3 Basic Sizing Using the g_m/I_D Methodology

This chapter introduces the reader to the concepts underpinning the g_m/I_D sizing methodology. To simplify this initial treatment, we focus on elementary circuits with only a few transistors and leave the sizing of more complex circuits for later chapters.

3.1 Sizing an Intrinsic Gain Stage (IGS)

The starting point of our discussion is the intrinsic gain stage (IGS), shown in Figure 3.1. The IGS can be viewed as an idealized version of a common-source stage with active load, which is frequently used in linear amplifiers. It also represents the small-signal half-circuit model of an actively-loaded differential pair, which is typically used as the input stage of a differential amplifier. The term "intrinsic" reflects the fact that no external components (other than the load capacitance C_L) are considered. For simplicity, we also assume that the stage is driven by an ideal voltage source and defer more realistic scenarios to later chapters (and Example 3.13).

In the circuit of Figure 3.1, the drain current is set up using an ideal current source (I_D) and the input bias voltage (V_{BIAS}) is assumed to be adjusted such that the transistor operates in saturation at some desired output quiescent point (e.g. $V_{OUT} = V_{DD}/2 = 0.6$ V). We will leave out details on how the bias voltage is generated, since such discussions are much more meaningful in the context of larger circuits. For basic biasing considerations, the reader may refer to introductory textbooks on CMOS transistor stages [1].

3.1.1 Circuit Analysis

The first step in any systematic design flow is to perform a suitable circuit analysis. For this purpose, we briefly review the frequency response of the IGS, considering its small-signal model shown in Figure 3.2. This model contains the gate capacitances C_{gs}, C_{gb} and $C_{gd,}$ whichwere already introduced in Section 2.3.8, as well as the drain-to-bulk junction capacitance C_{db}.

To simplify further, we will initially neglect the junction capacitance and assume $C_{db} << C_L$. Additionally, it is usually true that C_{gd} is much smaller than C_L. Under

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

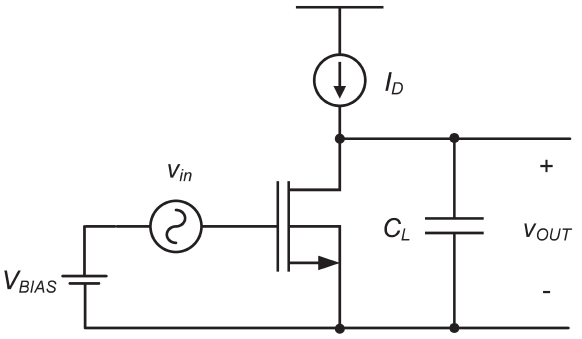

Figure 3.1 Circuit schematic of an intrinsic gain stage.

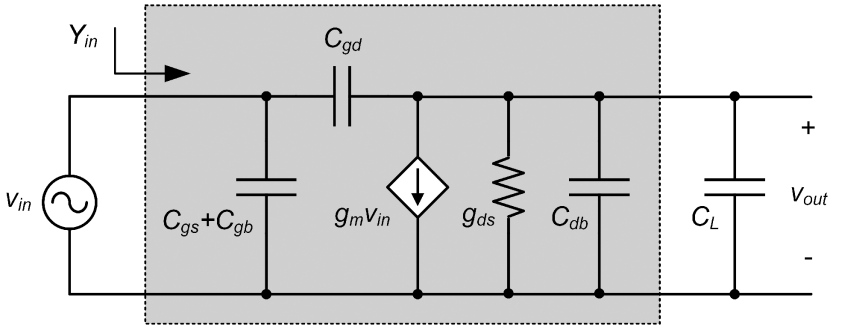

Figure 3.2 Small-signal model of the IGS (the transistor's bulk is assumed to be connected to the source).

these conditions, the frequency response is well-approximated by the following expression:

$$A_v(j\omega) = \frac{v_{out}}{v_{in}} \cong \frac{A_{v0}}{1 + j\dfrac{\omega}{\omega_c}}, \tag{3.1}$$

where A_{v0} is the low-frequency (LF) voltage gain (equal in magnitude to the intrinsic gain of the transistor):

$$A_{v0} = -\frac{g_m}{g_{ds}} = -A_{intr} \tag{3.2}$$

and ω_c is the circuit's angular corner frequency, given by

$$\omega_c = \frac{g_{ds}}{C_L}. \tag{3.3}$$

Figure 3.3 shows a straight-line approximation of the frequency response magnitude. The gain is constant at low frequencies and falls with –20 dB/decade beyond

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

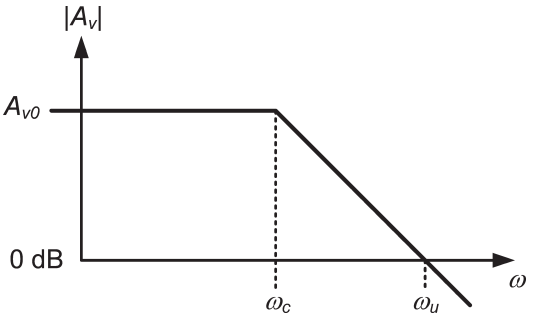

Figure 3.3 Magnitude response of the IGS.

the corner frequency. The low- and high-frequency asymptotes meet at ω_c. Another significant point is where the high-frequency asymptote crosses the horizontal axis, which defines the so-called angular unity gain frequency ω_u:

$$\omega_u \cong \frac{g_m}{C_L}. \tag{3.4}$$

Since the considered IGS is a first-order system, ω_u is (approximately) equal to the product of ω_c and A_{v0}. The frequency $f_u = \omega_u/2\pi$ is therefore commonly called the gain-bandwidth product (GBW).

Another important figure of merit of the IGS is how much output capacitance it drives relative to the capacitance it presents at its input. To quantify this ratio, we define the stage's fan-out (FO) as:

$$FO = \frac{C_L}{C_{gs} + C_{gb} + C_{gd}} = \frac{C_L}{C_{gg}}. \tag{3.5}$$

While the fan-out is known as a significant metric in the sizing of digital logic gates, it tends to be an underappreciated concept in analog design. However, as we shall see throughout this book, some analog design problems can be elegantly framed using FO.

Using FO to express the stage loading in a normalized manner is motivated by the fact that the load is often just another transistor stage (or general circuit) whose input capacitance is linked to the global specifications in a similar way as the stage under consideration. In many circuits, the capacitances of all components scale up and down together when different combinations of gain, bandwidth and noise specifications are being explored in the overall design space. For example, we will see in Chapter 4 that FO plays a significant role in the tradeoff between GBW, supply current and noise performance of the IGS.

The fan-out metric defined in (3.5) also coincides with a frequency ratio that is important from a modeling perspective:

$$\frac{\omega_T}{\omega_u} = \frac{f_T}{f_u} = \frac{g_m / C_{gg}}{g_m / C_L} = FO. \tag{3.6}$$

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

As we have argued in Chapter 2, the quasi-static transistor model used in Figure 3.2 becomes inaccurate as the frequencies of interest approach about 1/10th of f_T in moderate or strong inversion. Hence, for the IGS model to hold near f_u, the fan-out should ideally be larger than 10.

As a final note, it should be mentioned that the exact input admittance of the IGS is neither purely capacitive, nor exactly equal to $j\omega C_{gg}$. Using the circuit model in Figure 3.2, and applying the Miller theorem [1], it follows that:

$$Y_{in}(j\omega) = j\omega\left(C_{gs} + C_{gb}\right) + j\omega C_{gd}\left(1 - A_v(j\omega)\right). \tag{3.7}$$

This expression reduces asymptotically to $j\omega C_{gg}$ as the circuit loses its voltage gain at high frequencies (or if the drain is ac-grounded). The fan-out metric that we defined in (3.5) should therefore be viewed as an asymptotic metric that was intentionally simplified to be useful without getting into the complexities of accurate admittance modeling.

3.1.2 Sizing Considerations

With the expressions derived in the previous section, we are now ready to size the IGS per a given set of specifications. Specifically, for a given load capacitance C_L, we typically want to establish a methodology that enables us to determine the drain current, the device width and gate length for a given unity-gain frequency (f_u) target. For this purpose, we can consider the generic sizing flow that was already introduced in Chapter 1 (repeated here for convenience):

1. Determine g_m (from design specifications).
2. Pick L:
   - short channel → high speed, small area;
   - long channel → high intrinsic gain, improved matching, …
3. Pick g_m/I_D:
   - large g_m/I_D → low power, large signal swing (low V_{Dsat});
   - small g_m/I_D → high speed, small area.
4. Determine I_D (from g_m and g_m/I_D).
5. Determine W (from I_D/W).

Step 1 of this flow is straightforward for the given IGS circuit, since g_m is fixed per (3.4) and must be equal to $\omega_u C_L$. However, additional constraints are needed to decide on the best choice for the channel length L and the transconductance efficiency g_m/I_D. For example, suppose that we want to minimize the drain current, which is given by:

$$I_D = \frac{g_m}{g_m / I_D}. \tag{3.8}$$

Based on this expression alone, it would make sense to operate the transistor at the maximum possible g_m/I_D, which is obtained in weak inversion. Unfortunately, the

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

problem with operating the device at high g_m/I_D is that the transistor's f_T is small, leading to large gate capacitance and a fan-out that may be smaller than the recommended bound of 10.

The take-home from this initial discussion is that finding the optimal inversion level is not straightforward, and that all applicable design constraints must be considered jointly. In the remainder of this chapter and the next, we will therefore gradually introduce a variety of design constraints, and subsequently show how they impact the choice of L and g_m/I_D. However, an important prerequisite for this exploration is the ability to size a device once g_m/I_D and L have been chosen (by some design/optimization criterion). This procedure is thus the topic of the next section.

3.1.3 Sizing for Given L and g_m/I_D

If L and g_m/I_D are known, the sizing procedure simplifies to the flow illustrated in Figure 3.4. As discussed earlier, g_m follows from the design specifications, while I_D can be directly computed using g_m and g_m/I_D via (3.8). The last unknown parameter is the device width, which follows from the ratio of drain current and drain current density $J_D = I_D/W$ (expressed in A/μm):

$$W = \frac{I_D}{J_D} = \frac{I_D}{I_D/W}. \tag{3.9}$$

We refer to the step of computing the device width as "de-normalization," as it marks the transition from normalized quantities like g_m/I_D and I_D/W to absolute geometries.

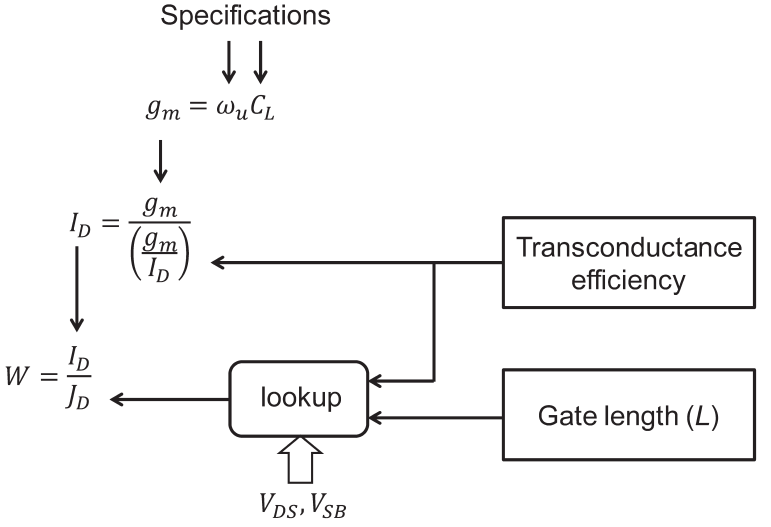

Figure 3.4 Sizing for given L and g_m/I_D.

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

Equation (3.9) assumes that the drain current scales strictly proportional to the transistor width, which requires that the device is large enough to make narrow-width effects negligible. Fortunately, this is a condition that is met in most analog circuits. Appendix 3 takes a closer look at this assumption.

Since there is a one-to-one mapping between transconductance efficiency and current density for given L, V_{DS} and V_{SB} (see Section 2.3.1), we can find J_D using the lookup table data for our technology and thereby complete the sizing. The mechanics of this approach are illustrated through the following example.

Example 3.1 A Basic Sizing Example

Size the circuit of Figure 3.1 so that that f_u = 1 GHz when C_L = 1 pF. Assume L = 60 nm, g_m/I_D = 15 S/A (moderate inversion), V_{DS} = 0.6 V and V_{SB} = 0 V (default values). Find the low-frequency voltage gain and the Early voltage of the transistor. Validate the results through SPICE simulations.

SOLUTION

Following the above-discussed procedure, we start by computing the transconductance using (3.4):

```
fu = 1e9;
CL = 1e-12;
gm = 2*pi*fu*CL;
```

Since g_m/I_D is given and equal to 15 S/A, we can find I_D via (3.8):

```
ID = gm/gm_ID
```

This yields I_D = 419 μA. To find the width W, we divide I_D by the drain current density J_D per (3.9). Conceptually, to find J_D, we can consider a plot like that of Figure 2.14(b) (for L = 60 nm), and look for the drain current density at g_m/I_D = 15 S/A. Equivalently, this is accomplished using the following lookup command:

```
JD = lookup(nch,'ID_W','GM_ID',gm_ID,'VDS',VDS,'VSB',VSB,'L',L);
```

where V_{DS}, V_{SB} and L correspond to the drain-to-source, source-to-bulk voltage and gate length of the transistor. We find that J_D is equal to 10.05 μA/μm. Hence, dividing I_D/J_D yields W = 41.72 μm and we have thus completed the sizing procedure.

We see from this flow that the drain current I_D is known as soon as we fix the inversion level (via g_m/I_D). Fixing the gate length then also determines W. These steps not only fix I_D and W, but also V_{GS}, A_{v0} and f_T, since these are the outcome of similar lookup operations involving the same variables. For example, we can find V_{GS} using the lookupVGS companion function (see Appendix 2 for a description):

```
VGS = lookupVGS(nch,'GM_ID',gm_ID,'VDS',VDS,'VSB',VSB,'L',L);
```

This yields V_{GS} = 0.4683 V. Using similar commands, we can also find the low-frequency voltage gain and the device's transit frequency:

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

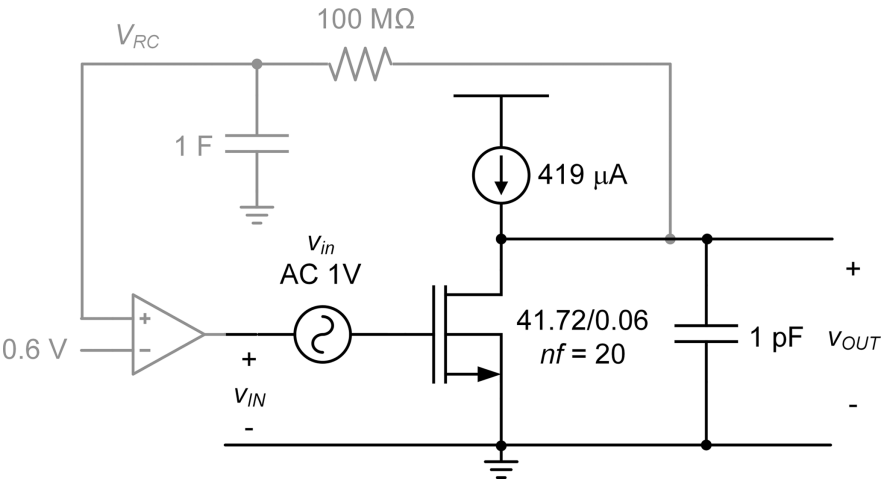

Figure 3.5 Schematic for SPICE simulation.[1]

```
Av0 = -lookup(nch,'GM_GDS','GM_ID',gm_ID,'VDS',VDS,'VSB',VSB,'L',L)
fT  = lookup(nch,'GM_CGG','GM_ID',gm_ID,'VDS',VDS,'VSB',VSB, ...
'L',L)/2/pi
```

This yields A_{vo} = –10.25 and f_T = 26.46 GHz. Note that f_T is much larger than the desired unity gain frequency. This implies that the FO is sufficiently large (FO = 26.46 > 10), and that the equivalent circuit of Figure 3.2 is valid. Finally, we compute the Early voltage V_A using:

$$V_A = \frac{I_D}{g_{ds}} = \frac{\left(\frac{g_m}{g_{ds}}\right)}{\left(\frac{g_m}{I_D}\right)}$$

This yields the rather low value of V_A = 0.683 V, due to the strong DIBL effect for L = 60 nm (see Sections 2.3.5 and 2.3.6).

We turn now to verification, and run SPICE simulations using the setup shown in Figure 3.5. The transistor is partitioned into 20 fingers (nf = 20), each 2.086 µm wide (see Appendix 3 for a discussion on finger partitioning). The quiescent point gate voltage is set using an auxiliary feedback circuit (drawn in gray) that computes V_{GS} such that V_{DS} = 0.6 V. For any meaningful frequency above DC, the feedback loop is open, and the circuit is evaluated across frequency as intended. In a practical implementation of this circuit, the gate voltage is sometimes set via an actual feedback circuit (similar to what is shown), or computed using a replica circuit [1]. More commonly, the transistor

[1] The AC amplitude of 1 V is conveniently chosen so that the transfer function follows directly from the output voltage. Note that one is free to choose the test amplitude, since the circuit is perfectly linear in the performed small-signal AC analysis.

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

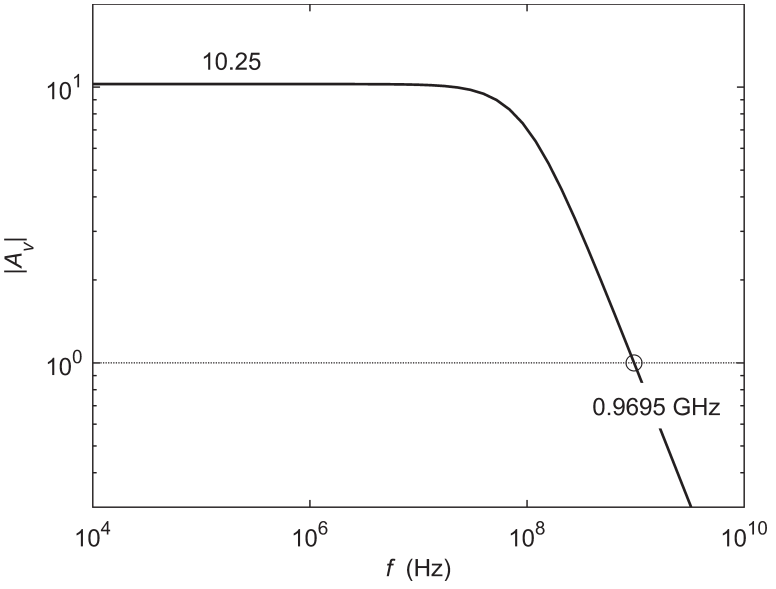

Figure 3.6 Magnitude response obtained from a SPICE AC analysis.

is operated within a differential pair, where the bias current is drawn from the source terminal, obviating the need for an explicit computation of V_{GS} (see Section 3.3).

We simulate this circuit and first inspect the DC operating point output:

V_{GS} = 468.119 mV,
V_{DS} = 600.468 mV,
I_D = 419.004 uA,
g_m = 6.28284 mS,
g_{ds} = 612.939 μS.

We thus have g_m/I_D = 6.28 mS/419 μA = 14.99 S/A, which is very close to the desired value. Also, the quiescent point gate voltage and the simulated g_m/g_{ds} ratio agree with the predicted values.

Next, we run a small-signal AC analysis and obtain the plot shown in Figure 3.6. Again, the result is very close to our expectation. This is not surprising, because we already saw from the operating point data that g_m is set almost exactly as desired. The gain-bandwidth error of approximately 3% can be explained by the fact that we have neglected the extrinsic capacitances (C_{gd} and C_{db}) when using (3.4); we will address this issue in Section 3.1.7.

The above example gave us a first feel for g_m/I_D-based design, taking advantage of pre-computed lookup tables. We arrived at the desired result without any iterations and "tweaking" in SPICE.

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

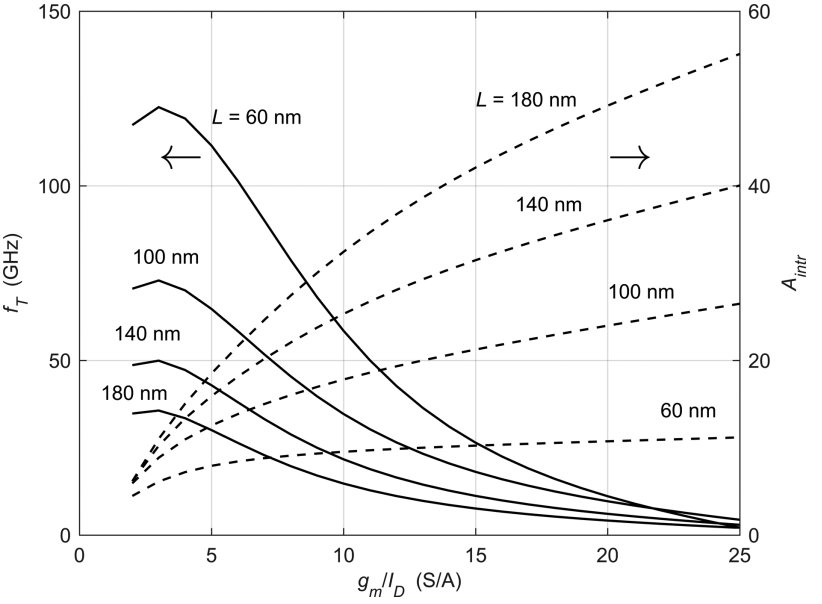

Figure 3.7 Transit frequency f_T and intrinsic gain A_{intr} versus g_m/I_D, considering four equally spaced gate lengths from 60 to 180 nm (V_{DS} = 0.6 V, V_{SB} = 0 V).[2]

3.1.4 Basic Tradeoff Exploration

In the previous section, we assumed that L and g_m/I_D were known, and this led to a straightforward sizing procedure. We will now begin to explore tradeoffs that will constrain and ultimately define the choice of these parameters in a practical design. We will focus in this section on first-order metrics (gain and bandwidth), and leave the inclusion of more advanced specifications (noise, linearity and mismatch) to Chapter 4.

We already know from Chapters 1 and 2 that g_m/I_D and L affect parameters that tradeoff with one another: the transit frequency and the intrinsic gain. This point is illustrated in Figure 3.7, which shows f_T and $A_{intr} = g_m/g_{ds}$ plotted against g_m/I_D. The data for this plot was obtained using:

```
Avo = lookup(nch,'GM_GDS','GM_ID',gmID,'L',L);
fT  = lookup(nch,'GM_CGG','GM_ID',gmID,'L',L)/(2*pi);
```

where gmID and L are vectors defining the shown sweep range.

The key observations from this plot are summarized as follows:

- The transit frequency is largest in strong inversion (small g_m/I_D), and gradually decays as we approach weak inversion (large g_m/I_D). Unfortunately, this means that we can either make the transistor fast or efficient, but not both.

[2] Note that the values of V_{DS} and V_{SB} won't affect the general tradeoffs shown in this figure (as long as the device remains saturated). See also Section 2.3.1.

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

- The intrinsic gain is large for long channels, but long channels have an adverse effect on the transit frequency. We can either achieve large gain (using large L) or high transit frequency (using short L), but not both.

Our job as circuit designers is to manage these tradeoffs in accordance with the overall design goal, which may vary widely (see the examples in Chapters 5 and 6). To untangle this problem further without losing generality, we will now consider three examples that each include one constraint.

The first example assumes that g_m/I_D is fixed and that we are free to choose L. This allows us to see the connection between channel length, voltage gain and other design parameters more clearly. In practice, the case of constant g_m/I_D may reflect a scenario where the circuit is limited by distortion. We will see in Chapter 4 that linearity requirements place upper bounds on g_m/I_D. Another scenario is a low-voltage, high-dynamic range circuit where stringent bounds on signal swing and V_{Dsat} may exist. For example, requiring V_{Dsat} = 150 mV, means $g_m/I_D = 2/V_{Dsat}$ = 13.33 S/A. We will see an example of this in Chapter 6.

Example 3.2 Sizing at Constant g_m/I_D

Consider an IGS with C_L = 1 pF and g_m/I_D of 15 S/A. Find combinations of L, W and I_D, that achieve f_u = 100 MHz and compute the corresponding low-frequency gain and fan-out. Assume V_{DS} = 0.6 V and V_{SB} = 0 V.

SOLUTION

Graphically, the problem boils down to tracing the thick vertical line in the plot of Figure 3.8(a) and to collect the intersecting transit frequencies and intrinsic gains for every gate length. The result is plotted in Figure 3.8(b), showing opposing trends in $|A_{v0}|$ and f_T as L is increased.

To compute the drain current that meets the desired f_u, we follow the same approach used in Section 3.1.3:

```
gm = 2*pi*fu*CL;
ID = gm/15;
```

This yields I_D = 41.89 μA. Note that since g_m/I_D is constant and g_m is fixed by the specifications, the drain current is also constant, regardless of the chosen L.

To find the device widths, we need the drain current densities, which depend on L. In the calculation below, we consider the entire L vector stored in the lookup tables (nch.L = [(0.06:0.01:0.2) (0.25:0.05:1)]):

```
JD = lookup(nch, 'ID_W', 'GM_ID', 15, 'L', nch.L);
W  = ID./JD;
```

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

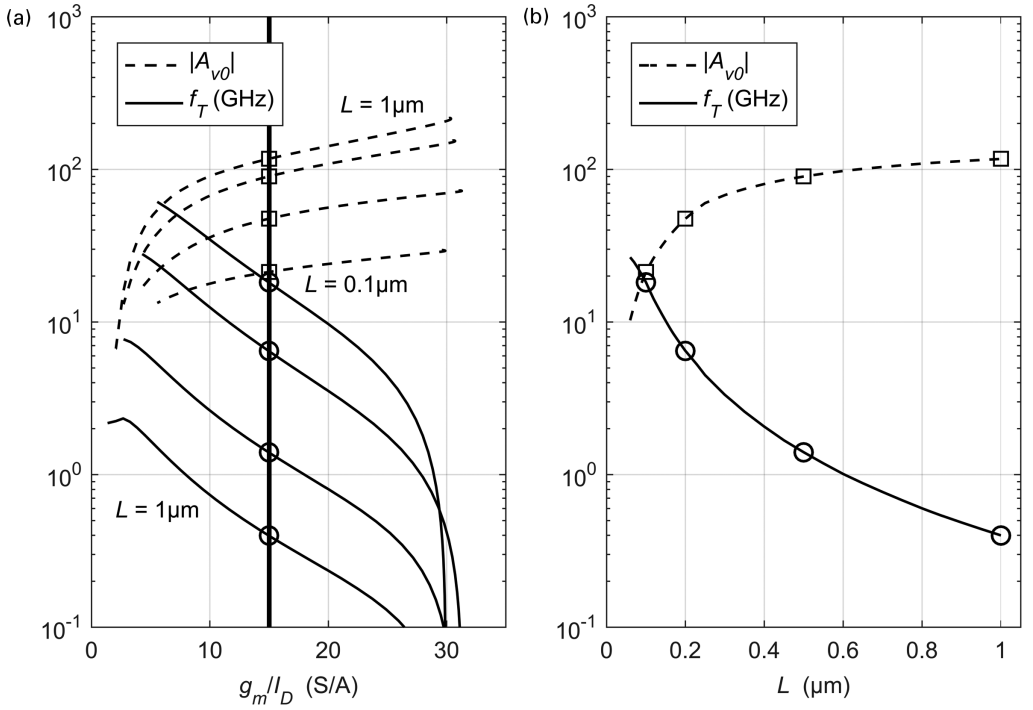

Figure 3.8 (a) Plot of LF gain and transit frequency versus g_m/I_D. The bold line marks the given value of g_m/I_D = 15 S/A. (b) Plot of $|A_{v0}|$ and f_T that intersect with the bold line in (a).

This completes the sizing and we can now compute the fan-out for each choice of L. We can do this using:

```
fT = lookup(nch, 'GM_CGG', 'GM_ID', 15,'L', nch.L)/(2*pi);
FO = fT/fu;
```

and find that f_T ranges from 0.4 to 26.5 GHz. Since we do not consider transit frequencies less than 10 times f_u (100 MHz), we find an index vector M for the usable gate length range and the corresponding maximum value:

```
M  = fT >= 10*fu;
Lmax = max(nch.L(M));
```

This yields L_{max} = 0.60 µm. Note that this value can also be read from Figure 3.8(b). The f_T curve crosses 1 GHz near L = 0.6 µm.

Figure 3.9 plots the above-computed data for the voltage gain, device width and fan-out for $L \le L_{max}$. We see that the largest achievable voltage gain is about 100 for $L = L_{max}$. Note that this number would reduce if we increased the f_u requirement, since it would lead to a higher f_T requirement and a correspondingly shorter channel. Conversely, larger L and hence larger voltage gains would be possible if we relaxed the f_u requirement. However, we can see

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

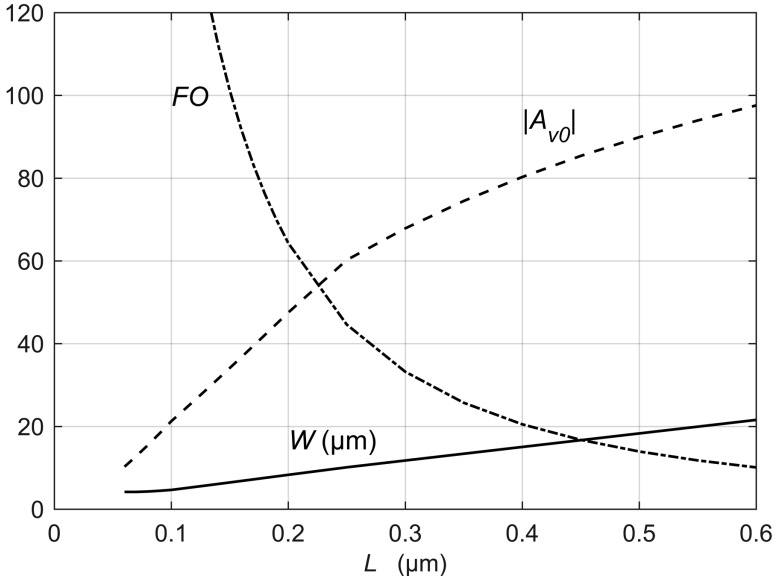

Figure 3.9 LF voltage gain, device width and fan-out versus gate length. Parameters: f_u = 100 MHz, C_L = 1 pF, g_m/I_D = 15 S/A, V_{DS} = 0.6 V, V_{SB} = 0 V.

from Figure 3.8(b) that the gain curve saturates for long channels, and not much improvement is possible.

The next scenario that we will consider is sizing at constant transit frequency. This case is of practical relevance, for instance, in amplifiers that target a fixed gain-bandwidth product. Another example is a cascode configuration, where we often want to size the common-gate device such that the non-dominant pole lies at or above some given frequency. Such cases will be studied in more detail in Chapter 6.

Example 3.3 Sizing at Constant f_T

Consider an IGS with C_L = 1 pF and f_u of 1 GHz. Find combinations of L and g_m/I_D, that achieve (i) maximum low-frequency gain and (ii) minimum current consumption. Assume FO = 10, V_{DS} = 0.6 V and V_{SB} = 0 V. Validate the results using SPICE simulations.

SOLUTION

As before, we start with a plot of $|A_{v0}|$ and f_T versus g_m/I_D, shown in Figure 3.10(a). This time, however, we look for the intersects of f_T and the bold gray line, marking the target of 10 GHz. This yields corresponding g_m/I_D values that we can use to find A_{v0} across a range of gate lengths. The result of this collection is shown in Figure 3.10(b). We see that there is a value pair of L and g_m/I_D that maximizes the gain.

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

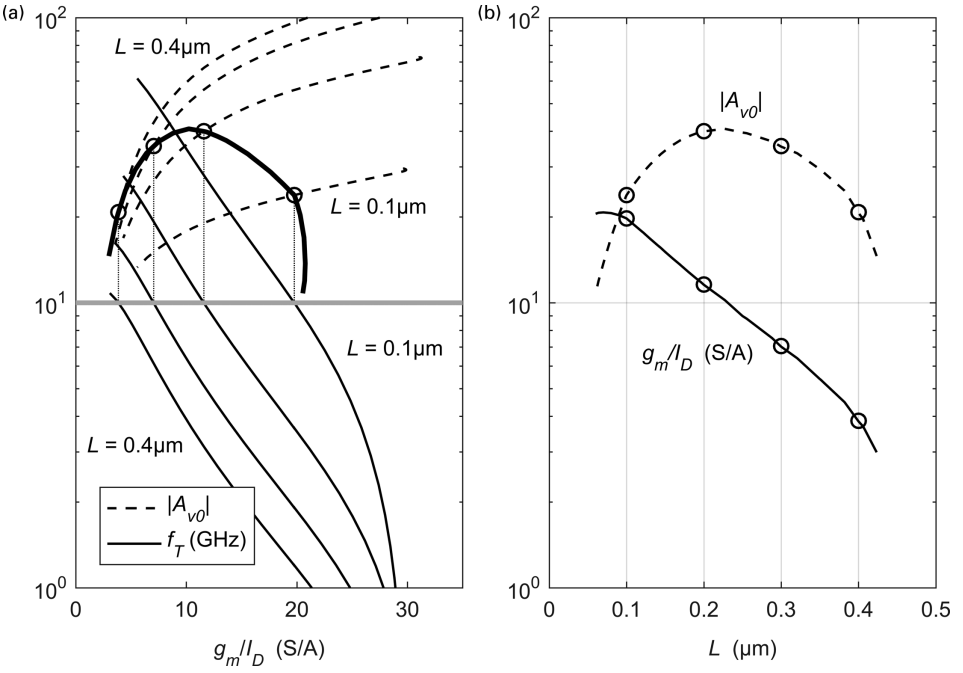

Figure 3.10 (a) Plot of transit frequency and intrinsic gain versus g_m/I_D. The bold black line marks the $|A_{v0}|$ values for which $f_T = 10$ GHz (bold gray line). (b) Plot of the corresponding $|A_{v0}|$ and g_m/I_D versus gate length L.

To investigate further, Figure 3.11 shows an expanded view of (a) the LF gain and (b) the gate length versus g_m/I_D. Plot (b) shows that L must decrease to keep f_T unchanged as we increase g_m/I_D. This explains the decline of $|A_{v0}|$ for large g_m/I_D beyond the maximum (marked by a circle). Before the maximum is reached, the dependence of A_{v0} on g_m/I_D dominates and leads to a positive slope. This is explained by the strong positive slope in $|A_{v0}|$ seen in Figure 3.10(a) for $g_m/I_D < 10$ S/A.

Interestingly, the plots of Figure 3.11 also show a trend reversal near their tail end and as we reach the minimum L offered by the technology. This is due to the reduced slope in the I-V characteristic for large g_m/I_D and near-minimum L (see Section 2.3.1). Physically, this effect is related to reduced gate control, or equivalently, increased drain-induced barrier lowering (DIBL) for short channels. The largest g_m/I_D value on these curves is marked with an asterisk and represents the minimum current design.

The design data for maximum gain (option (i)) and minimum current (option (ii)) are summarized in Table 3.1. It is important to note that that these design points were obtained only through the manipulation of parameter ratios in a normalized space. Both f_u and C_L have not yet entered the design process and become only important once we want to compute the device widths, which is the next step in this problem.

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

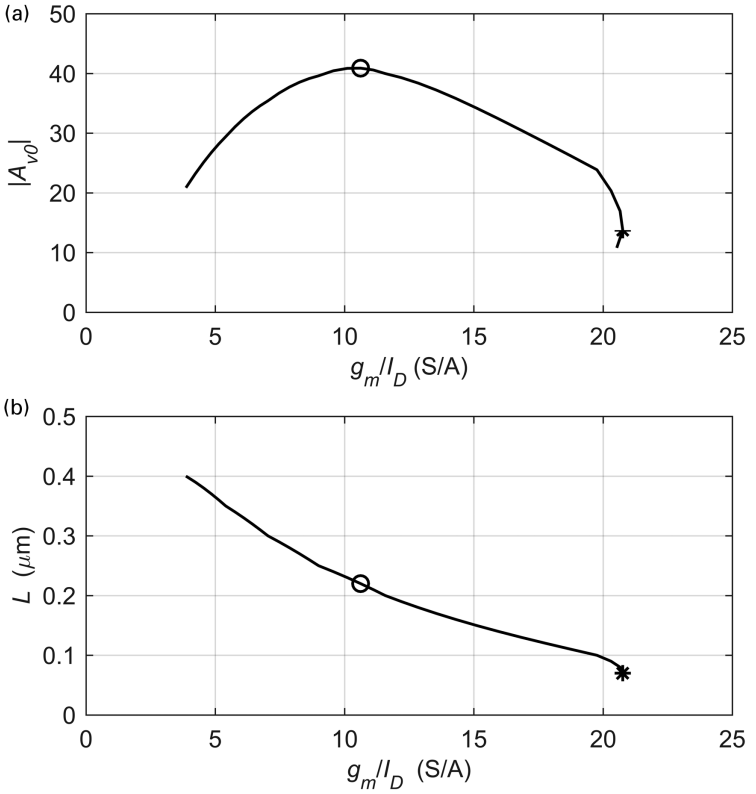

Figure 3.11 (a) LF gain and (b) gate length versus g_m/I_D. The circles mark the maximum gain points (option (i)) and the asterisks mark the design with minimum current (option (ii)). Parameters: f_T = 10 GHz, V_{DS} = 0.6 V, V_{SB} = 0 V.

Table 3.1 Design parameters that maximize the gain (option (i)) or minimize the drain current (option (ii)).

	$\|A_{v0}\|$	g_m/I_D (S/A)	L (nm)	V_{GS} (V)
Option (i)	40.88	10.62	220	0.5786
Option (ii)	13.75	20.76	70	0.4103

Since we know both L and g_m/I_D at this stage, all further calculations are carried out as in Example 3.1. Given FO =10, we know that $f_u = f_T/FO$ = 1 GHz and arrive at the sizing parameters listed in Table 3.2.

To compare the two sizing options, Figure 3.12 plots W and $|A_{v0}|$ versus the drain current I_D (the maximum gain and minimum current designs are marked with a circle and asterisk, respectively). Note that for the minimum current design, increasing the drain current only slightly would lead to a significant increase in voltage gain. Operating the design at the absolute minimum current (and correspondingly short L) may therefore not be desirable.

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

Table 3.2 Sizing parameters that maximize the gain (option (i)) or minimize the drain current (option (ii)).

	g_m (mS)	W (μm)	I_D (μA)
Option (i)	6.283	45.92	591.6
Option (ii)	6.283	114.2	302.7

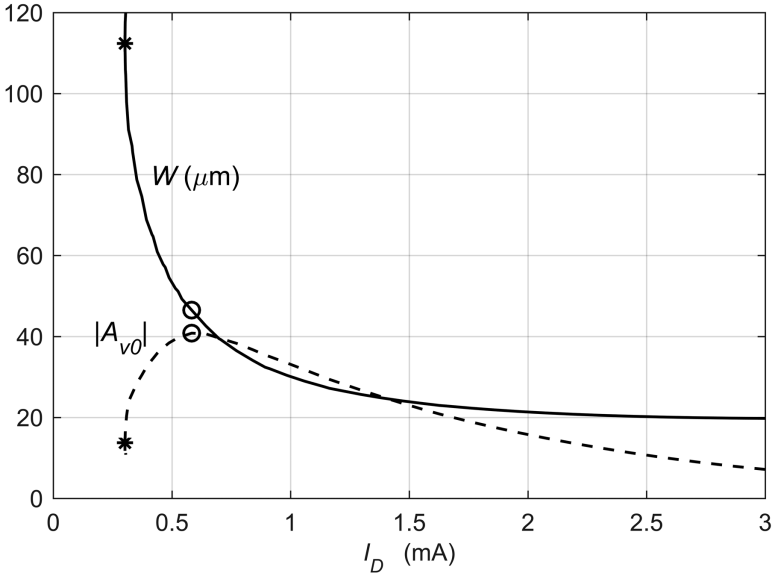

Figure 3.12 LF gain and device width versus drain current. The circles mark the maximum gain design option (i) and the asterisks mark the design with minimum current (ii).

Table 3.3 Simulation result summary.

	$\lvert A_{v0}\rvert$ SPICE	Error (%)	f_u (GHz) SPICE	Error (%)
Option (i)	41.0	+2.9	0.967	−3.3
Option (ii)	13.75	0	0.933	−6.7

To conclude, we validate the maximum gain and minimum current design options in SPICE and obtain the results listed in Table 3.3. We observe that the low-frequency voltage gains are almost exactly as predicted, whereas the unity gain frequencies are somewhat smaller than the design target. As already mentioned in Example 3.1, the small discrepancy is due to the parasitic drain capacitance that we have not yet considered in our design flow.

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

As a final example, we consider a scenario where we fix the circuit's low-frequency voltage gain. Such a situation may arise in the design of operational amplifiers, where we typically want to achieve a certain target for the feedback circuit's loop gain. We will encounter this situation in some of the examples in Chapter 6.

Example 3.4 Sizing at Constant $|A_{v0}|$

Consider an IGS with C_L = 1 pF and constant $|A_{v0}|$ = 50. Find combinations of L and g_m/I_D, that achieve (i) maximum unity gain frequency, and (ii) minimum current consumption for a design that achieves 80% of the maximum unity gain frequency. Assume FO = 10, V_{DS} = 0.6 V and V_{SB} = 0 V. Validate the results using SPICE simulations.

SOLUTION

The flow of our solution is very close to Example 3.3. We start with a plot of $|A_{v0}|$ and f_T versus g_m/I_D, shown in Figure 3.13(a), but this time we look for the

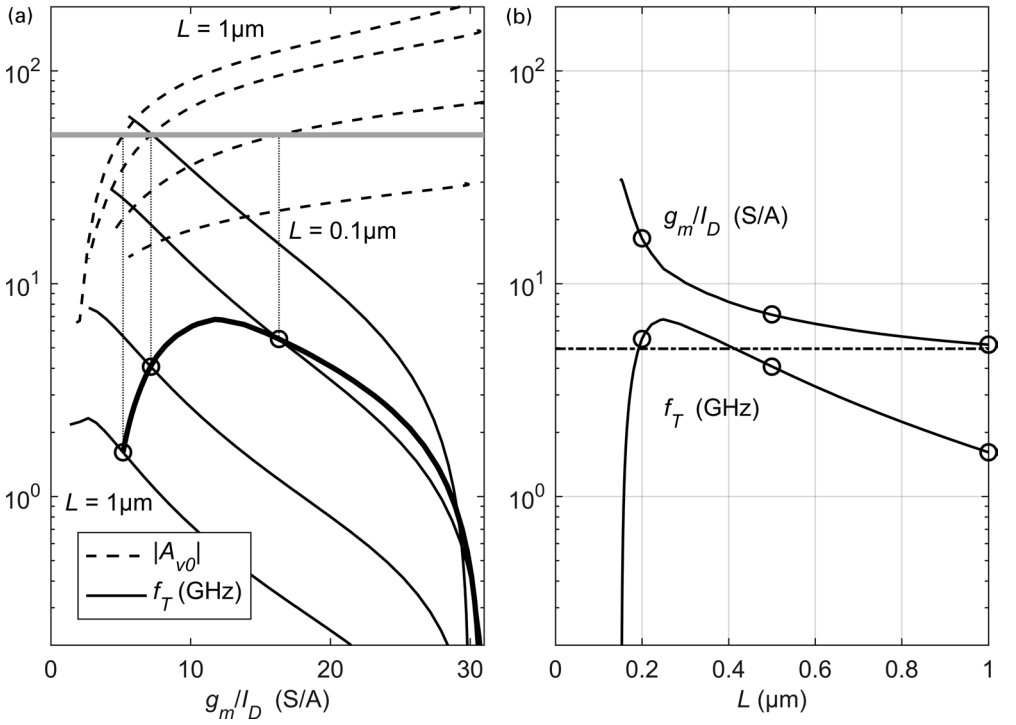

Figure 3.13 (a) Plot of transit frequency and LF gain versus g_m/I_D. The bold black line marks the f_T values for which $|A_{v0}|$ = 50 (bold gray line). (b) Plot of the corresponding f_T and g_m/I_D versus gate length L.

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

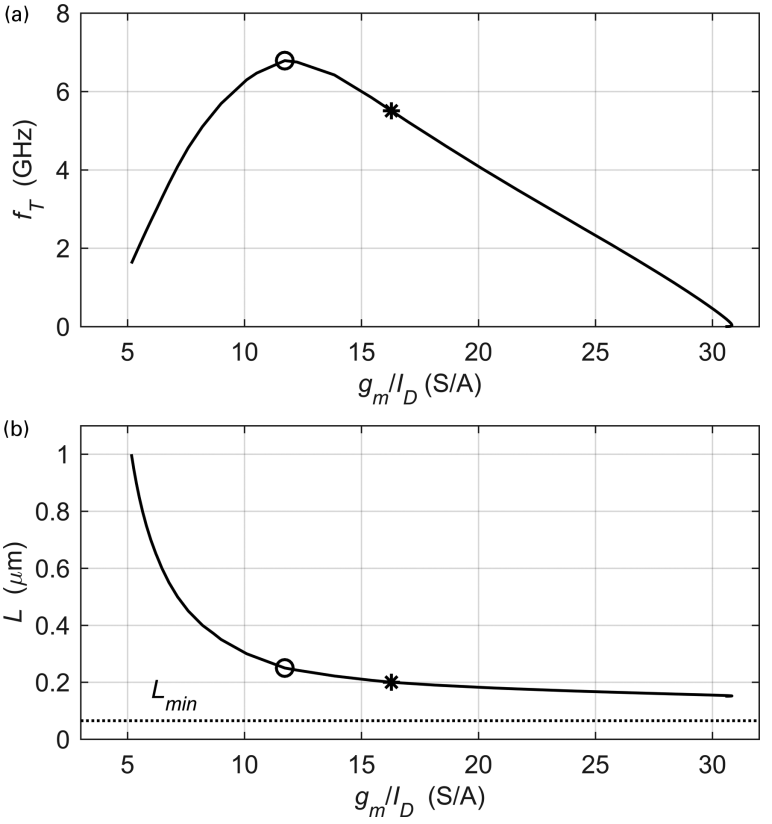

Figure 3.14 (a) Transit frequency and (b) gate length versus g_m/I_D. The circles mark the maximum f_T design (option (i)) and the asterisks the minimum current design at 20% loss in f_u (option ii). Parameters: $|A_{v0}| = 50$, $V_{DS} = 0.6$ V, $V_{SB} = 0$ V.

intersect between the LF gain curves and the target value of 50 (bold gray line). This yields corresponding g_m/I_D values that we can use to find f_T across a range of gate lengths (see Figure 3.13(b). We see that there is a value pair of L and g_m/I_D that maximizes f_T. Note that this point maximizes the unity gain frequency as well, since $f_u = f_T/FO = f_T/10$.

To investigate further, Figure 3.14 shows an expanded view of (a) the transit frequency and (b) the gate length versus g_m/I_D. Plot (b) shows that L must increase significantly to keep $|A_{v0}|$ constant as we reduce g_m/I_D toward strong inversion. This explains the sharp decline of f_T to the left of the maximum that we see in Figure 3.13. To the right of the maximum, f_T decreases due to the reduced inversion level as we increase g_m/I_D. At the tail end of the curve (onset of weak inversion), we find that the minimum channel length for which an LF gain of 50 can be achieved is about 150 nm. This lower bound explains the sharp drop of the f_T curve in Figure 3.13(b) as the designs to the left of the curve are unfeasible.

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

Table 3.4 Design and sizing parameters that maximize the unity gain frequency (option (i)) or minimize the drain current for a 20% f_u reduction (option (ii)).

	g_m/I_D (S/A)	L (nm)	f_u (MHz)	I_D (μA)	W (μm)	V_{GS} (V)
Option (i)	11.7	250	679	363.8	41.95	0.5533
Option (ii)	16.3	200	543	209.5	54.24	0.4818

Table 3.5 Simulation result summary.

	$\lvert A_{v0}\rvert$ SPICE	Error (%)	f_u (MHz) SPICE	Error (%)
Option (i)	49.9	−0.2	661	−2.7
Option (ii)	50.0	0	525	−3.3

At the maximum transit frequency point, we find $f_u = f_T/10 = 679$ MHz and $L = 0.25$ μm; this is the solution for option (i). Although there are two options for designs with 20% reduction in f_u, the point to the right of the maximum (marked with an asterisk) achieves minimum current. Both options require the same g_m, but the point with larger g_m/I_D will require a smaller drain current. The channel length for option (ii) is thus 0.2 μm. Since we now know L, g_m/I_D and f_T for both designs, the sizing parameters are readily computed (see Table 3.4).

Finally, we validate the maximum f_u and minimum I_D designs in SPICE and obtain the results listed in Table 3.5. Once again, we observe good agreement with the Matlab prediction.

In all the prior examples, we have assumed a unity gain frequency constraint in the range of 100 MHz to 1 GHz. However, there exist applications that operate at much lower frequencies. For example, biomedical and sensor interface circuits often operate in the kilohertz range. In such circuits, the transit frequency typically won't appear as a significant constraint that will affect the sizing.

Consider for example the tradeoff between f_T and g_m/I_D for a fixed value of LF gain in Example 3.4. From Figure 3.14(a), we know that we can move toward larger values of g_m/I_D when the f_T requirements are reduced, and this can help us save current. If the f_T constraint is removed entirely, the preferred design point would be at the tail end of the curve, namely at the maximum possible g_m/I_D. In this case, the transistor will operate in weak inversion and g_m/I_D is essentially constant. Finding and optimizing the device width in this specific scenario requires a different approach, which we will investigate in the following section.

3.1.5 Sizing in Weak Inversion

As an introduction to sizing in weak inversion, consider the amplifier circuit used in the ultra-low power sensor of [2] as an example. The node is designed to dissipate

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

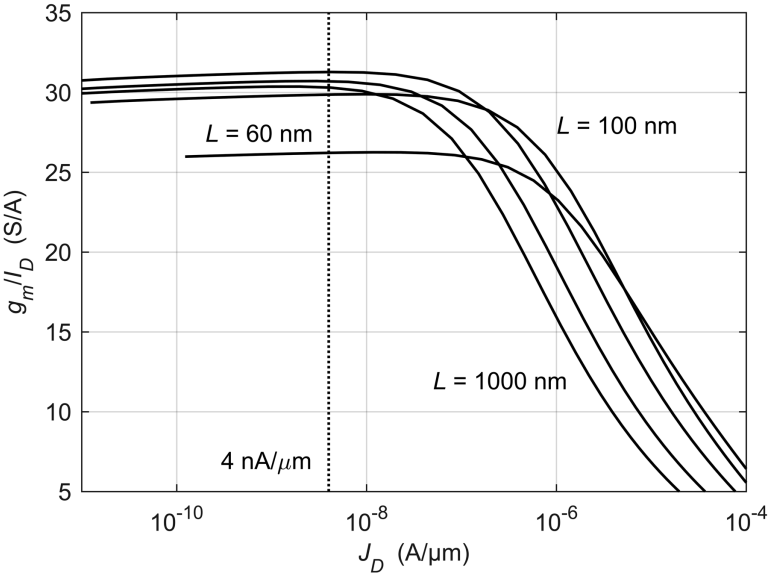

Figure 3.15 Transconductance efficiency (g_m/I_D) versus drain current density (J_D) of an n-channel device with gate lengths L = 60, 100, 200, 500 and 1000 nm. (V_{DS} = 0.6 V, V_{SB} = 0V).

only 3 nW, so that it can be powered from a thin-film battery for about 10 years. The differential front-end amplifier of the sensor node consumes only 1 nW from a 0.6 V supply and thus the differential pair transistors have a bias current of only about 0.8 nA each. With such low currents, the current densities will be correspondingly low. Even if we assume that the MOSFET has a width of 200 nm, which is the minimum value allowed in a typical 65-nm process, the current density will not exceed 0.8 nA/0.20 μm, which is 4.0 nA/μm. Considering the plot of g_m/I_D versus current density in Figure 3.15, we can therefore conclude that the transistor must operate in weak inversion.

In weak inversion, g_m is equal to $I_D/(nU_T)$. In this expression, the subthreshold slope parameter n is only a weak function of the channel length, provided that L is somewhat larger than minimum length (avoiding significant DIBL, as explained in Section 2.3.1). This is evident from Figure 3.15, which shows that $(g_m/I_D)_{\max}$ is nearly constant for L = 100…1000 nm. Therefore, the required drain current in weak inversion is to first order gate-length independent.

If no other constraint applies, it would seem natural to design for minimum area, i.e. minimum device width and length. However, this is usually not a good choice in analog circuits. Minimum length implies low intrinsic gain, and small gate area ($W{\cdot}L$) leads to poor device matching and flicker noise (see Chapter 4). This means that for a real-world problem, we typically need to invoke some of these additional constraints to complete the device sizing.

To investigate further, let us evaluate the intrinsic gain as a function of channel length, shown in Figure 3.16(a). We use the current density instead of g_m/I_D on the

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

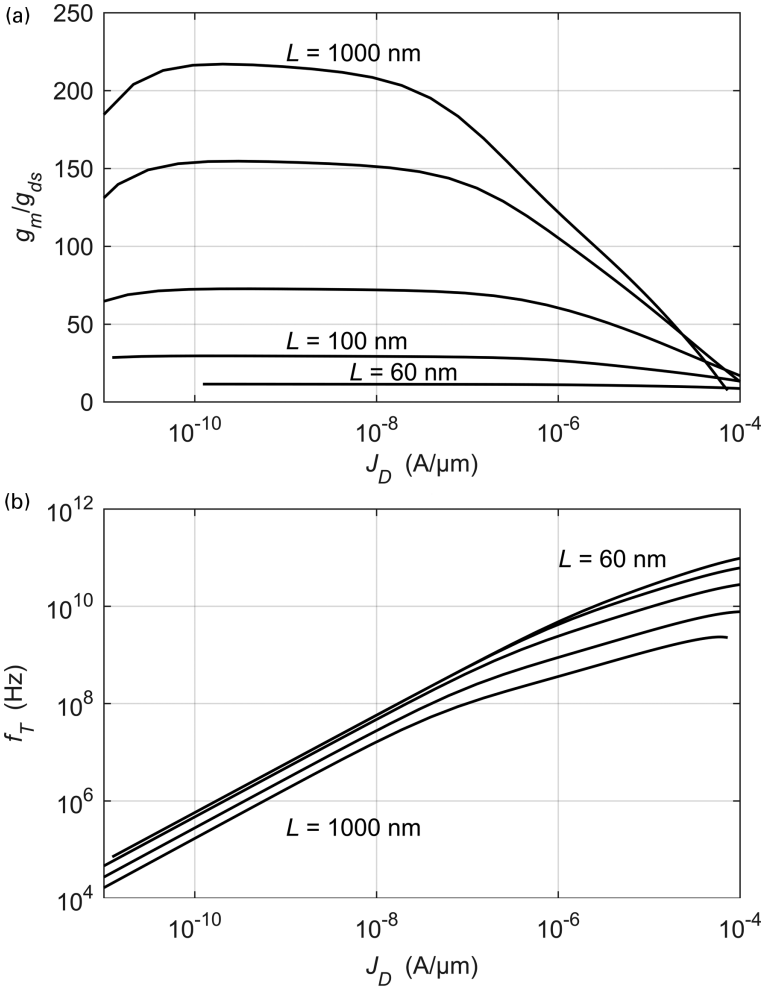

Figure 3.16 Intrinsic gain (a) and transit frequency (b) versus drain current density (J_D) of an n-channel device with gate lengths L = 60, 100, 200, 500 and 1000 nm. (V_{DS} = 0.6 V, V_{SB} = 0 V).

x-axis since the latter is essentially constant in the design region of interest. Also, we show the device f_T in Figure 3.16(b), to get a feel for the resulting numbers. As expected, the intrinsic gain increases significantly with longer channel lengths. The device f_T drops as we increase the channel length, but even at a current density of 4.0 nA/μm it is still above 1 MHz, and thus significantly larger than f_u in the kilohertz range. With this specific example at hand, we could make the channel length even larger than 1000 nm and continue to benefit from increasing intrinsic gain. At which point should we stop?

From a theoretical standpoint, it would make sense to continue to increase L until the given current density (4.0 nA/μm in our example) corresponds to the weak inversion "knee" (see trend in Figure 3.15). Moving any further would mean that

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

g_m/I_D and thus g_m drops. For the extreme example considered here (0.8 nA of drain current), this design point is reached at extremely large L, for which the transistor model may no longer be accurate. The designer will therefore stop at a point where a reasonable amount of gain is achieved and the f_T is still significantly higher than the operating frequencies. With a gate length of 500 nm, we already have an intrinsic gain of almost 150 per Figure 3.16(a).

As far as the device width is concerned, there is no incentive to increase W beyond the minimum value of 200 nm, unless we include further constraints on matching or flicker noise (see Chapter 4). We consider an example below that takes a minimum-width constraint into account.

Example 3.5 Sizing in Weak Inversion Given a Width Constraint

Size an ultra-low power IGS with C_L = 1 pF and I_D = 0.8 nA. Determine L to achieve a low-frequency voltage gain of about 150, and assume that the minimum desired device width is 5 μm. Compute V_{GS}, f_T and the circuit's unity gain frequency. Verify the design using a SPICE simulation. Assume V_{DS} = 0.6 V and V_{SB} = 0 V.

SOLUTION

We begin by computing the current density:

```
JD = ID/W
```

This yields 0.16 nA/μm. To find the gate length, we look at Figure 3.16(a), which indicates that L = 500 nm will suffice to achieve a gain of 150 at the computed current density. We can now use the lookup function to find g_m/I_D and f_T:

```
gm_ID = lookup(nch, 'GM_ID', 'ID_W', JD,'L', 0.5)
fT = lookup(nch, 'GM_CGG', 'ID_W', JD, 'L',0.5)/2/pi
```

This leads to g_m/I_D = 30.5 S/A and f_T = 445 kHz. The transconductance and f_u now follow mechanically:

```
gm = gm_ID*ID;
fu = gm/(2*pi*CL)
```

Finally, to compute the gate-to-source voltage we use:[3]

```
VGS = lookupVGS(nch, 'ID_W', JD, 'L', 0.5, 'METHOD', 'linear')
```

Table 3.6 summarizes the results. The above calculations and SPICE simulation data are in good agreement. Note that the circuit's fan-out is very large (FO = 445 kHz/3.82 kHz = 116).

[3] The default interpolation method of lookupVGS is "pchip." In the shown computation, the interpolation method is changed to "linear" to avoid numerical issues that can arise near the endpoints of the stored lookup data.

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

Table 3.6 Result summary.

	$\lvert A_{v0}\rvert$	g_m/I_D (S/A)	f_T (kHz)	g_m (nS)	f_u (kHz)	V_{GS} (mV)
Calculation	~150	30.5	445	24.4	3.82	128
SPICE	145	30.5	442	24.4	3.87	128

From the above example, we see that sizing in weak inversion is relatively straightforward, since g_m/I_D is essentially constant and does not play a role in the optimization. The gate length follows directly from the gain requirement. Furthermore, we do not need to worry about fan-out from a modeling perspective, since the ratios of f_T and f_u will almost always be larger than the desired bound of 10.

3.1.6 Sizing Using the Drain Current Density

Throughout Section 3.1.4 we used the sizing flow of Figure 3.4, which takes g_m/I_D as the main "knob" to define the transistor's inversion level. The advantage of this approach is that g_m/I_D spans a well-defined range that is nearly independent of technology, thus serving as a very intuitive numerical proxy for the inversion level and the associated tradeoff between gain, speed and efficiency. In addition, once g_m/I_D is known, one can immediately compute the current that is required for a certain g_m.

However, as we saw in Section 3.1.5, a problem with the sizing flow of Figure 3.4 is that g_m/I_D no longer uniquely defines the drain current density once the transistor enters weak inversion. In other words, a wide range of current densities (and thus designs) map to nearly the same value of g_m/I_D. In Section 3.1.5, our solution to this problem was to first choose the current density, and then work with the resulting g_m/I_D to complete the design (see Example 3.5). In this section, we explore a generalization of this approach as shown in Figure 3.17. Since there is a one-to-one mapping between J_D and g_m/I_D, we can always begin by selecting J_D, then look up g_m/I_D and complete the design as usual. As already mentioned in Section 1.2.5, this approach is similar to the EKV-based sizing method of [2] which works with a normalized representation of the current density.

The flow of Figure 3.17 is applicable or advantageous in the following scenarios:

- We know a priori that the circuit will operate in weak inversion. In this case, we have no choice but to design based on current density (see Example 3.5).
- We have no a-priori information or intuition about the device's inversion level and hence want to search across all possibilities, from weak inversion to strong inversion.

Since the latter scenario occurs infrequently in real-world design, we advocate the g_m/I_D-driven flow of Figure 3.4 for the practitioner working on moderate to high-speed circuits, and only resort to current density-based design when the

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

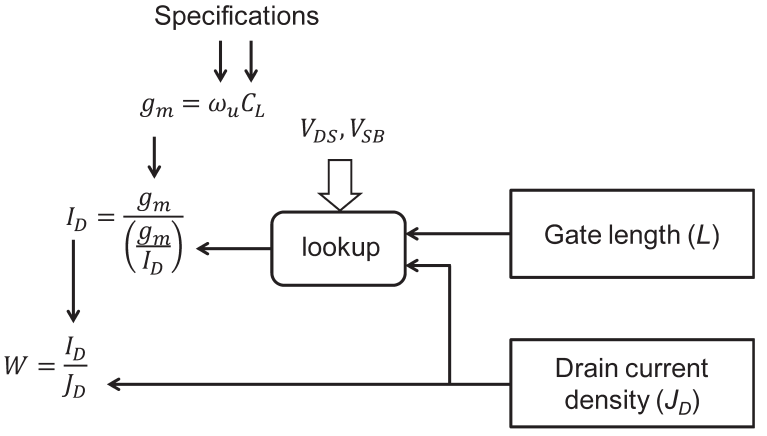

Figure 3.17 Current density-based sizing, appropriate for all inversion levels.

circuit is best operated in weak inversion. It is usually straightforward to make this decision.

Despite this stated preference, we dedicate the remainder of this section to a tradeoff exploration using the current density as the design variable. This lets us appreciate the sizing tradeoffs seen in the previous sections from another angle, and allows the reader to see the overall design space in a unified representation.

We begin by examining Figure 3.18. Plots (a), (b) and (c) show contours of constant intrinsic gain (A_{intr}), constant transconductance efficiency (g_m/I_D), and constant transit frequency (f_T), respectively. All quantities are plotted against L and J_D, which are the "knobs" in the design flow of Figure 3.17. The contours in (a), which are based on the same data as in Figure 3.15, exhibit a large "empty" area delineated by the 30 S/A locus (weak inversion). In this region, the choice of J_D and L does not impact g_m/I_D significantly. To enter moderate inversion, J_D must be made larger than 10 to 100 nA/μm. Figure 3.18(b) shows the corresponding evolution of the intrinsic gain. In weak inversion, L sets A_{intr}, corroborating the data of Figure 3.16(a), while in moderate and especially in strong inversion, the drain current density becomes the defining variable. Lastly, (c) conveys the same trend as seen in Figure 3.16(b). The constant transit frequency loci are nearly parallel and almost vertical (independent of J_D) in weak inversion, but more dependent on the current density in strong inversion.

It is now interesting to locate and contrast our two previous design examples in the plots of Figure 3.18. The strong-inversion design of Example 3.3 is marked by a circle in each subplot, while the weak-inversion design of Example 3.5 is marked using a plus sign.

We start with the weak inversion design. Inside the region delineated by the 30 S/A contour in Figure 3.18(a), g_m/I_D does not change significantly. Since g_m is fixed by the specifications, the drain current remains almost constant, too. However, we

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

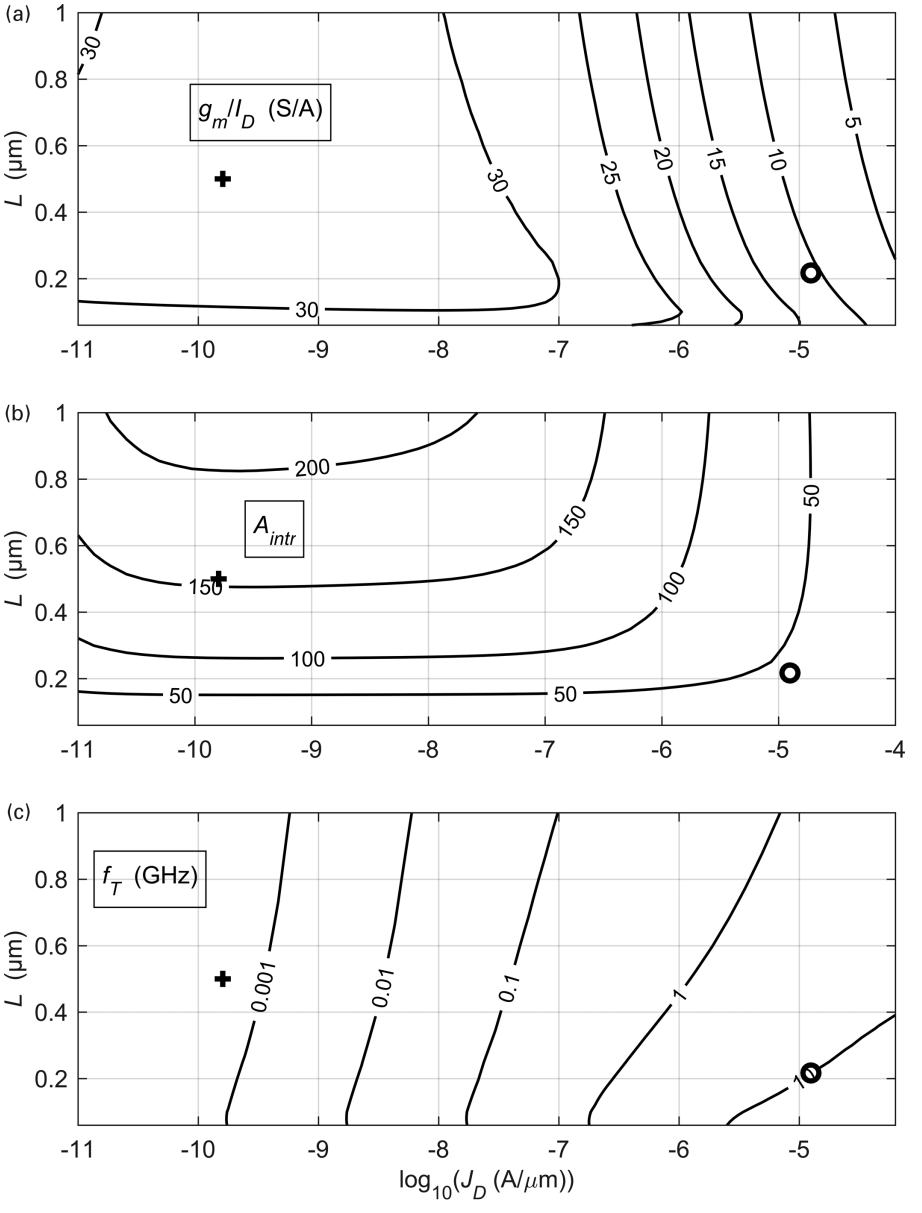

Figure 3.18 Constant contour plots representing (a) the transconductance efficiency, (b) intrinsic gain and (c) transit frequency versus the drain current density J_D and the gate length L (V_{DS} = 0.6 V, V_{SB} = 0 V). The data from Example 3.3 and Example 3.5 are marked by "o" and "+", respectively.

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

can still vary the gain moving up or down (adjusting L) keeping the current density constant. Thus, the gate length essentially sets the intrinsic gain, and the current density sets the device width.

The circumstances are very different for the strong inversion design of Example 3.5. Here, J_D and L pairs meeting the desired transit frequency (10 GHz) lie on the shown f_T contour with the label "10" in Figure 3.18(c). Picture now this contour of valid J_D and L pairs in the gain plot of Figure 3.18(b). Some of the gain contours in Figure 3.18(b) cross the target f_T contour twice. As we increase gain, the intersections come closer together. The largest gain that we can achieve is defined by the gain contour that shares only one point with the f_T contour. In other words, as we move away from the tangential point, the gain associated with the intersect pairs can only decrease. We saw this already in Figure 3.11 where for a given g_m/I_D the gain is largest.

The Matlab code given below implements a search for the maximum voltage gain, assuming V_{DS} = 0.6 V and V_{SB} = 0 V. We first set up the gate length and drain current density vectors that define the axes of Figure 3.18:

```
JDx  = logspace(-10,-4,100);
Ly   = .06:.01:1;
[X Y] = meshgrid(JDx,Ly);
```

We then make use of the Matlab contour function[4] to get drain current densities (J_{D1}) and gate lengths (L_1) that achieve a desired f_T value:

```
fTx  = lookup(nch,'GM_CGG','ID_W',JDx,'L',Ly)/(2*pi);
[a1 b1] = contour(X,Y,fTx,fT*[1 1]);
JD1 = a1(1,2:end)';
L1  = a1(2,2:end)';
```

We now look for the J_D and L pair that maximizes the intrinsic gain and evaluate the corresponding g_m/I_D and V_{GS}.

```
Av = diag(lookup(nch,'GM_GDS', 'ID_W', JD1, 'L', L1));
[a2 b2] = max(Av);
Avo = a2;
L  = L1(b2);
JD = JD1(b2);
gm_ID = lookup(nch, 'GM_ID', 'ID_W', JD, 'L', L);
VGS = lookupVGS(nch, 'GM_ID', gm_ID, 'L', L);
```

Table 3.7 summarizes the results obtained across a wide range of f_T. Note that for the lowest f_T values, we converge to the longest channel length available in our lookup tables (1 μm).

In the following example, we consider another illustration of the current density-based sizing approach of Figure 3.17, and reinforce some of the above-discussed observations.

[4] The Matlab function contour(x, y, F(x, y), C*[1 1]) finds x1 and y1 vectors that make F(x1, y1) equal to C.

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

Table 3.7 Summary of maximum gain parameters.

f_T (GHz)	g_m/I_D (S/A)	J_D (μA/μm)	max($\lvert A_{v0}\rvert$)	L (μm)	V_{GS} (V)
0.02	29.93	0.0125	206.4	1.00	0.2861
0.05	28.78	0.0369	196.0	1.00	0.3228
0.10	26.54	0.0986	179.0	0.99	0.3572
0.20	24.85	0.1874	155.1	0.79	0.3773
0.50	22.21	0.4746	124.0	0.57	0.4058
1.00	**20.00**	**0.9371**	**102.0**	**0.44**	**0.4300**
2.00	16.42	2.105	81.6	0.37	0.4718
5.00	13.19	5.397	57.6	0.27	0.5242
10.00	10.62	12.90	40.9	0.22	0.5787
20.00	8.75	28.48	27.4	0.17	0.6310
50.00	6.66	75.64	15.5	0.11	0.7050

Example 3.6 Sizing Using Contours in the J_D and L Plane

Consider an IGS with C_L = 1 pF and f_u = 100 MHz, which requires $f_T \geq 1$ GHz. Per Table 3.7, the largest voltage gain magnitude that we can achieve is 102. In this example, consider reduced gain values of $|A_{v0}| = 50$ and 80 and size the circuit for maximum fan-out and minimum current for each case (four design options total). Use a current density sweep to find the required g_m/I_D and L values. Assume V_{DS} = 0.6 V and V_{SB} = 0 V.

SOLUTION

For illustrative purposes, Figure 3.19 combines the contours of Figure 3.18 into one plot, focusing on the case of f_T = 1 GHz (bold line). The solid thin black line marks the contour with the maximum $|A_{v0}| = 102$. Note that these two curves have only one intersection point, as discussed previously. Also shown are gray contours for constant g_m/I_D. The contour for 20 S/A passes through the point with maximum gain, as expected from the result in Table 3.7.

Assume now a reduced value of $|A_{v0}| = 80$. The J_D and L pairs tracing this gain locus are illustrated by the dashed contour and are obtained using:

```
Av = lookup(nch,'GM_GDS','ID_W',JD,'L',L);
[a3 b3] = contour(X,Y,Av,80*[1 1]);
JD3 = a3(1,2:end)';
L3  = a3(2,2:end)';
```

The gain locus has two intersections with the bold f_T = 1 GHz locus, and the pertaining design values are given in Table 3.8. Note that the upper intersect (A) lies in

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

strong inversion, while the lower (B) is in weak inversion (within ~20% of maximum g_m/I_D).

To the right to the bold line in Figure 3.19, the transit frequency increases, leading to $FO > 10$. We select these points by running:

```
M  = FO >= FOmin;
FO  = diag(lookup(nch,'GM_CGG', 'ID_W', JD3, 'L', L3))/(2*pi*fu);
FO4 = FO(M);
JD4 = JD3(M);
L4  = L3(M);
```

As a last step, we find the corresponding transconductance efficiencies, drain currents and device widths:

```
gm_ID4 = diag(lookup(nch,'GM_ID','ID_W',JD4,'L',L4));
gm = 2*pi*fu*CL;
ID = gm./gm_ID4;
W  = ID./JD4;
```

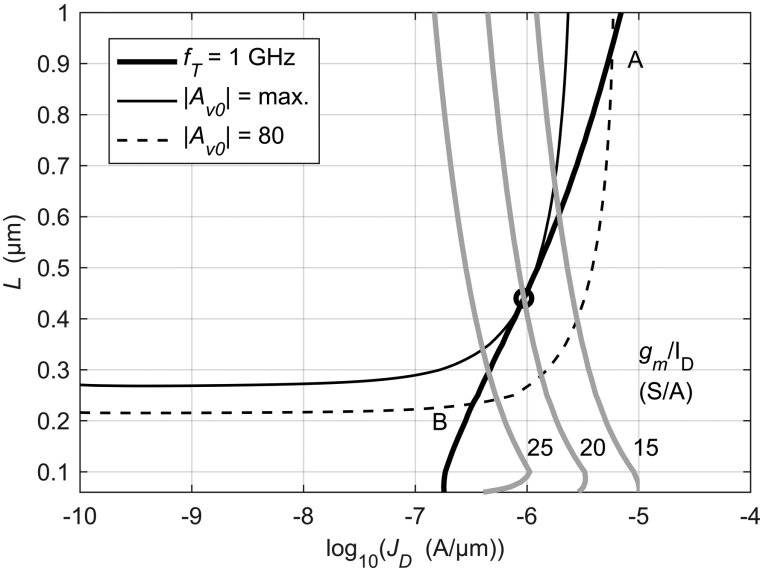

Figure 3.19 Contours for constant transit frequency (1 GHz), constant gain magnitude (102 and 80) and constant g_m/I_D (15, 20 and 25 S/A).

Table 3.8 Data for intersects A and B in Figure 3.19.

	g_m/I_D (S/A)	L (µm)	V_{GS} (V)
Point A	8.78	0.930	0.6120
Point B	27.02	0.233	0.3720

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

The results are plotted in Figure 3.20 considering not only $|A_{v0}| = 80$ but also 50. Notice that we meet the design requirements for a wide range of transconductance efficiencies, across which the currents change by about 3–5x. This stands in contrast with a circuit operating at maximum gain, which narrows the space to one single point. Note that the range of feasible g_m/I_D values widens for the design with lower gain.

In Table 3.9, we summarize the sizing data considering designs aiming at maximum fan-out and compare these to SPICE simulations. We see from the table

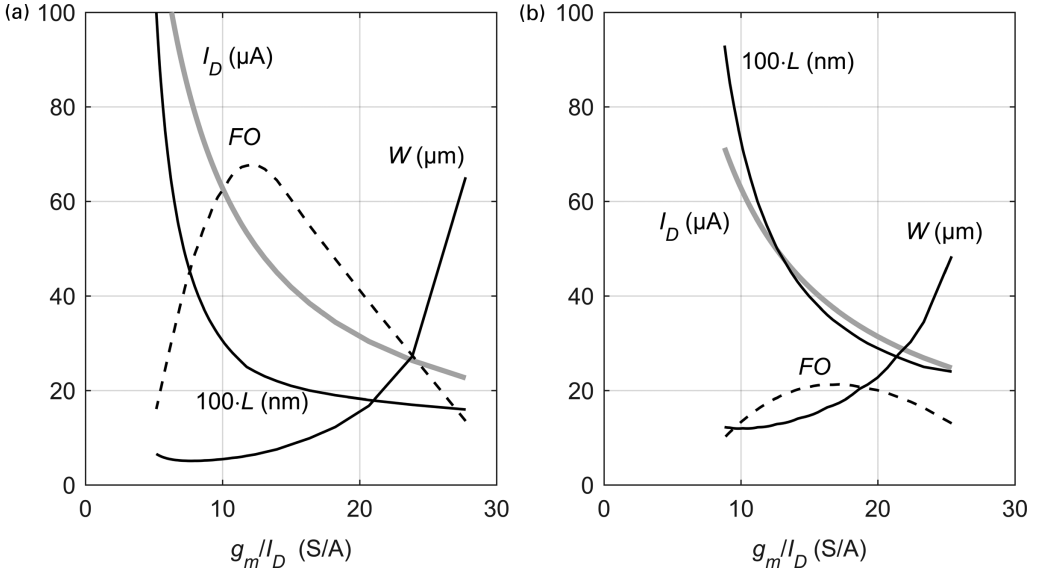

Figure 3.20 Plot of drain currents, device widths, gate lengths and fan-outs (a) $|A_{v0}| = 50$, and (b) $|A_{v0}| = 80$. ($V_{DS} = 0.6$ V and $V_{SB} = 0$ V).

Table 3.9 Result summary.

	(a)	SPICE verification	(b)	SPICE verification
$\|A_{v0}\|$	50	50.5	80	80.45
L (μm)	0.240	–	0.340	–
W (μm)	6.479	–	17.41	–
I_D (μA)	50.74	–	36.59	–
FO	68.0	–	21.2	–
g_m/I_D (S/A)	12.38	12.31	17.17	17.03
$f_u\ (MHz)$	100	99.0	100	98.0
V_{GS} (V)	0.541	0.544	0.465	0.465
C_{gg} (fF)	14.7	14.6	47.0	46.2

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

Table 3.10 Sizing data for minimum drain current.

$\lvert A_{v0}\rvert$	50	80
g_m/I_D (S/A)	28.53	27.02
L (μm)	0.158	0.233
I_D (μA)	22.02	23.26
W (μm)	88.90	71.02

that the calculated and SPICE-simulated numbers are very close. Since maximum fan-out minimizes the total gate capacitance $C_{gg} = C_L/FO$, this can be a desirable design choice.

As a final step, Table 3.10 summarizes the data for minimum I_D, which is achieved for maximum g_m/I_D and thus $FO = 10$ (minimum allowed value). As we see from the result, this yields design points in weak inversion. We observe that the currents and device widths don't change appreciably between the two cases. This is qualitatively consistent with our previous observations on weak inversion design in Section 3.1.5.

3.1.7 Inclusion of Extrinsic Capacitances

In this section, we take a closer look at the impact of extrinsic capacitances (which were ignored so far), and devise methods for including them in the sizing process. Extrinsic capacitances include the device's junction capacitances (C_{sb} and C_{db}), as well as the gate-to-drain fringe capacitance (C_{gd}). These capacitances are called extrinsic, since unlike the intrinsic gate capacitance (C_{gs}), they are not essential to the operation of the MOSFET. In other words, the device would still function if these where eliminated.

Figure 3.21(a) shows the small-signal model of the IGS with the relevant extrinsic capacitances C_{db} and C_{gd} included. Relative to Figure 3.2, we omitted C_{gs} and C_{gb} since these capacitances are shorted by the input voltage source, which is still assumed to be ideal for the sake of simplicity. With this assumption, the circuit is well approximated by the model in Figure 3.21(b), where the total drain capacitance $C_{dd} = C_{db} + C_{gd}$ appears in parallel to the load. This is an approximation since C_{gd} contributes a feedforward current that is being neglected. However, it can be easily shown that this current is relevant only beyond the transit frequency of the transistor, which is outside the range of our analysis.

Especially for high-speed designs with large g_m and correspondingly large device width, the added capacitance C_{dd} can result in a noticeable error in the gain-bandwidth product. For instance, in the high-speed, low-power design of Example 3.3, option (ii), we saw an error of 6.7% between the Matlab calculation and the SPICE simulation. The situation worsens once additional devices (such as

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

active loads in Section 3.2.1) are connected to the output. We call this self-loading and will refer to the sum of these unwanted capacitances C_{self}.

We would like to establish a sizing process in which C_{dd} (and other self-loading parasitics) are accounted for. A fundamental issue with factoring C_{dd} into the design is that we don't know its value before the sizing is completed (i.e. the device width W is known). One way to deal with this problem is to use a three-step design procedure:

1. Design the IGS while ignoring the total drain capacitance C_{dd} (as done in all prior examples).
2. Find C_{dd} of the design obtained in step 1, call this value C_{dd1}.
3. Now scale the device width and current by the following factor to arrive at the final design:

$$S = \frac{1}{1 - \frac{C_{dd1}}{C_L}}. \tag{3.10}$$

To see why this will work (perfectly), consider the scaling behavior of the device and how it affects the unity gain frequency. When we scale the device current and

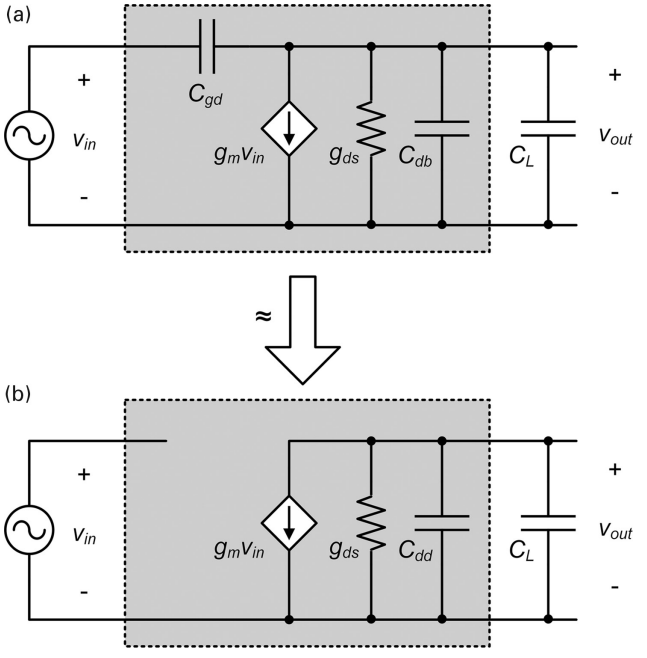

Figure 3.21 (a) Small-signal model of the IGS including the drain junction capacitance C_{db} and the gate-drain fringe capacitance C_{gd}. (b) Approximate model that adds $C_{dd} = C_{db} + C_{gd}$ in parallel to C_L.

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

width together, the current density and g_m/I_D stay constant, and both C_{dd} and g_m scale linearly with S. Therefore:

$$\omega_u = \frac{Sg_m}{C_L + SC_{dd1}} = \frac{g_m}{C_L}. \tag{3.11}$$

Solving (3.11) for the scaling factor S yields (3.10).

While this method works very well for the simple examples considered so far, it is restricted to cases where the parasitic capacitance scales linearly with the device transconductance that sets the unity gain frequency. As we shall see in Chapter 6, this is not the case for all circuits encountered in practice. We therefore devise a second solution that finds the correct sizing iteratively, and without any analytical assumptions (such as (3.11)). This approach is outlined as follows:

1. Start by assuming $C_{dd} = 0$.
2. Size the circuit to meet the GBW spec for $C + C_{dd}$. (For the first iteration, this means that we are ignoring C_{dd}.)
3. Estimate C_{dd} for the obtained design (using the device width from step 2).
4. Go to step 2 using the new C_{dd} estimate.
5. Repeat until convergence.

We will now illustrate this approach using an example.

Example 3.7 Iterative Sizing to Account for Self-Loading

Repeat Example 3.3, minimal power option (ii), with $C_L = 1$ pF and $f_u = 1$ GHz. Account for C_{dd} in the sizing process.

SOLUTION

In the quoted example, we showed that to maximize the intrinsic gain, g_m/I_D and L should equal 20.76 S/A and 70 nm, respectively. SPICE verifications showed good agreement as far as the gain is concerned, but the GBW was 6.7% less than expected. The error is due to self-loading. To account for C_{dd}, we pre-compute:

```
JD   = lookup(nch,'ID_W','GM_ID',gm_ID,'L',L);
Cdd_W = lookup(nch,'CDD_W','GM_ID',gm_ID,'L',L);
```

and then run the iterative loop below:

```
Cdd = 0;
for m = 1:5,
    gm = 2*pi*GBW*(CL + Cdd);
    ID(m,1) = gm./gm_ID;
    W(m,1)  = ID(m,1)./JD;
    Cdd = W*CDD_W;
end
```

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

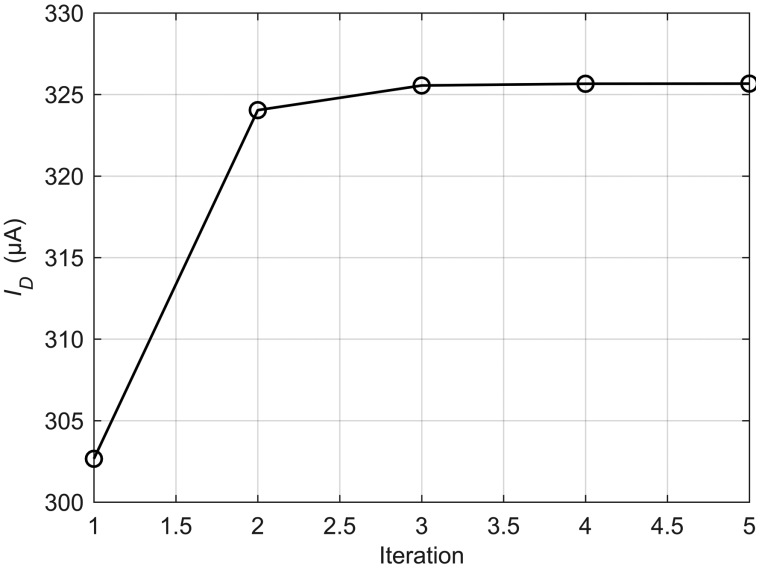

Figure 3.22 Progression of I_D as the extrinsic capacitance is factored in.

Figure 3.22 shows the evolution of the drain current as the number of iterations increases. After step 1, we get a first estimate of C_{dd} that we add to C_L in step 2. The current increases from 302.7 μA to 324.0 μA. We reach 325.6 μA after step 3 and observe that additional iterations lead to negligible increments. The final cumulative increase of the drain current is 7.60%. The width increases by the same percentage, from 114.2 to 122.8 μm.

The reader can verify that the scale factor S in (3.10) is equal to 1.076, predicting the same amount of current and width increase analytically.

Having laid the grounds of the g_m/I_D sizing methodology, we now turn the IGS progressively into a more realistic amplifier stage. In the following sections, we introduce active loading (Section 3.2.1) and resistive loading (Section 3.2.2). Finally, we consider a differential pair arrangement (Section 3.3).

3.2 Practical Common-Source Stages

To turn the IGS into a more practical common-source (CS) stage, we replace the ideal bias current source by a saturated p-channel transistor (Figure 3.23(a)) or by a resistor (Figure 3.23(b)). The following subsections discuss the sizing of these circuits.

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

3.2.1 Active Load

The presence of M_2 in Figure 3.23(a) introduces an extra conductance g_{ds2} in the circuit's small-signal model. Since the conductance lies in parallel with g_{ds1} (see Figure 3.2), it decreases the low-frequency voltage gain but does not impact the gain-bandwidth product. However, the additional extrinsic capacitance introduced by the active load can have a noticeable impact on the GBW if it is a significant fraction of C_L. We will consider this capacitance in some of the examples that follow below.

The drain current of M_2 is fixed per the unity gain frequency that the signal path device M_1 must realize (approximately g_{m1}/C_L). The only degrees of freedom for M_2 are its inversion level $(g_m/I_D)_2$ and gate length L_2. To develop guidelines on selecting these parameters, we begin by inspecting the circuit's low-frequency gain expression:

$$A_{v0} = -\frac{g_{m1}}{g_{ds1} + g_{ds2}} = -\frac{\left(\frac{g_m}{I_D}\right)_1}{\left(\frac{g_{ds}}{I_D}\right)_1 + \left(\frac{g_{ds}}{I_D}\right)_2} = -\frac{\left(\frac{g_m}{I_D}\right)_1}{\frac{1}{V_{EA1}} + \frac{1}{V_{EA2}}}. \tag{3.12}$$

When the Early voltages of the n- and p-channel transistors are equal, the CS stage exhibits a gain loss of 50% relative to the IGS. We can reduce this loss by increasing V_{EA2}, or equivalently, by reducing $(g_{ds}/I_D)_2$. Thus, to investigate further, Figure 3.24 illustrates the evolution of $(g_{ds}/I_D)_2$ versus gate lengths, with $(g_m/I_D)_2$ swept from strong inversion toward weak inversion.

We see that minimizing $(g_{ds}/I_D)_2$ requires the transconductance efficiency of the load transistor to be small, nearly 5 S/A, regardless of the gate length. This corresponds to strong inversion and per (2.34) yields a large saturation voltage V_{Dsat2} of nearly 0.4 V. With only 1.2V supply voltage, this loss in headroom is typically not acceptable, and one may instead pick a compromise with $(g_m/I_D)_2$ = 10…12 S/A,

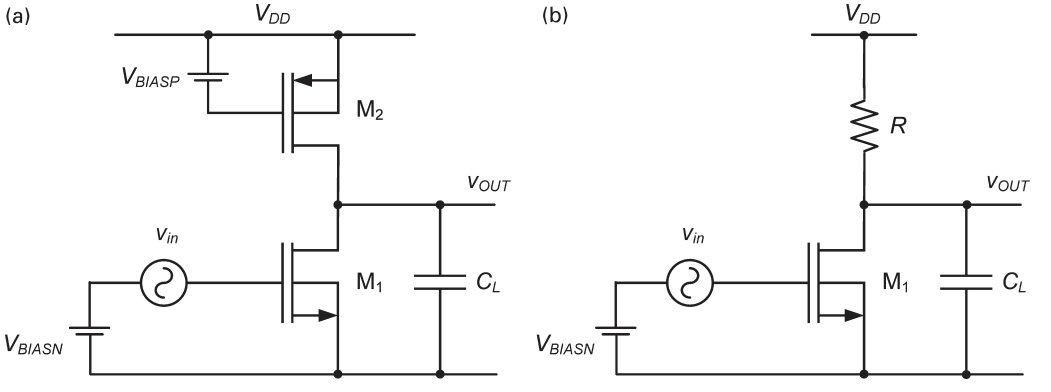

Figure 3.23 (a) Common source stage with (a) active p-channel load and (b) resistive load.

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

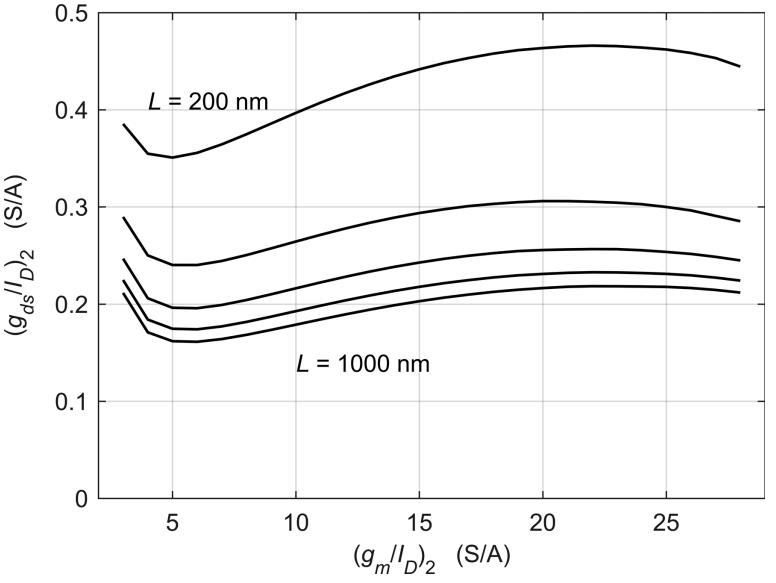

Figure 3.24 Plot of $(g_{ds}/I_D)_2$ versus g_m/I_D of a p-channel load device for various channel lengths (V_{DS} = 0.6 V, V_{SB} = 0 V).

leading to V_{Dsat2} = 0.20…0.17 V. Picking the "optimum" $(g_m/I_D)_2$ requires further knowledge about the specific design tradeoffs, and the extent to which gain is preferred over signal swing. In Section 4.1.1, we re-visit this question from a perspective of output dynamic range (ratio of maximum signal power and noise power). There, we find that g_m/I_D values near 10 S/A are indeed a good choice, and we therefore pick this value for the remainder of the discussion.

The final decision to be made concerns the gate length L_2. The choice is not obvious, since the benefits of large L_2 (for high gain) are offset by larger widths needed to sustain the same current. We will therefore study the tradeoffs in the following example.

Example 3.8 Sizing a CS Stage with Active Load

Consider the IGS of Example 3.3 with f_u = 1 GHz, C_L = 1 pF and FO = 10, but now add a p-channel load as shown in Figure 3.23(a). The supply voltage V_{DD} is equal to 1.2 V and the quiescent output voltage is $V_{OUT} = V_{DD}/2$. Assuming $(g_m/I_D)_2$ = 10 S/A, evaluate the impact of L_2 on all other device geometries and the required drain current. Compare the results for L_2 = 0.3, 0.5, and 1 μm with the data from Example 3.3 and perform a SPICE validation for L_2 = 0.5 μm.

SOLUTION

We begin by defining a suitable sweep range for L_1 and then compute the corresponding $(g_m/I_D)_1$ vector per the required transit frequency ($f_T = f_u \cdot FO$). This now also lets us compute $(g_{ds}/I_D)_1$.

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

```
L1 = .06: .001: .4;
gm_ID1  = lookup(nch,'GM_ID','GM_CGG',2*pi*fu*FO,'L',L1);
gds_ID1 = diag(lookup(nch,'GDS_ID','GM_ID',gm_ID1,'L',L1));
```

Next, we perform a similar sweep, which lets us compute $(g_{ds}/I_D)_1$ and A_{v0} using (3.12). From the obtained vector, we then select the maximum gain value, called $|A_{v0max}|$ below.

```
gm_ID2 = 10;
L2 = [.06 .1*(1:10)];
for k = 1:length(L2)
   gds_ID2 = lookup(pch,'GDS_ID','GM_ID',gm_ID2,'L',L2(k))
   Av0(:,k) = gm_ID1./(gds_ID1 + gds_ID2);
end
[a b] = max(Av0);
gain  = a';
```

The remaining steps perform the usual de-normalization discussed in Section 3.1.3, while simultaneously accounting for self-loading as done in Section 3.1.7.

```
Cself = 0;
for k = 1:10,
   gm = 2*pi*fu*(CL + Cself);
   ID = gm./gm_ID1(b);
   W1 = ID./diag(lookup(nch,'ID_W','GM_ID',gm_ID1(b),...
   'L',L1(b)));
   Cdd1 = W1.*diag(lookup(nch,'CDD_W','GM_ID',gm_ID1(b),,...
   'L',L1(b)));
   W2 = ID./lookup(pch,'ID_W','GM_ID',gm_ID2,'L',L2);
   Cdd2 = W2.*lookup(pch,'CDD_W','GM_ID',gm_ID2,'L',L2);
   Cself = Cdd1 + Cdd2;
end
```

Figure 3.25 shows the transistor geometries, maximum gain values and the drain current when L_2 is swept from 0.1 to 1 μm. We see that there is no reason to push L_2 beyond 0.5 μm, since the gain doesn't increase appreciably and the width of the load transistor grows rapidly. Also, notice the significant increase of the drain current for large L_2. To first order, I_D is fixed by M_1. However, as the self-loading increases with larger L_2, a larger g_{m1} is needed to maintain the desired unity gain frequency.

Table 3.11 compares sizes and performance data of the IGS from Example 3.3 and the two-transistor stage. We see that the gain loss caused by the active load is beyond 40%, even when L_2 is largest.

Table 3.12 investigates the impact of small changes in $(g_m/I_D)_2$ while keeping L_2 constant at 0.5 μm. We observe that the geometries of M_1 are nearly unaffected, while the width of the active load (W_2) varies strongly. We see that I_D increases for large W_2, due to the increased self-loading.

To conclude, we compare the data for L_2 = 0.5 μm to a SPICE simulation in Table 3.13. The agreement is better than in the earlier examples, since we included self-loading effects in the sizing process.

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

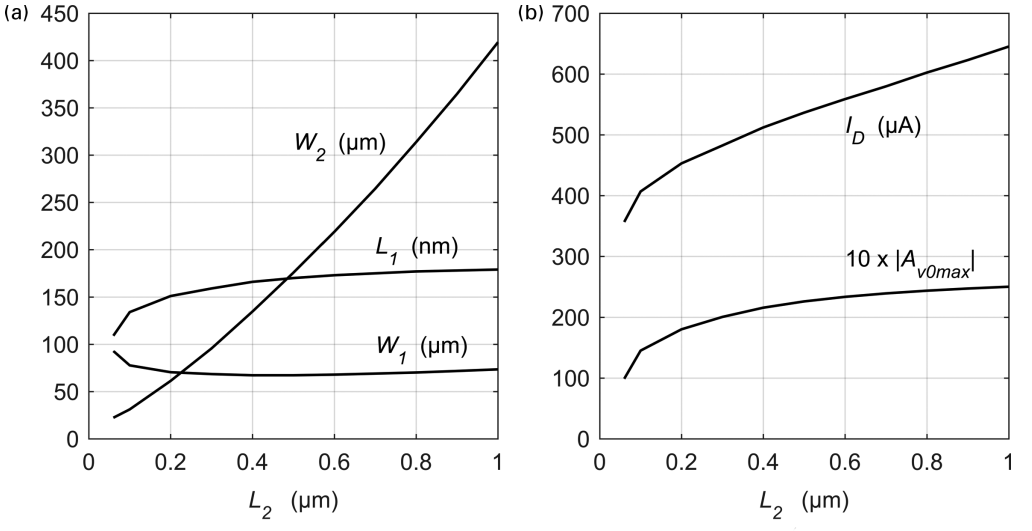

Figure 3.25 (a) Device geometries and (b) drain current and maximum gain of the actively loaded CS stage versus the gate length of the active p-channel load.

Table 3.11 Result comparison. IGS and p-channel load CS stages designed for maximum LF gain and $f_u = 1$ GHz, $C_L = 1$ pF and $FO = 10$.

	IGS	**CS stage with p-channel load**		
		$L_2 = 1$ µm	$L_2 = 0.5$ µm	$L_2 = 0.3$ µm
$\lvert A_{v0max}\rvert$	40.82	25.0	22.6	20.1
$(g_m/I_D)_1$	10.62	12.92	13.55	14.38
I_D (µA)	592	645	537	483
L_1 (nm)	220	179	170	159
W_1 (µm)	45.9	73.4	67.2	68.5
W_2 (µm)	–	419.3	176	95.6
C_{self} (pF)	–	0.327	0.157	0.104

Table 3.12 Impact of small changes in $(g_m/I_D)_2$ for fixed $L_2 = 0.5$ µm.

$(g_m/I_D)_2$ (S/A)	$\lvert A_{v0max}\rvert$	$(g_m/I_D)_1$ (S/A)	L_1 (nm)	W_1 (µm)	W_2 (µm)	I_D (µA)
9	22.85	13.48	171	65.32	135.1	526.8
10	22.61	13.55	170	67.21	175.6	536.6
11	22.37	13.63	169	69.52	225.6	548.8

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

Table 3.13 Design values and SPICE verification data.

	Design values	SPICE verification
L_1 (nm)	170	–
L_2 (nm)	500	–
W_1 (μm)	67.21	–
W_2 (μm)	175.6	–
$(g_m/I_D)_1$ (S/A)	13.55	13.55
$(g_m/I_D)_2$ (S/A)	10.00	9.99
I_D (μA)	536.6	536.7
$\|A_{v0}\|$	22.61	22.60
V_{GS1} (V)	0.5213	0.5212
V_{GS2} (V)	0.5857	–
f_u (MHz)	1000	1013
C_{self} (pF)	0.157	0.151

As a final example, we take now a closer look at the large-signal characteristic of an actively loaded CS stage. The large-signal characteristic is important as it defines the output signal swing that the stage can accommodate. The example will show that we can obtain an accurate characteristic using our Matlab lookup tables.

Example 3.9 Large-Signal Characteristic of a CS Stage with Active Load

Construct the transfer characteristic (v_{OUT} versus v_{IN}) of the CS stage from Table 3.13. Compare the small-signal voltage gain to the slope of the transfer characteristic at the quiescent point (V_{OUT} = 0.6 V) and assess the available output voltage swing. Compare the results to a SPICE simulation.

SOLUTION

To construct the transfer characteristic, we sweep v_{OUT} across the voltage range of interest and find the corresponding drain current I_{D2} of the p-channel load. We then determine the gate voltages of M_1 that make I_{D1} equal to I_{D2}. In the code below, I_{D2} is a vector, and I_{D1} a matrix.

```
VDS1 = .05: .01: 1.15;
ID2 = Wp*lookup(pch,'ID_W','VGS',VGS2,'VDS',VDD-VDS1,'L',L2);
ID1 = Wn*lookup(nch,'ID_W','VGS',nch.VGS,'VDS',VDS1,'L',L1)';
for m = 1:length(VDS1),
  VGS1(:,m) = interp1(ID1(m,:),nch.VGS,ID2(m));
end
```

The resulting transfer characteristic is shown in Figure 3.26. The voltage gain extracted from the slope of the transfer characteristic at the quiescent point is equal

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

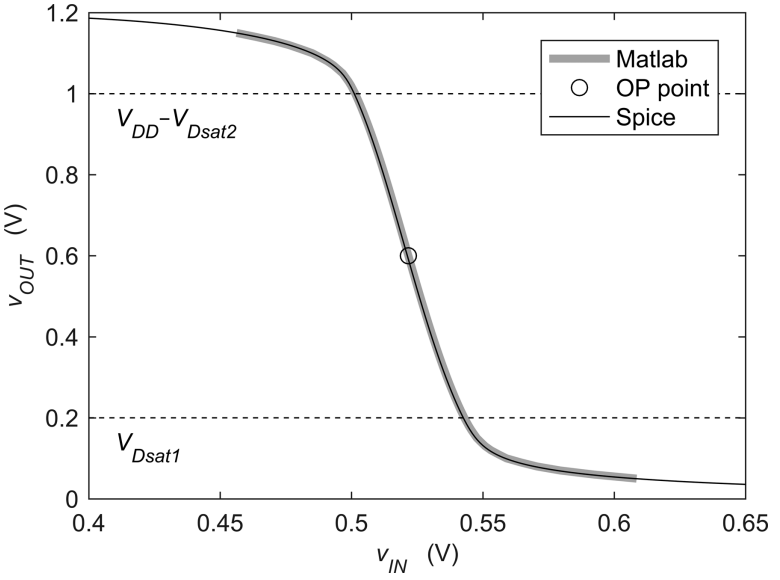

Figure 3.26 Transfer characteristic of the p-channel loaded IGS.

to −22.60, the same value as in Table 3.13. The horizontal lines in the plot mark the (approximate) output voltages for which the transistors leave saturation. The saturation voltages are found using (2.34) ($V_{Dsat} = 2/(g_m/I_D)$), which yields 2/(10 S/A) = 200 mV for both transistors in this example. Hence, the maximum output voltage swing that can be handled by this circuit is approximately ±400 mV. As the circuit nears these peak excitations, it becomes progressively less linear. We will quantify these nonlinear effects in Chapter 4.

The thin solid line in Figure 3.26 was obtained from a SPICE simulation. We observe a good agreement with the curve computed using the lookup data in Matlab.

3.2.2 Resistive Load

For the resistively loaded CS stage in Figure 3.23(b), the expression for the LF voltage gain is obtained by replacing g_{ds2} of the p-channel load in (3.12) with $1/R$. This yields:

$$A_{v0} = -\frac{g_{m1}}{g_{ds1} + \dfrac{1}{R}} = -\frac{\left(\dfrac{g_m}{I_D}\right)_1}{\left(\dfrac{g_{ds}}{I_D}\right)_1 + \dfrac{1}{I_D R}} = -\frac{\left(\dfrac{g_m}{I_D}\right)_1}{\dfrac{1}{V_{EA1}} + \dfrac{1}{I_D R}}. \tag{3.13}$$

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

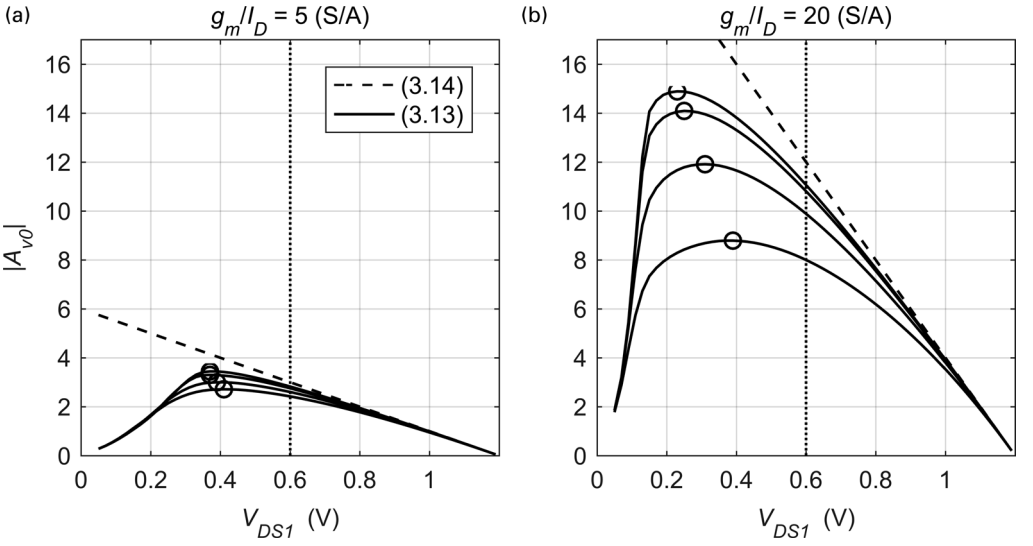

Figure 3.27 LF gain of the resistively loaded CS stage versus drain-to-source voltage. (a) g_m/I_D = 5 S/A and (b) g_m/I_D = 20 S/A. V_{DD} = 1.2 V and L_1 = 0.1, 0.2, 0.5, and 1.0 μm (bottom up).

We see from this expression that the DC voltage drop across the load resistor (I_DR) plays a key role in setting the gain. In fact, for scenarios where the $I_DR << V_{EA1}$, the expression simplifies to:

$$A_{v0} \cong -\left(\frac{g_m}{I_D}\right)_1 I_D R. \tag{3.14}$$

To maximize the gain, one should make I_DR as large as possible, but this reduces the available voltage swing, and additionally lowers V_{EA1}, due to the decrease in V_{DS1}. The latter effect gives rise to a maximum in the achievable gain.

Figure 3.27 plots (3.13) and (3.14) when V_{DS1} is swept over almost the entire supply voltage range, considering various gate lengths and two inversion levels: (a) strong inversion (5 S/A) and (b) moderate inversion (20 S/A). Circles mark the maxima, which are well below the gains obtained with the IGS and actively loaded CS stage. Note also that the corresponding drain-to-source voltages are well below $V_{DD}/2$ (dotted vertical line), which is non-ideal if large signal swing is desired. In the case of strong inversion, it would be impractical to operate the circuit at the maximum gain values, since V_{DS1} is very close to V_{DSat1} (400 mV) allowing essentially no signal swing. The situation is much improved in moderate inversion. Lastly, we see that for large L_1 (top curves) and large V_{DS1}, which leads to large V_{EA1}, the true voltage gain given by (3.13) is close to the simplified expression of (3.14).

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

Example 3.10 Sizing a CS Stage with Resistive Load

Consider a resistively loaded CS stage with C_L = 1 pF, f_u = 1 GHz, V_{DD} = 1.2 V and FO = 10. Find the values of V_{DS} and L that maximize the LF voltage gain and compute the corresponding load resistor value. Account for self-loading during sizing and validate the design using a SPICE simulation.

SOLUTION

We begin by defining a sweep range for the gate length and drain-to-source voltage (LL and UDS). Next, we find all g_m/I_D and g_{ds}/I_D values within this space for which f_T is equal to the required 10 GHz.

```
UDS = .1*(2:6);       %horizontal
LL  = .06: .01: .2;   %vertical
for k = 1:length(UDS)
  gmID(:,k)=lookup(nch,'GM_ID','GM_CGG',2*pi*fT,...
  'VDS',UDS(k),'L',LL);
  gdsID(:,k)=lookup(nch,'GDS_ID','GM_CGG',2*pi*fT,...
  'VDS',UDS(k),'L',LL);
end
```

Now we can compute the LF gain using (3.13) and determine the gate length, drain-to-source voltage and transconductance efficiency that maximize the expression:

```
AvoR = gmID./(gdsID + 1./(VDD-UDS(ones(length(LL),1),:)))
[a b] = max(AvoR);
[c d] = max(a);
AvoRmax = c
L    = LL(b(d))
VDS  = UDS(d)
gm_ID = gmID(b(d),d)
```

The maximum gain magnitude is found to be 8.49, along with L = 0.11 μm, V_{DS} = 0.4 V and g_m/I_D = 18.04. The remaining task is to de-normalize as usual, and iterate to accommodate the self-loading:

```
JD   = lookup(nch,'ID_W','GM_ID',gm_ID,'VDS',VDS,'L',L);
Cdd_W = lookup(nch,'CDD_W','GM_ID',gm_ID,'VDS',VDS,'L',L);
Cdd = 0;
for k = 1:5,
   gm = 2*pi*fT/10*(C+Cdd);
   ID = gm/gm_ID;
   W  = ID/JD;
   Cdd = W*Cdd_W;
end
```

We find I_D = 368.4 μA and W = 87.92 μm. Since we know the drain current and the voltage drop across the load resistor, we can readily compute $R = (V_{DD} - V_{DS})$

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

Table 3.14 Result summary.

	Design Values	SPICE
L_1 (nm)	110	–
W (μm)	87.92	–
V_{DS} (V)	0.400	–
g_m/I_D (S/A)	18.04	18.03
I_D (μA)	368.4	368.3
$\lvert A_{v0}\rvert$	8.487	8.493
V_{GS} (V)	0.4646	–
f_u (MHz)	1000	1000

divided by I_D = 2.172 kΩ. Table 3.14 summarizes the design values along with the SPICE-simulated numbers, which are in close agreement.

3.3 Differential Amplifier Stages

Differential amplifiers are key building blocks in analog design. At the core of a differential amplifier, we typical find a differential pair, as shown in Figure 3.28. The basic operation of this circuit is well described and analyzed in classical circuit design textbooks, such as [3], [4]. However, these treatments are typically based on the square-law, and have therefore become inaccurate.

The goal of this section is twofold. First, we want to analyze the large-signal behavior of the differential pair using the basic EKV model introduced in Chapter 2. This analysis sets the stage for the distortion analysis that will follow in Chapter 4. Secondly, we want to review the sizing of amplifiers that are based on the differential pair. A key difference compared to the IGS and CS stages discussed previously is that the transconductance is set by the tail current source I_0, rather than a voltage applied at the gate.

We begin with the large-signal analysis of the differential pair. To derive EKV-based expressions, we first rewrite (2.19) for M_1 and M_2:

$$\begin{aligned} v_{P1} - v_S &= U_T\left(2\left(q_1 - 1\right) + log\left(q_1\right)\right) \\ v_{P2} - v_S &= U_T\left(2\left(q_2 - 1\right) + log\left(q_2\right)\right). \end{aligned} \tag{3.15}$$

We then subtract the two equations to isolate the incremental differential input, using (2.21) to relate the gate and pinch-off voltages:

$$v_{id} = v_{i1} - v_{i2} = nU_T\left[2\left(q_1 - q_2\right) + \log\left(\frac{q_1}{q_2}\right)\right]. \tag{3.16}$$

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

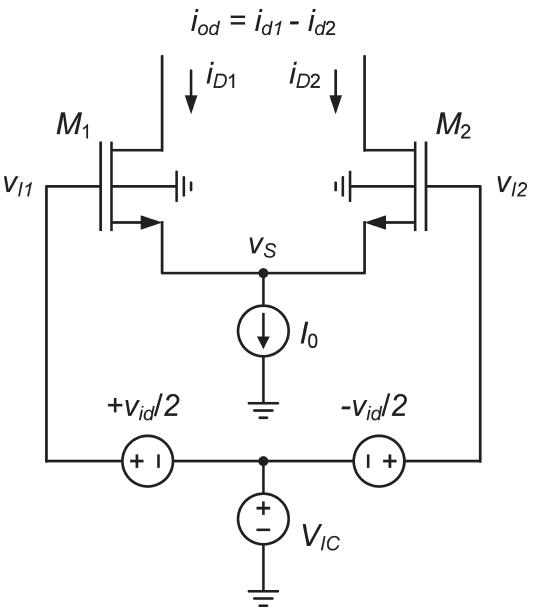

Figure 3.28 The differential pair.

To find the tail node voltage excursion v_S, we add the two instances of (3.15) and replace the quiescent point term ($V_{P1} + V_{P2}$) by:

$$2\frac{V_{IC} - V_T}{n} = 2V_P = 2(U_T\left(2\left(q_0 - 1\right) + log\left(q_0\right) - V_S\right). \tag{3.17}$$

Here, q_0 designates the normalized charge density at the quiescent point. We then solve for v_S from the sum, and find the incremental tail voltage:

$$v_s = v_S - V_S = U_T\left[2q_0 - \left(q_1 + q_2\right) + \log\left(\frac{q_0}{\sqrt{q_1 q_2}}\right)\right]. \tag{3.18}$$

So far, the differential input voltage v_{id} and the incremental tail voltage v_s are functions of q_1 and q_2. To switch to the normalized EKV drain currents i_1 and i_2, we invert (2.22):

$$q_1 = 0.5\left[\sqrt{1 + 4i_1} - 1\right] \quad and \quad q_2 = 0.5\left[\sqrt{1 + 4i_2} - 1\right]. \tag{3.19}$$

We can now express the differential input voltage as a function of the normalized quantities i_1/i_0 or i_2/i_0, calling i_0 the sum of i_1 and i_2. Substituting i_{D1}/I_0 and i_{D2}/I_0 for i_1/i_0 or i_2/i_0 lets us numerically plot the transfer and source voltage curves of Figure 3.29, using the Matlab code below:

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

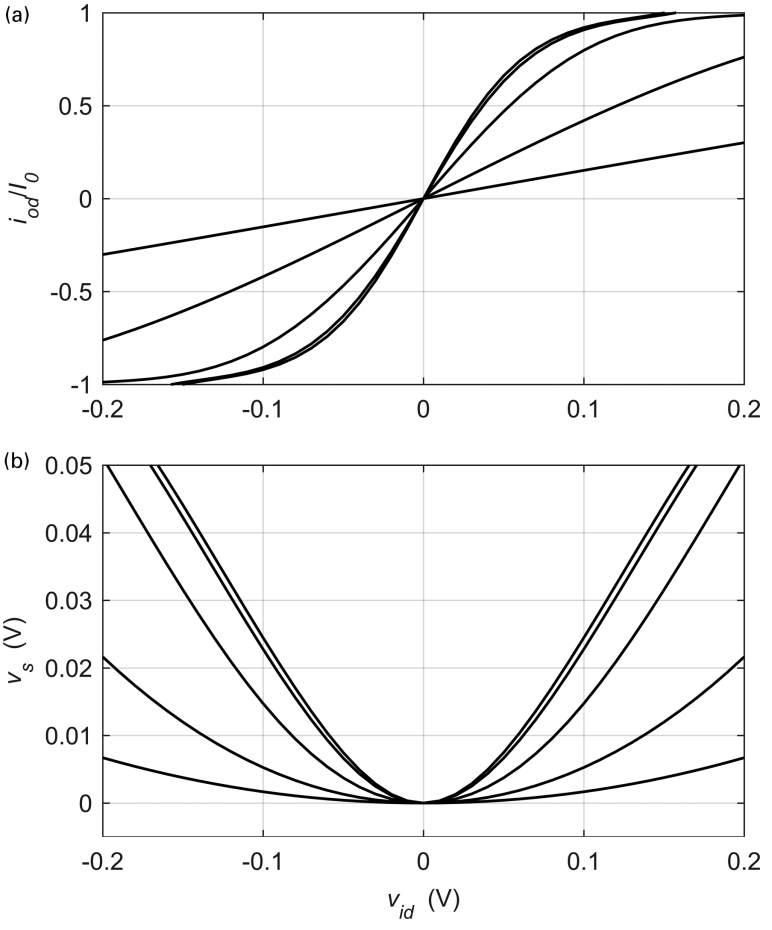

Figure 3.29 Large-signal characteristic of the differential pair as a function of differential input voltage considering five tail currents, (a) normalized differential output current and (b) Incremental component of the source node voltage. The five shown curves correspond to tenfold multiples of the normalized tail current i_0, from 0.01 (weak inversion, steepest curves) to 100 (strong inversion).

```
% data ==================
n  = 1.2;                  % subthreshold slope
io = 2*logspace(-2,2,5);   % normalized tail current I0/IS
vid = .01*(-20:20);        % input diff voltage range (V)
% compute ================
UT = .026;
m  = (.05:.05:1.95); b = find(m==1);
for k = 1:length(io),
   i2  = .5*io(k)*m;
   q2  = .5*(sqrt(1+4*i2)-1);
   q1  = .5*(-1 + sqrt(1 + 4*(io(k)-q2.^2-q2)));
   vg = n*UT*(2*(q2-q1) + log(q2./q1));
   IoD_I0(k,:) = 2*interp1(vg,i2,vid,'spline')/io(k) - 1 ;
   q  = q2(b);
   vs = UT*(2*q - (q1+q2) + log(q./sqrt(q1.*q2)));
```

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

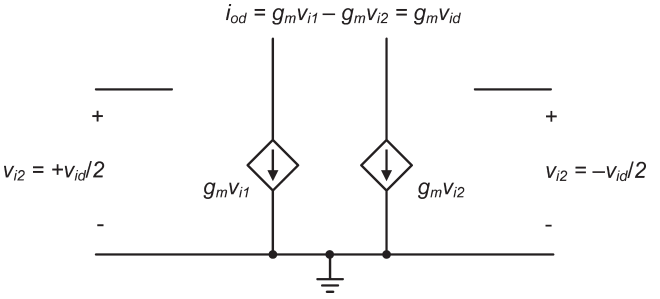

Figure 3.30 Simplified small-signal model of the differential pair. (All parasitic elements are omitted.)

```
   VS(k,:) = interp1(vg,vs,vid,'spline');
end
```

The five shown curves correspond to tenfold multiples of the normalized tail current i_0, from 0.01 (weak inversion, steepest curves) to 100 (strong inversion). They confirm the well-known fact that the tail node voltage does not move significantly if the differential input is small. Lastly, note that the shown plots required only one parameter, the subthreshold slope factor n. We will see in Section 4.2 that this greatly simplifies the treatment of nonlinear distortion.

Given that the tail node does not move for small inputs allows us to turn it into an AC ground for small-signal modeling purposes, leading to the model shown Figure 3.30. Each half circuit sees half of the differential input, but after taking the difference at the output, the transconductance that links the (differential) input voltage and (differential) output current is simply g_m, just like in an IGS.

Given the similarity between the small-signal model of the IGS and the differential pair, everything that we learned in the previous sections is also applicable to amplifiers that are based on the differential pair. As mentioned initially, the main difference lies in how the bias point is established. The following example will repeat Example 3.1 for a differential pair to illustrate this.

Example 3.11 Sizing a Differential Pair with Ideal Current Source Loads

Size the differential amplifier shown in Figure 3.31 to achieve $f_u = 1$ GHz with a load capacitance $C_L = 1$ pF. Assume $g_m/I_D = 15$ S/A and $L = 60$ nm. The input common mode is $V_{IC} = 0.7$ V and the shown common-mode feedback (CMFB) circuit forces $V_{OC} = (V_{O1} + V_{O2})/2 = 1$ V at the operating point. Plot the large-signal transfer characteristic from the differential input to the differential drain current. Verify the results using SPICE simulations.

SOLUTION

We first compute the transconductance and drain currents of M_1 and M_2 as in Example 3.1, using (3.4) and (3.8):

```
gm = 2*pi*fu*CL;
ID = gm/gm_ID;
```

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

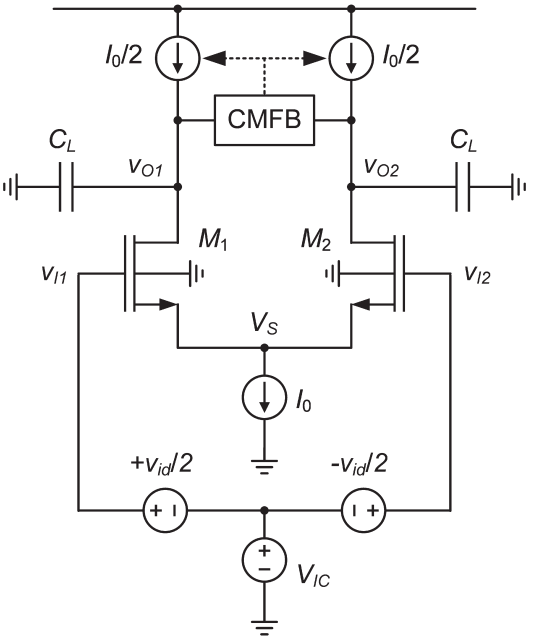

Figure 3.31 Differential pair amplifier. The common mode feedback (CMFB) block forces the output common-mode voltage to the desired value.

To find the width, we divide I_D by the current density J_D keeping in mind that the tail node (V_S) of the amplifier is not grounded. To find the actual source-to-bulk voltage, called V_{SB} in the lookup function calls below, we compute the gate-to-source voltage V_{GS} using the lookupVGS function:

```
VGS = lookupVGS(nch, 'GM_ID',gmID,'VDB',VDB,'VGB',VGB);
```

and derive V_{SB} from the difference $V_{GB} - V_{GS}$. We then find the current density using:

```
JD = lookup(nch,'ID_W','GM_ID',gmID,'VSB',VSB,'VDS',VDB-VSB)
```

Since f_u and L are the same as in Example 3.1, the drain currents of M_1 and M_2 again equal 419 μA. The device widths differ slightly, however; they are equal to 41.06 μm instead of 41.72 μm. The decrease in W is the result of (1) a larger V_{GS} (0.475 instead of 0.468 V) due to backgate bias and (2) a smaller V_{DS} (0.5254 instead of 0.6 V). The latter difference impacts the result due to DIBL, which is a strong effect with L = 60 nm. The source-to-bulk voltage V_{SB} (equal to the tail node voltage V_S) is 225.4 mV.

The circuit in Figure 3.31 is now simulated in SPICE with the given parameters (V_{OC} = 1 V, V_{IC} = 0.7 V, W = 41.06 μm, L = 60 nm, $I_0/2$ = 419 μA). The DC operating point simulation yields: g_m/I_D = 6.31 mS/419 μA = 15.06 S/A and V_S = 226 mV. The frequency response is shown below, indicating a good match with the targeted unity gain frequency of 1 GHz.

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

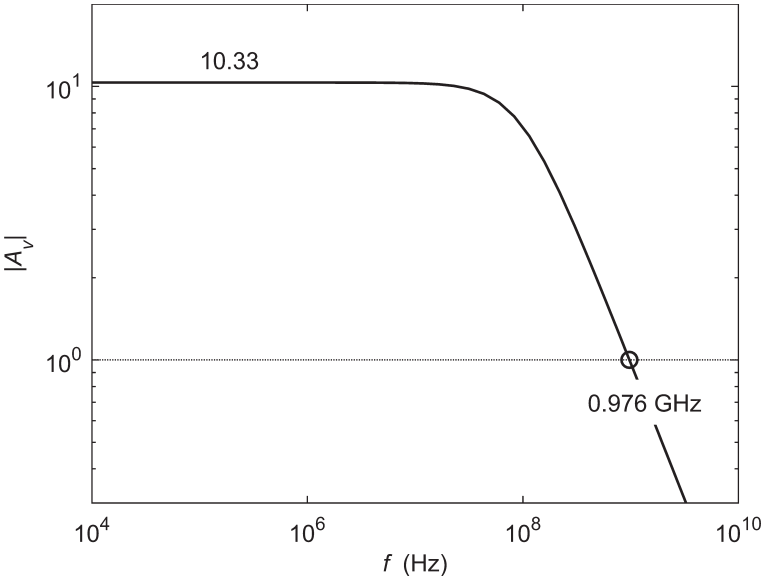

Figure 3.32 Simulated frequency response of the differential pair amplifier.

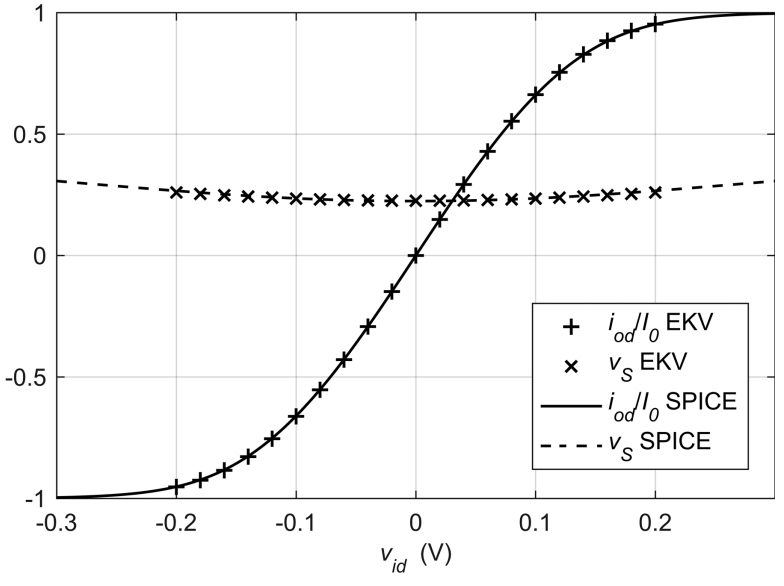

Figure 3.33 Normalized differential output current and tail node voltage (v_S) versus differential input voltage (v_{id}).

Lastly, we are to compare the transfer characteristic obtained by SPICE to the characteristic predicted by the basic EKV model. Therefore, we extract the basic EKV parameters considering the source and drain voltages used in the amplifier. This is done by running:

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

```
jd = lookup(nch,'ID_W','VGS',nch.VGS,'VSB',VSB, ...
'VDS', VDB-VSB, 'L',0.06);
y = XTRACT2(nch.VGS,jd);
```

Once the EKV parameters are obtained, the construction of the i_{od}/I_o and v_S versus v_{id} characteristics follows from (3.16) and (3.18). Note that only the subthreshold slope factor n and the specific current I_S are required.

Figure 3.33 compares the normalized differential current for the SPICE simulation and the analytical result. The error is less than 0.3% over the differential input range, from –0.2 to +0.2 V. In the same diagram, we also show the plots for the source voltage V_S. Because the model does not include mobility degradation, the V_S of the EKV model departs slightly from the SPICE result as v_{id} grows. The error between the two curves is about 8 mV as v_{id} approaches ±0.2 V.

As another example, we consider the classical differential amplifier with current-mirror load [3]. This circuit is more practical than the idealized amplifier in the previous example, and it is also used within larger circuits covered in this book (see low-dropout regulator (LDO) example in Chapter 5).

Example 3.12 Sizing a Differential Amplifier with Current-Mirror Load

Design the amplifier in Figure 3.34 to achieve f_u = 100 MHz, assuming C_L = 1 pF, V_{IC} = 0.7 V and V_{DD} = 1.2 V. To simplify this task, you may re-use results obtained earlier in this chapter (consider Example 3.6, Section 3.2.1 and Figure 3.25).

SOLUTION

The first step is to choose suitable transconductance efficiencies and gate lengths for the differential pair and the current-mirror load. To simplify this task, we can look at Example 3.6, Table 3.9(b), where we considered a CS stage with the same f_u, FO = 21.2, L = 0.34 μm and g_m/I_D = 17 S/A. Furthermore, in Section 3.2.1 we examined the addition of a p-channel load and concluded that its g_m/I_D should be about 10 S/A due to gain, V_{Dsat} and noise considerations. Finally, Figure 3.25 showed that there is diminishing return in the voltage gain for gate lengths larger than 0.5 μm. Our decision is therefore to use this length for our current mirror load.

Next, we proceed to the evaluation of the source and drain voltages of the input pair transistors. For $M_{1a,b}$, V_{SB1} is equal to zero due to the source-bulk tie shown in the schematic. V_{DS1} is unknown and we must find the source and drain voltages of $M_{1a,b}$ to compute it. Since we know the gate voltages of $M_{1a,b}$ (equal to the given V_{IC}), we can find the source voltage by subtracting V_{GS1} from V_{IC}. To find V_{GS1}, we use:

```
VGS1 = lookupVGS(nch,'GM_ID',gmID1,'L',L1);
```

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

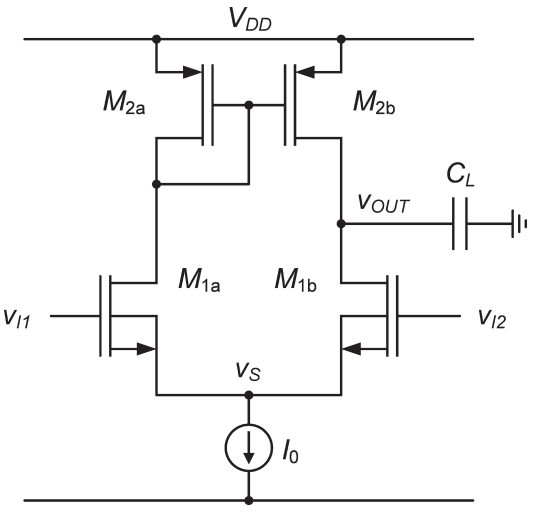

Figure 3.34 Differential amplifier with current-mirror load.

Here, we assumed the default value for V_{DS1} (0.6 V). This is fine, since the transistor is saturated and the gate length won't be minimum. We find that V_{GS1} is equal to 0.4661 V. Note that we can re-run this calculation later once an improved estimate of V_{DS1} is available. To find the drain voltage, we subtract V_{GS2} from V_{DD}.

```
VGS2 = lookupVGS(pch,'GM_ID',gmID2,'L',L2);
```

taking again the default value for V_{DS2}. We refine the result running the command below, where V_{GS2} serves as the new estimate for V_{DS2}.

```
VGS2 = lookupVGS(pch,'GM_ID',gmID2,'VDS',VGS2,'L',L2);
```

This yields V_{GS2} = 0.5858 V and thus V_{DS1} = 0.3812 V, which indicates that M_{1b} is saturated. Rerunning the commands a second time, we'll see that V_{GS1} increased slightly from 0.4661 to 0.4670 V. We also find that V_S = 0.2330 V.

To compute the LF voltage gain of the amplifier, we make use of the result from Figure 3.13. This is valid since the voltage gain of a differential amplifier with current-mirror load is the same as that of an actively loaded CS stage [4]. Hence:

```
gdsID1=lookup(nch,'GDS_ID','GM_ID',gmID1,'VDS',VDS1,'L',L1);
gdsID2=lookup(pch,'GDS_ID','GM_ID',gmID2,'VDS',VGS2,'L',L2);
Av0 = gmID1/(gdsID1 + gdsID2)
```

We find A_{v0} = 31.06. Finally, to determine the drain currents and the transistor widths, we repeat the same steps taken for the CS stage:

```
gm1 = 2*pi*fu*CL;
ID  = gm1/gmID1;
JD1 = lookup(nch,'ID_W','GM_ID',gmID1,'VDS',VDS1,'L',L1);
```

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

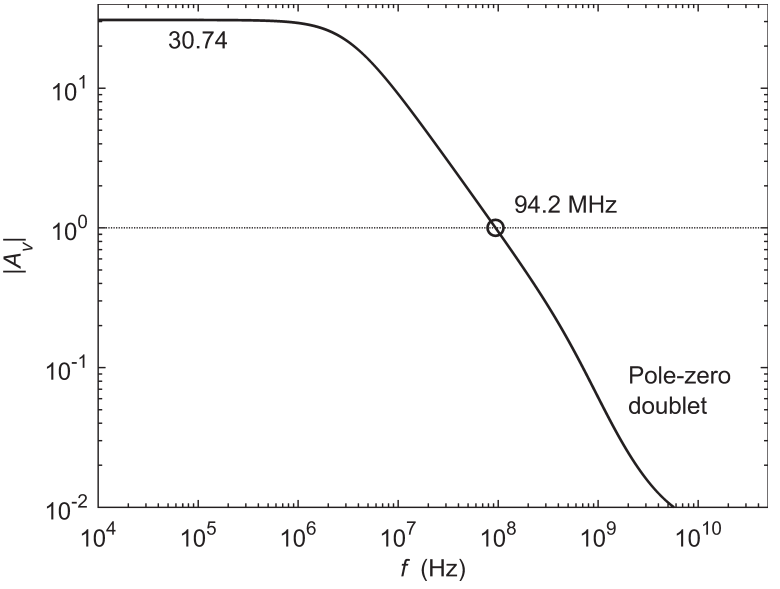

Figure 3.35 SPICE-simulated frequency response of the differential pair with current mirror load.

```
W1  = ID/JD1;
JD2 = lookup(pch,'ID_W','GM_ID',gmID2,'VDS',VGS2,'L',L2);
W2  = ID/JD2;
```

We find that the drain currents of $M_{1a,b}$ and $M_{2a,b}$ are 36.96 μA, W_1 is 18.07 μm and W_2 = 12.13 μm. The design is now completely specified and we can simulate it in SPICE. Figure 3.35 shows the obtained frequency response, which matches well with the computed numbers.

The simulation showed a tail node voltage of 0.2329 V and an output voltage of 0.6142 V, which agrees with the predicted values (0.2330 V and 0.6142 V). For the voltage gain, we measure 30.74, which is also close to the predicted value of 31.06. The unity gain frequency f_u is a little less than 100 MHz, because we neglected self-loading. Scaling the current and all widths by the factor S in (3.10) would correct this.

Even though the simulated unity gain frequency matches our expectations, it is worth paying attention to the pole-zero doublet caused by the current mirror [3] (see Figure 3.35). The pole of the doublet lies at approximately $f_T/2$ of the p-channel load devices,[5] which amounts to 580 MHz. This has little impact on our design with f_u = 100 MHz, but becomes a problem in faster circuits. One way to push the doublet to higher frequencies is to reduce L_2.

[5] The pole location is defined by the current mirror node resistance $1/g_{m2}$ and two times C_{gg2}, since two devices are loading the node. The result is $\omega_T/2$.

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

As a final example, we consider a differential amplifier with a resistive input driver. This is a very common scenario, for example in cascaded gain stages. A key aspect in this problem is the inclusion of the Miller effect in the bandwidth calculation, which we have so far not considered.

Example 3.13 Sizing a Differential Amplifier with Resistive Input Driver and Resistive Loads

Design the amplifier in Figure 3.36 to achieve a differential voltage gain of $|A_{v0}| = 4$ assuming $C_L = 50$ fF, $R_D = 1$ kΩ, $R_S = 10$ kΩ, $V_{SC} = 0.7$ V and $V_{DD} = 1.2$ V. Size the transistors assuming $L = 100$ nm and $g_m/I_D = 15$ S/A. Estimate the dominant and non-dominant pole frequency of the circuit and validate the design using SPICE simulations.

SOLUTION

The first step is to compute g_m from the voltage gain requirement. For this purpose, we first need to find the intrinsic gain of the transistors:

```
gm_gds = lookup(nch, 'GM_GDS', 'GM_ID', gm_ID, 'L', L);
```

Using this information, we can readily compute g_m using:

```
gm = 1/RL*(1/Av0 - 1./gm_gds).^-1;
```

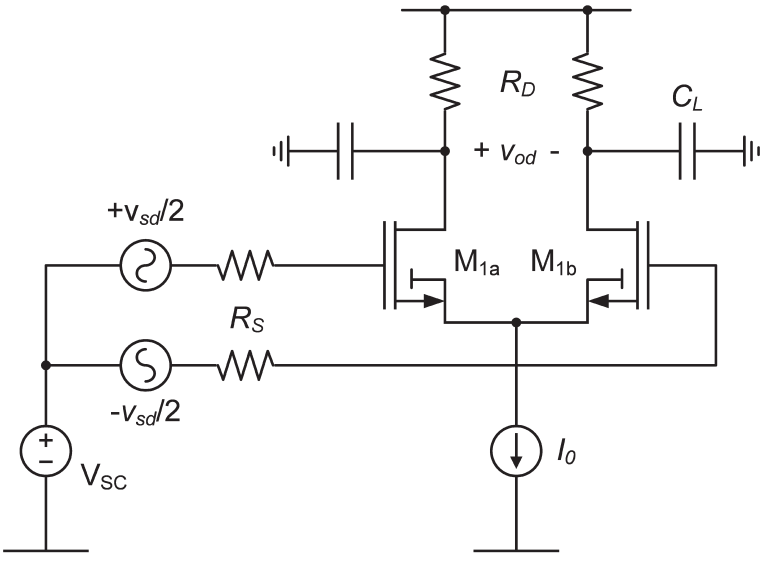

Figure 3.36 Differential amplifier with resistive load and resistive input driver.

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

This yields g_m = 4.93 mS, I_0 = 2 x 328 μA and W = 36.46 μm (using the usual denormalization). To calculate the circuit's pole frequencies, we can invoke the following textbook expressions, which are based on the dominant pole approximation [3]:

$$\omega_{p1} = \frac{1}{b_1} \quad \omega_{p2} = \frac{b_1}{b_2}, \tag{3.20}$$

where:

$$\begin{aligned} b_1 &= R_s\left[C_{gs} + C_{gd}\left(1 + |A_{v0}|\right)\right] + R_L\left(C_{Ltot} + C_{gd}\right) \\ b_2 &= R_s R_L\left[C_{gs}C_{Ltot} + C_{gs}C_{gd} + C_{Ltot}C_{gd}\right] \end{aligned} \tag{3.21}$$

and $C_{Ltot} = C_L + C_{db}$. Note the multiplication of C_{gd} by the term $(1 + |A_{v0}|)$, which is due to the Miller effect.

Since we have already sized the transistors, we can readily compute the pole frequencies using the following capacitance estimates:

```
Cgs = W.*lookup(nch, 'CGS_W', 'GM_ID', gm_ID, 'L', L)
Cgd = W.*lookup(nch, 'CGD_W', 'GM_ID', gm_ID, 'L', L)
Cdd = W.*lookup(nch, 'CDD_W', 'GM_ID', gm_ID, 'L', L)
Cdb = Cdd-Cgd
CLtot = CL+Cdb;
```

Table 3.15 summarizes the obtained design values along with their SPICE validation (using an operating point analysis, followed by a pole-zero analysis). We observe close agreement between the calculated and simulated numbers. The largest discrepancy is in the non-dominant pole frequency, caused mainly by the approximate nature of the employed analytical expression. There are also some small discrepancies in the capacitance values. These are mainly due to assuming

Table 3.15 Result summary.

	Design values	SPICE
L (nm)	100	–
W (μm)	36.46	–
g_m/I_D (S/A)	15	15.09
g_m (mS)	4.93	4.95
$\|A_{v0}\|$	4	4.06
C_{gs} (fF)	27.94	28.01
C_{gd} (fF)	12.13	12.20
C_{db} (fF)	10.76	9.2
f_{p1} (MHz)	166	167
f_{p2} (GHz)	5.50	6.10

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press

V_{DS} = 0.6 V in the Matlab calculation, while the SPICE simulation indicates V_{DS} = 0.66 V. These errors can be minimized by re-computing the capacitances with a better estimate of V_{DS}, but it would not be worth the effort in this example.

3.4 Summary

This chapter established a methodology for the systematic sizing of simple gain stages. The starting point for this development was the intrinsic gain stage (IGS), a single-transistor circuit with an ideal current source and capacitive load. To find the current I_D, we divide the required transconductance g_m by g_m/I_D. To find the device width, we "de-normalize" by dividing I_D with the drain current density J_D obtained from lookup tables.

Several examples served to illustrate how to pick appropriate gate lengths and transconductance efficiencies to match a given set of objectives, for example, maximum voltage gain, minimal current consumption, etc.

For circuits that are not constrained by speed requirements, weak inversion is a natural choice. In weak inversion, g_m/I_D is nearly constant and thus cannot serve as a distinguishing design knob. In this case, the current density J_D can be used as a substitute.

In high-speed circuits, extrinsic capacitances at the transistors' drain nodes can lead to noticeable self-loading and bandwidth reduction. We showed two iterative sizing methods that address this issue.

After establishing these basics, we extended the scope of the discussion and considered the sizing of common-source stages with active and resistive loads, as well as differential pair stages. We showed that the basic flows and considerations established for the IGS still apply.

3.5 References

[1] B. Murmann, *Analysis and Design of Elementary MOS Amplifier Stages*. NTS Press, 2013.

[2] P. Harpe, H. Gao, R. van Dommele, E. Cantatore, and A. van Roermund, "A 3nW Signal-Acquisition IC Integrating an Amplifier with 2.1 NEF and a 1.5fJ/conv-step ADC," in *ISSCC Dig. Tech. Papers*, 2015, pp. 382–383.

[3] P. R. Gray, P. Hurst, S. H. Lewis, and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*, 5th ed. Wiley, 2009.

[4] T. Chan Caruosone, D. A. Johns, and K. W. Martin, *Analog Integrated Circuit Design*, 2nd ed. Wiley, 2011.

https://doi.org/10.1017/9781108125840.003 Published online by Cambridge University Press