$V_{out} \geq V_b - V_{TH4}$. 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 $\mu_n C_{ox} = 60\ \mu$A/V^2, $\mu_p C_{ox} = 30\ \mu$A/V^2, $\lambda_n = 0.1$ V^{-1}, $\lambda_p = 0.2$ 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.

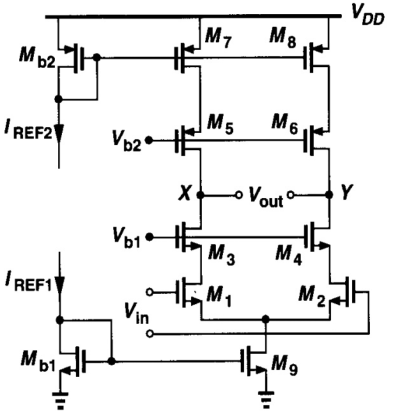

Figure 9.10

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, recal that $\lambda \propto 1/L$, and hence $g_m r_O \propto \sqrt{WL/I_D}$. We can therefore increase the width or length o *decrease* the bias current of the transistors. In practice, speed or noise requirements may dictate th bias current, leaving only the dimensions as the variables. Of course, the width of each transistor mus 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 th signal path, it is desirable to keep their capacitances to a minimum. The PMOS devices, M_5-M_8, o the other hand, affect the signal to a much lesser extent and can therefore have larger dimensions. Doubling the (effective) length and width of each of these transistors in fact *doubles* their $g_m r_O$ becaus g_m remains constant while r_O increases by a factor of 2. Choosing $(W/L)_{5-8} = 1111\ \mu$m/1.0 μm an hence $\lambda_p = 0.1$ V^{-1}, we obtain $A_v \approx 4000$. Thus, the PMOS dimensions can be somewhat smalle Note that with such large dimensions for PMOS transistors, we may revisit our earlier distribution c the overdrive voltages, possibly reducing that of M_9 by 100 to 200 mV and allocating more to th 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 s 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 $V_{GS3} + V_{OD1} + V_{OD9} = 1.$ 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 proce variations. Also, the increase in the threshold voltages due to body effect must be taken into accoun

[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

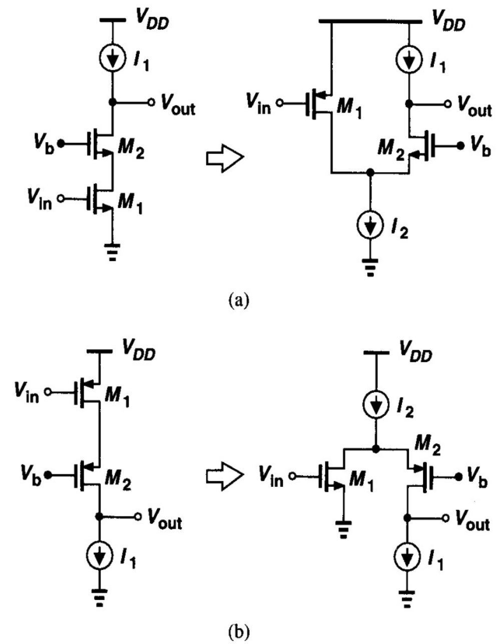

Figure 9.11 Folded cascode circuits.

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