold mismatch of ΔV, i.e., $V_{TH1} = V_{TH}$ and $V_{TH2} = V_{TH} + \Delta V$, what is the error in the output voltage?

(d) Repeat part (c) for M_3 and M_4.

11.4. In Fig. 11.15, if the open-loop gain of the op amp A_1 is not sufficiently large, then $|V_X - V_Y|$ exceeds V_e, where V_e is the maximum tolerable error. Calculate the minimum value of A_1 in terms of V_e such that the condition $|V_X - V_Y| < V_e$ is satisfied.

11.5. In the circuit of Fig. 11.15, assume Q_2 and Q_4 have a finite current gain β. Calculate the error in the output voltage.

11.6. Calculate the output noise voltage of the circuit shown in Fig. 11.29 due to the thermal and flicker noise of M_1 and M_2.

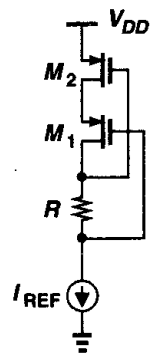


Figure 11.38

11.7. Consider the self-biased cascode shown in Fig. 11.38. Determine the minimum and maximum values of RI_{REF} such that both M_1 and M_2 remain in saturation.

11.8. The circuit of Fig. 11.3(a) sometimes turns on even with no explicit start-up mechanism. Identify the capacitive path(s) that couple the transition on V_{DD} to the internal nodes and hence provide the start-up current.

11.9. Sketch the temperature coefficient of V_{BE} [Eq. (11.13)] versus temperature. Some iteration may be necessary.

11.10. Determine the derivative of Eq. (11.13) with respect to temperature and sketch the result versus T. This quantity reveals the curvature of the voltage.

11.11. Suppose in Fig. 11.9 the amplifier has an output resistance R_{out}. Calculate the error in V_{out}.

11.12. The circuit of Fig. 11.9 is designed with $R_3 = 1$ kΩ and a current of 50 μA through it. Calculate $R_1 = R_2$ and n for a zero TC.

11.13. In the circuit of Fig. 11.15, Q_1 and Q_2 are biased at 100 μA and Q_3 and Q_4 at 50 μA. If $R_1 = 1$ kΩ, calculate R_2 and $(W/L)_{1-4}$ such that the circuit operates with $V_{DD} = 3$ V. Which op amp topology can be used here?

11.14. Since the bandgap of silicon exhibits a small temperature coefficient, Eq. (11.36) suggests that $\partial V_{REF}/\partial T \propto (4+m)k/q$, a relatively large value, whereas we derived V_{REF} such that it has a zero TC. Explain the flaw in this argument.

11.15. A differential pair with resistive loads is designed such that its voltage gain, $g_m R_D$, has a zero TC at room temperature. If only the temperature dependence of the mobility is considered, determine the required temperature behavior of the tail current. Design a circuit that roughly approximates this behavior.

11.16. In Problem 11.15, assume the tail current is constant but the load resistors exhibit a finite TC. What resistor temperature coefficient cancels the variation of the mobility at room temperature?

11.17. Equation (11.36) suggests that a zero-TC voltage cannot be generated if the supply voltage is as low as, say, 1 V. Figure 11.39 shows a bandgap reference that can operate with low supply voltages [8]. If $R_2 = R_3$, derive an expression for V_{out}.

11.18. Repeat Problem 11.17, if the op amp has an offset voltage V_{OS}.

11.19. Figure 11.40 illustrates a "single-junction" bandgap design [9]. Here, switches S_1 and S_2 are driven by complementary clocks.

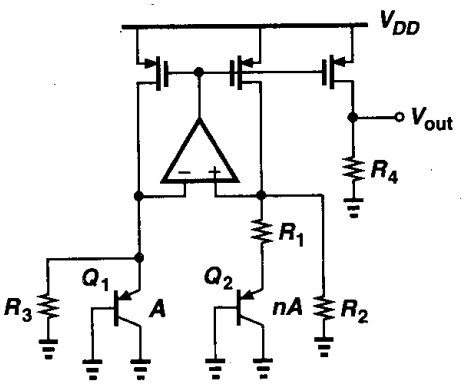


Figure 11.39

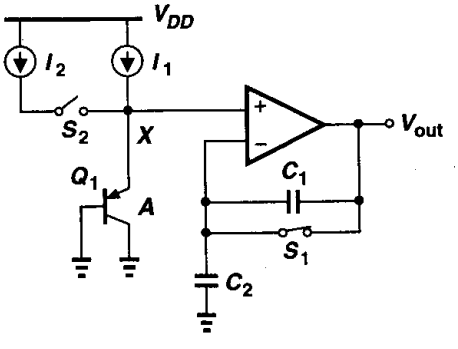


Figure 11.40

(a) What is V_{out} when S_1 is on and S_2 is off?
(b) What is the change in V_{out} when S_1 turns off and S_2 turns on?
(c) How are I_1, I_2, C_1, and C_2 chosen to produce a zero-TC output when S_1 is off?

11.20. Suppose in Fig. 11.40, I_2/I_1 deviates from its nominal value by a small error ϵ. Calculate V_{out} when S_1 is off.

11.21. The circuit of Fig. 11.20 is designed with $(W/L)_{1-4} = 50/0.5$, $I_{D1} = I_{D2} = 50\ \mu$A, $R_1 = 1\ \text{k}\Omega$, and $R_2 = 2\ \text{k}\Omega$. Assume $\lambda = \gamma = 0$ and Q_3 is identical to Q_1.
(a) Determine n and $(W/L)_5$ such that V_{out} has a zero TC at room temperature.
(b) Neglecting the noise contribution of Q_1-Q_3, calculate the output thermal noise.

11.22. Consider the circuit of Fig. 11.21. Assume $K = 4$, $f_{CK} = 50$ MHz, and a power budget of 1 mW. Determine the aspect ratio of M_1-M_4 and the value of C_S such that $g_{m1} = 1/(500\ \Omega)$.

References

1. R. A. Blauschild et al., "A New NMOS Temperature-Stable Voltage Reference," *IEEE J. of Solid-State Circuits*, vol. 13, pp. 767–774, Dec. 1978.
2. Y. P. Tsividis and R. W. Ulmer, "A CMOS Voltage Reference," *IEEE J. of Solid-State Circuits*, vol. 13, pp. 774–778, Dec. 1978.

3. D. Hilbiber, "A New Semiconductor Voltage Standard," *ISSCC Dig. of Tech. Papers,* pp. 32–33, Feb. 1964.
4. K. E. Kujik, "A Precision Reference Voltage Source," *IEEE J. of Solid-State Circuits,* vol. 8, pp. 222–226, June 1973.
5. G. C. M. Meijer, P. C. Schmall and K. van Zalinge, "A New Curvature-Corrected Bandgap Reference," *IEEE J. of Solid-State Circuits,* vol. 17, pp. 1139–1143, Dec. 1982.
6. M. Gunawan et al., "A Curvature-Corrected Low-Voltage Bandgap Reference," *IEEE J. of Solid-State Circuits,* vol. 28, pp. 667–670, June 1993.
7. T. Brooks and A. L. Westwisk, "A Low-Power Differential CMOS Bandgap Reference," *ISSCC Dig. of Tech. Papers,* pp. 248–249, Feb. 1994.
8. H. Banba et al., "A CMOS Bandgap Reference Circuit with Sub-1-V Operation," *IEEE J. of Solid-State Circuits,* vol. 34, pp. 670–674, May 1999.
9. B. Gilbert, "Monolithic Voltage and Current References: Themes and Variations," pp. 269–352 in *Analog Circuit Design,* J. H. Huijsing, R. J. van de Plassche, and W. M. C. Sansen, Editors, Boston: Kluwer Academic Publishers, 1996.

Chapter 12

Introduction to Switched-Capacitor Circuits

Our study of amplifiers in previous chapters has dealt only with cases where the input signal is continuously available and applied to the circuit and the output signal is continuously observed. Called "continuous-time" circuits, such amplifiers find wide application in audio, video, and high-speed analog systems. In many situations, however, we may sense the input only at periodic instants of time, ignoring its value at other times. The circuit then processes each "sample," producing a valid output at the end of each period. Such circuits are called "discrete-time" or "sampled-data" systems.

In this chapter, we study a common class of discrete-time systems called "switched-capacitor (SC) circuits." Our objective is to provide the foundation for more advanced topics such as filters, comparators, ADCs, and DACs. Most of our study deals with switched-capacitor amplifiers but the concepts can be applied to other discrete-time circuits as well. Beginning with a general view of SC circuits, we describe sampling switches and their speed and precision issues. Next, we analyze switched-capacitor amplifiers, considering unity-gain, noninverting, and multiply-by-two topologies. Finally, we examine a switched-capacitor integrator.

12.1 General Considerations

In order to understand the motivation for sampled-data circuits, let us first consider the simple continuous-time amplifier shown in Fig. 12.1(a). Used extensively with bipolar op amps, this circuit presents a difficult issue if implemented in CMOS technology. Recall that, to achieve a high voltage gain, the open-loop output resistance of CMOS op amps is maximized, typically approaching hundreds of kilo-ohms. We therefore suspect that R_2 heavily drops the open-loop gain, degrading the precision of the circuit. In fact, with the aid of the simple equivalent circuit shown in Fig. 12.1(b), we can write

$$-A_v\left(\frac{V_{out}-V_{in}}{R_1+R_2}R_1+V_{in}\right)-R_{out}\frac{V_{out}-V_{in}}{R_1+R_2}=V_{out}, \tag{12.1}$$

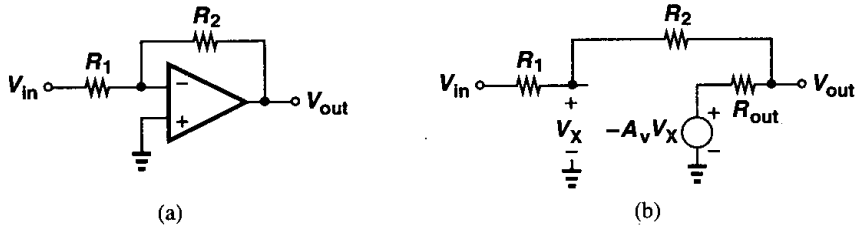


Figure 12.1 (a) Continuous-time feedback amplifier, (b) equivalent circuit of (a).

and hence

$$\frac{V_{out}}{V_{in}} = -\frac{R_2}{R_1} \cdot \frac{A_v - \dfrac{R_{out}}{R_2}}{1 + \dfrac{R_{out}}{R_1} + A_v + \dfrac{R_2}{R_1}}. \tag{12.2}$$

Equation (12.2) implies that, compared to the case where $R_{out} = 0$, the closed-loop gain suffers from inaccuracies in both the numerator and the denominator. Also, the input resistance of the amplifier, approximately equal to R_1, loads the preceding stage while introducing thermal noise.

Example 12.1

Using the feedback techniques described in Chapter 8, calculate the closed-loop gain of the circuit of Fig. 12.1(a) and compare the result with Eq. (12.2).

Solution

With the aid of the approach described in Example 8.9, the reader can prove that

$$\frac{V_{out}}{V_{in}} = \frac{-R_2^2 A_v}{R_2^2 + R_1 R_{out} + R_2 R_{out} + (1 + A_v) R_1 R_2} \tag{12.3}$$

$$= -\frac{R_2}{R_1} \cdot \frac{A_v}{\dfrac{R_2}{R_1} + \dfrac{R_{out}}{R_2} + \dfrac{R_{out}}{R_1} + 1 + A_v}. \tag{12.4}$$

The two results are approximately equal if $R_{out}/R_2 \ll A_v$, a condition required to ensure the transmission through R_2 is negligible.

In the circuit of Fig. 12.1(a), the closed-loop gain is set by the ratio of R_2 and R_1. In order to avoid reducing the open-loop gain of the op amp, we postulate that the resistors can be replaced by capacitors [Fig. 12.2(a)]. But, how is the bias voltage at node X set? We may add a large feedback resistor as in Fig. 12.2(b), providing dc feedback while negligibly affecting the ac behavior of the amplifier in the frequency band of interest. Such an arrangement is indeed practical if the circuit senses *only* high-frequency signals. But suppose, for example,

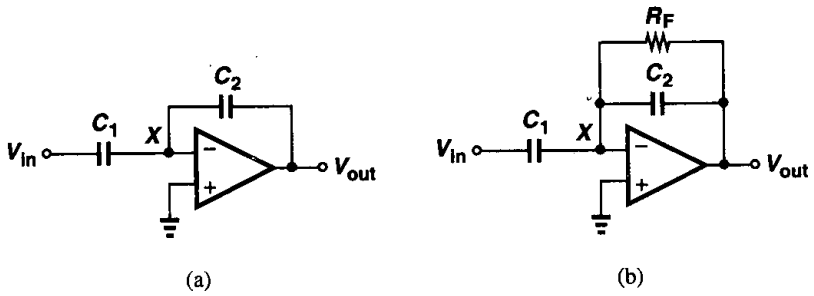


Figure 12.2 (a) Continuous-time feedback amplifier using capacitors, (b) use of resistor to define bias point.

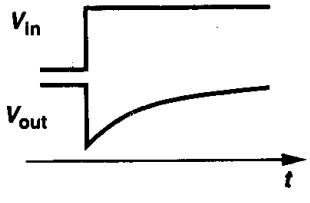


Figure 12.3 Step response of the amplifier of Fig. 12.2(b).

the circuit is to amplify a voltage step. Illustrated in Fig. 12.3, the response contains a step change due to the initial amplification by the circuit consisting of C_1, C_2, and the op amp, followed by a "tail" resulting from the loss of charge on C_2 through R_F. From another point of view, the circuit may not be suited to amplify *wideband* signals because it exhibits a high-pass transfer function. In fact, the transfer function is given by

$$\frac{V_{out}}{V_{in}}(s) \approx -\frac{R_F \dfrac{1}{C_2 s}}{R_F + \dfrac{1}{C_2 s}} \div \frac{1}{C_1 s} \tag{12.5}$$

$$= -\frac{R_F C_1 s}{R_F C_2 s + 1}, \tag{12.6}$$

indicating that $V_{out}/V_{in} \approx -C_1/C_2$ only if $\omega \gg (R_F C_2)^{-1}$.

The above difficulty can be remedied by increasing $R_F C_2$, but in many applications the required values of the two components become prohibitively large. We must therefore seek other methods of establishing the bias while utilizing capacitive feedback networks.

Let us now consider the switched-capacitor circuit depicted in Fig. 12.4, where three switches control the operation: S_1 and S_3 connect the left plate of C_1 to V_{in} and ground, respectively, and S_2 provides unity-gain feedback. We first assume the open-loop gain of the op amp is very large and study the circuit in two phases. First, S_1 and S_2 are on and S_3 is off, yielding the equivalent circuit of Fig. 12.5(a). For a high-gain op amp, $V_B = V_{out} \approx 0$, and hence the voltage across C_1 is approximately equal to V_{in}. Next, at $t = t_0$, S_1 and S_2

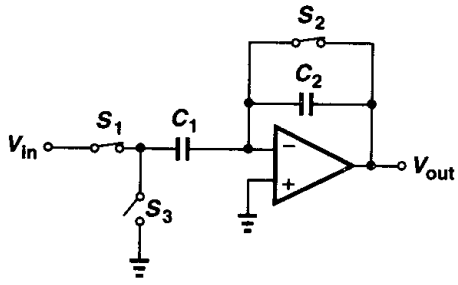


Figure 12.4 Switched-capacitor amplifier.

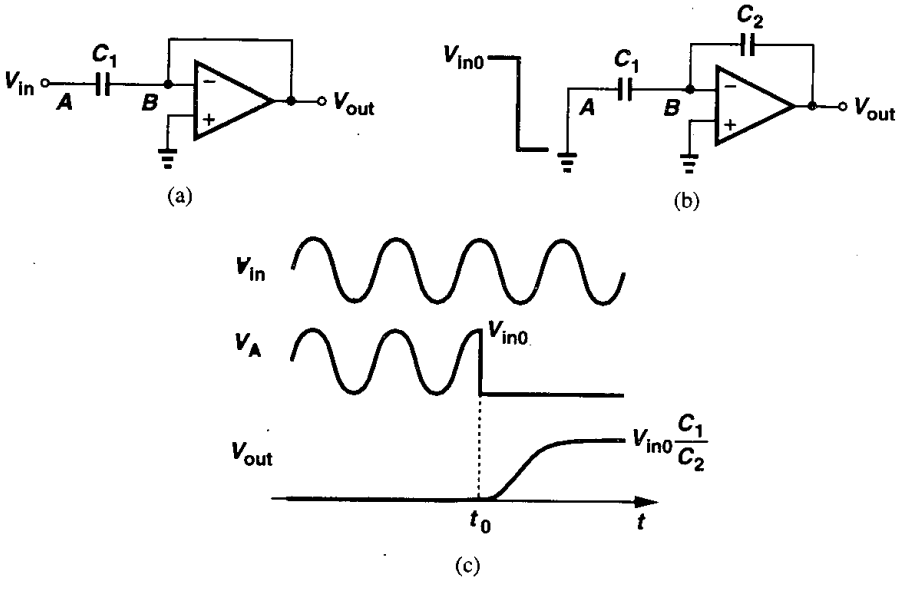


Figure 12.5 Circuit of Fig. 12.4 in (a) sampling mode, (b) amplification mode.

turn off and S_3 turns on, pulling node A to ground. Since V_A changes from V_{in0} to 0, the output voltage must change from zero to $V_{in0}C_1/C_2$.

The output voltage change can also be calculated by examining the transfer of charge. Note that the charge stored on C_1 just before t_0 is equal to $V_{in0}C_1$. After $t = t_0$, the negative feedback through C_2 drives the op amp input differential voltage and hence the voltage across C_1 to zero (Fig. 12.6). The charge stored on C_1 at $t = t_0$ must then be transferred to C_2, producing an output voltage equal to $V_{in0}C_1/C_2$. Thus, the circuit amplifies V_{in0} by a factor of C_1/C_2.

Several attributes of the circuit of Fig. 12.4 distinguish it from continuous-time implementations. First, the circuit devotes some time to "sample" the input, setting the output to zero and providing no amplification during this period. Second, after sampling, for $t > t_0$, the circuit ignores the input voltage V_{in}, amplifying the sampled voltage. Third, the circuit

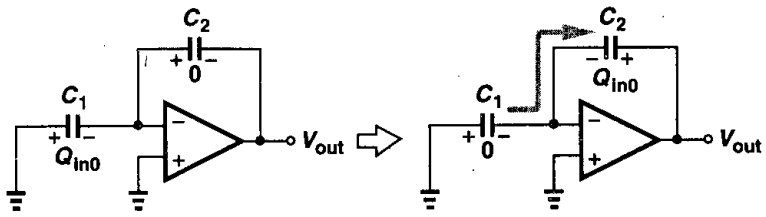


Figure 12.6 Transfer of charge from C_1 to C_2.

configuration changes considerably from one phase to another, as seen in Fig. 12.5(a) and (b), raising concern about its stability.

What is the advantage of the amplifier of Fig. 12.4 over that in Fig. 12.1? In addition to sampling capability, we note from the waveforms depicted in Fig. 12.5 that after V_{out} settles, the current through C_2 approaches zero. That is, the feedback capacitor does not reduce the open-loop gain of the amplifier if the output voltage is given enough time to settle. In Fig. 12.1, on the other hand, R_2 continuously loads the amplifier.

The switched-capacitor amplifier of Fig. 12.4 lends itself to implementation in CMOS technology much more easily than in other technologies. This is because discrete-time operations require switches to perform sampling as well as a high input impedance to sense the stored quantities with no corruption. For example, if the op amp of Fig. 12.4 incorporates bipolar transistors at its input, the base current drawn from the inverting input in the amplification phase [Fig. 12.5(b)] creates an error in the output voltage. The existence of simple switches and a high input impedance have made CMOS technology the dominant choice for sampled-data applications.

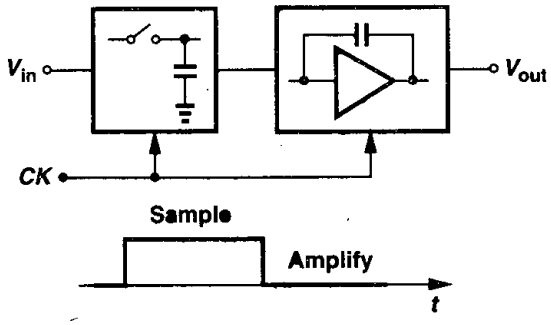


Figure 12.7 General view of switched-capacitor amplifier.

The foregoing discussion leads to the conceptual view illustrated in Fig. 12.7 for switched-capacitor amplifiers. In the simplest case, the operation takes place in two phases: sampling and amplification. Thus, in addition to the analog input, V_{in}, the circuit requires a clock to define each phase.

Our study of SC amplifiers proceeds according to these two phases. First, we analyze various sampling techniques. Second, we consider SC amplifier topologies.

12.2 Sampling Switches

12.2.1 MOSFETS as Switches

A simple sampling circuit consists of a switch and a capacitor [Fig. 12.8(a)]. A MOS transistor can serve as a switch [Fig. 12.8(b)] because (a) it can be on while carrying zero

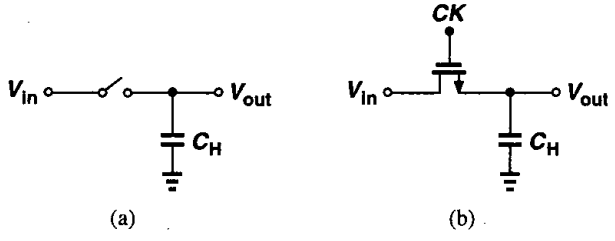


Figure 12.8 (a) Simple sampling circuit, (b) implementation of the switch by a MOS device.

current, and (b) its source and drain voltages are not "pinned" to the gate voltage, i.e., if the gate voltage varies, the source or drain voltage need not follow that variation. By contrast, bipolar transistors lack both of these properties, typically necessitating complex circuits to perform sampling.

To understand how the circuit of Fig. 12.8(b) samples the input, first consider the simple cases depicted in Fig. 12.9, where the gate command, CK, goes high at $t = t_0$. In Fig. 12.9(a), we assume that $V_{in} = 0$ for $t \geq t_0$ and the capacitor has an initial voltage equal to V_{DD}.

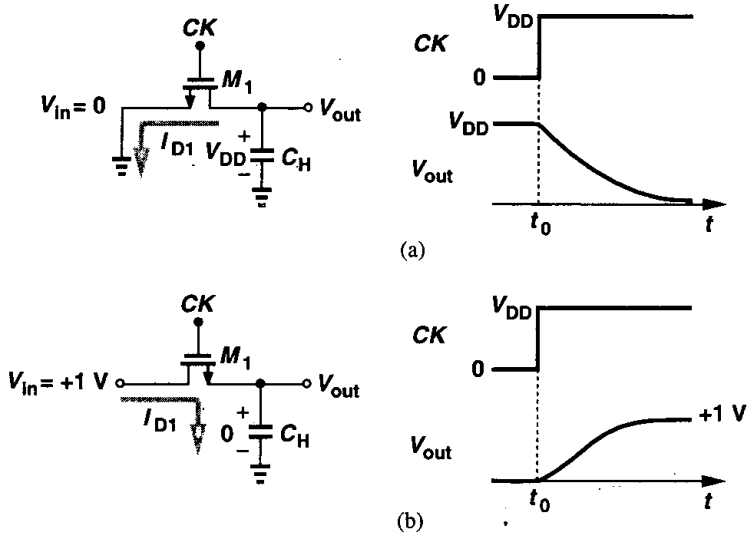


Figure 12.9 Response of a sampling circuit to different input levels and initial conditions.

Thus, at $t = t_0$, M_1 senses a gate-source voltage equal to V_{DD} while its drain voltage is also equal to V_{DD}. The transistor therefore operates in saturation, drawing a current of $I_{D1} = (\mu_n C_{ox}/2)(W/L)(V_{DD} - V_{TH})^2$ from the capacitor. As V_{out} falls, at some point $V_{out} = V_{DD} - V_{TH}$, driving M_1 into the triode region. The device nevertheless continues to discharge C_H until V_{out} approaches zero. We note that for $V_{out} \ll 2(V_{DD} - V_{TH})$, the transistor can be viewed as a resistor equal to $R_{on} = [\mu_n C_{ox}(W/L)(V_{DD} - V_{TH})]^{-1}$.

Now consider the case in Fig. 12.9(b), where $V_{in} = +1$ V, $V_{out}(t = t_0) = 0$ V, and $V_{DD} = 3$ V. Here, the terminal of M_1 connected to C_H acts as the source, and the transistor turns on with $V_{GS} = +3$ V, but $V_{DS} = +1$ V. Thus, M_1 operates in the triode region, charging C_H until V_{out} approaches +1 V. For $V_{out} \approx +1$ V, M_1 exhibits an on-resistance of $R_{on} = [\mu_n C_{ox}(W/L)(V_{DD} - V_{in} - V_{TH})]^{-1}$.

The above observations reveal two important points. First, a MOS switch can conduct current in either direction simply by exchanging the role of its source and drain terminals. Second, as shown in Fig. 12.10, when the switch is on, V_{out} follows V_{in} and when the switch

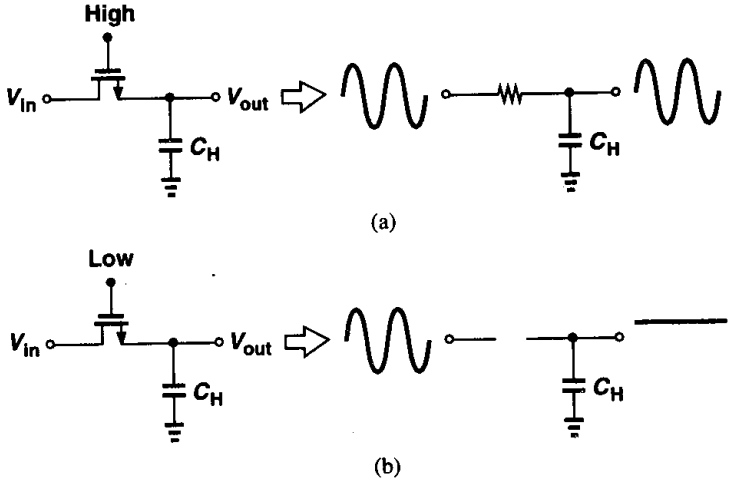


Figure 12.10 Track and hold capabilities of a sampling circuit.

is off, V_{out} remains constant. Thus, the circuit "tracks" the signal when CK is high and "freezes" the instantaneous value of V_{in} across C_H when CK goes low.

Example 12.2

In the circuit of Fig. 12.9(a), calculate V_{out} as a function of time. Assume $\lambda = 0$.

Solution

Before V_{out} drops below $V_{DD} - V_{TH}$, M_1 is saturated and we have:

$$V_{out}(t) = V_{DD} - \frac{I_{D1}t}{C_H} \tag{12.7}$$

$$= V_{DD} - \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{DD} - V_{TH})^2\frac{t}{C_H}. \tag{12.8}$$

After

$$t_1 = \frac{2V_{TH}C_H}{\mu_n C_{ox}\frac{W}{L}(V_{DD} - V_{TH})^2}, \tag{12.9}$$

M_1 enters the triode region, yielding a time-dependent current. We therefore write:

$$C_H\frac{dV_{out}}{dt} = -I_{D1} \tag{12.10}$$

$$= -\frac{1}{2}\mu_n C_{ox}\frac{W}{L}\left[2(V_{DD} - V_{TH})V_{out} - V_{out}^2\, big\right] \quad t > t_1. \tag{12.11}$$

Rearranging (12.11), we have

$$\frac{dV_{out}}{[2(V_{DD} - V_{TH}) - V_{out}]V_{out}} = -\frac{1}{2}\mu_n\frac{C_{ox}}{C_H}\frac{W}{L}dt, \tag{12.12}$$

which, upon separation into partial fractions, is written as

$$\left[\frac{1}{V_{out}} + \frac{1}{2(V_{DD} - V_{TH}) - V_{out}}\right]\frac{dV_{out}}{V_{DD} - V_{TH}} = -\mu_n\frac{C_{ox}}{C_H}\frac{W}{L}dt. \tag{12.13}$$

Thus,

$$\ln V_{out} - \ln[2(V_{DD} - V_{TH}) - V_{out}] = -(V_{DD} - V_{TH})\mu_n\frac{C_{ox}}{C_H}\frac{W}{L}(t - t_1), \tag{12.14}$$

that is,

$$\ln\frac{V_{out}}{2(V_{DD} - V_{TH}) - V_{out}} = -(V_{DD} - V_{TH})\mu_n\frac{C_{ox}}{C_H}\frac{W}{L}(t - t_1). \tag{12.15}$$

Taking the exponential of both sides and solving for V_{out}, we obtain

$$V_{out} = \frac{2(V_{DD} - V_{TH})\exp\left[-(V_{DD} - V_{TH})\mu_n\frac{C_{ox}}{C_H}\cdot\frac{W}{L}(t - t_1)\right]}{1 + \exp\left[-(V_{DD} - V_{TH})\mu_n\frac{C_{ox}}{C_H}\cdot\frac{W}{L}(t - t_1)\right]}. \tag{12.16}$$

In the circuit of Fig. 12.9(b), we assumed $V_{in} = +1$ V (Fig. 12.11). Now suppose $V_{in} = V_{DD}$. How does V_{out} vary with time? Since the gate and drain of M_1 are at the same potential, the transistor is saturated and we have:

$$C_H\frac{dV_{out}}{dt} = I_{D1} \tag{12.17}$$

$$= \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{DD} - V_{out} - V_{TH})^2, \tag{12.18}$$

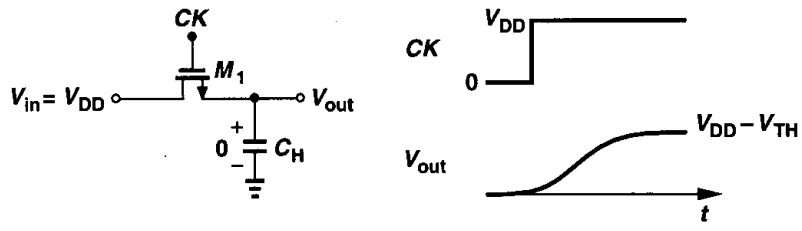


Figure 12.11 Maximum output level in an NMOS sampler.

where channel-length modulation is neglected. It follows that

$$\frac{dV_{out}}{(V_{DD} - V_{out} - V_{TH})^2} = \frac{1}{2}\mu_n \frac{C_{ox}}{C_H}\frac{W}{L}dt, \tag{12.19}$$

and hence

$$\frac{1}{V_{DD} - V_{out} - V_{TH}}\bigg|_0^{Vout} = \frac{1}{2}\mu_n \frac{C_{ox}}{C_H}\frac{W}{L}t\bigg|_0^t, \tag{12.20}$$

where body effect is neglected and $V_{out}(t = 0)$ is assumed zero. Thus,

$$V_{out} = V_{DD} - V_{TH} - \frac{1}{\frac{1}{2}\mu_n \frac{C_{ox}}{C_H}\frac{W}{L}t + \frac{1}{V_{DD} - V_{TH}}}. \tag{12.21}$$

Equation (12.21) implies that as $t \to \infty$, $V_{out} \to V_{DD} - V_{TH}$. This is because as V_{out} approaches $V_{DD} - V_{TH}$, the overdrive voltage of M_1 vanishes, reducing the current available for charging C_H to negligible values. Of course, even for $V_{out} = V_{DD} - V_{TH}$, the transistor conducts some subthreshold current and, given enough time, eventually brings V_{out} to V_{DD}. Nonetheless, as mentioned in Chapter 3, for typical operation speeds, it is reasonable to assume that V_{out} does not exceed $V_{DD} - V_{TH}$.

The foregoing analysis demonstrates a serious limitation of MOS switches: if the input signal level is close to V_{DD}, then the output provided by an NMOS switch cannot track the input. From another point of view, the on-resistance of the switch increases considerably as the input and output voltages approach $V_{DD} - V_{TH}$. We may then ask: what is the maximum input level that the switch can pass to the output faithfully? In Fig. 12.11, for $V_{out} \approx V_{in}$, the transistor must operate in deep triode region and hence the upper bound of V_{in} equals $V_{DD} - V_{TH}$. As explained below, in practice V_{in} must be quite lower than this value.

Example 12.3

In the circuit of Fig. 12.12, calculate the minimum and maximum on-resistance of M_1. Assume $\mu_n C_{ox} = 50\ \mu A/V^2$, $W/L = 10/1$, $V_{TH} = 0.7$ V, $V_{DD} = 3$ V, and $\gamma = 0$.

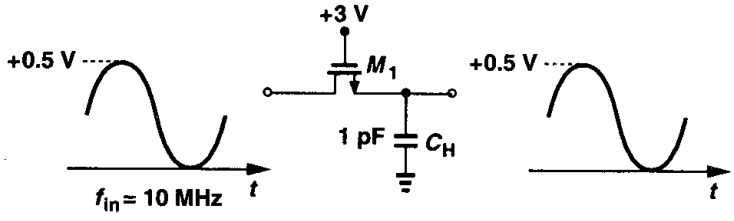


Figure 12.12

Solution

We note that in the steady state, M_1 remains in the triode region because the gate voltage is higher than both V_{in} and V_{out} by a value greater than V_{TH}. If $f_{in} = 10$ MHz, we predict that V_{out} tracks V_{in} with a negligible phase shift due to the on-resistance of M_1 and C_H. Assuming $V_{out} \approx V_{in}$, we need not distinguish between the source and drain terminals, obtaining

$$R_{on1} = \frac{1}{\mu_n C_{ox} \frac{W}{L}(V_{DD} - V_{in} - V_{TH})}. \tag{12.22}$$

Thus, $R_{on1,max} \approx 1.11$ kΩ and $R_{on1,min} \approx 870$ Ω. By contrast, if the maximum input level is raised to 1.5 V, then $R_{on1,max} = 2.5$ kΩ.

MOS devices operating in deep triode region are sometimes called "zero-offset" switches to emphasize that they exhibit no dc shift between the input and output voltages of the simple sampling circuit of Fig. 12.8(b).[1] This is evident from examples of Fig. 12.9, where the output eventually becomes equal to the input. Nonexistent in bipolar technology, the zero offset property proves crucial in precise sampling of analog signals.

We have thus far considered only NMOS switches. The reader can verify that the foregoing principles apply to PMOS switches as well. In particular, as shown in Fig. 12.13, a PMOS transistor fails to operate as a zero-offset switch if its gate is grounded and its drain terminal senses an input voltage of $|V_{THP}|$ or less. In other words, the on-resistance of the device rises rapidly as the input and output levels drop to $|V_{THP}|$ above ground.

12.2.2 Speed Considerations

What determines the speed of the sampling circuits of Fig. 12.8? We must first define the speed here. Illustrated in Fig. 12.14, a simple, but versatile measure of speed is the time required for the output voltage to go from zero to the maximum input level after the switch turns on. Since V_{out} would take infinite time to become equal to V_{in0}, we consider the output settled when it is within a certain "error band," ΔV, around the final value. For example, we say the output settles to 0.1% accuracy after t_S seconds, meaning that in Fig. 12.14, $\Delta V / V_{in0} = 0.1\%$. Thus, the speed specification must be accompanied by an accuracy

[1] We assume the circuit following the sampler draws no input dc current.

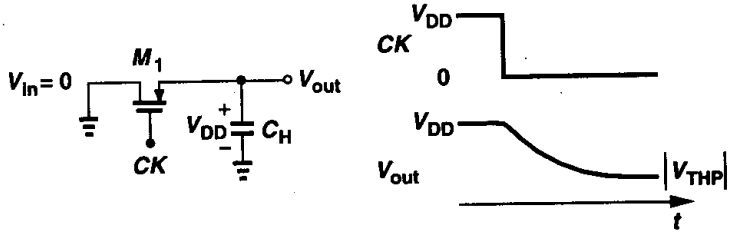


Figure 12.13 Sampling circuit using PMOS switch.

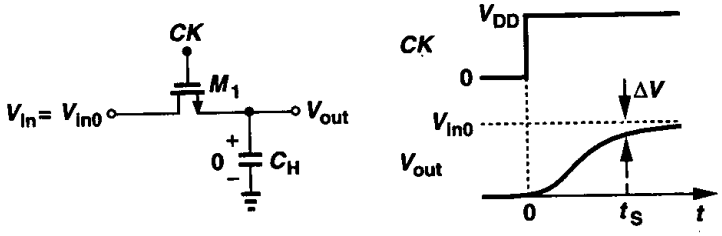


Figure 12.14 Definition of speed in a sampling circuit.

specification as well. Note that after $t = t_S$, we can consider the source and drain voltages to be approximately equal.

From the circuit of Fig. 12.14, we surmise that the sampling speed is given by two factors: the on-resistance of the switch and the value of the sampling capacitor. Thus, to achieve a higher speed, a large aspect ratio and a small capacitor must be used. However, as illustrated in Fig. 12.12, the on-resistance also depends on the input level, yielding a greater time constant for more positive inputs (in the case of NMOS switches). From Eq. (12.22), we plot the on-resistance of the switch as a function of the input level [Fig. 12.15(a)], noting the sharp rise as V_{in} approaches $V_{DD} - V_{TH}$. For example, if we restrict the variation of R_{on} to a range of 4 to 1, then the maximum input level is given by

$$\frac{1}{\mu_n C_{ox}\dfrac{W}{L}(V_{DD} - V_{in,max} - V_{TH})} = \frac{4}{\mu_n C_{ox}\dfrac{W}{L}(V_{DD} - V_{TH})}. \tag{12.23}$$

That is,

$$V_{in,max} = \frac{3}{4}(V_{DD} - V_{TH}). \tag{12.24}$$

This value falls around $V_{DD}/2$, translating to severe swing limitations. Note that the device threshold voltage directly limits the voltage swings.[2]

[2]By contrast, the output swing of cascode stages is typically limited by overdrive voltages rather than the threshold voltage.

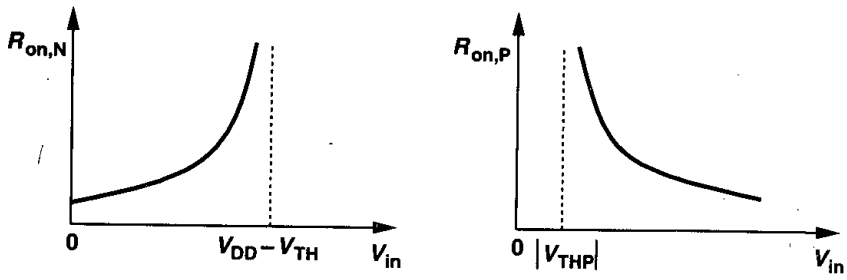


Figure 12.15 On-resistance of (a) NMOS and (b) PMOS devices as a function of input voltage.

In order to accommodate greater voltage swings in a sampling circuit, we first observe that a PMOS switch exhibits an on-resistance that *decreases* as the input voltage becomes more positive [Fig. 12.15(b)]. It is then plausible to employ "complementary" switches so as to allow rail-to-tail swings. Shown in Fig. 12.16(a), such a combination requires complementary clocks, producing an equivalent resistance:

$$R_{on,eq} = R_{on,N} \| R_{on,P} \tag{12.25}$$

$$= \frac{1}{\mu_n C_{ox}\left(\frac{W}{L}\right)_N (V_{DD} - V_{in} - V_{THN})} \Bigg\| \frac{1}{\mu_p C_{ox}\left(\frac{W}{L}\right)_P (V_{in} - |V_{THP}|)} \tag{12.26}$$

$$= \frac{1}{\mu_n C_{ox}\left(\frac{W}{L}\right)_N (V_{DD} - V_{THN}) - \left[\mu_n C_{ox}\left(\frac{W}{L}\right)_N - \mu_p C_{ox}\left(\frac{W}{L}\right)_P\right] V_{in} - \mu_p C_{ox}\left(\frac{W}{L}\right)_P |V_{THP}|} \tag{12.27}$$

Interestingly, if $\mu_n C_{ox}(W/L)_N = \mu_p C_{ox}(W/L)_P$, then $R_{on,eq}$ is independent of the input level.[3] Figure 12.16(b) plots the behavior of $R_{on,eq}$ in the general case, revealing much less variation than that corresponding to each switch alone.

For high-speed input signals, it is critical that the NMOS and PMOS switches in Fig. 12.16(a) turn off simultaneously so as to avoid ambiguity in the sampled value. If, for example, the NMOS device turns off Δt seconds earlier than the PMOS device, then the output voltage tends to track the input for the remaining Δt seconds, but with a large, input-dependent time constant (Fig. 12.17). This effect gives rise to distortion in the sampled value. For moderate precision, the simple circuit shown in Fig. 12.18 provides complementary clocks by duplicating the delay of inverter I_1 through the gate G_2.

[3] In reality, V_{THN} and V_{THP} vary with V_{in} through body effect but we ignore this variation here.

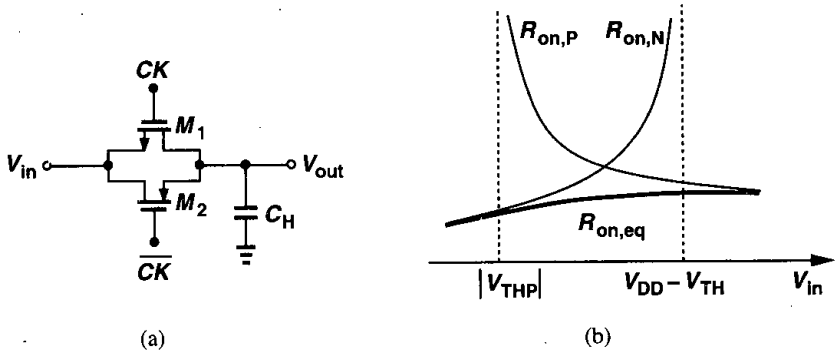


Figure 12.16 (a) Complementary switch, (b) on-resistance of the complementary switch.

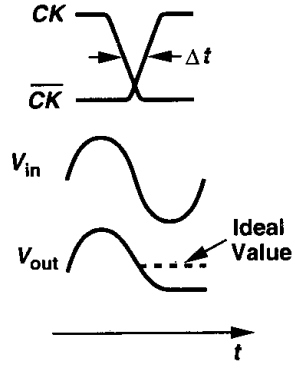


Figure 12.17 Distortion generated if complementary switches do not turn off simultaneously.

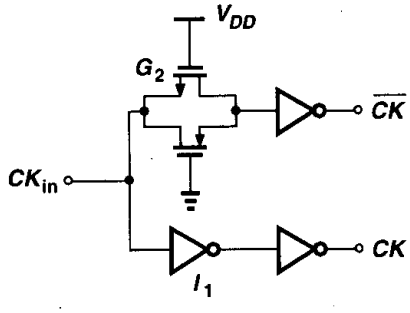


Figure 12.18 Simple circuit generating complementary clocks.

12.2.3 Precision Considerations

Our foregoing study of MOS switches indicates that a larger W/L or a smaller sampling capacitor results in a higher speed. In this section, we show that these methods of increasing the speed degrade the precision with which the signal is sampled.

Three mechanisms in MOS transistor operation introduce error at the instant the switch turns off. We study each effect individually.

Channel Charge Injection Consider the sampling circuit of Fig. 12.19 and recall that for a MOSFET to be on, a channel must exist at the oxide-silicon interface. Assuming $V_{in} \approx V_{out}$, we use our derivations in Chapter 2 to express the total charge in the inversion layer as

$$Q_{ch} = WLC_{ox}(V_{DD} - V_{in} - V_{TH}), \tag{12.28}$$

where L denotes the effective channel length. When the switch turns off, Q_{ch} exits through the source and drain terminals, a phenomenon called "channel charge injection."

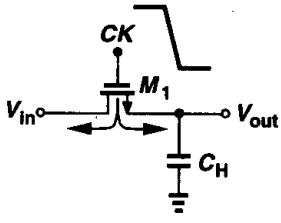


Figure 12.19 Charge injection when a switch turns off.

The charge injected to the left side of Fig. 12.19 is absorbed by the input source, creating no error. On the other hand, the charge injected to the right side is deposited on C_H, introducing an error in the voltage stored on the capacitor. For example, if half of Q_{ch} is injected onto C_H, the resulting error equals

$$\Delta V = \frac{WLC_{ox}(V_{DD} - V_{in} - V_{TH})}{2C_H}. \tag{12.29}$$

Illustrated in Fig. 12.20, the error for an NMOS switch appears as a negative "pedestal" at the output. Note that the error is directly proportional to WLC_{ox} and inversely proportional to C_H.

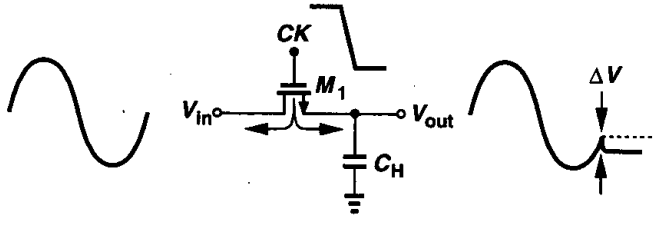


Figure 12.20 Effect of charge injection.

An important question that arises now is: why did we assume in arriving at (12.29) that exactly *half* of the channel charge is injected onto C_H? In reality, the fraction of charge that exits through the source and drain terminals is a relatively complex function of various parameters such as the impedance seen at each terminal to ground and the transition time of the clock [1, 2]. Investigations of this effect have not yielded any rule of thumb that can predict the charge splitting in terms of such parameters. Furthermore, in many cases,

these parameters, e.g., the clock transition time, are poorly controlled. Also, most circuit simulation programs model charge injection quite inaccurately. As a worst-case estimate, we can assume that the entire channel charge is injected onto the sampling capacitor.

How does charge injection affect the precision? Assuming all of the charge is deposited on the capacitor, we express the sampled output voltage as

$$V_{out} \approx V_{in} - \frac{WLC_{ox}(V_{DD} - V_{in} - V_{TH})}{C_H}, \tag{12.30}$$

where the phase shift between the input and output is neglected. Thus,

$$V_{out} = V_{in}\left(1 + \frac{WLC_{ox}}{C_H}\right) - \frac{WLC_{ox}}{C_H}(V_{DD} - V_{TH}), \tag{12.31}$$

suggesting that the output deviates from the ideal value through two effects: a non-unity gain equal to $1 + WLC_{ox}/C_H$,[4] and a constant offset voltage $-WLC_{ox}(V_{DD} - V_{TH})/C_H$ (Fig. 12.21). In other words, since we have assumed channel charge is a *linear* function of the input voltage, the circuit exhibits only gain error and dc offset.

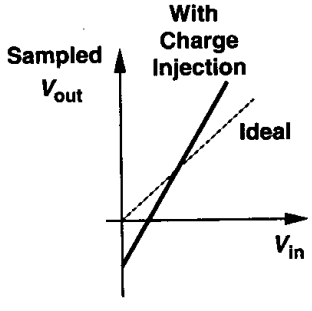


Figure 12.21 Input/output characteristic of sampling circuit in the presence of charge injection.

In the foregoing discussion, we tacitly assumed that V_{TH} is constant. However, for NMOS switches (in an n-well technology), body effect must be taken into account.[5] Since $V_{TH} = V_{TH0} + \gamma(\sqrt{2\phi_B + V_{SB}} - \sqrt{2\phi_B})$, and $V_{BS} \approx -V_{in}$, we have

$$V_{out} = V_{in} - \frac{WLC_{ox}}{C_H}\left(V_{DD} - V_{in} - V_{TH0} - \gamma\sqrt{2\phi_B + V_{in}} + \gamma\sqrt{2\phi_B}\right), \tag{12.32}$$

$$= V_{in}\left(1 + \frac{WLC_{ox}}{C_H}\right) + \gamma\frac{WLC_{ox}}{C_H}\sqrt{2\phi_B + V_{in}} - \frac{WLC_{ox}}{C_H}\left(V_{DD} - V_{TH0} + \gamma\sqrt{2\phi_B}\right). \tag{12.33}$$

[4]The voltage gain is *greater* than unity because the pedestal becomes smaller as the input level rises.

[5]Even for PMOS switches, the n-well is connected to the most positive supply voltage because the source and drain terminals of the switch may interchange during sampling.

It follows that the nonlinear dependence of V_{TH} upon V_{in} introduces nonlinearity in the input/output characteristic.

In summary, charge injection contributes three types of errors in MOS sampling circuits: gain error, dc offsets, and nonlinearity. In many applications, the first two can be tolerated or corrected whereas the last cannot.

It is instructive to consider the speed-precision trade-off resulting from charge injection. Representing the speed by a simple time constant τ and the precision by the error ΔV due to charge injection, we define a figure of merit as $F = (\tau \Delta V)^{-1}$. Writing

$$\tau = R_{on} C_H \tag{12.34}$$

$$= \frac{1}{\mu_n C_{ox}(W/L)(V_{DD} - V_{in} - V_{TH})} C_H, \tag{12.35}$$

and

$$\Delta V = \frac{WLC_{ox}}{C_H}(V_{DD} - V_{in} - V_{TH}), \tag{12.36}$$

we have

$$F = \frac{\mu_n}{L^2}. \tag{12.37}$$

Thus, to the first order, the trade-off is independent of the switch width and the sampling capacitor.

Clock Feedthrough In addition to channel charge injection, a MOS switch couples the clock transitions to the sampling capacitor through its gate-drain or gate-source overlap capacitance. Depicted in Fig. 12.22, the effect introduces an error in the sampled output voltage. Assuming the overlap capacitance is constant, we express the error as

$$\Delta V = V_{CK} \frac{WC_{ov}}{WC_{ov} + C_H}, \tag{12.38}$$

where C_{ov} is the overlap capacitance per unit width. The error ΔV is independent of the input level, manifesting itself as a constant offset in the input/output characteristic. As with charge injection, clock feedthrough leads to a trade-off between speed and precision as well.

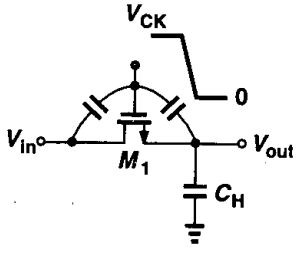


Figure 12.22 Clock feedthrough in a sampling circuit.

***kT/C* Noise** Recall from Example 7.1 that a resistor charging a capacitor gives rise to a total rms noise voltage of $\sqrt{kT/C}$. As shown in Fig. 12.23, a similar effect occurs in sampling circuits. The on-resistance of the switch introduces thermal noise at the output and, when the switch turns off, this noise is stored on the capacitor along with the instantaneous value of the input voltage. It can be proved that the rms voltage of the sampled noise in this case is still approximately equal to $\sqrt{kT/C}$ [3, 4].

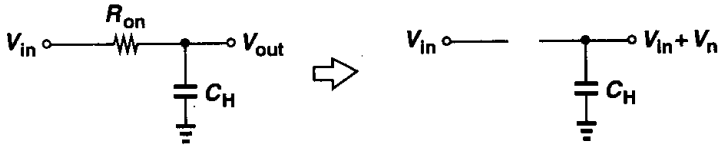


Figure 12.23 Thermal noise in a sampling circuit.

The problem of kT/C noise limits the performance in many high-precision applications. In order to achieve a low noise, the sampling capacitor must be sufficiently large, thus loading other circuits and degrading the speed.

12.2.4 Charge Injection Cancellation

The dependence of charge injection upon the input level and the trade-off expressed by (12.37) make it necessary to seek methods of cancelling the effect of charge injection so as to achieve a higher F. We consider a few such techniques here.

To arrive at the first technique, we postulate that the charge injected by the main transistor can be *removed* by means of a second transistor. As shown in Fig. 12.24, a "dummy" switch, M_2, driven by $\overline{CK}$ is added to the circuit such that after M_1 turns off and M_2 turns on, the channel charge deposited by the former on C_H is absorbed by the latter to create a channel. Note that both the source and drain of M_2 are connected to the output node.

How do we ensure that the charge injected by M_1, Δq_1, is equal to that absorbed by M_2, Δq_2? Suppose half of the channel charge of M_1 is injected onto C_H, i.e.,

$$\Delta q_1 = \frac{W_1 L_1 C_{ox}}{2}(V_{CK} - V_{in} - V_{TH1}). \tag{12.39}$$

Since $\Delta q_2 = W_2 L_2 C_{ox}(V_{CK} - V_{in} - V_{TH2})$, if we choose $W_2 = 0.5W_1$ and $L_2 = L_1$, then $\Delta q_2 = \Delta q_1$. Unfortunately, the assumption of equal splitting of charge between source and drain is generally invalid, making this approach less attractive.

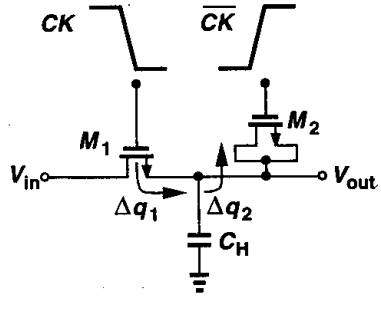


Figure 12.24 Addition of dummy device to reduce charge injection and clock feedthrough.

Interestingly, with the choice $W_2 = 0.5W_1$ and $L_2 = L_1$, the effect of clock feedthrough is suppressed. As depicted in Fig. 12.25, the total charge in V_{out} is zero because

$$-V_{CK}\frac{W_1C_{ov}}{W_1C_{ov}+C_H+2W_2C_{ov}} + V_{CK}\frac{2W_2C_{ov}}{W_1C_{ov}+C_H+2W_2C_{ov}} = 0. \quad (12.40)$$

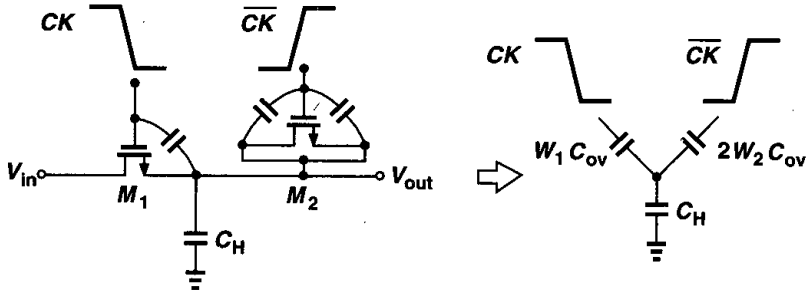


Figure 12.25 Clock feedthrough suppression by dummy switch.

Another approach to lowering the effect of charge injection incorporates both PMOS and NMOS devices such that the opposite charge packets injected by the two cancel each other (Fig. 12.26). For Δq_1 to cancel Δq_2, we must have $W_1L_1C_{ox}(V_{CK} - V_{in} - V_{THN}) = W_2L_2C_{ox}(V_{in} - |V_{THP}|)$. Thus, the cancellation occurs for only one input level. Even for clock feedthrough, the circuit does not provide complete cancellation because the gate-drain overlap capacitance of NFETs is not equal to that of PFETs.

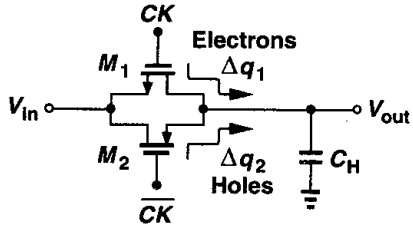


Figure 12.26 Use of complementary switches to reduce charge injection.

Our knowledge of the advantages of differential circuits suggests that the problem of charge injection may be relieved through differential operation. As shown in Fig. 12.27, we surmise that the charge injection appears as a common-mode disturbance. But, writing $\Delta q_1 = WLC_{ox}(V_{CK} - V_{in1} - V_{TH1})$, and $\Delta q_2 = WLC_{ox}(V_{CK} - V_{in2} - V_{TH2})$, we recognize that $\Delta q_1 = \Delta q_2$ only if $V_{in1} = V_{in2}$. In other words, the overall error is not suppressed for differential signals. Nevertheless, this technique both removes the constant offset and lowers the nonlinear component. This can be understood by writing

$$\Delta q_1 - \Delta q_2 = WLC_{ox}[(V_{in2} - V_{in1}) + (V_{TH2} - V_{TH1})] \quad (12.41)$$

$$= WLC_{ox}\left[V_{in2} - V_{in1} + \gamma\left(\sqrt{2\phi_F + V_{in2}} - \sqrt{2\phi_F + V_{in1}}\right)\right]. \quad (12.42)$$

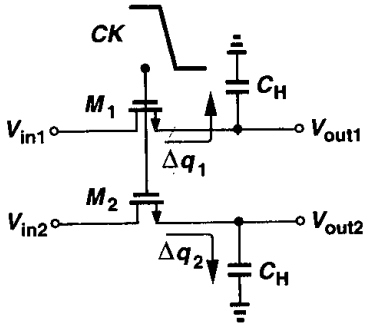


Figure 12.27 Differential sampling circuit.

Since for $V_{in1} = V_{in2}$, $\Delta q_1 - \Delta q_2 = 0$, the characteristic exhibits no offset. Also, the nonlinearity of body effect now appears in both square-root terms of (12.42), leading to only odd-order distortion (Chapter 13).

The problem of charge injection continues to limit the speed-precision envelope in sampled-data systems. Many cancellation techniques have been introduced but each leading to other trade-offs. One such technique, called "bottom-plate sampling," is widely used in switched-capacitor circuits and is described later in this chapter.

12.3 Switched-Capacitor Amplifiers

As mentioned in Section 12.1 and exemplified by the circuit of Fig. 12.4, CMOS feedback amplifiers are more easily implemented with a capacitive feedback network than a resistive one. Having examined sampling techniques, we are now ready to study a number of switched-capacitor amplifiers. Our objective is to understand the underlying principles as well as the speed-precision trade-offs encountered in the design of each circuit.

Before studying SC amplifiers, it is helpful to briefly look at the physical implementation of capacitors in CMOS technology. A simple capacitor structure is shown in Fig. 12.28(a), where the "top plate" is realized by a polysilicon layer and the "bottom plate" by a heavily

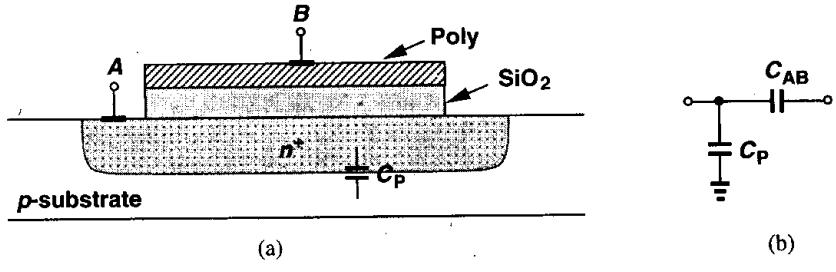


Figure 12.28 (a) Monolithic capacitor structure, (b) circuit model of (a) including parasitic capacitance to the substrate.

doped n^+ region. The dielectric is the thin oxide layer used in MOS devices as well.[6] An important concern in using this structure is the parasitic capacitance between each plate and the substrate. In particular, the bottom plate suffers from substantial junction capacitance to the underlying p region—typically about 10 to 20% of the oxide capacitance. For this reason, we usually model the capacitor as in Fig. 12.28(b). Monolithic capacitors are described in more detail in Chapters 17 and 18.

12.3.1 Unity-Gain Sampler/Buffer

While a unity-gain amplifier can be realized with no resistors or capacitors in the feedback network [Fig. 12.29(a)], for discrete-time applications, it still requires a sampling circuit. We may therefore conceive the circuit shown in Fig. 12.29(b) as a sampler/buffer. However, the input-dependent charge injected by S_1 onto C_H limits the accuracy here.

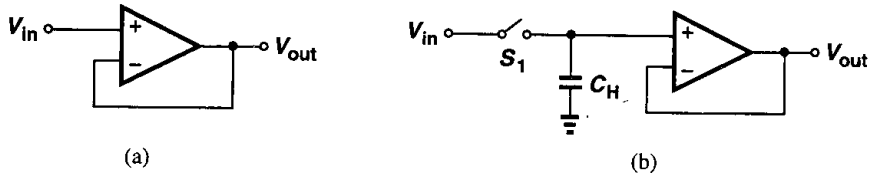


Figure 12.29 (a) Unity-gain buffer, (b) sampling circuit followed by unity-gain buffer.

Now consider the topology depicted in Fig. 12.30(a), where three switches control the sampling and amplification modes. In the sampling mode, S_1 and S_2 are on and S_3 is off, yielding the topology shown in Fig. 12.30(b). Thus, $V_{out} = V_X \approx 0$, and the voltage across C_H tracks V_{in}. At $t = t_0$, when $V_{in} = V_0$, S_1 and S_2 turn off and S_3 turns on, placing the capacitor around the op amp and entering the circuit into the amplification mode [Fig. 12.30(c)]. Since the op amp's high gain requires that node X still be a virtual ground and since the charge on the capacitor must be conserved, V_{out} rises to a value approximately equal to V_0. This voltage is therefore "frozen" and it can be processed by subsequent stages.

With proper timing, the circuit of Fig. 12.30(a) can substantially alleviate the problem of channel charge injection. As Fig. 12.31 illustrates in "slow motion," in the transition from the sampling mode to the amplification mode, S_2 turns off slightly *before* S_1 does. We carefully examine the effect of the charge injected by S_2 and S_1. When S_2 turns off, it injects a charge packet Δq_2 onto C_H, producing an error equal to $\Delta q_2/C_H$. However, this charge is quite independent of the input level because node X is a virtual ground. For example, if S_2 is realized by an NMOS device whose gate voltage equals V_{CK}, then $\Delta q_2 = WLC_{ox}(V_{CK} - V_{TH} - V_X)$. Although body effect makes V_{TH} a function of V_X, Δq_2 is relatively constant because V_X is quite independent of V_{in}.

The constant magnitude of Δq_2 means that channel charge of S_2 introduces only an offset (rather than gain error or nonlinearity) in the input/output characteristic. As described

[6] The oxide in this type of capacitor is typically thicker than the MOS gate oxide because silicon dioxide grows faster on a heavily-doped material.

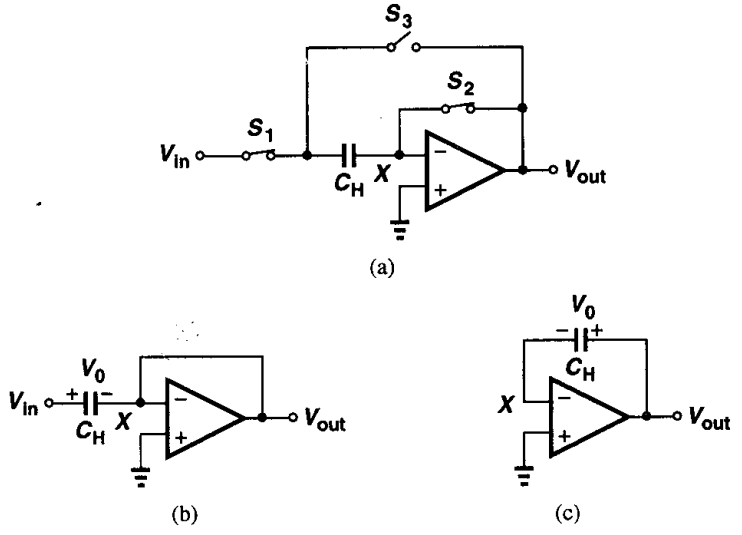


Figure 12.30 (a) Unity-gain sampler, (b) circuit of (a) in sampling mode, (c) circuit of (a) in amplification mode.

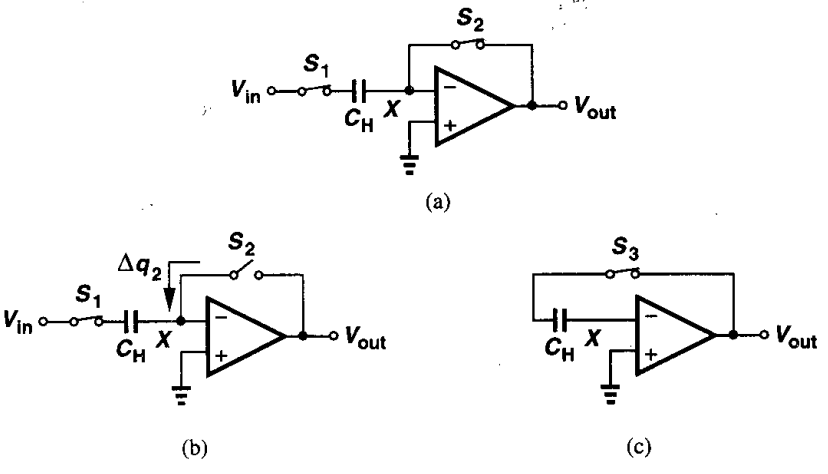


Figure 12.31 Operation of the unity-gain sampler in slow motion.

below, this offset can easily be removed by differential operation. But, how about the charge injected by S_1 onto C_H? Let us set V_{in} to zero and suppose S_1 injects a charge packet Δq_1 onto node P [Fig. 12.32(a)]. If the capacitance connected from X to ground (including the input capacitance of the op amp) is zero, V_P and V_X jump to infinity. To simplify the analysis, we assume a total capacitance equal to C_X from X to ground [Fig. 12.32(b)], and we will see shortly that its value does not affect the results. In Fig. 12.32(b), each of C_H

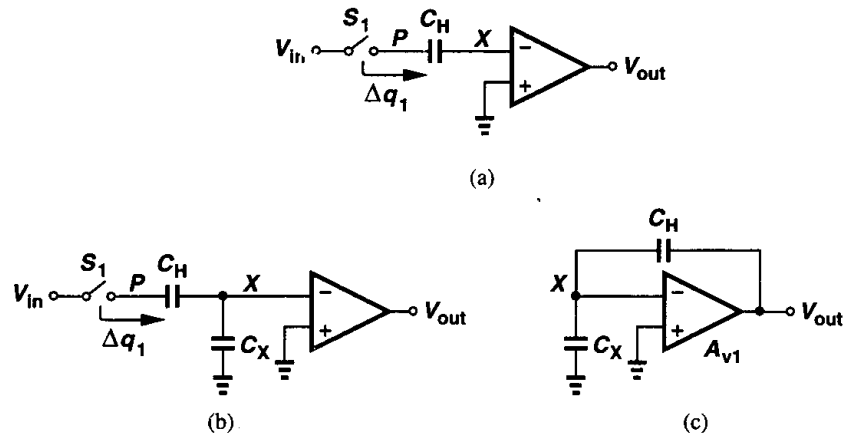


Figure 12.32 Effect of charge injected by S_1 with (a) zero and (b) finite op amp input capacitance, (c) transition of circuit to amplification mode.

and C_X carries a charge equal to Δq_1. Now, as shown in Fig. 12.32(c), we place C_H around the op amp, seeking to obtain the resulting output voltage.

To calculate the output voltage, we must make an important observation: the total charge at node X cannot change after S_2 turns off because no path exists for electrons to flow into or out of this node. Thus, if before S_1 turns off, the total charge on the right plate of C_H and the top plate of C_X is zero, it must still add up to zero after S_1 injects charge because no *resistive* path is connected to X. The same holds true after C_H is placed around the op amp.

Now consider the circuit of Fig. 12.32(c), assuming the total charge at node X is zero. We can write $C_X V_X - (V_{out} - V_X)C_H = 0$, and $V_X = -V_{out}/A_{v1}$. Thus, $-(C_X + C_H)V_{out}/A_{v1} - V_{out}C_H = 0$, i.e., $V_{out} = 0$. Note that this result is independent of Δq_1, capacitor values, or the gain of the op amp, thereby revealing that the charge injection by S_1 introduces no error *if S_2 turns off first.*

In summary, in Fig. 12.30(a), after S_2 turns off, node X "floats," maintaining a constant total charge regardless of the transitions at other nodes of the circuit. As a result, after the feedback configuration is formed, the output voltage is not influenced by the charge injection due to S_1. From another point of view, node X is a virtual ground at the moment S_2 turns off, freezing the instantaneous input level across C_H and yielding a charge equal to $V_0 C_H$ on the left plate of C_H. After settling with feedback, node X is again a virtual ground, forcing C_H to still carry $V_0 C_H$ and hence the output voltage to be approximately equal to V_0.

The effect of the charge injected by S_1 can be studied from yet another perspective. Suppose in Fig. 12.32(c), the output voltage is finite and positive. Then, since $V_X = V_{out}/(-A_{v1})$, V_X must be finite and negative, requiring negative charge on the top plate of C_X. For the total charge at X to be zero, the charge on the left plate of C_H must be positive and that on its right plate negative, giving $V_{out} \leq 0$. Thus, the only valid solution is $V_{out} = 0$.

The third switch in Fig. 12.30(a), S_3, also merits attention. In order to turn on, S_3 must establish an inversion layer at its oxide interface. Does the required channel charge come from C_H or from the op amp? We note from the foregoing analysis that after the feedback circuit has settled, the charge on C_H equals $V_0 C_H$, unaffected by S_3. The channel charge of this switch is therefore entirely supplied by the op amp, introducing no error.

Our study of Fig. 12.30(a) thus far suggests that, with proper timing, the charge injected by S_1 and S_3 is unimportant and the channel charge of S_2 results in a constant offset voltage. Fig. 12.33 depicts a simple realization of the clock edges to ensure S_1 turns off after S_2 does.

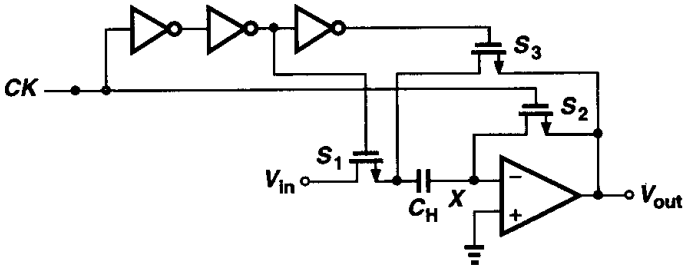


Figure 12.33 Generation of proper clock edges for unity-gain sampler.

The input-independent nature of the charge injected by the reset switch allows complete cancellation by differential operation. Illustrated in Fig. 12.34, such an approach employs a differential op amp along with two sampling capacitors so that the charge injected by S_2 and S_2' appears as a *common-mode* disturbance at nodes X and Y. This is in contrast to the behavior of the differential circuit shown in Fig. 12.27, where the input-dependent charge injection still leads to nonlinearity. In reality, S_2 and S_2' exhibit a finite charge injection mismatch, an issue resolved by adding another switch, S_{eq}, that turns off slightly after S_2 and S_2' (and before S_1 and S_1'), thereby equalizing the charge at nodes X and Y.

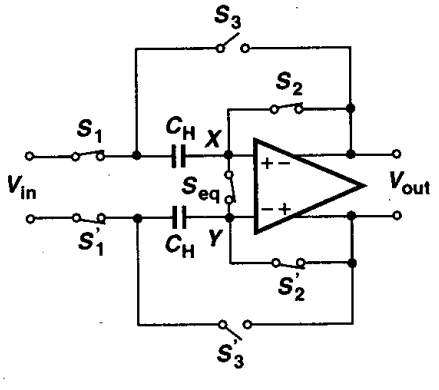


Figure 12.34 Differential realization of unity-gain sampler.

Precision Considerations The circuit of Fig. 12.30(a) operates as a unity-gain buffer in the amplification mode, producing an output voltage approximately equal to the voltage stored across the capacitor. How close to unity is the gain here? As a general case, we assume the op amp exhibits a finite input capacitance C_{in} and calculate the output voltage when the circuit goes from the sampling mode to the amplification mode (Fig. 12.35). Owing to the finite gain of the op amp, $V_X \neq 0$ in the amplification mode, giving a charge

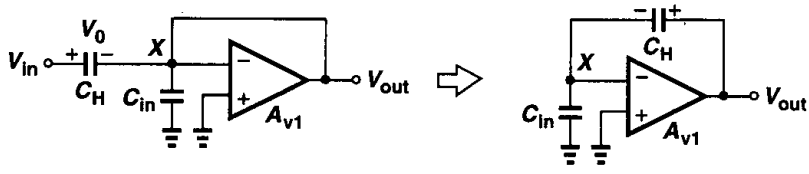


Figure 12.35 Equivalent circuit for accuracy calculations.

equal to $C_{in}V_X$ on C_{in}. The conservation of charge at X requires that $C_{in}V_X$ come from C_H, raising the charge on C_H to $C_HV_0 + C_{in}V_X$.[7] It follows that the voltage across C_H equals $(C_HV_0 + C_{in}V_X)/C_H$. We therefore write $V_{out} - (C_HV_0 + C_{in}V_X)/C_H = V_X$ and $V_X = -V_{out}/A_{v1}$. Thus,

$$V_{out} = \frac{V_0}{1 + \frac{1}{A_{v1}}\left(\frac{C_{in}}{C_H} + 1\right)} \tag{12.43}$$

$$\approx V_0\left[1 - \frac{1}{A_{v1}}\left(\frac{C_{in}}{C_H} + 1\right)\right]. \tag{12.44}$$

As expected, if $C_{in}/C_H \ll 1$, then $V_{out} \approx V_0/(1 + A_{v1}^{-1})$. In general, however, the circuit suffers from a gain error of approximately $-(C_{in}/C_H + 1)/A_{v1}$, suggesting that the input capacitance must be minimized even if speed is not critical. Recall from Chapter 9 that to increase A_{v1}, we may choose a large width for the input transistors of the op amp, but at the cost of higher input capacitance. An optimum device size must therefore yield minimum gain error rather than maximum A_{v1}.

Example 12.4

In the circuit of Fig. 12.35, $C_{in} = 0.5$ pF and $C_H = 2$ pF. What is the minimum op amp gain that guarantees a gain error of 0.1%?

Solution

Since $C_{in}/C_H = 0.25$, we have $A_{v1,min} = 1000 \times 1.25 = 1250$.

[7] The charge on C_H *increases* because positive charge transfer from the left plate of C_H to the top plate of C_{in} leads to a more positive voltage across C_H.

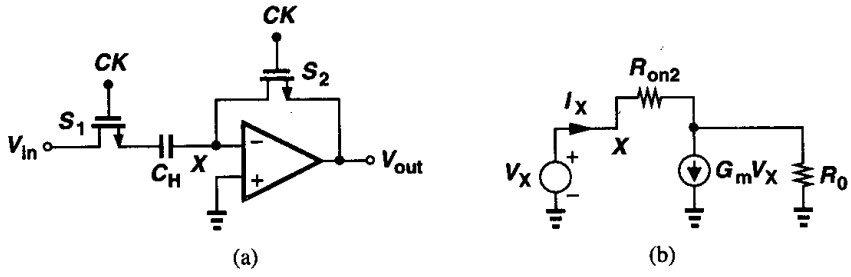


Figure 12.36 (a) Unity-gain sampler in sampling mode, (b) equivalent circuit of (a).

Speed Considerations Let us first examine the circuit in the sampling mode [Fig. 12.36(a)]. What is the time constant in this phase? The total resistance in series with C_H is given by R_{on1} and the resistance between X and ground, R_X. Using the simple op amp model shown in Fig. 12.36(b), where R_0 denotes the open-loop output impedance of the op amp, we have

$$(I_X - G_m V_X)R_0 + I_X R_{on2} = V_X, \tag{12.45}$$

that is,

$$R_X = \frac{R_0 + R_{on2}}{1 + G_m R_0}. \tag{12.46}$$

Since typically $R_{on2} \ll R_0$ and $G_m R_0 \gg 1$, we have $R_X \approx 1/G_m$. For example, in a telescopic op amp employing differential to single-ended conversion, G_m equals the transconductance of each input transistor.

The time constant in the sampling mode is thus equal to

$$\tau_{sam} = \left(R_{on1} + \frac{1}{G_m}\right) C_H. \tag{12.47}$$

The magnitude of τ_{sam} must be sufficiently small to allow settling in the test case of Fig. 12.14 to the required precision.

Now let us consider the circuit as it enters the amplification mode. Shown in Fig. 12.37 along with both the op amp input capacitance and the load capacitance, the circuit must begin with $V_{out} \approx 0$ and eventually produce $V_{out} \approx V_0$. If C_{in} is relatively small, we can assume that the voltages across C_L and C_H do not change instantaneously, concluding that if $V_{out} \approx 0$ and $V_{CH} \approx V_0$, then $V_X = -V_0$ at the beginning of the amplification mode. In other words, the input difference sensed by the op amp initially jumps to a large value, possibly causing the op amp to slew. But, let us first assume the op amp can be modeled by a linear model and determine the output response.

To simplify the analysis, we represent the charge on C_H by an explicit series voltage source, V_S, that goes from zero to V_0 at $t = t_0$ while C_H carries no charge itself (Fig. 12.38). The objective is to obtain the transfer function $V_{out}(s)/V_S(s)$ and hence the step response.

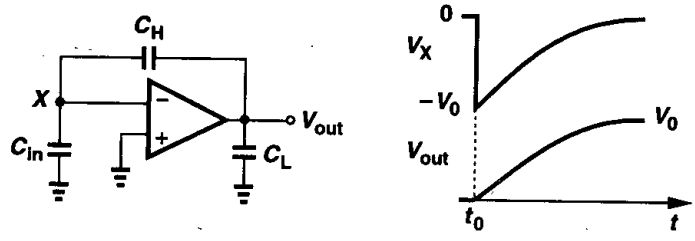


Figure 12.37 Time response of unity-gain sampler in amplification mode.

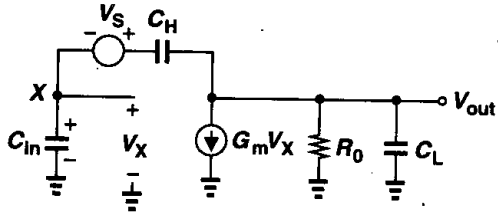


Figure 12.38 Equivalent circuit of unity-gain circuit in amplification mode.

We have

$$V_{out}\left(\frac{1}{R_0}+C_Ls\right)+G_mV_X=(V_S+V_X-V_{out})C_Hs. \tag{12.48}$$

Also, since the current through C_{in} equals $V_XC_{in}s$,

$$V_X\frac{C_{in}s}{C_Hs}+V_X+V_S=V_{out}. \tag{12.49}$$

Calculating V_X from (12.49) and substituting in (12.48), we arrive at the transfer function:

$$\frac{V_{out}}{V_S}(s)=R_0\frac{(G_m+C_{in}s)C_H}{R_0(C_LC_{in}+C_{in}C_H+C_HC_L)s+G_mR_0C_H+C_H+C_{in}}. \tag{12.50}$$

Note that for $s=0$, (12.50) reduces to a form similar to (12.43). Since typically $G_mR_0C_H \gg C_H, C_{in}$, we can simplify (12.50) as

$$\frac{V_{out}}{V_S}(s)=\frac{(G_m+C_{in}s)C_H}{(C_LC_{in}+C_{in}C_H+C_HC_L)s+G_mC_H}. \tag{12.51}$$

Thus, the response is characterized by a time constant equal to

$$\tau_{amp}=\frac{C_LC_{in}+C_{in}C_H+C_HC_L}{G_mC_H}, \tag{12.52}$$

which is independent of the op amp output resistance. This is because a higher R_0 leads to a greater loop gain, eventually yielding a constant closed-loop speed. If $C_{in} \ll C_L, C_H$, then (12.52) reduces to C_L/G_m, an expected result because with negligible C_{in}, the output resistance of the unity-gain buffer is equal to $1/G_m$.

We now study the slewing behavior of the circuit, considering a telescopic op amp as an example. Upon entering the amplification mode, the circuit may experience a large step at the inverting input (Fig. 12.37). As shown in Fig. 12.39, the tail current of the op amp's input differential pair is then steered to one side and its mirror current charges the capacitance seen at the output. Since M_2 is off during slewing, C_{in} is negligible and the slew rate is approximately equal to I_{SS}/C_L. The slewing continues until V_X is sufficiently close to the gate voltage of M_1, after which point the settling progresses with the time constant given in (12.52).

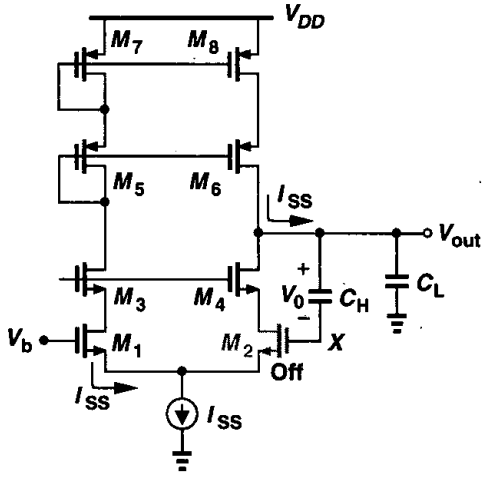


Figure 12.39 Unity-gain sampler during slewing.

Our foregoing studies reveal that the input capacitance of the op amp degrades both the speed and the precision of the unity-gain sampler/buffer. For this reason, the bottom plate of C_H in Fig. 12.30 is usually driven by the input signal or the output of the op amp and the top plate is connected to node X (Fig. 12.40), minimizing the parasitic capacitance seen from node X to ground. This technique is called "bottom-plate sampling." Driving the bottom plate by the input or the output also avoids the injection of substrate noise to node X (Chapter 18).

It is instructive to compare the performance of the sampling circuits shown in Figs. 12.29(b) and 12.30(a). In Fig. 12.29(b), the sampling time constant is smaller because it depends on only the on-resistance of the switch. More importantly, in Fig. 12.29(b), the amplification after the switch turns off is almost instantaneous, whereas in Fig. 12.30 it requires a finite settling time. However, the critical advantage of the unity-gain sampler is the input-independent charge injection.

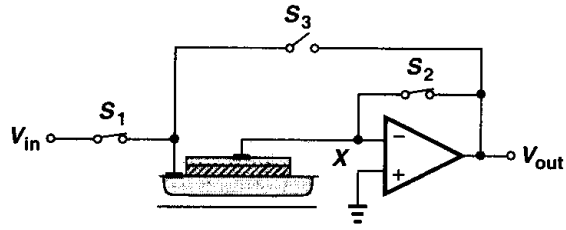


Figure 12.40 Connection of capacitor to the unity-gain sampler.

12.3.2 Noninverting Amplifier

In this section, we revisit the amplifier of Fig. 12.4, studying its speed and precision properties. Repeated in Fig. 12.41(a), the amplifier operates as follows. In the sampling mode, S_1 and S_2 are on and S_3 is off, creating a virtual ground at X and allowing the voltage across C_1 to track the input voltage [Fig. 12.41(b)]. At the end of the sampling mode, S_2 turns off first, injecting a constant charge, Δq_2, onto node X. Subsequently, S_1 turns off and S_3 turns on [Fig. 12.41(c)]. Since V_P goes from V_{in0} to 0, the output voltage changes from 0 to approximately $V_{in0}(C_1/C_2)$, providing a voltage gain equal to C_1/C_2. We call the circuit a "noninverting amplifier" because the final output has the same polarity as V_{in0} and the gain can be greater than unity.

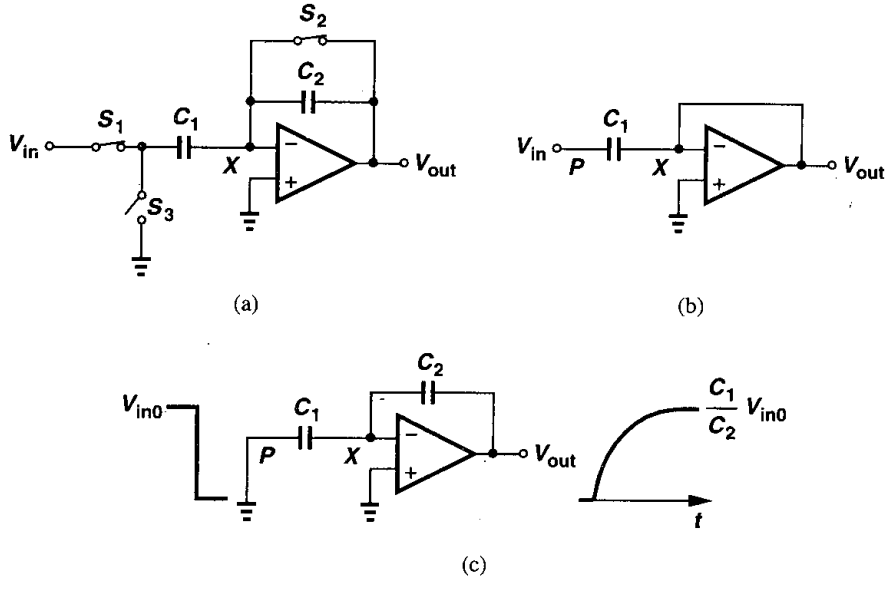


Figure 12.41 (a) Noninverting amplifier, (b) circuit of (a) in sampling mode, (c) transition of circuit to amplification mode.

As with the unity-gain circuit of Fig. 12.30(a), the noninverting amplifier avoids input-dependent charge injection by proper timing, namely, turning S_2 off before S_1 (Fig. 12.42). After S_2 is off, the total charge at node X remains constant, making the circuit insensitive to charge injection of S_1 or charge "absorption" of S_3. Let us first study the effect of S_1 carefully. As illustrated in Fig. 12.43, the charge injected by S_1, Δq_1, changes the voltage at node P by approximately $\Delta V_P = \Delta q_1 / C_1$, and hence the output voltage by $-\Delta q_1 / C_2$. However, after S_3 turns on, V_P drops to zero. Thus, the *overall* change in V_P is equal to $0 - V_{in0} = -V_{in0}$, producing an overall change in the output equal to $-V_{in0}(-C_1/C_2) = V_{in0}C_1/C_2$.

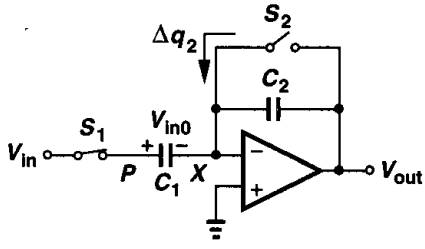


Figure 12.42 Transition of noninverting amplifier to amplification mode.

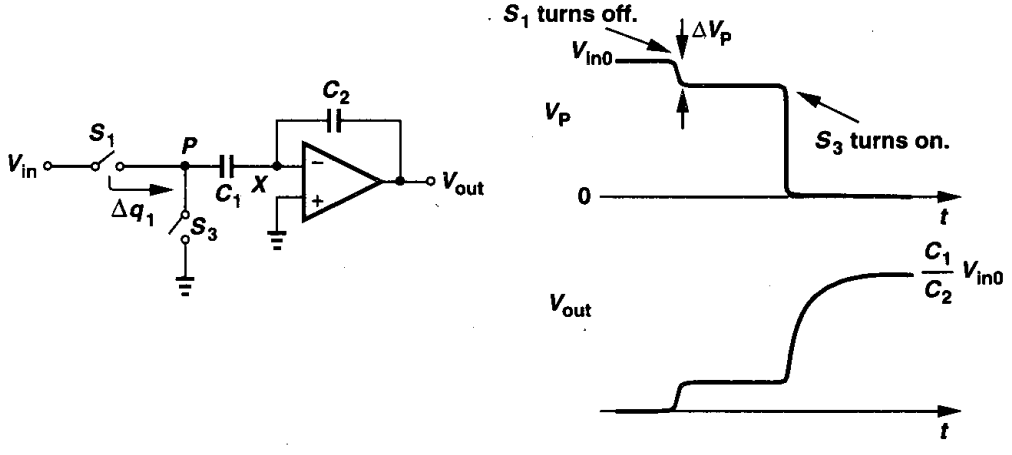


Figure 12.43 Effect of charge injected by S_1.

The key point here is that V_P goes from a fixed voltage, V_0, to another, 0, with an intermediate perturbation due to S_1. Since the output voltage of interest is measured after node P is connected to ground, the charge injected by S_1 does not affect the final output. From another perspective, as shown in Fig. 12.44, the charge on the right plate of C_1 at the instant S_2 turns off is approximately equal to $-V_{in0}C_1$. Also, the total charge at node X must remain constant after S_2 turns off. Thus, when node P is connected to ground and the circuit settles, the voltage across C_1 and hence its charge are nearly zero, and the charge $-V_{in0}C_1$ must reside on the left plate of C_2. In other words, the output voltage is approximately equal to $V_{in0}C_1/C_2$ regardless of the intermediate excursions at node P.

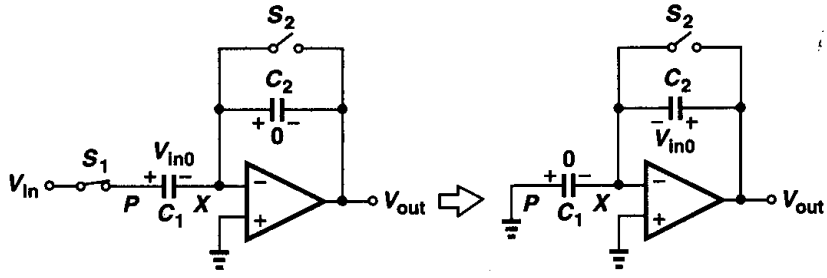


Figure 12.44 Charge redistribution in noninverting amplifier.

The foregoing discussion indicates that two other phenomena have no effect on the final output. First, from the time S_2 turns off until the time S_1 turns off, the input voltage may change significantly (Fig. 12.45) without introducing any error. In other words, the sampling instant is defined by the turn-off of S_2. Second, when S_3 turns on, it requires some channel charge but since the final value of V_P is zero, this charge is unimportant. Neither of these effects introduces error because the total charge at node X is conserved and V_P is eventually set by a fixed (zero) potential. To emphasize that V_P is initially and finally determined by fixed voltages, we say node P is "driven" or node P switches from a low-impedance node to another low-impedance node. Here the term low-impedance distinguishes node P, at which charge is not conserved, from "floating" nodes such as X, where charge is conserved.

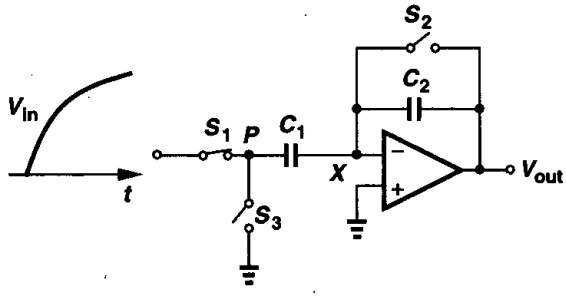


Figure 12.45 Effect of input change after S_2 turns off.

In summary, proper timing in Fig. 12.41(a) ensures that node X is perturbed by only the charge injection of S_2, making the final value of V_{out} free from errors due to S_1 and S_3. The constant offset due to S_2 can be suppressed by differential operation (Fig. 12.46).

Example 12.5

In the differential circuit of Fig. 12.46, suppose the equalizing switch is not used and S_2 and S_2' exhibit a threshold voltage mismatch of 10 mV. If $C_1 = 1$ pF, $C_2 = 0.5$ pF, $V_{TH} = 0.6$ V, and for all switches $WLC_{ox} = 50$ fF, calculate the dc offset measured at the output assuming all of the channel charge of S_2 and S_2' is injected onto X and Y, respectively.

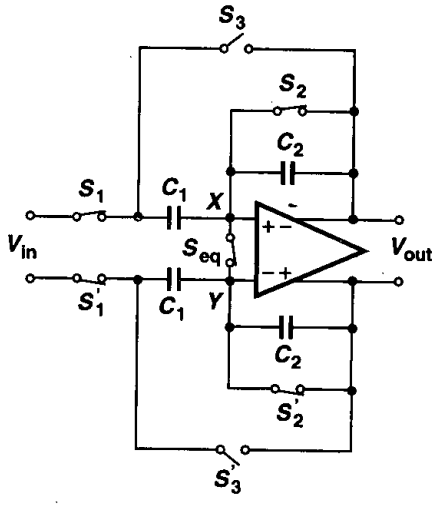


Figure 12.46 Differential realization of noninverting amplifier.

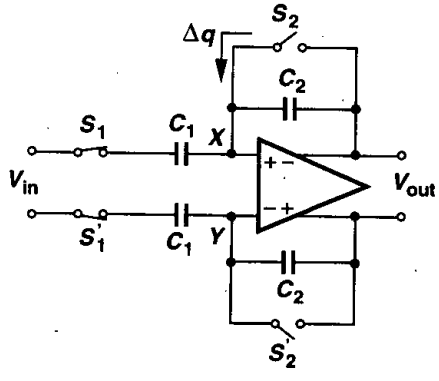


Figure 12.47

Solution

Simplifying the circuit as in Fig. 12.47, we have $V_{out} \approx \Delta q/C_2$, where $\Delta q = WLC_{ox}\Delta V_{TH}$. Note that C_1 does not appear in the result because X is a virtual ground, i.e., the voltage across C_1 changes only negligibly. Thus, the injected charge resides primarily on the left plate of C_2, giving an output error voltage equal to $\Delta V_{out} = WLC_{ox}\Delta V_{TH}/C_2 = 1$ mV.

Precision Considerations As mentioned above, the circuit of Fig. 12.41(a) provides a nominal voltage gain of C_1/C_2. We now calculate the actual gain if the op amp exhibits a finite open-loop gain equal to A_{v1}. Depicted in Fig. 12.48 along with the input capacitance of the op amp, the circuit amplifies the input voltage change such that:

$$(V_{out} - V_X)C_2 s = V_X C_{in} s + (V_X - V_{in})C_1 s. \tag{12.53}$$

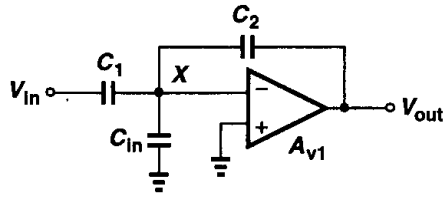


Figure 12.48 Equivalent circuit of noninverting amplifier during amplification.

Since $V_{out} = -A_{v1}V_X$, we have

$$\left|\frac{V_{out}}{V_{in}}\right| = \frac{C_1}{C_2 + \dfrac{C_2 + C_1 + C_{in}}{A_{v1}}}. \tag{12.54}$$

For large A_{v1},

$$\left|\frac{V_{out}}{V_{in}}\right| \approx \frac{C_1}{C_2}\left(1 - \frac{C_2 + C_1 + C_{in}}{C_2} \cdot \frac{1}{A_{v1}}\right), \tag{12.55}$$

implying that the amplifier suffers from a gain error of $(C_2 + C_1 + C_{in})/(C_2 A_{v1})$. Note that the gain error increases with the nominal gain C_1/C_2.

Comparing (12.44) with (12.55), we note that with $C_H = C_2$ and for a nominal gain of unity, the noninverting amplifier exhibits greater gain error than does the unity-gain sampler. This is because the feedback factor equals $C_2/(C_1 + C_{in} + C_2)$ in the former and $C_H/(C_H + C_{in})$ in the latter. For example, if C_{in} is negligible, the unity-gain sampler's gain error is half that of the noninverting amplifier.

Speed Considerations The smaller feedback factor in Fig. 12.48 suggests that the time response of the amplifier may be slower than that of the unity-gain sampler. This is indeed true. Consider the equivalent circuit shown in Fig. 12.49(a). Since the only difference between this circuit and that in Fig. 12.38 is the capacitor C_1, which is connected from node X to an ideal voltage source, we expect that (12.52) gives the time constant of this amplifier as well if C_{in} is replaced by $C_{in} + C_1$. But for a more rigorous analysis, we substitute V_{in}, C_1, and C_{in} in Fig. 12.49(a) by a Thevenin equivalent as in Fig. 12.49(b), where $\alpha = C_1/(C_1 + C_{in})$, and $C_{eq} = C_1 + C_{in}$, and note that

$$V_X = (\alpha V_{in} - V_{out})\frac{C_{eq}}{C_{eq} + C_2} + V_{out}. \tag{12.56}$$

Thus,

$$\left[(\alpha V_{in} - V_{out})\frac{C_{eq}}{C_{eq} + C_2} + V_{out}\right] G_m + V_{out}\left(\frac{1}{R_0} + C_L s\right) = (\alpha V_{in} - V_{out})\frac{C_{eq}C_2}{C_{eq} + C_2}s, \tag{12.57}$$

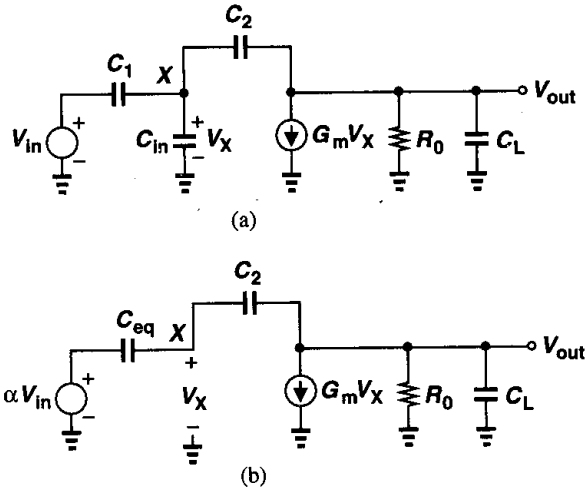


Figure 12.49 (a) Equivalent circuit of noninverting amplifier in amplification mode, (b) circuit of (a) with V_{in}, C_1, and C_{in} replaced by a Thevenin equivalent.

and hence

$$\frac{V_{out}}{V_{in}}(s) = \frac{-C_{eq}\dfrac{C_1}{C_1 + C_{in}}(G_m - C_2 s)R_0}{C_2 G_m R_0 + C_{eq} + C_2 + R_0[C_L(C_{eq} + C_2) + C_{eq}C_2]s}. \tag{12.58}$$

Note that for $s = 0$, (12.58) reduces to (12.54). For a large $G_m R_0$, we can simplify (12.58) to

$$\frac{V_{out}}{V_{in}}(s) \approx \frac{-C_{eq}\dfrac{C_1}{C_1 + C_{in}}(G_m - C_2 s)R_0}{R_0(C_L C_{eq} + C_L C_2 + C_{eq}C_2)s + G_m R_0 C_2}, \tag{12.59}$$

obtaining a time constant of

$$\tau_{amp} = \frac{C_L C_{eq} + C_L C_2 + C_{eq}C_2}{G_m C_2}, \tag{12.60}$$

which is the same as the time constant of Fig. 12.37 if C_{in} is replaced by $C_{in} + C_1$. Note the direct dependence of τ_{amp} upon the nominal gain, C_1/C_2.

It is instructive to examine the amplifier's time constant for the special case $C_L = 0$. Equation (12.60) yields $\tau_{amp} = (C_1 + C_{in})/G_m$, a value *independent* of the feedback capacitor. This is because, while a larger C_2 introduces heavier loading at the output, it also provides a greater feedback factor.

The reader may wonder why Eq. (12.58) yields a negative gain for the circuit that we have called a "noninverting" amplifier. This equation simply means if the left plate of C_1 is

stepped *down*, then the output goes *up*. This does not contradict the operation of the original circuit (Fig. 12.41), where the *change* in V_P is equal to $-V_{in}$.

12.3.3 Precision Multiply-by-Two Circuit

The circuit of Fig. 12.41(a) can operate with a relatively high closed-loop gain, but it suffers from speed and precision degradation due to the low feedback factor. In this section, we study a topology that provides a nominal gain of two while achieving a higher speed and lower gain error [5]. Shown in Fig. 12.50(a), the amplifier incorporates two equal capacitors, $C_1 = C_2 = C$. In the sampling mode, the circuit is configured as in Fig. 12.50(b),

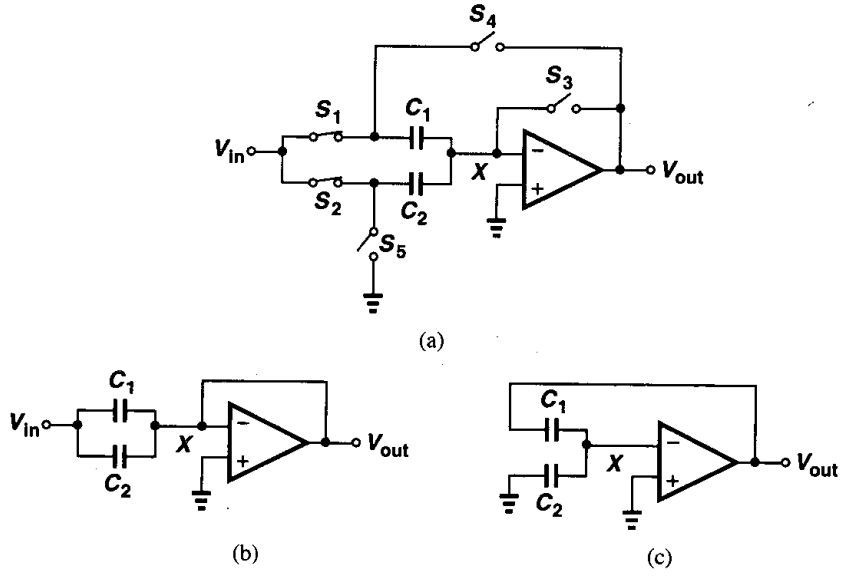


Figure 12.50 (a) Multiply-by-two circuit, (b) circuit of (a) in sampling mode, (c) circuit of (a) in amplification mode.

establishing a virtual ground at X and allowing the voltage across C_1 and C_2 to track V_{in}. In the transition to the amplification mode, S_3 turns off first, C_1 is placed around the op amp, and the left plate of C_2 is switched to ground [Fig. 12.50(c)]. Since at the moment S_3 turns off, the total charge on C_1 and C_2 equals $2V_{in0}C$ (if the charge injected by S_3 is neglected), and since the voltage across C_2 approaches zero in the amplification mode, the final voltage across C_1 and hence the output voltage are approximately equal to $2V_{in0}$. This can also be seen from the slow motion illustration of Fig. 12.51.

The reader can show that the charge injected by S_1 and S_2 and absorbed by S_4 and S_5 is unimportant and that injected by S_3 introduces a constant offset. The offset can be suppressed by differential operation.

The speed and precision of the multiply-by-two circuit are expressed by (12.60) and (12.55), respectively, but the advantage of the circuit is the higher feedback factor for a given closed-loop gain. Note, however, that the input capacitance of the multiply-by-two circuit in the sampling mode is higher.

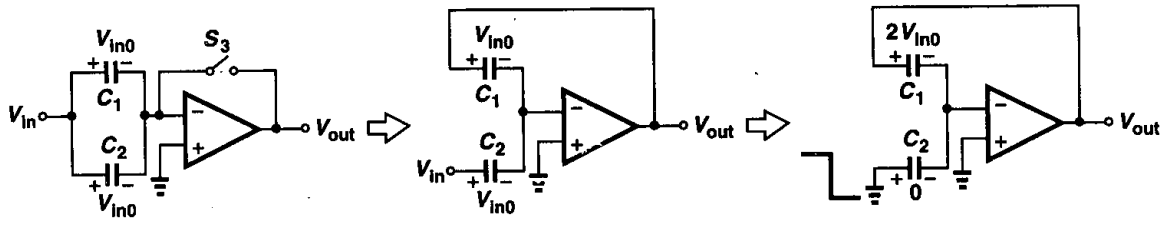


Figure 12.51 Transition of multiply-by-two circuit to amplification mode in slow motion.

12.4 Switched-Capacitor Integrator

Integrators are used in many analog systems. Examples include filters and oversampled analog-to-digital converters. Fig. 12.52 depicts a continuous-time integrator, whose output can be expressed as

$$V_{out} = -\frac{1}{RC_F}\int V_{in}dt, \tag{12.61}$$

if the op amp gain is very large. For sampled-data systems, we must devise a discrete-time counterpart of this circuit.

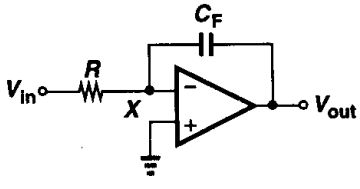


Figure 12.52 Continuous-time integrator.

Before studying SC integrators, let us first point out an interesting property. Consider a resistor connected between two nodes [Fig. 12.53(a)], carrying a current equal to $(V_A - V_B)/R$. The role of the resistor is to take a certain amount of charge from node A every second and move it to node B. Can we perform the same function by a capacitor? Suppose in the circuit of Fig. 12.53(b), capacitor C_S is alternately connected to nodes A and B at a clock rate f_{CK}. The *average* current flowing from A to B is then equal to the charge moved

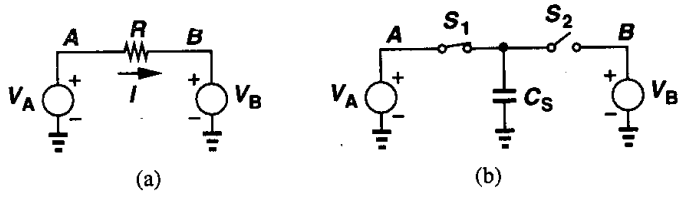


Figure 12.53 (a) Continuous-time and (b) discrete-time resistors.

in one clock period:

$$\overline{I_{AB}} = \frac{C_S(V_A - V_B)}{f_{CK}^{-1}} \tag{12.62}$$

$$= C_S f_{CK}(V_A - V_B). \tag{12.63}$$

We can therefore view the circuit as a "resistor" equal to $(C_S f_{CK})^{-1}$. Recognized by James Clark Maxwell, this property formed the foundation for many modern switched-capacitor circuits.

Let us now replace resistor R in Fig. 12.52 by its discrete-time equivalent, arriving at the integrator of Fig. 12.54(a). We note that in every clock cycle, C_1 absorbs a charge equal

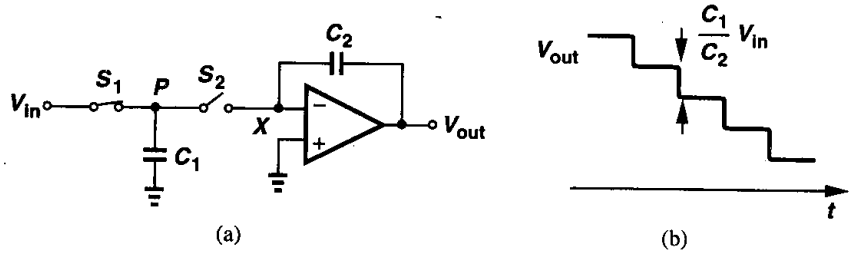


Figure 12.54 (a) Discrete-time integrator, (b) response of circuit to a constant input voltage.

to $C_1 V_{in}$ when S_1 is on and deposits the charge on C_2 when S_2 is on (node X is a virtual ground). For example, if V_{in} is constant, the output changes by $V_{in}C_1/C_2$ every clock cycle [Fig. 12.54(b)]. Approximating the staircase waveform by a ramp, we note that the circuit behaves as an integrator.

The final value of V_{out} in Fig. 12.54(a) after every clock cycle can be written as

$$V_{out}(kT_{CK}) = V_{out}[(k-1)T_{CK}] - V_{in}[(k-1)T_{CK}] \cdot \frac{C_1}{C_2}, \tag{12.64}$$

where the gain of the op amp is assumed large. Note that the small-signal settling time constant as charge is transferred from C_1 to C_2 is given by (12.52).

The integrator of Fig. 12.54(a) suffers from two important drawbacks. First, the input-dependent charge injection of S_1 introduces nonlinearity in the charge stored on C_1 and hence the output voltage. Second, the nonlinear capacitance at node P resulting from the source/drain junctions of S_1 and S_2 leads to a nonlinear charge-to-voltage conversion when C_1 is switched to X. This can be understood with the aid of Fig. 12,55, where the charge stored on the total junction capacitance, C_j, is *not* equal to $V_{in0}C_j$, but rather equal to

$$q_{cj} = \int_0^{Vin0} C_j dV. \tag{12.65}$$

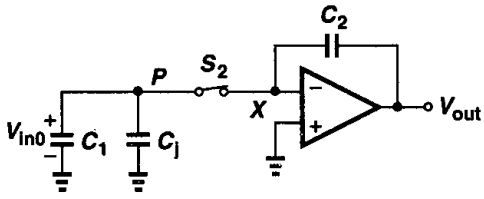


Figure 12.55 Effect of junction capacitance nonlinearity in SC integrator.

Since C_j is a function of voltage, q_{cj} exhibits a nonlinear dependence on V_{in0}, thereby creating a nonlinear component at the output after the charge is transferred to the integration capacitor.

An integrator topology that resolves both of the foregoing issues is shown in Fig. 12.56(a). We study the circuit's operation in the sampling and integration modes. As shown in Fig. 12.56(b), in the sampling mode S_1 and S_3 are on and S_2 and S_4 are off, allowing the voltage across C_1 to track V_{in} while the op amp and C_2 hold the previous value. In the transition to the integration mode, S_3 turns off first, injecting a constant charge onto C_1, S_1 turns off next, and subsequently S_2 and S_4 turn on [Fig. 12.56(c)]. The charge stored on C_1 is therefore transferred to C_2 through the virtual ground node.

Since S_3 turns off first, it introduces only a constant offset, which can be suppressed by differential operation. Moreover, because the left plate of C_1 is "driven" (Section 12.3.2), the charge injection or absorption of S_1 and S_2 contributes no error. Also, since node X is a virtual ground, the charge injected or absorbed by S_4 is constant and independent of V_{in}.

How about the nonlinear junction capacitance of S_3 and S_4? We observe that the voltage across this capacitance goes from near zero in the sampling mode to virtual ground in the

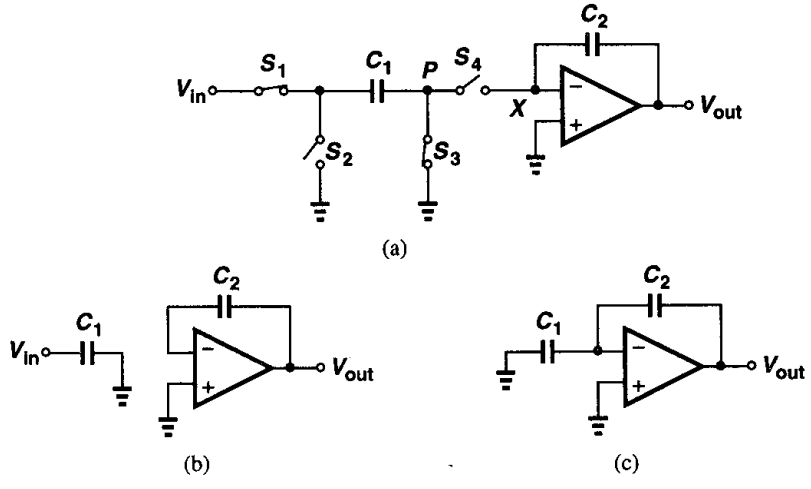


Figure 12.56 (a) Parasitic-insensitive integrator, (b) circuit of (a) in sampling mode, (c) circuit of (a) in integration mode.

integration mode. Since the voltage across the nonlinear capacitance changes by a very small amount, the resulting nonlinearity is negligible.

12.5 Switched-Capacitor Common-Mode Feedback

Our study of common-mode feedback in Chapter 9 suggested that sensing the output CM level by means of resistors lowers the differential voltage gain of the circuit considerably. We also observed that sensing techniques using MOSFETs that operate as source followers or variable resistors suffer from a limited linear range. Switched-capacitor CMFB networks provide an alternative that avoids both of these difficulties (but the circuit must be refreshed periodically.)

In switched-capacitor common-mode feedback, the outputs are sensed by capacitors rather than resistors. Figure 12.57 depicts a simple example, where equal capacitors C_1 and C_2 reproduce at node X the average of the changes in each output voltage. Thus, if V_{out1} and V_{out2} experience, say, a positive CM change, then V_X and hence I_{D5} increase, pulling V_{out1} and V_{out2} down. The output CM level is then equal to V_{GS2} plus the voltage across C_1 and C_2.

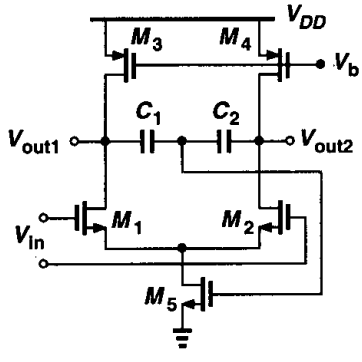


Figure 12.57 Simple SC common-mode feedback.

How is the voltage across C_1 and C_2 defined? This is typically carried out when the amplifier is in the sampling (or reset) mode and can be accomplished as shown in Fig. 12.58. Here, during CM level definition, the amplifier differential input is zero and switch S_1 is on. Transistors M_6 and M_7 operate as a linear sense circuit because their gate voltages are nominally equal. Thus, the circuit settles such that the ouput CM level is equal to $V_{GS6,7} + V_{GS5}$. At the end of this mode, S_1 turns off, leaving a voltage equal to $V_{GS6,7}$ across C_1 and C_2. In the amplification mode, M_6 and M_7 may experience a large nonlinearity but they do not impact the performance of the main circuit because S_1 is off.

In applications where the output CM level must be defined more accurately than in the above example, the topology shown in Fig. 12.59 may be used. Here, in the reset mode, one plate of C_1 and C_2 is switched to V_{CM} while the other is connected to the gate of M_6. Each capacitor therefore sustains a voltage equal to $V_{CM} - V_{GS6}$. In the amplification mode, S_2 and S_3 are on and the other switches are off, yielding an output CM level equal to $V_{CM} - V_{GS6} + V_{GS5}$. Proper definition of I_{D3} and I_{D4} with respect to I_{REF} can guarantee that $V_{GS5} = V_{GS6}$ and hence the output CM level is equal to V_{CM}.

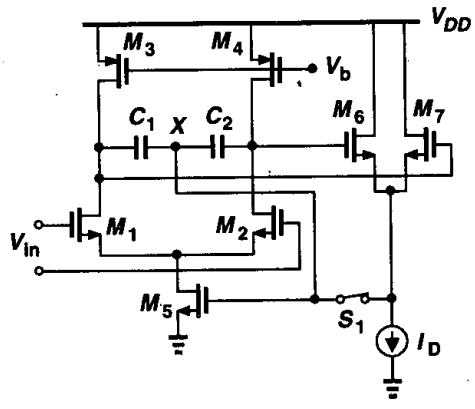


Figure 12.58 Definition of the voltage across C_1 and C_2.

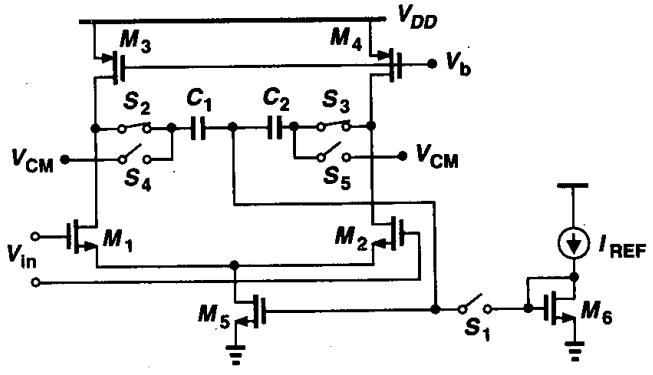


Figure 12.59 Alternative topology for definition of output CM level.

With large output swings, the speed of the CMFB loop may in fact influence the settling of the differential output [6]. For this reason, part of the tail current of the differential pairs in Figs. 12.58 and 12.59 can be provided by a *constant* current source so that M_5 makes only small adjustments to the circuit.

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.

12.1. The circuit of Fig. 12.2(b) is designed with $C_1 = 2$ pF and $C_2 = 0.5$ pF.

(a) Assuming $R_F = \infty$ but the op amp has an output resistance R_{out}, derive the transfer function $V_{out}(s)/V_{in}(s)$.

(b) If the op amp is ideal, determine the minimum value of R_F that guarantees a gain error of 1% for an input frequency of 1 MHz.

12.2. Suppose in Fig. 12.5(a), the op amp is characterized by a transconductance G_m and an output resistance R_{out}.
(a) Determine the transfer function V_{out}/V_{in} in this mode.
(b) Plot the waveform at node B if V_{in} is a 100-MHz sinusoid with a peak amplitude of 1 V, $C_1 = 1$ pF, $G_m = 1/(100\ \Omega)$, and $R_{out} = 20\ \text{k}\Omega$.

12.3. In Fig. 12.5(b), node A is in fact connected to ground through a switch (Fig. 12.4). If the switch introduces a series resistance R_{on} and the op amp is ideal, calculate the time constant of the circuit in this mode. What is the total energy dissipated in the switch as the circuit enters the amplification mode and V_{out} settles to its final value?

12.4. The circuit of Fig. 12.9(a) is designed with $(W/L)_1 = 20/0.5$ and $C_H = 1$ pF.
(a) Using Eqs. (12.9) and (12.16), calculate the time required for V_{out} to drop to +1 mV.
(b) Approximating M_1 by a linear resistor equal to $[\mu_n C_{ox}(W/L)_1(V_{DD}-V_{TH})]^{-1}$, calculate the time required for V_{out} to drop to +1 mV and compare the result with that obtained in part (a).

12.5. The circuit of Fig. 12.11 cannot be characterized by a single time constant because the resistance charging C_H (equal to $1/g_{m1}$ if $\gamma = 0$) varies with the output level. Assume $(W/L)_1 = 20/0.5$ and $C_H = 1$ pF.
(a) Using Eq. (12.21), calculate the time required for V_{out} to reach 2.1 V.
(b) Sketch the transconductance of M_1 versus time.

12.6. In the circuit of Fig. 12.8(b), $(W/L)_1 = 20/0.5$ and $C_H = 1$ pF. Assume $\lambda = \gamma = 0$ and $V_{in} = V_0 \sin \omega_{in} t + V_m$, where $\omega_{in} = 2\pi \times (100\ \text{MHz})$.
(a) Calculate R_{on1} and the phase shift from the input to the output if $V_0 = V_m = 10$ mV.
(b) Repeat part (a) if $V_0 = 10$ mV but $V_m = 1$ V. The variation of the phase shift translates to distortion.

12.7. Describe an efficient SPICE simulation that yields the plot of $R_{on,eq}$ for the circuit of Fig. 12.16.

12.8. The sampling network of Fig. 12.16 is designed with $(W/L)_1 = 20/0.5$, $(W/L)_2 = 60/0.5$, and $C_H = 1$ pF. If $V_{in} = 0$ and the initial value of V_{out} is +3 V, estimate the time required for V_{out} to drop to +1 mV.

12.9. In the circuit of Fig. 12.19, $(W/L)_1 = 20/0.5$ and $C_H = 1$ pF. Calculate the maximum error at the output due to charge injection. Compare this error with that resulting from clock feedthrough.

12.10. The circuit of Fig. 12.60 samples the input on C_1 when CK is high and connects C_1 and C_2 when CK is low. Assume $(W/L)_1 = (W/L)_2$ and $C_1 = C_2$.
(a) If the initial voltages across C_1 and C_2 are zero and $V_{in} = 2$ V, plot V_{out} versus time for many clock cycles. Neglect charge injection and clock feedthrough.

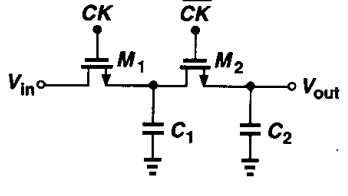


Figure 12.60

(b) What is the maximum error in V_{out} due to charge injection and clock feedthrough of M_1 and M_2? Assume the channel charge of M_2 splits equally between C_1 and C_2.

(c) Determine the sampled kT/C noise at the output after M_2 turns off.

12.11. For $V_{in} = V_0 \sin \omega_0 t + V_0$, where $V_0 = 0.5$ V and $\omega_0 = 2\pi \times (10$ MHz), plot the output waveforms of the circuits shown in Fig. 12.29(b) and 12.30(a). Assume a clock frequency of 50 MHz.

12.12. In Fig. 12.45, S_1 turns off Δt seconds after S_2 and S_3 turns on Δt seconds after S_1 turns off. Plot the output waveform, taking into account the charge injection and clock feedthough of S_1-S_3. Assume all of the switches are NMOS devices.

12.13. The circuit of Fig. 12.48 is designed with $C_1 = 2$ pF, $C_{in} = 0.2$ pF and $A_v = 1000$. What is the maximum nominal gain, C_1/C_2, that the circuit can provide with a gain error of 1%?

12.14. In Problem 12.13, what is the maximum nominal gain if $G_m = 1/(100\ \Omega)$ and the circuit must achieve a time constant of 2 ns in the amplification mode? Assume $C_{in} = 0.2$ pF and calculate C_1 and C_2.

12.15. The integrator of Fig. 12.54 is designed with $C_1 = C_2 = 1$ pF and a clock frequency of 100 MHz. Neglecting charge injection and clock feedthrough, sketch the output if the input is a 10-MHz sinusoid with a peak amplitude of 0.5 V. Approximating C_1, S_1, and S_2 by a resistor, estimate the output amplitude.

12.16. Consider the switched-capacitor amplifier depicted in Fig. 12.61, where the common-mode feedback is not shown. Assume $(W/L)_{1-4} = 50/0.5$, $I_{SS} = 1$ mA, $C_1 = C_2 = 2$ pF, $C_3 = C_4 = 0.5$ pF, and the output CM level is 1.5 V. Neglect the transistor capacitances.

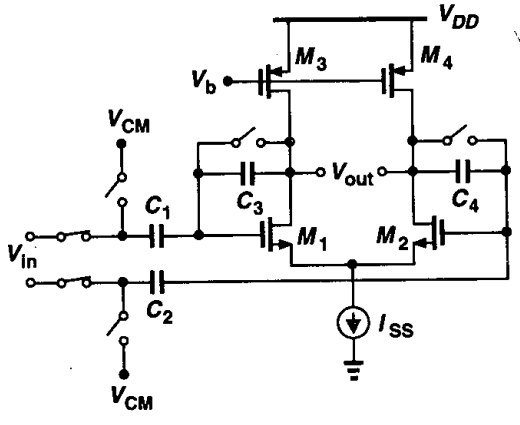


Figure 12.61

(a) What is the maximum allowable output voltage swing in the amplification mode?

(b) Determine the gain error of the amplifier.

(c) What is the small-signal time constant in the amplification mode?

12.17. Repeat Problem 12.16(c) if the gate-source capacitance of M_1 and M_2 is not neglected.

12.18. A differential circuit incorporating a well-designed common-mode feedback network exhibits the open-loop input-output characteristic shown in Fig. 12.62(a). In some circuits, however, the characteristic appears as in Fig. 12.62(b). Explain how this effect occurs.

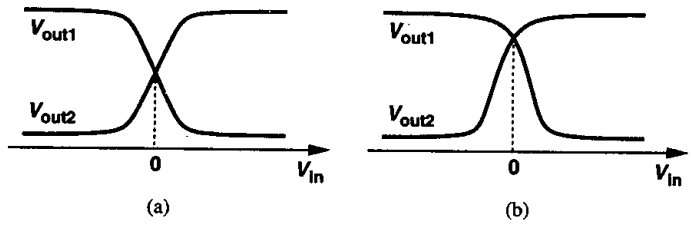


Figure 12.62

12.19. In the common-mode feedback network of Fig. 12.58, assume $W/L = 50/0.5$ for all transistors, $I_{D5} = 1$ mA, and $I_{D6,7} = 50$ μA. Determine the allowable range of the input common-mode level.

12.20. Repeat Problem 12.19 if $(W/L)_{6,7} = 10/0.5$.

12.21. Suppose in the common-mode feedback network of Fig. 12.58, S_1 injects a charge of Δq onto the gate of M_5. How much do the gate voltage of M_5 and the output common-mode level change due to this error?

12.22. In the circuit of Fig. 12.63, each op amp is represented by a Norton equivalent and characterized by G_m and R_{out}. The output currents of two op amps are summed at node Y [7]. (The circuit is shown in the amplification mode.) Note that the main amplifier and the auxiliary amplifier are identical and the error amplifier senses the voltage variation at node X and injects a

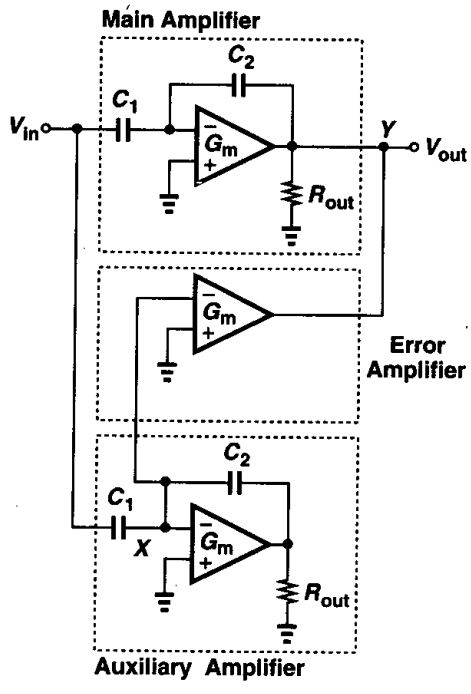


Figure 12.63

proportional current into node Y. The output impedance of the error amplifier is much greater than R_{out}. Assume $G_m R_{out} \gg 1$.

(a) Calculate the gain error of the circuit.

(b) Repeat part **(a)** if the auxiliary and error amplifiers are eliminated and compare the results.

References

1. G. Wegmann, E. A. Vittoz, and F. Rahali, "Charge Injection in Analog MOS Switches," *IEEE J. Solid-State Circuits*, vol. SC-22, pp. 1091–1097, Dec. 1987.
2. B. J. Sheu and C. Hu, "Switch-Induced Error Voltage on a Switched Capacitor," *IEEE J. Solid-State Circuits*, vol. SC-19, pp. 519–525, April 1984.
3. R. Gregorian and G. C. Temes, *Analog MOS Integrated Circuits for Signal Processing*, New York: John Wiley and Sons, 1986.
4. J. H. Fischer, "Noise Sources and Calculation Techniques for Switched Capacitor Filters," *IEEE J. Solid-State Circuits*, vol. 17, pp. 742–752, Aug. 1982.
5. B. S. Song, M. F. Tompsett, and K. R. Lakshmikumar, "A 12-Bit 1-Msample/s Capacitor-Averaging Pipelined A/D Converter," *IEEE J. Solid-State Circuits*, vol. SC-23, pp. 1324–1333, Dec. 1988.
6. B. Razavi, *Principles of Data Conversion System Design*, New York: IEEE Press, 1995.
7. P. C. Yu and H.-S. Lee, "A High-Swing 2-V CMOS Op Amp with Replica-Amp Gain Enhancement," *IEEE J. Solid-State Circuits*, vol. 28, pp. 1265–1272, Dec. 1993.