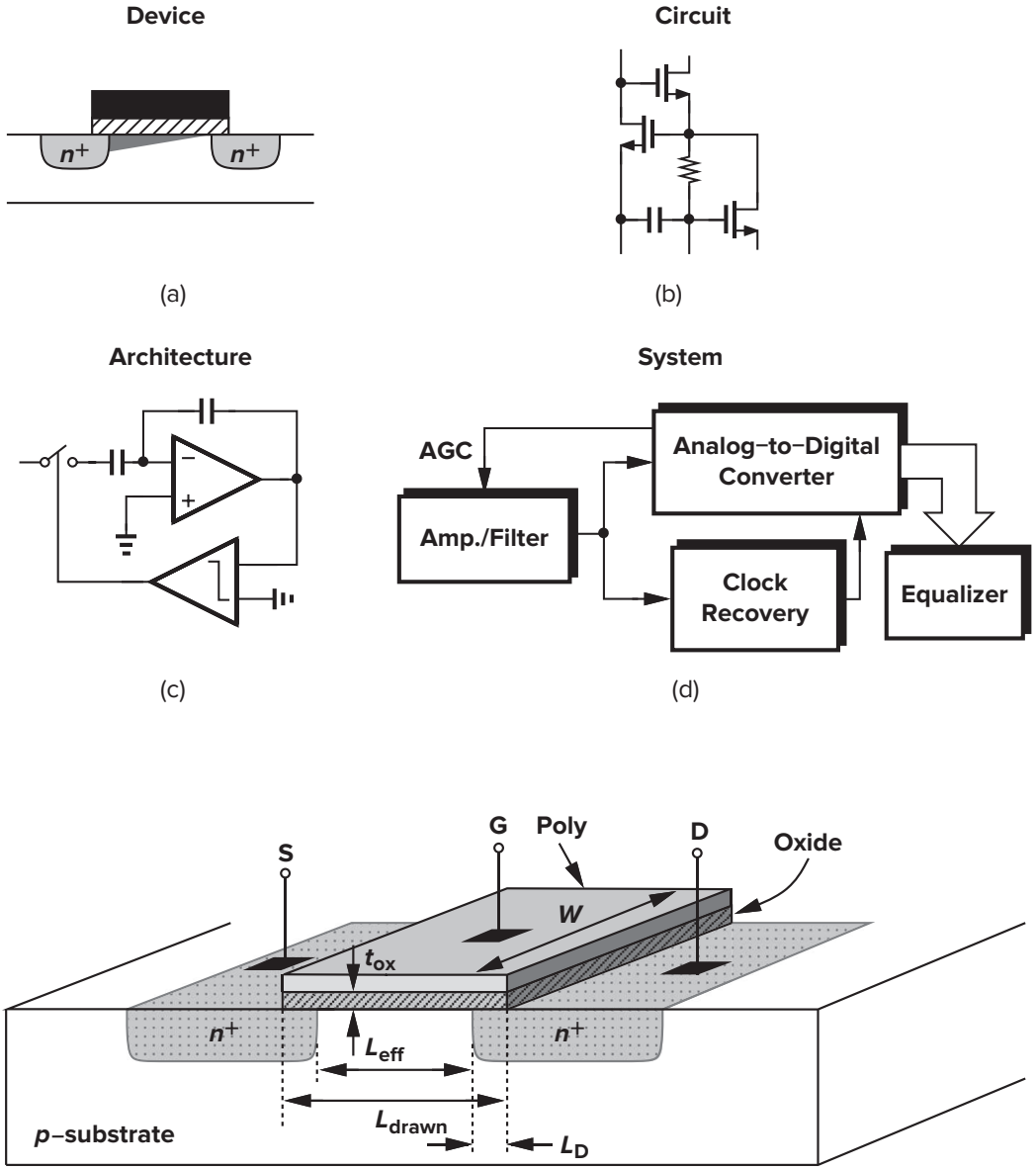
Notes for *Design of Analog CMOS Integrated Circuits*

《模拟 CMOS 集成电路设计》笔记

Yi Ding
(University of Chinese Academy of Sciences, Beijing 100049, China)

丁毅
(中国科学院大学，北京 100049)

2024.11 – ...



Preface

to be completed

Table 1: Learning Plan

Task	Week	Date	Planned Pages (actual)
☑	1	2025.01.13 - 2025.01.19	007 - 046 (007-044)
☑	2	2025.01.20 - 2025.01.26	047 - 086 (045-064)
☑	3	2025.01.27 - 2025.02.02	087 - 126 (065-104)
	4	2025.02.03 - 2025.02.09	127 - 166 (105-000)
	5	2025.02.10 - 2025.02.16	167 - 206 (000-000)
	6	2025.02.17 - 2025.02.23	207 - 246 (000-000)

序言

待完成

Contents

Chapter 1 Introduction to Analog Design

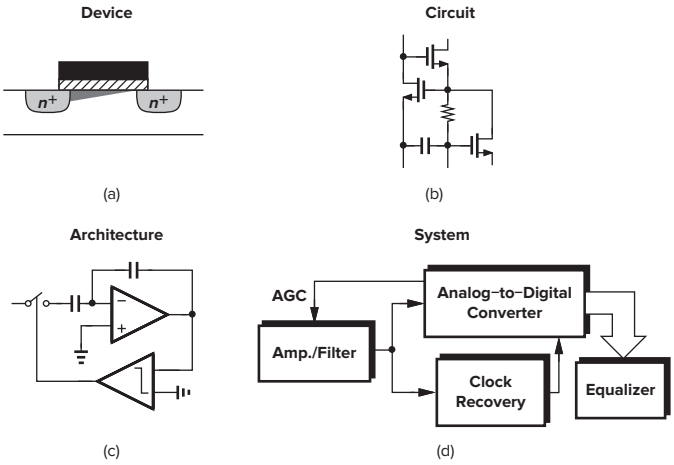


Figure 1.1: Abstraction levels in analog design:
(a) device level;
(b) circuit level;
(c) architecture level;
(d) system level

Chapter 2 Basic MOS Device Physics

In this chapter, we study the physics of MOSFETs at an elementary level, covering the bare minimum that is necessary for basic analog design. The ultimate goal is still to develop a circuit model for each device by formulating its operation.

After studying many analog circuits in Chapters 3 through 14 and gaining motivation for a deeper understanding of devices, we return to the subject in Chapter 17 and deal with other aspects of MOS operation including more advanced properties and second-order effects.

2.1 General Considerations

Figure 2.1 shows a simplified structure of an n-type MOSFET (NMOS) device.

(1) p-type substrate: also called **bulk** or **body**;
(2) a heavily-doped (conductive) piece of polysilicon (called **poly**) operating as the gate;
(3) a thin layer of silicon dioxide (SiO_2) (called **oxide**) insulats the gate from the substrate;
(4) $L_{\text{eff}} = L_{\text{drawn}} - 2L_D$, where L_{eff}[①] is the effective channel length (typically 10 nm in 2015), L_{drawn} is the total length, and L_D is the amount of side diffusion.
(5) t_{ox}: gate oxide thickness (typical 15 Å in 2015)

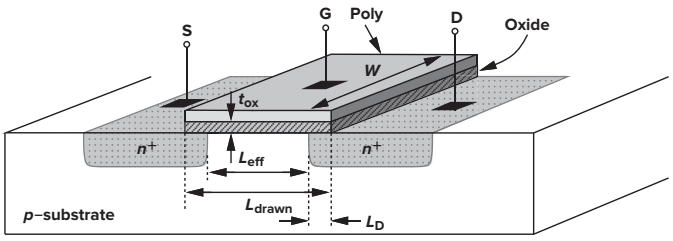


Figure 2.1: Simplified structure of an NMOS device

In practice, NMOS and PMOS devices must be fabricated on the same wafer, i.e., the same substrate. For this reason, one device type can be placed in a **local substrate**, usually called a **well**. In today's CMOS processes, the PMOS device is fabricated in an n-well (on the p-type substrate), depicted in Figure 2.2.

Figure 2.2 indicates that, while all NFETs share the same substrate, each PFET can have an independent n-well. This flexibility of PFETs is exploited in some analog circuits.

Some modern CMOS processes offer a **deep n-well** (an n-well that contains an NMOS device and its p-type bulk), so that the NMOS device can be isolated from the other NMOS devices.

The circuit symbols used to represent NMOS and PMOS transistors are shown in Figure 2.3. The letter "B" stands for **bulk** or **body**, i.e., the substrate of the device. The source of the PMOS device is positioned on top as a visual aid because it has a higher potential than its gate.

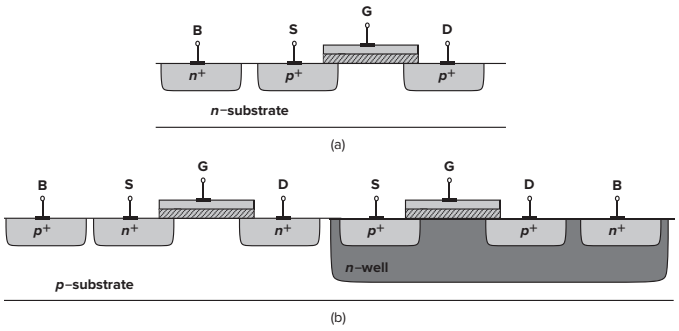


Figure 2.2: CMOS processes. (a) A simple PMOS device; (b) NMOS and PMOS devices on the same substrate

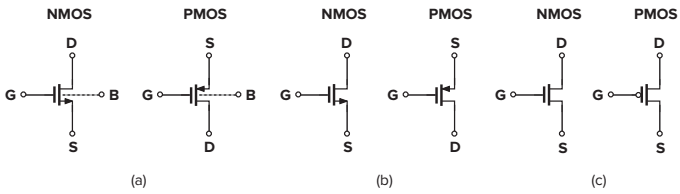


Figure 2.3: Circuit symbols for MOSFETs.
(a) NMOS and PMOS devices;
(b) omit bulk connections (GND for NMOS and VDD for PMOS);
(c) digital representation of NMOS and PMOS devices

In this book, we prefer those in Figure 2.3(b) to gain a clear view of the device.

2.2 MOS I/V Characteristics

2.2.1 Threshold Voltage

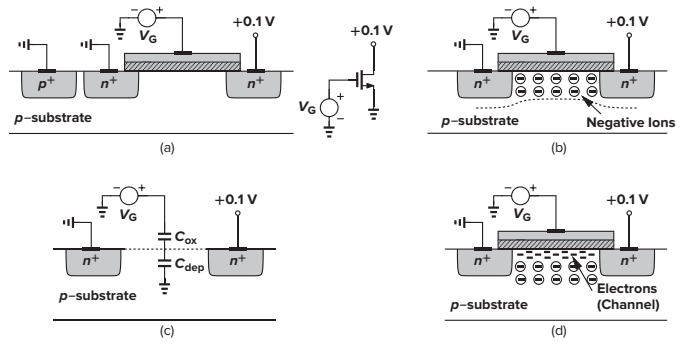


Figure 2.4: Formation of inversion layer in an NMOS.
(a) A MOSFET driven by a gate voltage;
(b) formation of depletion region;
(c) onset of inversion;
(d) formation of inversion layer.

[①] In the remainder of this book, we denote the effective length L_{eff} by L unless otherwise stated.

When the interface potential reaches a sufficiently positive value (V_{GS}), a **channel** of charge carriers is formed under the gate oxide between S and D, and the transistor is "turned on". We say the interface is **inverted**, and the channel is called **inversion layer**.

In reality, the turn-on phenomenon is a gradual function of the gate voltage, making it difficult to define V_{TH} unambiguously. In semiconductor physics, the V_{TH} of an NFET is usually defined as the gate voltage for which the interface is "as much n-type as the substrate is p-type", given by

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

where

(1) Φ_{MS}: work function difference between the polysilicon gate and the silicon substrate;
(2) Q_{dep}: the charge in the depletion region (see Figure 2.4);
(3) $C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}} = \frac{3.9\varepsilon_0}{t_{ox}}$: the oxide capacitance per unit;
(4) Φ_F: Fermi potential difference between the surface and bulk, given by $\Phi_F = \frac{kT}{q_e} \ln \frac{N_{\text{sub}}}{n_i}$, where N_{sub} is the doping density of the substrate, n_i is the density of electrons in intrinsic silicon, and q_e is the electron charge.

Since $C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}} = \frac{3.9\varepsilon_0}{t_{ox}}$ appears frequently in device and circuit equations, it could be helpful to remember that $t_{ox}C_{ox} = 3.9 \times (8.854 \times 10^{-12})\ \text{F/m} = 345\ \text{Å} \cdot \text{fF}/\mu\text{m}^2$ remains constant.

In practice, V_{TH} is usually adjusted by implantation of dopants into the channel area during device fabrication. In essence, to deplete the layer, if a thin layer of p^+ (n^-) is created, the threshold voltage of an NMOS device increases (decreases), depicted in Figure 2.5.

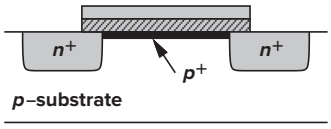


Figure 2.5: Implantation of p^+ dopants in an NMOS to alter (increase) the threshold voltage.

2.2.2 I/V Characteristics

Now considering an NMOS with $V_{GS} \geqslant V_{TH}$ (see Figure 2.6), and assume $Q_d = \frac{\mathrm{d}\rho_{\text{charge}}}{\mathrm{d}L} = \frac{\rho_{\text{charge}}}{L}$ is the mobile charge density along the direction of current, regard gate-channel as a capacitor, yielding

$$Q_d = WC_{ox}\left[V_{GS} - V_{TH} - V(x)\right] \tag{2.2}$$

where $V(x)$ is the channel potential at x. The current I_D is given by

$$I_D = -Q_d \cdot v = -WC_{ox}\left[V_{GS} - V_{TH} - V(x)\right] \cdot \mu_n \frac{\mathrm{d}V(x)}{\mathrm{d}x}$$

subject to $V(0) = 0$ and $V(L) = V_{DS}$. Performing integration yields I_D as a function of V_{DS}

$$I_D = \mu_n C_{ox} \frac{W}{L}\left[\left(V_{GS} - V_{TH}\right)V_{DS} - \frac{V_{DS}^2}{2}\right] \tag{2.3}$$

$$I_{D,\max} = \frac{1}{2}\mu_n C_{ox} \frac{W}{L}\left(V_{GS} - V_{TH}\right)^2 \tag{2.4}$$

Note that L is the effective channel length. We call $(V_{GS} - V_{TH})$ the **overdrive voltage** and $\frac{W}{L}$ the **aspect ratio**. We say the device is in the **triode region** (or linear region) if $V_{DS} \leqslant V_{GS} - V_{TH}$, and in the **saturation region** if $V_{DS} > V_{GS} - V_{TH}$.

Remark that the integration in (2.3) assumes that μ_n and V_{TH} are independent of x, V_D and V_G.

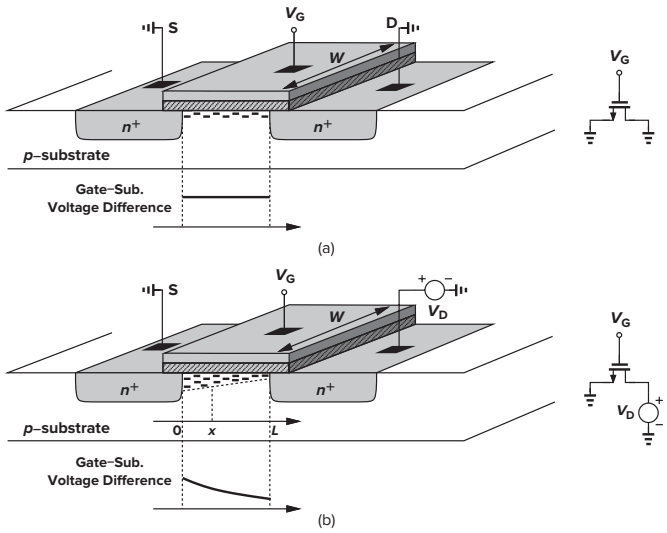


Figure 2.6: I/V characteristics of an NMOS device.
(a) $V_D = V_S$;
(b) $V_D > V_S$

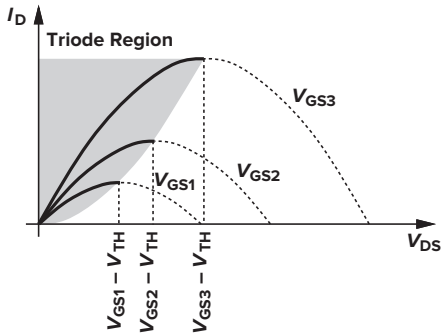


Figure 2.7: Triode region of an NMOS device.

In the deep triode region, i.e., $V_{DS} \ll 2(V_{GS} - V_{TH})$, we have

$$I_D \approx \mu_n C_{ox} \frac{W}{L}\left(V_{GS} - V_{TH}\right)V_{DS} \tag{2.5}$$

$$R_{\text{on}}|_{V_{DS}\to 0^+} = \frac{1}{\mu_n C_{ox}\frac{W}{L}\left(V_{GS} - V_{TH}\right)} \tag{2.6}$$

Therefore, as long as $V_{DS} \ll 2(V_{GS} - V_{TH})$, a MOSFET can operate as a voltage-controlled resistor [actually for $|V_{DS}| \ll 2(V_{GS} - V_{TH})$].

In the satyration region [$V_{DS} > (V_{GS} - V_{TH})$], the channel is **pinched off** (the inversion layer stops at $x < L$), leading to a relatively constant current I_D with respect to V_{DS}, depicted in Figure 2.6 and given by

$$I_D = \frac{1}{2}\mu_n C_{ox}\frac{W}{L'}(V_{GS} - V_{TH})^2 \tag{2.7}$$

$$V_{GS} = V_{TH} + \sqrt{\frac{2I_D}{\mu_n C_{ox}\frac{W}{L'}}} \tag{2.8}$$

We say the device exhibits a **square-law** behavior. As the electrons approach the pinch-off point where $Q_d \to 0$, their velocity rises tremendously ($v = \frac{I}{Q_d}$) so that they simply shoot through the depletion region and arrive at the drain terminal.

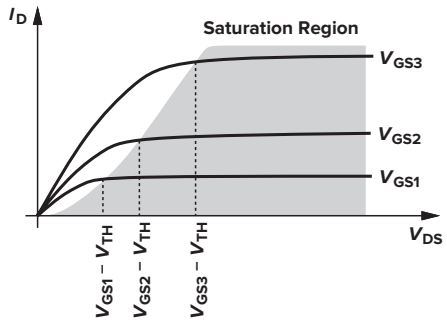


Figure 2.8: Saturation of drain current

For PMOS devices, the equations (2.3) and (2.7) are respectively written as

$$I_D = -\mu_p C_{ox}\frac{W}{L}\left[(V_{GS} - V_{TH})V_{DS} - \frac{V_{DS}^2}{2}\right] \tag{2.9}$$

$$I_D = -\frac{1}{2}\mu_p C_{ox}\frac{W}{L'}(V_{GS} - V_{TH})^2 \tag{2.10}$$

Note that V_{GS}, V_{DS}, V_{TH}, and $(V_{GS} - V_{TH})$ are negative for a PMOS transistor that is turned on. We can also rewrite the equations as

$$I_{SD} = \mu_p C_{ox}\frac{W}{L}\left[(V_{SG} - |V_{TH}|)V_{SD} - \frac{V_{SD}^2}{2}\right] \tag{2.11}$$

$$I_{SD} = \frac{1}{2}\mu_p C_{ox}\frac{W}{L'}(V_{SG} - |V_{TH}|)^2 \tag{2.12}$$

2.2.3 MOS Transconductance

Define the transconductance g_m as the change in I_D with respect to V_{GS}, that is

$$g_m = \frac{\partial I_D}{\partial V_{GS}} \tag{2.13}$$

In the triode region, we have

$$g_m = \mu_n C_{ox}\frac{W}{L}V_{DS} \tag{2.14}$$

In the saturation region

$$g_m = \mu_n C_{ox}\frac{W}{L}(V_{GS} - V_{TH}) \tag{2.15}$$

$$= \sqrt{2\mu_n C_{ox}\frac{W}{L}I_D} \tag{2.16}$$

$$= \frac{2I_D}{V_{GS} - V_{TH}} \tag{2.17}$$

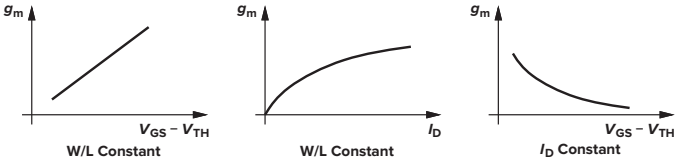


Figure 2.9: Approximate MOS transconductance as a function of overdrive and drain current.

Simply add a negative sign to obtain the transconductance of a PMOS device.

2.3 Second-Order Effects

2.3.1 Body Effect

Our analysis has so far entailed the assumption that the substrate (bulk, body) is connected to the source terminal. By changing the substrate voltage, we can alter the threshold voltage of the device because the gate charge must mirror Q_d before an inversion layer is formed, which is called the **body effect**. It can be derived that

$$\text{NMOS:}\quad V_{TH} = V_{TH0} + \gamma\left(\sqrt{2\Phi_F + V_{SB}} - \sqrt{2\Phi_F}\right)$$

$$\text{PMOS:}\quad V_{TH} = V_{TH0} + \gamma\left(\sqrt{2\Phi_F + V_{BS}} - \sqrt{2\Phi_F}\right)$$

$$\frac{\partial V_{TH}}{\partial V_{SB}} = \frac{g_{mb}}{g_m} = \eta$$

where

(1) $\gamma = \frac{\sqrt{2q_e\varepsilon_{si}N_{sub}}}{C_{ox}}$: the **body effect coefficient**, typically 0.4 $V^{1/2}$;

(2) $\Phi_F = \frac{kT}{q_e}\ln\frac{N_{sub}}{n_i}$: the Fermi potential difference between the surface and bulk.

2.3.2 Channel-Length Modulation

I_D in the saturation region, given by (2.7) and (2.10), is actually a function of V_{DS}, which is called **channel-length modulation**. Writing the actual length L' as $L' = L - \Delta L$, i.e., $\frac{1}{L'} \approx \frac{1+\frac{\Delta L}{L}}{L}$, we have (in saturation)

$$I_D \approx \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{GS} - V_{TH})^2(1 + \lambda V_{DS}) \tag{2.18}$$

where $\lambda \propto \frac{1}{L}$ is the **channel-length modulation coefficient**. λ represents the relative variation in length for a given increment in V_{DS}. Thus, for longer channels, λ is smaller. Note that there is no channel-length modulation in triode region.

To consider channel modulation, simply change $\frac{1}{L}$ to $\frac{1+\lambda V_{DS}}{L}$ to revise the previous equations for better accuracy.

Nanometer transistors suffer from various imperfections and markedly depart from square-law behavior. Shown below are the actual I-V characteristics of an NFET with $\frac{W}{L} = \frac{5\ \mu m}{40\ nm}$ for $V_{GS} = 0.3\ \text{V}, ..., 0.8\ \text{V}$. Also plotted are the characteristics of a square-law device of the same dimensions. Despite our best efforts to match the latter device to the former, we still observe significant differences.

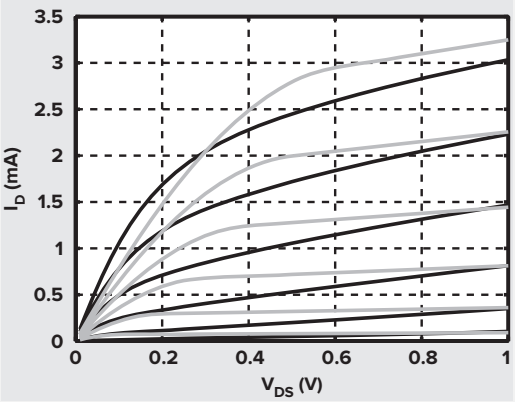


Figure 2.10: Actual I-V characteristics of an NFET and a square-law device.

2.3.3 Subthreshold Conduction

In reality, for $V_{GS} \approx V_{TH}$, a **week inversion layer** will still exist. Even for $V_{GS} < V_{TH}$, I_D is not zero, but exhibiting an expotional dependence on V_{GS}, called **subthreshold conduction**. Assuming V_{DS} is large enough ($V_{DS} > 100$ mV), the drain current is given by

$$I_D = I_0 \exp \frac{V_{GS}}{\xi V_T} = \alpha \frac{W}{L} \exp \frac{V_{GS}}{\xi V_T} \tag{2.19}$$

where

(1) $I_0 \propto \frac{W}{L}$: the drain current at $V_{GS} = V_{TH}$;
(2) $\xi > 1$: the noideality factor (typically 1.5);

As shown in Figure 2.11, we extrapolate the transfer characteristics I_D-V_{GS} on a logarithmic scale and consider their intercept voltage as the threshold voltage.

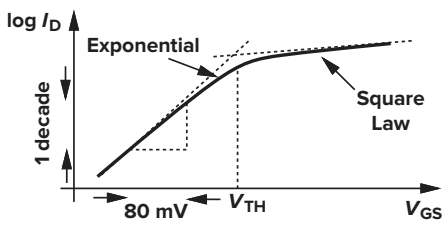


Figure 2.11: MOS subthreshold characteristics

When the corresponding transconductances g_m become equal for the same drain current, we say the transistor switch the inversion region. More specifically

$$g_m = \frac{I_D}{\xi V_T} = \frac{2I_D}{(V_{GS} - V_{TH})_{\text{switch}}} \tag{2.20}$$

$$(V_{GS} - V_{TH})_{\text{switch}} = 2\,\xi V_T \approx 80\ \text{mV} \tag{2.21}$$

We say the device operates in **week inversion** for $V_{GS} \leqslant (V_{TH} + 80\ \text{mV})$, and similarly in **strong inversion** for $V_{GS} > (V_{TH} + 80\ \text{mV})$. Equation (2.19) indicates that V_{GS} must decrease by roughly 80 mV for I_D to decrease by one decade (at room temperature), resulting a significant leak current (or power dissipation) for low threshold voltage devices.

To determine the operation region, for a given I_D, we need to obtain V_{GS} from both the suquare-law and the exponential-law and select the lower value:

$$V_{GS} = \min \left\{ \sqrt{\frac{2I_D}{\mu_n C_{ox} \frac{W}{L}}} + V_{TH},\ \xi V_T \ln \frac{I_D}{\alpha \frac{W}{L}} \right\} \tag{2.22}$$

2.4 MOS Device Models

2.4.1 MOS Device Layout

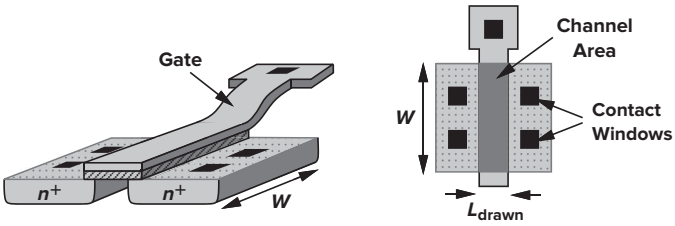


Figure 2.12

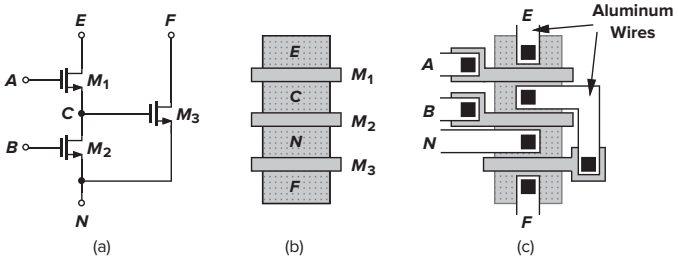


Figure 2.13: Layout of an NMOS device.
(a) bird's-eye view;
(b) top view (vertical view).

2.4.2 MOS Device Capacitances

To predict the high-frequency behavior, depicted in Figure 2.15, we expect that a capacitance exists between every two terminals of a MOSFET (C_{DS} is negligible). And these capacitances may depend on the bias conditions.

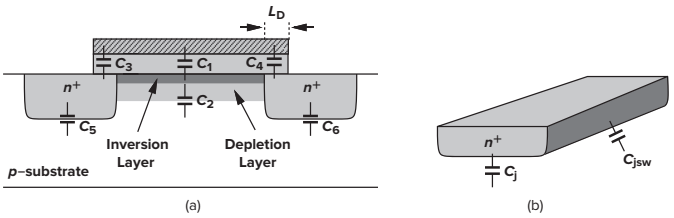


Figure 2.14: Decomposition of MOS capacitances
(a) MOS device capacitances
(b) S/D junction capacitance into bottom-plate and sidewall components

As shown in the figure above, we have:

(1) $C_1 = WLC_{ox}$: the oxide capacitance between the gate and the channel;

(2) $C_2 = WL\sqrt{\frac{q_e \varepsilon_{si} N_{sub}}{4\Phi_F}}$: the depletion capacitance between the channel and the substrate;

(3) $C_3 = C_4 = WC_{ov}$: the overlap capacitance, where C_{ov} is the overlap capacitance per unit width;

(4) $C_j = \frac{C_{j0}}{(1+\frac{V_R}{V_{\text{built-in}}})^{m_j}}$: the junction bottom-plate capacitance per unit area, where m typically in the range of 0.3 and 0.4;

(5) $C_{jsw} = \frac{C_{jsw0}}{(1+\frac{V_R}{V_{\text{built-in}}})^{m_{jsw}}}$: the junction sidewall capacitance per unit area;

(6) $C_5 = C_6 = S_{bp}C_j + S_{sw}C_{jsw}$: the junction capacitance

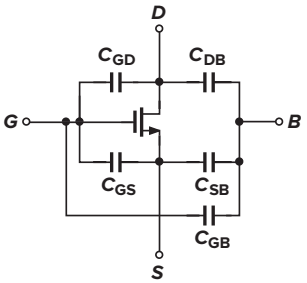


Figure 2.15: MOS capacitances

The terminal capacitances can be concluded as follows:

$$C_{DB} = C_{SB} = C_5 = S_{bp}C_j + S_{sw}C_{jsw}$$

$$C_{GB} = \begin{cases} C_1 \text{ series } C_2, & \text{off} \\ 0, & \text{else} \end{cases}$$

$$C_{GD} = \begin{cases} C_3 = WC_{ov}, & \text{else} \\ \frac{1}{2}C_1 + C_3 = \frac{1}{2}WLC_{ox} + WC_{ov}, & \text{deep triode} \end{cases}$$

$$C_{GS} = \begin{cases} C_3 = WC_{ov}, & \text{off} \\ \frac{2}{3}C_1 + C_3 = \frac{2}{3}WLC_{ox} + WC_{ov}, & \text{saturation} \\ \frac{1}{2}C_1 + C_3 = \frac{1}{2}WLC_{ox} + WC_{ov}, & \text{deep triode} \end{cases}$$

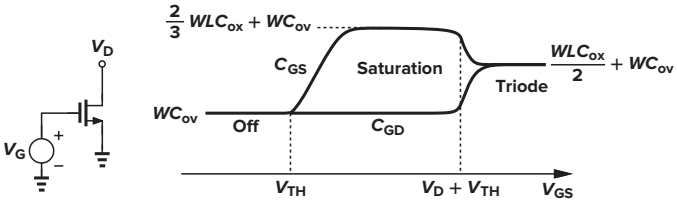


Figure 2.16: Variation of C_{GS} and C_{DS} versus V_{GS}

If the bulk terminal is tied to source, then C_{SB} is shorted, C_{DB} becomes C_{DS}, and C_{GB} is added to the original C_{GS}, yielding

$$C_{DS} = C_5 = S_{bp}C_j + S_{sw}C_{jsw}$$

$$C_{GD} = \begin{cases} C_3 = WC_{ov}, & \text{else} \\ \frac{1}{2}C_1 + C_3 = \frac{1}{2}WLC_{ox} + WC_{ov}, & \text{deep triode} \end{cases}$$

$$C_{GS} = \begin{cases} (C_1 \text{ series } C_2) + C_3, & \text{off} \\ \frac{2}{3}C_1 + C_3 = \frac{2}{3}WLC_{ox} + WC_{ov}, & \text{saturation} \\ \frac{1}{2}C_1 + C_3 = \frac{1}{2}WLC_{ox} + WC_{ov}, & \text{deep triode} \end{cases}$$

As shown in the figure below, new generations of CMOS technology incorporate the **FinFET** structure. The transistor carries current from S to D on the surfaces of the fin. The FinFET exhibits less channel-length modulation and subthreshold leakage by sacrificing contacts land and some other parameters.

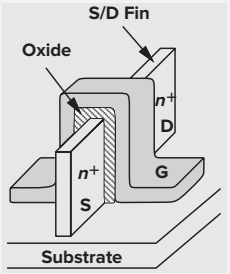


Figure 2.17: FinFET structure

2.4.3 MOS Small-Signal Model

$g_m = \frac{\partial I_D}{\partial V_{GS}}$ describe the small-signal behavior of MOSFETs when there is a perturbation in V_{GS}. Owing to channel-length modulation, I_D also varies with V_{DS}, and it can be modeled by a linear resistor r_O, given by

$$r_O = \frac{\partial V_{DS}}{\partial I_D} = \begin{cases} \left[\mu_n C_{ox}\frac{W}{L}\left(V_{GS} - V_{TH} - V_{DS}\right)\right]^{-1}, & \text{triode} \\ \left[\frac{1}{2}\mu_n C_{ox}\frac{W}{L}\left(V_{GS} - V_{TH}\right)^2 \cdot \lambda\right]^{-1}, & \text{saturation} \end{cases}$$

In saturation, we can rewrite r_O as:

$$r_O = \frac{1 + \lambda V_{DS}}{\lambda I_D} = \frac{1}{\lambda I_D} + \frac{V_{DS}}{I_D} \approx \frac{1}{\lambda I_D} \tag{2.23}$$

Now considering body effect with current source $g_{mb}V_{BS}$, we have

$$g_{mb} = \frac{\partial I_D}{\partial V_{BS}} = \begin{cases} 0, & \text{triode} \\ \eta\, g_m = \frac{\gamma}{2\sqrt{2\Phi_F - V_{BS}}} \cdot g_m, & \text{saturation} \end{cases}$$

where η is typically $\frac{1}{4}$. g_{mb} and g_m have the same polarity, indicating that rising V_{BS} and V_{GS} have the similar effect (on small-signal model).

Figure 2.18 illustrates all the effects above[2] And the complete small-signal model is shown in Figure 2.19[3].

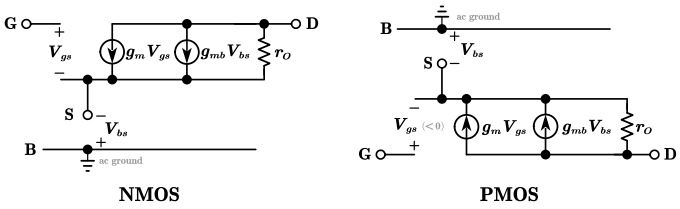


Figure 2.18: MOSFET small-signal models

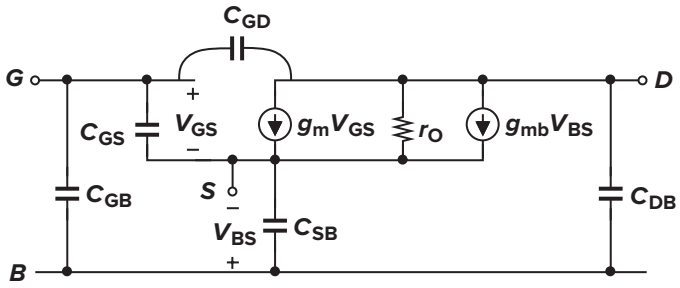


Figure 2.19: Complete MOS small-signal model

2.4.4 MOS SPICE Models (Level 1)

Table 2.1: Level 1 SPICE models for NMOS and PMOS devices

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

Table 2.2: Definition of the parameters in level 1 SPICE model

Parameter	Symbol	Definition	Unit
VTO	V_{TH0}	threshold voltage with zero V_{SB}	V
GAMMA	γ	body-effect coefficient	$V^{\frac{1}{2}}$
PHI	$2\Phi_F$	twice Fermi potential difference	V
TOX	t_{ox}	gate-oxide thickness	m
NSUB	N_{sub}	substrate doping	cm^{-3}
LD	L_D	source/drain side diffusion	m
UO	μ_n or μ_p	channel mobility	$cm^2 \cdot V^{-1} \cdot s^{-1}$
LAMBDA	λ	channel-length modulation coefficient	V^{-1}
CJ	C_j	source/drain bottom-plate junction capacitance per unit area	F/m^2
CJSW	C_{jsw}	source/drain sidewall junction capacitance per unit length	F/m
PB	$V_{\text{built-in}}$	source/drain junction built-in potential	V
MJ	m_j	exponent in CJ equation	unitless
MJSW	m_{jsw}	exponent in CJSW equation	unitless
CGDO	C_{ovGD}	gate-drain overlap capacitance per unit width	F/m
CGSO	C_{ovGS}	gate-source overlap capacitance per unit width	F/m
JS	$\frac{I_{0,\text{sub}}}{A}$	source/drain leakage current per unit area	A/m2

2.4.5 NMOS Versus PMOS Devices

In most CMOS technologies, PMOS devices are quite inferior to NMOS transistors. For example, due to the lower mobility of holes, $\mu_p C_{ox} \approx \mu_n C_{ox}$, yielding low current drive and transconductance.

Moreover, for given dimensions and bias currents, NMOS transistors exhibit a higher output resistance, providing more ideal current sources and higher gain in amplifiers. For these reasons, incorporating NFETs rather than PFETs wherever possible is preferred. And one exception is when flicker noise is critical.

2.4.6 Long-Channel Versus Short-Channel Devices

Most of our treatment in this chapter is valid for **long-channel** devices, i.e., transistors having a minimum length of a few micrometers. Many of the relationships must be revised for short-channel devices. These issues are studied in Chapter 17.

[2] Unless otherwise stated, in the following contents of this book, we assume that the bulk of all NFETs is tied to the most negtive supply (usually GND) and that of PFETs to the most positive supply (usually VDD).

[3] In reality, each terminal of a MOSFET exhibits a finite resistance resulting from the material and contacts, but proper layout can minimize these resistances (such as **folding**, see *Razavi CMOS* page 33)

Chapter 3 Single-Stage Amplifiers

Amplification is an essential function in most analog (and many digital) circuits. In this chapter, we study the low-frequency behavior of single-stage CMOS amplifiers.

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

3.1 Applications

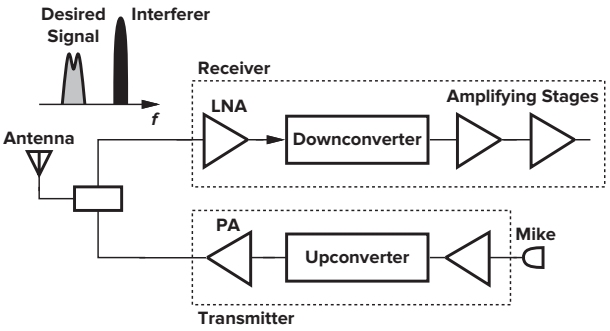


Figure 3.1: General RF transceiver

3.2 General Considerations

An ideal amplifier generates an output $y(t)$ that is a linear replica of the input $x(t)$ with a dc operating point α_0:

$$y(t) = \alpha_0 + \alpha_1 x(t) \tag{3.1}$$

As depicted in Figure 3.2, we can use a polynomial to approximate the nonideal characteristic:

$$y(t) = \alpha_0 + \alpha_1 x(t) + \alpha_2 x^2(t) + \cdots + \alpha_n x^n(t) \tag{3.2}$$

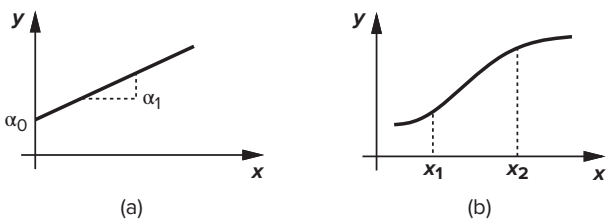


Figure 3.2: Input-output characteristic of a (a) linear and (b) nonlinear system

For an amplifier, in addition to gain and speed, such parameters as power dissipation, supply voltage, linearity, noise, or maximum voltage swings may be important. Furthermore, the input and output impedances determine how the circuit interacts with the preceding and subsequent stages.

Such trade-offs lead to the "analog design octagon" illustrated in Figure 3.3, and present many challenges in the design of high-performance amplifiers, requiring intuition and experience to arrive at an acceptable compromise.

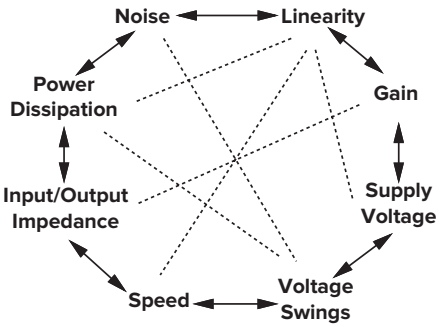


Figure 3.3: Analog design octagon

Figure 3.3 gives a preview of the amplifier topologies that will be studied in this chapter, indicating the much wider use of the common-source (CS) stage than other circuit configurations. For these amplifiers, we must set up proper bias conditions, and analyze the circuit's behavior as the input and output signals cause small or large departures from the bias input (small-signal and large-signal analyses, respectively). We deal with the latter task here and defer the former to Chapter 5.

3.3 Common-Source Stage

3.3.1 Diode-Connected MOSFET

A MOSFET can operate as a small-signal resistor if its gate and drain are shorted, called a **diode-connected MOS**. As depicted in Figure 3.4, the small-signal resistance is:

$$r = \frac{1}{g_m} \parallel r_O \quad \text{(ignoring body effect)} \tag{3.3}$$

we ignore the body effect in the above equation, but bear in mind that the body effect might be significant in some cases.

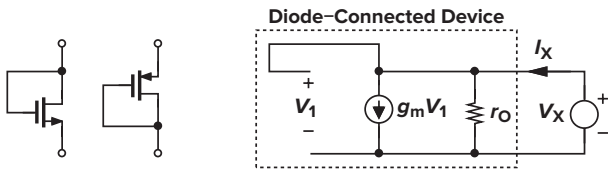


Figure 3.4: Diode-connected MOS

3.3.2 CS stages with various loads

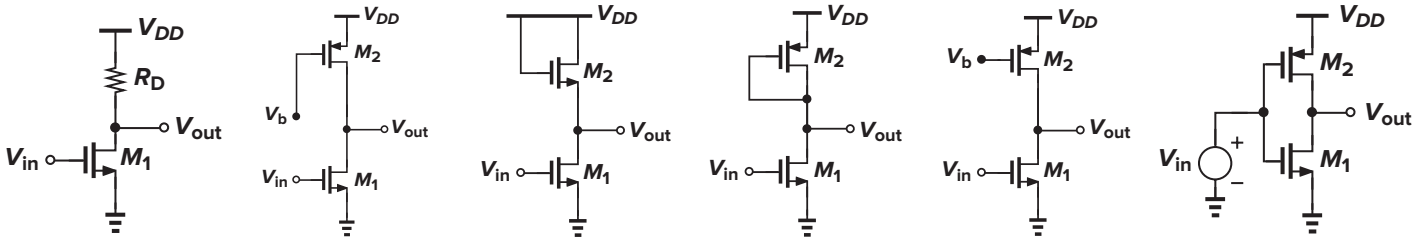
Table 3.1 and Figure 3.5 summarize the five types of common-source stages and their main characteristics. The output resistance of the CS stage can be derived from

$$\boldsymbol{R_{out} = R_L \parallel r_O} \tag{3.4}$$

where R_L denotes the load resistance.

Table 3.1: types of common-source stage

load	R_{out}	G_m	A_v (without λ)	swing u bound (in satu)	note
resistive	$R_D \parallel r_O$	g_m	$-g_m R_D$	$V_{DD} - I_D R_D$	no body effect
triode	$\frac{1}{g_{m2}} \parallel r_{O1}$	g_{m1}	$\frac{g_{m1}}{g_{m2}}$	$V_{DD} - g_{m2} I_D$	similar to resistive
d-c NMOS	$\left(\frac{1}{g_{m2}} \parallel \frac{1}{g_{mb2}} \parallel r_{O2}\right) \parallel r_{O1}$	g_{m1}	$-\frac{1}{1+\eta_2}\sqrt{\frac{a_1}{a_2}}$	$V_{DD} - V_{TH2}$	better linearity
d-c PMOS	$\left(\frac{1}{g_{m2}} \parallel r_{O2}\right) \parallel r_{O1}$	g_{m1}	$-\sqrt{\frac{\mu_n a_1}{\mu_p a_2}}$	$V_{DD} - V_{TH2}$	linearity, no b-e
current source	$r_{O2} \parallel r_{O1}$	g_{m1}	$-g_{m1}(r_{O1} \parallel r_{O2})$	$V_B - V_{TH2}$	large gain
active load	$r_{O2} \parallel r_{O1}$	$g_{m1} + g_{m2}$	$-(g_{m1} + g_{m2})(r_{O1} \parallel r_{O2})$	$V_{IN} - V_{TH2}$	PTV, supply noise



(a) resistive load (b) triode load (c) diode-c NMOS load (d) diode-c PMOS load (e) current source load (f) active load

Figure 3.5: Common-source stage with various loads

3.3.3 CS Stage with Source Degeneration

To improve the linearity in some cases, we add a **degeneration resistor** R_S to the source of the CS stage, as shown in Figure 3.8. The equivalent tansconductance G_m is:

$$G_m = \frac{\partial I_D}{\partial V_{in}} = \frac{\partial I_D}{\partial V_{GS}} \frac{\partial V_{GS}}{\partial V_{in}} \tag{3.5}$$

$$= g_m \frac{\partial (V_{in} - I_D R_S)}{\partial V_{in}} = g_m (1 - G_m R_S) \tag{3.6}$$

$$\Longrightarrow G_m = \frac{1}{\frac{1}{g_m} + R_S} \tag{3.7}$$

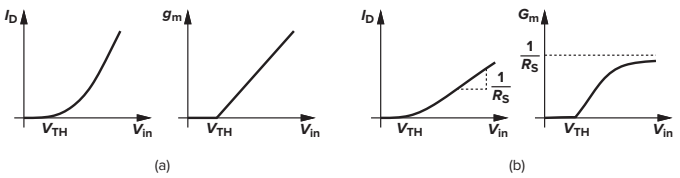


Figure 3.6: I_D and G_m of CS stage (a) without and (b) with source degeneration

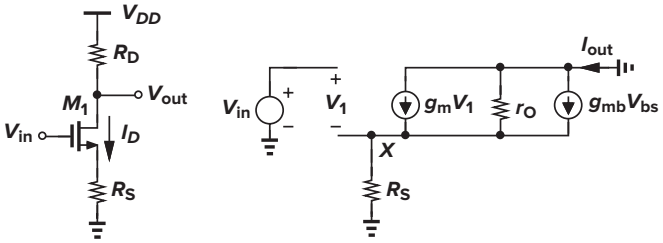
Equation (3.7) can also be derived from the small-signal model ignoring r_O and g_{mb}. With the second-order effects:

$$i_D = g_m v_{gs} - g_{mb} v_{bs} - \frac{i_D R_S}{r_O} \tag{3.8}$$

$$i_D = g_m (v_{in} - i_{out} R_S) - g_{mb} (i_{out} R_S) - \frac{i_D R_S}{r_O} \tag{3.9}$$

$$\Longrightarrow G_m = \frac{i_{out}}{v_{in}} = \frac{1}{\left(1 + \frac{R_S}{r_O}\right)\frac{1}{g_m} + (1+\eta) R_S} \tag{3.10}$$

$$\text{or} \quad G_m = \frac{g_m r_O}{[1 + (g_m + g_{mb}) R_S] r_O + R_S} \tag{3.11}$$



(a) circuit schematic (b) small-signal model

Figure 3.7: Common-source stage with source degeneration

3.3.4 Drain-Source Resistance Approximation

To simplify the analysis, we can approximately view the magnitude of the gain as the ratio of the **drain path resistance** to the **source path resistance** (denoted by R_{source}), given by:

$$|A_v| \approx \frac{R_{\text{drain}}}{R_{\text{source}}} \tag{3.12}$$

Remark that the drain path resistance and source path resistance is different from the output resistance and the resistance seen in source, respectively.

For instance, in Figure 3.8 (a), we have (*CMOS* P.53):

$$R_{\text{source path}} = \frac{1}{g_m} \parallel \frac{1}{g_{mb}} \parallel r_O + R_S \tag{3.13}$$

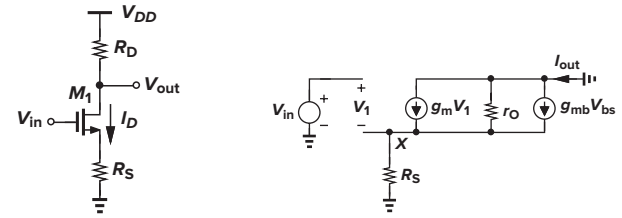
$$R_{\text{seen at source}} = \frac{1}{g_m} \parallel \frac{1}{g_{mb}} \parallel (r_O + R_D) + R_S \tag{3.14}$$

3.3.5 Higher output resistance with degeneration

With source degeneration shown in Figure 3.8, in which the load resistor R_D is excluded, we have:

$$R_O = [1 + (g_m + g_{mb})R_S]\, r_O + R_S \tag{3.15}$$

Note that the overall output resistance is the parallel combination of R_D and R_O, i.e., $R_{out} = R_D \parallel R_O$.



(a) large-signal circuit (b) small-signal model

Figure 3.8: Common-source stage with source degeneration

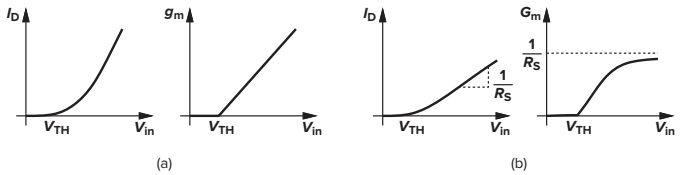
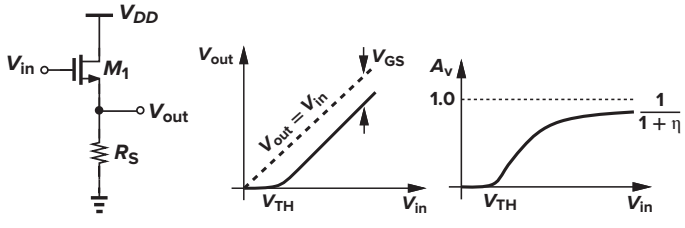


Figure 3.9: I_D and g_m (a) with and (b) without degeneration

Equation 3.15 can also be derived from the resistance looking into the source ($\frac{1}{g_m} \parallel \frac{1}{g_{mb}} \parallel r_O$), see *Razavi CMOS* page 65 for more details.

3.4 Common-Drain Stage

As illustrated in Figure 3.10, the source follower, namely common-drain stage, allows source to "follow" the gate voltage, while presenting a high input impedance, a moderate output impedance and a relatively low voltage gain.



(a) source follower (b) I/O characteristic (c) voltage gain

Figure 3.10: Source follower and its characteristics

The output resistance of the source follower is given by:

$$\boldsymbol{R_{out} = R_L \parallel \left(\frac{1}{g_m} \parallel \frac{1}{g_{mb}} \parallel r_O \right)} \tag{3.16}$$

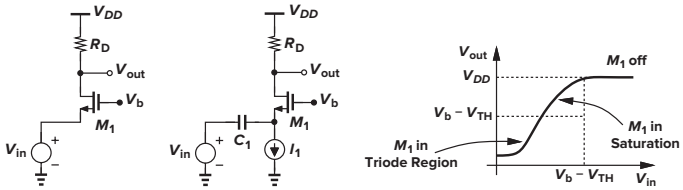
Using a current source as the load, or PMOS with separate n-wells can alleviate the nonlinearity.

However, due to the significant nonlinearity caused by body effect, low voltage gain, and voltage headroom limitation, the source follower is not widely used in practice. As a general rule, we avoid using source follower unless absolutely necessary (such as performing voltage-level shift).

By the way, source follower also introduces substantial noise that it is ill-suited for low-noise applications.

3.5 Common-Gate Stage

Figure 3.11 shows two input modes for the common-gate stage. The second one have the signal capacitively coupled to the circuit, yielding a higher G_m.



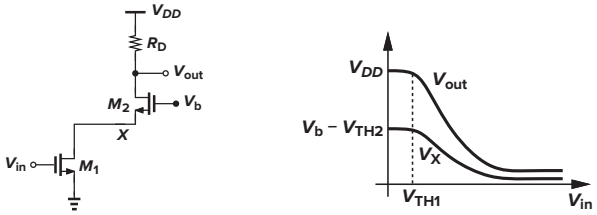
(a) direct input (b) capacitive input (c) I/O characteristics

Figure 3.11: CG stage and its I/O characteristics

Note that body effect increases the equivalent transconductance $[G_m \approx (g_m + g_{mb})R_D]$. And the relatively low input impedance proves useful in some applications.

3.6 Cascode Stage

The cascade of a CS stage and a CG stage is called a **cascode** topology, as shown in Figure 3.12 (also known as the **telescopic cascode**). We call M_1 the input device and M_2 the cascode device.



(a) cascode stage (b) I/O characteristics

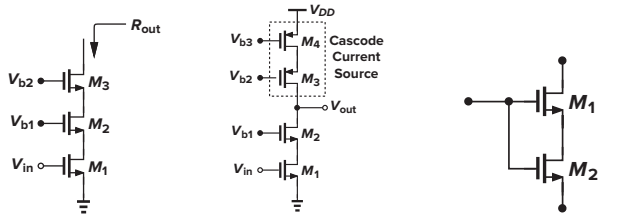
Figure 3.12: Cascode stage and its I/O characteristics

To bias M_1 and M_2 into saturation, we need to set:

$$V_B > V_{IN} + V_{GS2} - V_{TH1} \tag{3.17}$$

$$V_{OUT} > (V_{GS1} - V_{TH1}) + (V_{GS2} - V_{TH2}) \tag{3.18}$$

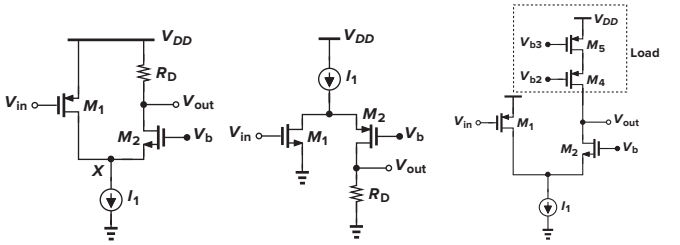
which limits the voltage swing.



(a) triple cascode (b) NMOS cascode with PMOS current source (c) poor man's cascode current source

Figure 3.13: Other cascode types

As shown in Figure 3.13 (c), M_2 is in triode region. And it is equivalent to a single transistor having twice L if M_1 and M_2 have identical dimensions.



(a) PMOS input (b) NMOS input (c) PMOS input with cascode load (current source)

Figure 3.14: Folded cascode with proper bias

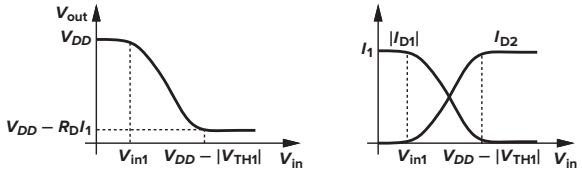


Figure 3.15: Large-signal characteristics of folded cascode

Note that if I_1 is excessively large, M_2 may enter the deep triode region, possibly driving I_1 into the triode region as well.

3.7 Chapter Summary

3.7.1 General voltage gain

Lemma: In a linear circuit we have

$$\boldsymbol{A_v = -G_m R_{out}} \tag{3.19}$$

where G_m denotes the transconductance of the circuit when the output is shorted (to ground) [Fig. 3.32(b)] and R_{out} represents the output resistance of the circuit when the input is set to zero.

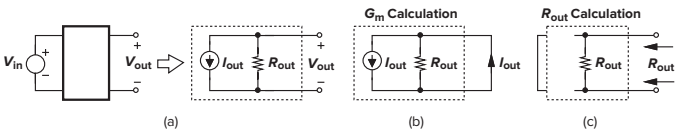


Figure 3.16: (a) Norton equivalent of a linear circuit; (b) G_m calculation; (c) R_{out} calculation

The lemma proves useful when G_m and R_{out} can be determined by inspection. Note the direction of I_{out}.

3.7.2 MOSFET Terminal Resistance

Refer to ZhiHu (https://zhuanlan.zhihu.com/p/20686774030) or YiDing's Website for the derivations.

$$R_{\text{drain}} = \left(1 + \frac{R_S}{R_{S0}}\right) R_{D0}, \quad R_{D0} = r_O$$

$$R_{\text{source}} = \left(1 + \frac{R_D}{R_{D0}}\right) R_{S0}, \quad R_{S0} = \frac{1}{g_m} \parallel \frac{1}{g_{mb}} \parallel r_O$$

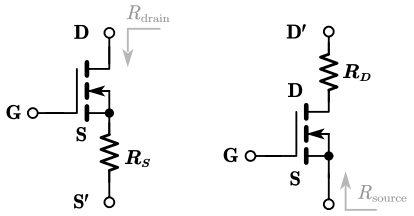


Figure 3.17: MOSFET terminal resistance

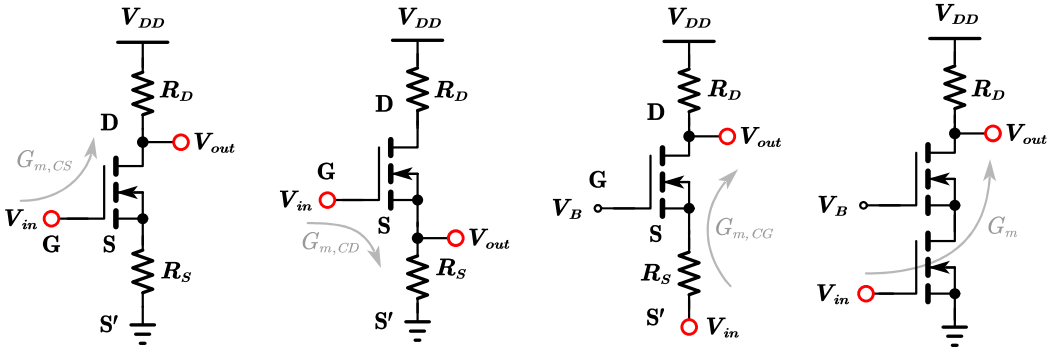
3.7.3 Single Stage G_m and A_v Calculation

Considering the stages shown in Figure 3.18, note that R_D and R_S are included for better universality.

The conclusion of G_m, R_{out} and A_v of the four single stages are summarized in Table 3.2. See ZhiHu (https://zhuanlan.zhihu.com/p/20721118064) or YiDing's Website for the derivations.

Table 3.2: G_m, R_{out} and A_v summary of single stages

Stage	G_m	R_{out}	$A_v = -G_m R_{out}$	Approximation	Condition
CS	$g_m \cdot \frac{r_O}{R_{\text{drain}}}$	$R_D \parallel R_{\text{drain}}$	$-g_m \cdot \frac{r_O R_D}{R_{\text{drain}}+R_D}$	$-g_m (R_D \parallel r_O)$	$R_{S0} \gg R_S$
CD	$-\frac{g_m}{1+\frac{R_D}{r_O}}$	$R_S \parallel R_{\text{source}}$	$g_m \cdot \frac{R_S R_{S0}}{R_S+\left(1+\frac{R_D}{r_O}\right)R_{S0}}$	$g_m (R_S \parallel R_{S0})$	$r_O \gg R_D$
CG	$-\frac{1}{R_S+R_{S0}}$	$R_D \parallel R_{\text{drain}}$	$\frac{R_D r_O}{R_{S0}R_D+(R_{S0}+R_S)r_O}$	$(g_m + g_{mb} + \frac{1}{r_O})(R_D \parallel r_O)$	$R_{S0} \gg R_S$
cascode (CG + CS)	$-\frac{g_{m1}r_{O1}}{r_{O1}+R_{S02}}$	$R_D \parallel R_{\text{drain2}}$	$\frac{(g_{m1}r_{O1})\cdot r_{O2}R_D}{(R_{S02}+r_{O1})r_{O2}+R_{S02}R_D}$	$(g_{m1}r_{O1})(g_{m2}+g_{mb2})(R_D \parallel r_{O2})$	$R_{S02} \gg r_{O1}$
cascode with PMOS load (2 NMOS + 2 PMOS)	same	$R_{\text{drain 3}} \parallel R_{\text{drain 2}}$	too long	$g_{m1} \cdot [(g_{m2}+g_{mb2})r_{O1}\, r_{O2}] \parallel [(g_{m3}+g_{mb3})\, r_{O3}\, r_{O4}]$	$(g_{m2}+g_{mb2}) \gg (\frac{1}{r_{O1}}+\frac{1}{r_{O2}})$ $(g_{m3}+g_{mb3}) \gg (\frac{1}{r_{O3}}+\frac{1}{r_{O4}})$



(a) CS (b) CD (c) CG (d) Cascode

Figure 3.18: Single stages

Chapter 4 Differential Amplifiers

The differential amplifier is one of the most important circuit inventions in analog IC. This chapter deals with the analysis and design of CMOS differential amplifiers, including the basic differential pair, cascode differential pair, and finally the Gilbert cell.

4.1 Single-Ended and Differential

A **single-ended signal** is defined as one that is measured with respect to a fixed reference point (usually GND). By contrast, a **differential signal** is defined as one that is measured bewteen two nodes that have "equal" and "opposite" signal excursions around a fixed potential.

It is usually helpful to view the CM level as the bias value, i.e., the value in the absence of any signal.

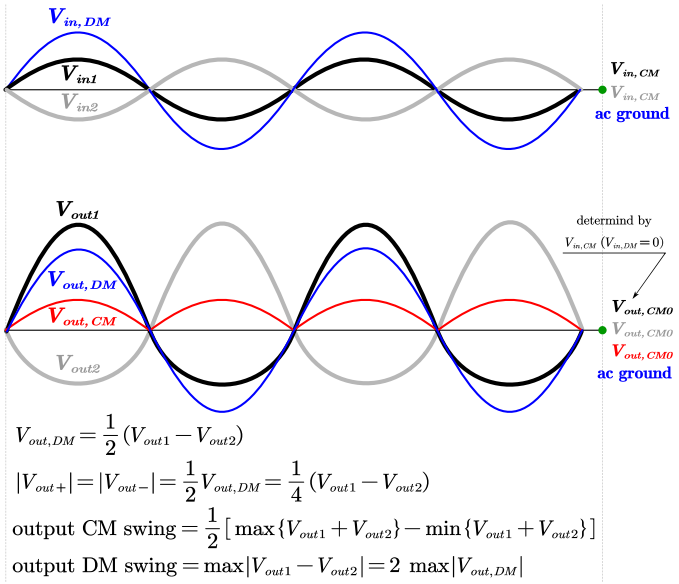


Figure 4.1: Relationship between input and output signals

4.2 Basic Differential Pair

4.2.1 Structure and Operation

The basic differential pair, composed of two identical CS stages, is shown Figure 4.2. Since the output voltage can not exceed V_{DD}, to avoid chopping the output signal, the input CM voltage is bounded as:

$$\begin{cases} V_{in,CM} \geqslant V_{GS1} + (V_{GS3} - V_{TH3}) \\ V_{in,CM} \leqslant \min\left\{V_{DD} - R_D\frac{I_{SS}}{2} + V_{TH},\ V_{DD}\right\} \end{cases} \tag{4.1}$$

Input CM voltage also affects the small-signal gain A_v and the output swing. To ensure M_1 and M_2 operate in saturation, the single-ended output range is given by:

$$V_{out1} \in [V_{in,CM} - V_{TH1},\ V_{DD}],\ \text{yielding} \tag{4.2}$$

$$\text{small-signal swing} < (V_{DD} - V_{in,CM} + V_{TH1}) \tag{4.3}$$

In essence, increasing $V_{in,CM}$ yields lower $V_{out,CM}$ and lower output DM swing. Therefore, it is desirable to choose a relatively low $V_{in,CM}$ (in basic differential pair), but not less than $V_{GS1} + (V_{GS3} - V_{TH3})$ (output DM swing is also limited by $V_{out,CM}$).

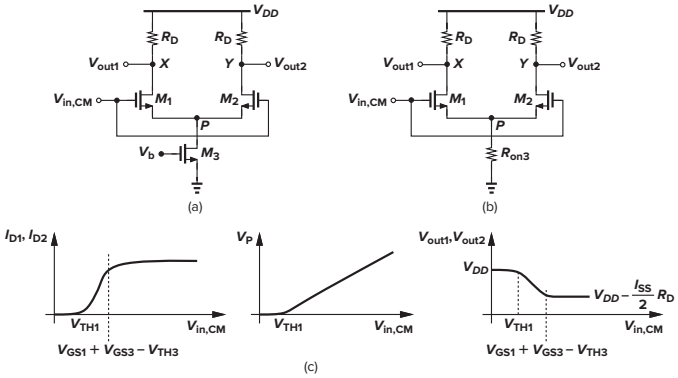


Figure 4.2: Basic differential pair
(a) circuit schematic
(b) M_3 in deep triode ($V_{in,CM} < V_{TH1}$)
(c) CM input-output characteristics

4.2.2 Large-Signal Analysis

Now we examine the large-signal behavior of the basic differential pair (assuming $\lambda = \gamma = 0$). In saturation, we have $V_{GS} = \sqrt{\frac{2I_D}{\mu_n C_{ox}\frac{W}{L}}} - V_{TH}$, yielding:

$$V_{in1} - V_{in2} = \sqrt{\frac{2I_{D1}}{\mu_n C_{ox}\frac{W}{L}}} - \sqrt{\frac{2I_{D2}}{\mu_n C_{ox}\frac{W}{L}}} + V_{TH} \tag{4.4}$$

$$\frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{in1} - V_{in2})^2 = I_{SS} - 2\sqrt{I_{D1}I_{D2}} \tag{4.5}$$

applying $4I_{D1}I_{D2} = I_{SS}^2 - (I_{D1} - I_{D2})^2$, obtianing:

$$(I_{D1} - I_{D2}) = \sqrt{\mu_n C_{ox}\frac{W}{L}I_{SS}\left(1 - \frac{\mu_n C_{ox}\frac{W}{L}}{4I_{SS}}(V_{in1} - V_{in2})^2\right)} \cdot (V_{in1} - V_{in2}) \tag{4.6}$$

In this case, G_m can be derived as:

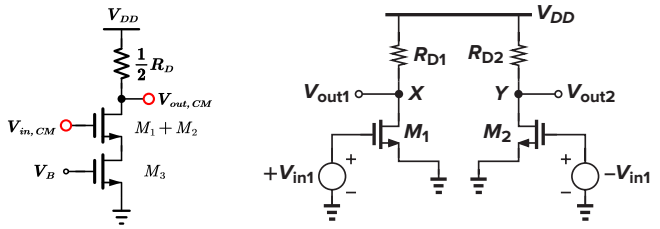
$$G_m = \frac{\partial \Delta I_D}{\partial \Delta V_{in}} = \frac{1}{2}\mu_n C_{ox}\frac{W}{L} \cdot \frac{\frac{4I_{SS}}{\mu_n C_{ox}\frac{W}{L}} - 2\Delta V_{in}^2}{\sqrt{\frac{4I_{SS}}{\mu_n C_{ox}\frac{W}{L}} - \Delta V_{in}^2}} \tag{4.7}$$

$$G_m|_{\Delta V_{in}=0} = \sqrt{\mu_n C_{ox}\frac{W}{L}I_{SS}} = g_m \tag{4.8}$$

4.2.3 Small-Signal Analysis (DM)

In small-signal analysis, we assume that the variation in the input signal is small enough (not changing the operation point). Thus, the input signal can be divided into two parts: CM signal and DM signal.

According to the symmetry of the circuit, we can combine the two resistances and two transistors into a single one, yielding the equivalent circuit as Figure 4.3 (a).



(a) A_{CM} calculation **(b) A_{DM} calculation**

Figure 4.3: Equivalent circuits of basic differential pair for small-signal gain calculation

It can be proved that:

$$A_{CM} = -g_{m12} \cdot \frac{r_{O12}(\frac{1}{2}R_D)}{R_{\text{drain12}+(\frac{1}{2}R_D)}} \tag{4.9}$$

$$R_{\text{drain12}} = \left(1 + \frac{r_{O3}}{R_{S012}}\right) r_{O12} \tag{4.10}$$

$$= \left[1 + 2\left(g_{m1} + g_{mb1} + \frac{1}{r_{O1}}\right) r_{O3}\right] \frac{r_{O1}}{2} \tag{4.11}$$

note that $g_{m12} = 2\,g_{m1}, g_{mb12} = 2\,g_{mb1}$ and $r_{O12} = \frac{1}{2}r_{O1}$. Assuming ... yields:

$$A_{CM} \approx \tag{4.12}$$

4.2.4 Small-Signal Analysis (CM)

To obtain A_{DM}, we first introduce a lemma. Then, applying the **half-circuit concept**, we obtain the equivalent circuit shown in Figure fig: equivalent circuits of basic differential pair for small-signal gain calculation (b). Therefore, we have:

$$A_{DM} = -g_{m1}\,(R_D \parallel r_{O1}) \tag{4.13}$$

Lemma: Consider the symmetric circuit shown below, where D1 and D2 represent any three-terminal active device. Suppose V_{in1} and V_{in2} change differentially, i.e., the former increasing ΔV_{in} and the latter decreasing ΔV_{in}. Then, if the circuit remains linear, V_P does not change.

Figure 4.4

Note that the lemma is valid even if the tail current source is not ideal.

4.2.5 Degenerated Differential Pair

4.2.6 Mismatch

Reference

[1] Behzad Razavi. *Fundamentals of Microelectronics*. University of California Press, 2nd edition, 2014.

[2] Behzad Razavi. *Design of Analog CMOS Integrated Circuits*. McGraw-Hill Education, 2nd edition, 2017.