MICROELECTRONICS I COURSE NOTES

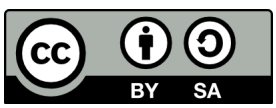
Spring 2020

Chris Winstead
Associate Professor
Electrical and Computer Engineering
chris.winstead@usu.edu

UtahStateUniversity

PUBLISHED BY UTAH STATE UNIVERSITY
DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

http://www.ece.usu.edu



Contents

List of Netlists

List of Examples

List of EveryCircuit Demonstrations

Introduction

Signal sources

Electronic circuits and systems can be loosely divided into two classes: those that process *signals*, which are electrical representations of information, and those that convey or convert electrical power. In this course we are primarily concerned with circuits that process signals, in the broadest possible sense. A signal may convey physical information, e.g. an audio signal produced from a microphone, or it may convey discrete or digital information as part of a computational process. Regardless of the context, we will always view signals as electrical information, either a current or voltage.

A **transducer** is a device that converts a physical signal into an electrical one. From the circuit perspective, we usually *model*[1] a transducer as either a voltage source or a current source. Every signal source has an associated internal impedance, represented by Thévenin or Norton equivalent circuits.

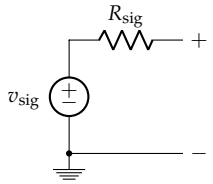


Figure 1: Thévenin equivalent voltage signal source.

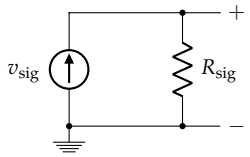


Figure 2: Norton equivalent current signal source.

[1] A **model** is a useful approximation of a physical device or system. We use models to simplify our understanding of complex electronic components.

Spectrum and Frequency Response

Most information-bearing signals are not constant. A signal $x(t)$ that changes over time can be represented as a superposition of sinusoidal signals with various magnitudes, frequencies and phases. The function that characterizes the magnitudes and phases at each frequency is called the signal's complex-valued *Fourier spectrum*, written $X(j\omega)$. The theory of spectral transforms is quite sophisticated, but in this course we mostly require a simplified version known as the *steady-state*. For our purposes, we may consider the *Laplace transform* $X(s)$ to be equivalent to the Fourier spectrum, with $s = j\omega$.

On a modern digital oscilloscope, a signal's spectrum can be viewed by selecting a **Fast Fourier Transform (FFT)** display,

which reports the signal's **magnitude spectrum**, equal to the complex magnitude $|X(j\omega)|$. Since $X(j\omega)$ is a complex function,

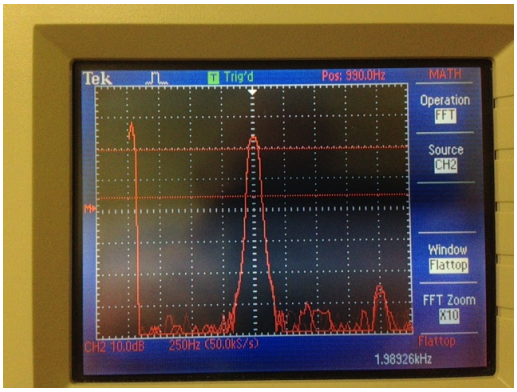


Figure 3: Example FFT display on an oscilloscope.

the magnitude is obtained as

$$|X(j\omega)|^2 = X(j\omega) \times X^*(j\omega).$$

The magnitude spectrum is typically expressed in units of **decibels**, and the complex phase $\angle X(j\omega)$ is usually expressed in **degrees** (0° to 360°).

The individual sinusoidal signal components are expressed as

$$v_a(t) = \mathcal{V}_\mathcal{A} \sin(\omega_0 t + \phi),$$

where $\mathcal{V}_\mathcal{A}$ is the **zero-to-peak amplitude**, ω_s is the signal frequency in radians per second, t is the time in seconds, and ϕ is the phase-shift in radians.

A pure or single-tone sinusoid has a magnitude spectrum represented by a single impulse function

$$V_A(\omega) = \frac{1}{2}\mathcal{V}_\mathcal{A}\delta(\omega - \omega_s).$$

The impulse height in decibels is $20\log_{10}(\mathcal{V}_\mathcal{A}/2)$. So a zero-to-peak magnitude of 1 V corresponds to -6 dB, 10 V corresponds to 14 dB, 100 V corresponds to 34 dB, and so on.

When signals pass through electronic circuits, their magnitude and phase are altered. The circuit's **transfer function** is the ratio of the output spectrum to the input spectrum. If a circuit's input is $X(s)$ and its output is $Y(s)$, then the transfer function is

$$H(s) = \frac{Y(s)}{X(s)}.$$

We may interpret the transfer function in terms of frequency by substituting $s = j\omega$. The transfer function's magnitude response is usually expressed in decibels as

$$\begin{aligned} |H(\omega)| \text{ (dB)} &= 20\log_{10}|H(\omega)| \\ &= 10\log_{10}|H(\omega)|^2. \end{aligned}$$

When expressed in decibels, the magnitude reveals useful information about the circuit. When $|H(\omega)| = 0$ dB, the output's amplitude is equal to the input. When $|H(\omega)|$ is **positive**, the output's amplitude is greater than the input, and the circuit is said to have **gain**. When $|H(\omega)|$ is **negative**, the output's amplitude is less than the input, and the circuit is said to **attenuate** the signal.

For a *linear circuit*, the transfer function is obtained using **complex impedances** for capacitors and inductors. A capacitor with capacitance C has impedance $1/(sC)$, and an inductor with

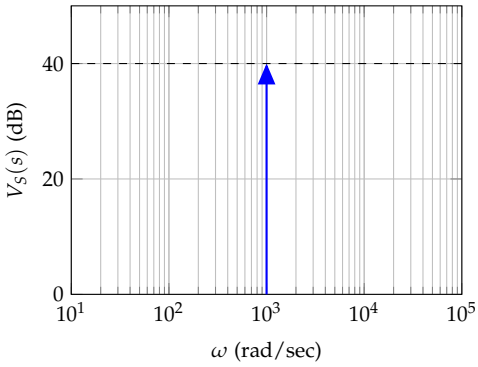


Figure 4: A sinusoid has a single Fourier component that appears as an impulse function on the spectral representation. In this case the magnitude is 40 dB, which corresponds to a time-domain zero-to-peak amplitude of 200 V.

Frequency Units:

- $\omega = 2\pi f$
- f is in Hz (cycles per second)
- ω is in radians per second.
- ω is "omega", not w.
- A magnitude V in dB is $20\log_{10}(V)$
- A power P in dB is $10\log_{10}(P)$.

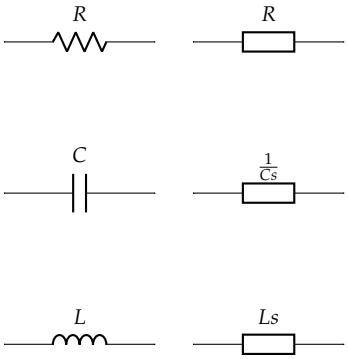


Figure 5: Passive linear components and their equivalent Laplace-domain impedances.

inductance L has impedance sL. Once components are replaced by their equivalent impedances, they can be analyzed as though they were resistors where the resistance values are polynomials in s.

Example 1 (Low-pass RC circuit).

The low-pass configuration is like a simple voltage divider. The impedances are $Z_1 = R$ and $Z_2 = 1/sC$. Then

$$V_{OUT}(s) = V_{IN}(s)\left(\frac{Z_2}{Z_1 + Z_2}\right)$$
$$\Rightarrow H(s) = \frac{V_{OUT}(s)}{V_{IN}(s)} = \frac{1/sC}{R + 1/sC}$$
$$= \frac{1}{1 + sRC}$$

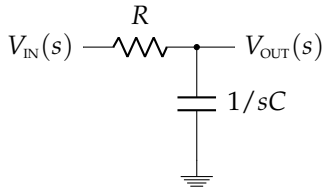


Figure 6: Low-pass configuration.

The low-pass transfer function is commonly represented as

$$H(j\omega) = \frac{1}{1 + j\omega/\omega_{3dB}}$$

. The magnitude response is then

$$|H(\omega)|^2 = \left(\frac{1}{1 + j\omega/\omega_{3dB}}\right)\left(\frac{1}{1 - j\omega/\omega_{3dB}}\right)$$
$$= \left(\frac{1}{1 + \omega^2/\omega_{3dB}^2}\right)$$

At **low frequencies** where $\omega \ll \omega_{3dB}$, the **magnitude response is flat**, approximately equal to one. At higher frequencies where $\omega \gg \omega_{3dB}$, the magnitude drops rapidly. At these high frequencies, since $\omega/\omega_{3dB} \gg 1$, we can make an approximation:

$$\omega/\omega_{3dB} + 1 \approx \omega/\omega_{3dB}$$
$$\Rightarrow |H(\omega)| \approx \omega_{3dB}/\omega \text{ (high frequencies above } \omega_{3dB})$$

When represented **in decibels**, we find that

$$|H(\omega)| \approx 20\log_{10}\left(\frac{\omega_{3dB}}{\omega}\right)$$
$$= 20\log_{10}\omega_{3dB} - 20\log_{10}\omega.$$

So as ω increases, the magnitude **decreases by 20 dB per decade**.

Example 2 (High-pass RC circuit).

The high-pass configuration has impedances are $Z_1 = 1/sC$ and $Z_2 = R$. Then

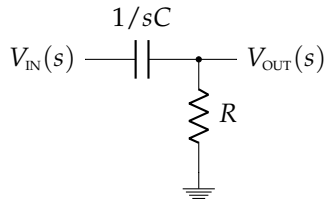


Figure 7: High-pass configuration.

$$V_{\text{OUT}}(s) = V_{\text{IN}}(s)\left(\frac{Z_2}{Z_1 + Z_2}\right)$$
$$\Rightarrow H(s) = \frac{V_{\text{OUT}}(s)}{V_{\text{IN}}(s)} = \frac{R}{R + 1/sC} = \frac{sRC}{1 + sRC}$$

The high-pass transfer function is commonly represented as

$$H(j\omega) = \frac{j\omega/\omega_{3\text{dB}}}{1 + j\omega/\omega_{3\text{dB}}}$$

. The magnitude response is then

$$|H(\omega)|^2 = \left(\frac{j\omega/\omega_{3\text{dB}}}{1 + j\omega/\omega_{3\text{dB}}}\right)\left(\frac{-j\omega/\omega_{3\text{dB}}}{1 - j\omega/\omega_{3\text{dB}}}\right) = \left(\frac{\omega^2/\omega_{3\text{dB}}^2}{1 + \omega^2/\omega_{3\text{dB}}^2}\right)$$

At **high frequencies** where $\omega \gg \omega_{3\text{dB}}$, the magnitude response is flat, approximately equal to one. At lower frequencies where $\omega \ll \omega_{3\text{dB}}$, the magnitude drops rapidly. Since $\omega/\omega_{3\text{dB}} \ll 1$, we can make an approximation:

$$\omega/\omega_{3\text{dB}} + 1 \approx 1$$
$$\Rightarrow |H(\omega)| \approx \omega/\omega_{3\text{dB}} \text{ (low frequencies below } \omega_{3\text{dB}})$$

When represented **in decibels**, we find that

$$|H(\omega)| \approx 20\log_{10}\left(\frac{\omega}{\omega_{3\text{dB}}}\right) = 20\log_{10}\omega - 20\log_{10}\omega_{3\text{dB}}.$$

So as ω increases from very low frequencies, the magnitude **increases by 20 dB per decade**.

A note on approximations: in these examples we used a very common method of **large-value approximation**. We will use this procedure many times. Suppose two quantities A and B differ greatly in value, so that $A \gg B$. The notion of "much greater than" is somewhat fuzzy, but in this course we will define it as more than a $10\times$ difference between two quantities.

If $A \gg B$ then:

$$A + B \approx A$$
$$\frac{1}{A} + \frac{1}{B} \approx \frac{1}{B}$$
$$\frac{A}{A + B} \approx 1$$
$$\frac{B}{A + B} \approx \frac{B}{A}$$

and so on...

Ideal Amplifier Models

An **ideal linear amplifier** is a circuit which receives an input signal X and produces an output signal $Y = AX$. In other words, the output is larger than the input by a constant multiple A, called the **gain**. The input/output signals can be either current or voltage, which introduces four possible amplifier configurations:

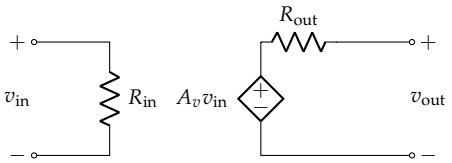


Figure 8: Ideal linear voltage amplifier model at low or mid-band frequencies.

Input	Output	Amplifier Type	Gain Name and Symbol
Voltage	Voltage	Voltage Amplifier	Gain A_v
Current	Current	Current Amplifier	Gain A_i
Voltage	Current	Transconductance Amplifier	Transconductance G_m
Current	Voltage	Transresistance Amplifier	Transresistance R_m

In order to use an amplifier, it has to be connected to its signal source on the input side and its **load** on the output side. This creates a **coupling interaction** between the amplifiers internal resistances and the neighboring signal resistances. In the voltage amplifier, we see a **voltage-divider** effect at both the input and output interfaces:

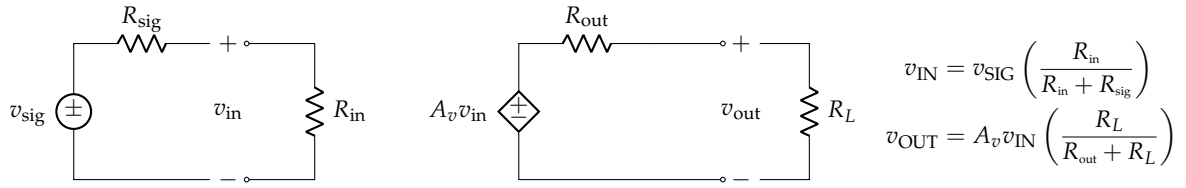


Figure 9: Coupling interactions in voltage amplifiers. Resistive voltage dividers appear at the input and output interfaces.

As a result the complete system is described by the gain equation in combination with the **coupling divider ratios**. To describe this effect, we distinguish the **open-circuit gain** from the **loaded gain**:

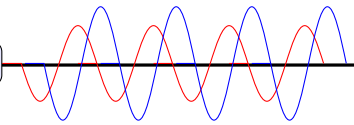
$$\text{open-circuit gain: } A_{vo} \triangleq \frac{v_{OUT}}{v_{IN}} = A_v$$

$$\text{loaded gain: } A_{vL} \triangleq \frac{v_{OUT}}{v_{IN}} = A_v \left(\frac{R_{in}}{R_{in}+R_{sig}}\right)\left(\frac{R_L}{R_{out}+R_L}\right).$$

To maximize the amplifier's gain, we want to eliminate the coupling ratios by making them very close to one. This is achieved when the amplifier has **large input resistance** and a **small output resistance**.

Maximum gain in voltage amp: $R_{out} \ll R_L$ and $R_{in} \gg R_{sig}$. In the limit, a truly ideal voltage amp has $R_{in} \to \infty$ and $R_{out} \to 0$.

EveryCircuit Demonstration 2 (Ideal Voltage Amplifier).



Exercise: Increase R_{in} to 10 kΩ and then 100 kΩ, and observe what happens to the amplitude of v_{in} compared to v_{sig} for these values. Then do the same for R_L. You should notice that the coupling effects disappear when $R_L \gg R_{out}$ and $R_{in} \gg R_{sig}$. Verify that your observations match the value of the loaded gain predicted by our analysis in this section.

The **ideal linear current amplifier** is very similar to the voltage amplifier, except that we get current dividers instead of voltage dividers at the input and output terminals. In a current divider, the opposite resistance appears in the numerator, so the conditions for achieving maximum are reversed.

For current amplifiers, the most ideal gain is called the **short-circuit gain** A_{is}, since we can eliminate the coupling ratios by setting R_L to zero, hence short-circuiting the output. The gain expressions for a current amplifier are:

$$\text{short-circuit gain: } A_{is} \triangleq \frac{i_{\text{OUT}}}{i_{\text{IN}}} = A_i$$

$$\text{loaded gain: } A_{iL} \triangleq \frac{i_{\text{OUT}}}{i_{\text{IN}}} = A_i \left(\frac{R_{\text{sig}}}{R_{\text{in}} + R_{\text{sig}}} \right) \left(\frac{R_{\text{out}}}{R_{\text{out}} + R_L} \right).$$

To maximize the current amplifier's gain, we want to eliminate the coupling ratios by making them very close to one. This is achieved when the amplifier has **small input resistance** and a **large output resistance**, the opposite of what we found for voltage amplifiers.

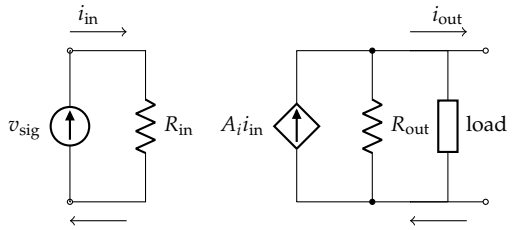
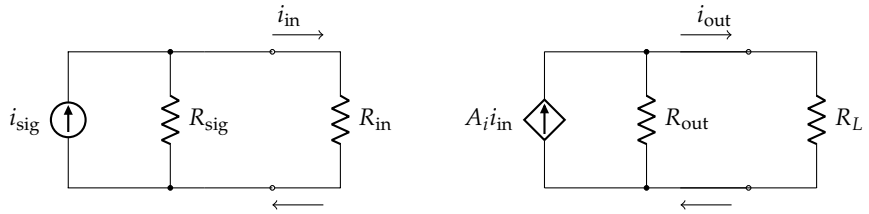


Figure 10: Ideal linear current amplifier model at low or mid-band frequencies.

Maximum gain in current amp: $R_{\text{out}} \gg R_L$ and $R_{\text{in}} \ll R_{\text{sig}}$. In the limit, a truly ideal current amp has $R_{\text{in}} \to 0$ and $R_{\text{out}} \to \infty$.



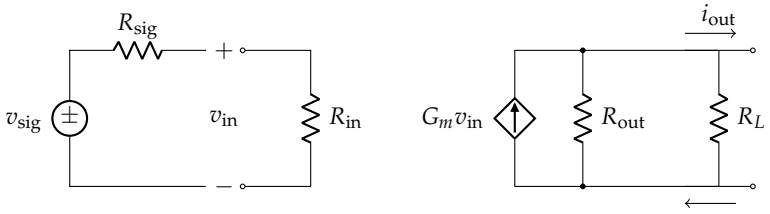
$$i_{\text{IN}} = v_{\text{SIG}} \left(\frac{R_{\text{sig}}}{R_{\text{in}} + R_{\text{sig}}} \right)$$

$$i_{\text{OUT}} = A_i i_{\text{IN}} \left(\frac{R_{\text{out}}}{R_{\text{out}} + R_L} \right)$$

Figure 11: Coupling interactions in current amplifiers. Resistive current dividers appear at the input and output interfaces.

The remaining amplifier types are mixtures of voltage and current amplifiers. The **transconductance amplifier** takes voltage input and delivers a current output. Then we see a voltage divider at the input interface and a current divider at the output interface/

Transconductance amplifiers are especially important since they are the basis of transistor device models.



$$v_{IN} = v_{SIG}\left(\frac{R_{in}}{R_{in}+R_{sig}}\right)$$
$$i_{OUT} = G_m v_{IN}\left(\frac{R_{out}}{R_{out}+R_L}\right)$$

Figure 12: Coupling interactions in transconductance amplifiers. A resistive voltage divider appears at the input interface and a current divider at the output interface.

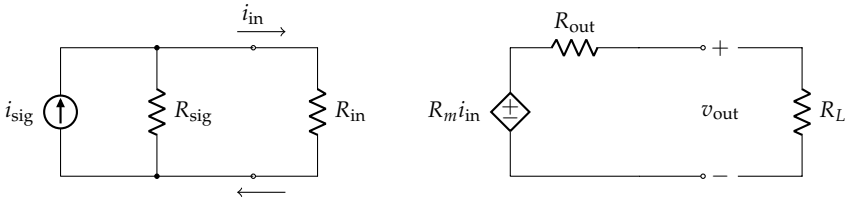
For the transconductance amplifier, the gain expressions are:

$$\text{short-circuit gain: } G_{ms} \triangleq \frac{i_{OUT}}{v_{IN}} = G_m$$
$$\text{loaded gain: } G_{mL} \triangleq \frac{i_{OUT}}{v_{IN}} = G_m\left(\frac{R_{in}}{R_{in}+R_{sig}}\right)\left(\frac{R_{out}}{R_{out}+R_L}\right).$$

Maximum gain in transconductance amp: $R_{out} \gg R_L$ and $R_{in} \gg R_{sig}$. In the limit, a truly ideal transconductance amp has $R_{in} \to \infty$ and $R_{out} \to \infty$.

To maximize the transconductance amplifier's gain, we want to eliminate the coupling ratios by making them very close to one. This is achieved when the amplifier has **large input resistance** and a **large output resistance**.

Lastly, The transresistance amplifier takes current input and delivers a voltage output. Then we see a current divider at the input interface and a voltage divider at the output interface/



$$i_{IN} = v_{SIG}\left(\frac{R_{sig}}{R_{in}+R_{sig}}\right)$$
$$v_{OUT} = R_m i_{IN}\left(\frac{R_L}{R_{out}+R_L}\right)$$

Figure 13: Coupling interactions in transresistance amplifiers. A resistive current divider appears at the input interface and a voltage divider at the output interface.

For the transresistance amplifier, the gain expressions are:

$$\text{open-circuit gain: } R_{mo} \triangleq \frac{v_{OUT}}{i_{IN}} = R_m$$
$$\text{loaded gain: } R_{mL} \triangleq \frac{v_{OUT}}{i_{IN}} = R_m\left(\frac{R_{sig}}{R_{in}+R_{sig}}\right)\left(\frac{R_L}{R_{out}+R_L}\right).$$

Maximum gain in transresistance amp: $R_{out} \ll R_L$ and $R_{in} \ll R_{sig}$. In the limit, a truly ideal transconductance amp has $R_{in} \to 0$ and $R_{out} \to 0$.

To maximize the transresistance amplifier's gain, we want to eliminate the coupling ratios by making them very close to one. This is achieved when the amplifier has **small input resistance** and a **small output resistance**, the opposite of what we found for transconductance amplifiers.

Real Amplifiers

Real amplifiers are affected by **nonlinear transfer characteristics** between the input and the output.

The ideal transfer characteristic is a straight line, extending from $-\infty$ to $+\infty$ with a constant slope. The gain of this amplifier is the slope of its transfer characteristic:

$$\text{Gain} \triangleq \frac{dv_{\text{OUT}}}{dv_{\text{IN}}}.$$

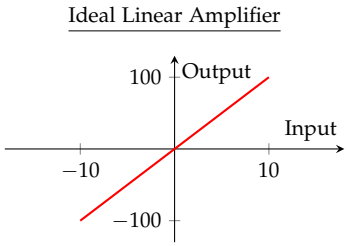


Figure 14: DC transfer characteristic of an ideal amplifier.

Real amplifiers do not exhibit such ideal behavior. A more realistic transfer characteristic is a curve that **saturates** at maximum and minimum values of v_{OUT}, with a non-constant slope in between.

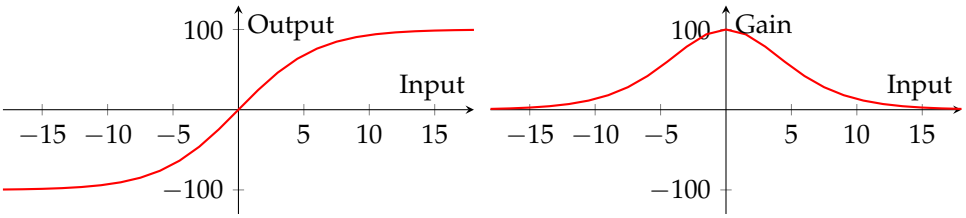


Figure 15: Non-linear transfer characteristic showing non-constant slope. The amplifier saturates when the gain falls below 1 V/V.

Since the slope varies, the gain is non-constant. This introduces **distortion** into the signal being amplified. Due to this nonlinear behavior, we are unable to use linear circuit methods to analyze the amplifier system. As a result, analysis and design can become very complex tasks. To simplify our understanding of nonlinear systems, we rely on the concepts of **linearization** and **small-signal analysis**.

A linearized model is a direct application of the first-order Taylor series approximation. For a non-linear function $f(x)$, the Taylor approximation is defined around an **offset** x_0 as

$$\begin{aligned} f(x) &\approx f(x_0) + (x - x_0) \left.\frac{\partial f}{\partial x}\right|_{x_0}. \\ &= f_0 + A(x - x_0) \end{aligned}$$

This approximation is only valid for small variations, i.e. when $|x - x_0|$ is small (the meaning of "small" here is fuzzy; the variation is considered small enough if the approximation is sufficiently accurate for our needs). The Taylor approximation can be interpreted as zooming-in on the original function, such that the zoomed portion is a nearly straight line:

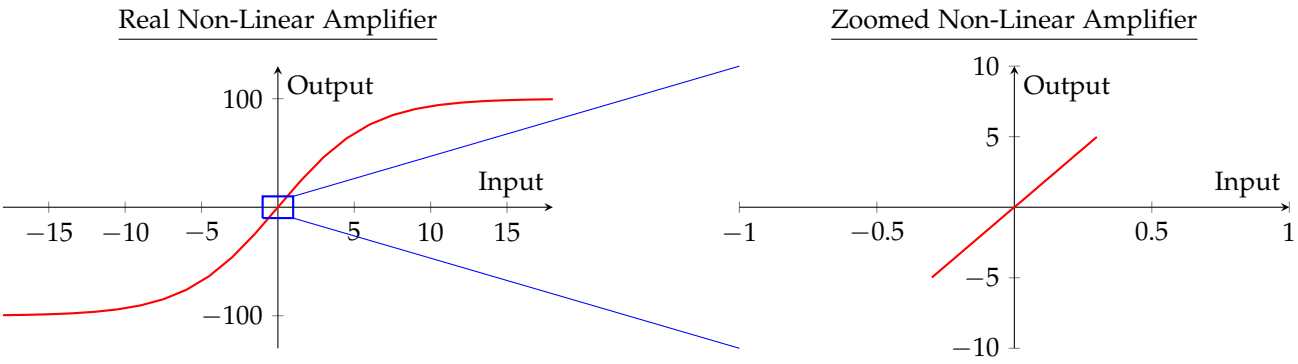


Figure 16: Zoomed transfer characteristic showing approximately linear behavior for small signal variations.

Small-signal equivalent circuit models

The Taylor linearization reveals an extremely useful aspect of linearized circuits: thanks to the principle of **superposition**, we can separate the circuit's behavior into two parts: the **DC offset** or **bias point** x_0, f_0 and the **small signal** variation $(x - x_0)$. In the circuit context, the transfer characteristic shows the **large-signal** relationship between two signals v_{IN} and v_{OUT}. We refer to these as the **total instantaneous signals**, i.e. the precise physical signal value at an instant in time.

For non-linear circuits, it is often difficult to analyze the total instantaneous signal, so we split it into a superposition of two parts:

DC offset – the central or average value of a signal; what you would measure on an oscilloscope as the signal's `MEAN`. We write DC offsets using **all capital letters**, as in V_{IN} or $VOUT$.

Small-signal – the amount by which the signal varies from the offset; what you would measure on an oscilloscope set to `AC Coupling`. We write small-signal quantities in all-lowercase, as in v_{in} or v_{out}.

Total instantaneous signal – the superposition of the offset and small signal; what you would measure on an oscilloscope set to `DC coupling`. We write the total instantaneous signal using lowercase letters with uppercase subscripts, as in v_{IN} or v_{OUT}.

The uppercase/lowercase notation is useful to keep track of our separate analysis domains, but is not entirely perfect. For example, we also use uppercase symbols to represent sinusoidal amplitudes, which can sometimes create ambiguity. To help distinguish these quantities, we will try and use calligraphic font for sinusoidal amplitudes, as in $\mathcal{V}_{\mathcal{A}}$.

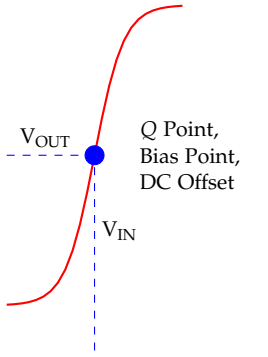
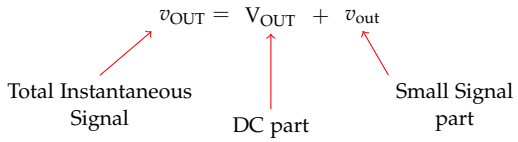


Figure 17: Offset point of a non-linear transfer characteristic.



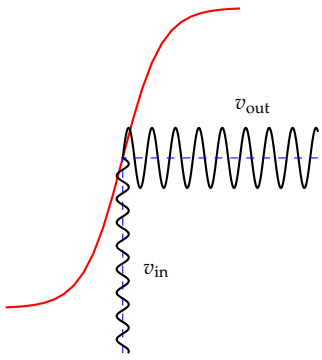


Figure 18: Small-signal activity overlaid on the nonlinear transfer characteristic.

Procedure for small-signal analysis

When analyzing a linearized circuit, we often want to analyze **just the signals**, without being distracted by their DC offsets. We want to know, for example, the AC amplitude and phase shift of signals at various points in a circuit. The principle of superposition allows us to extract the small-signal behavior by following these steps:

1. Solve the circuit's **DC operating point**. In many cases we may only need to find part of the DC solution in order to do the next step.
2. **Linearize** the circuit by applying a Taylor approximation centered at the DC operating point.
3. Replace any non-linear components with their linearized equivalents.
4. Set all **DC independent sources** (both current and voltage) **to zero**. Voltage sources become short-circuits, and current sources become open-circuits. Note: do not modify any time-varying or dependent sources.

Example 3 (Linearization of a sensor).

A temperature sensor provides a change of 2mV per °C, connected to a load of 10kΩ. The output changes by 10mV when T is changed by 10°C. What is the source resistance of the sensor?
The sensor model is linearized:

$$v_s = V_S + \left.\frac{dv_S}{dT}\right|_{T_0} \Delta T$$

where T_0 is the reference temperature and ΔT is the variation from that temperature. To consider only the variation in v_{OUT}, we isolate the small signal portion:

$$v_{out} = \left.\frac{dv_S}{dT}\right|_{T_0} \Delta T$$

The problem statement tells us that

$$\left.\frac{dv_S}{dT}\right|_{T_0} = 2\text{mV}/^\circ\text{C}$$

It also tells us that $v_{\text{out}} = 10\,\text{mV}$, so we can solve for R_S:

$$\begin{aligned} v_{\text{out}} &= v_s \frac{R_L}{R_L + R_S} \\ &= (2\,\text{mV}/^\circ\text{C})\,(10\,^\circ\text{C})\,\frac{R_L}{R_L + R_S} \\ \rightarrow R_S &= \frac{(2\,\text{mV}/^\circ\text{C})\,(10\,^\circ\text{C})\,R_L}{v_{\text{out}}} - R_L \\ &= 10\,\text{k}\Omega \end{aligned}$$

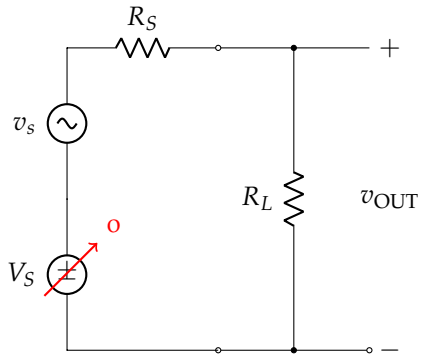


Figure 19: Linear temperature sensor model. The DC offset V_S is set to zero (i.e. shorted out) for small-signal analysis.

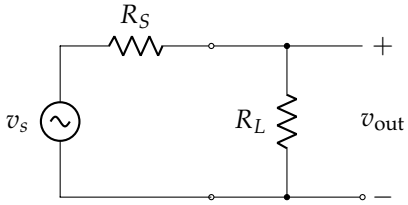


Figure 20: Small-signal equivalent temperature sensor model. The lower-case signals v_s and v_{out} represent the *variations* in the corresponding physical signals.

Example 4 (Linearization of a thermistor model)**.**

A **thermistor** is modeled by the Steinhart-Hart equation:

$$R = R_0 e^{-B\left(\frac{1}{T_0} - \frac{1}{T}\right)}$$

where R_0 and T_0 are reference measurements, T and T_0 are in Kelvin, and B is a device-specific parameter. For **small temperature changes** (e.g. changes in a room's temperature), we can approximate this using a linearized model centered around T_0:

$$\begin{aligned} R &\approx R_0 + \Delta T \left. \frac{d}{dT}\left(R_0 e^{-B\left(\frac{1}{T_0} - \frac{1}{T}\right)}\right)\right|_{T=T_0} \\ &= R_0 + \Delta T \left(R_0 e^{-B\left(\frac{1}{T_0} - \frac{1}{T}\right)}\right) \left. \frac{d}{dT}\left(-B\left(\frac{1}{T_0} - \frac{1}{T}\right)\right)\right|_{T=T_0} \\ &= R_0 - \Delta T R_0 \left(\frac{B}{T_0^2}\right) \end{aligned}$$

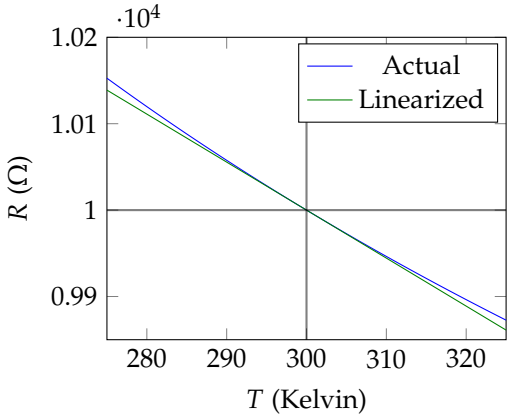


Figure 21: Linearized approximation of thermistor resistance for temperatures near 300 K

So if $T_0 = 300\,\text{K}$, $R_0 = 10\,\text{k}\Omega$ and $B = 50\,\text{K}^{-1}$, then for temperatures near 300 K we have

$$R \approx 10\,\text{k}\Omega - \Delta T \times 5.5\,\Omega$$

So we should see a difference of about $5.5\,\Omega/\text{K}$. To check the accuracy of this approximation, we can compare the actual (nonlinear) equation to the linearized result, as shown in Note that the accuracy is best for very small ΔT, and the error begins to grow as $|\Delta T|$ increases.

Equivalent small-signal resistance and impedance

In

$$\text{large-signal definition: } r_X = \left.\frac{\partial v_X}{\partial i_X}\right|_{\text{DC}}$$
$$\text{small-signal definition: } = \frac{v_x}{i_x}$$

From this definition we can define an analysis procedure for determining a circuit's small-signal equivalent resistance:

1. Linearize the circuit and obtain the small-signal equivalent model.
2. Set any independent signal sources to zero.
3. Insert a **test voltage source** v_x across the terminals of interest.
4. Solve the current i_x that flows through the test source v_x.
5. The equivalent resistance is $r_x = \frac{v_x}{i_x}$.

Note that this procedure only works for small-signal models. **Do not use the large-signal ratio v_X/i_X!**

Frequency Response of Amplifiers

Every circuit has a frequency response. At the very least, there is a hidden capacitance between every pair of nodes, called the **parasitic capacitance**. These capacitances introduce multiple **poles and zeros** into the circuit's frequency response.

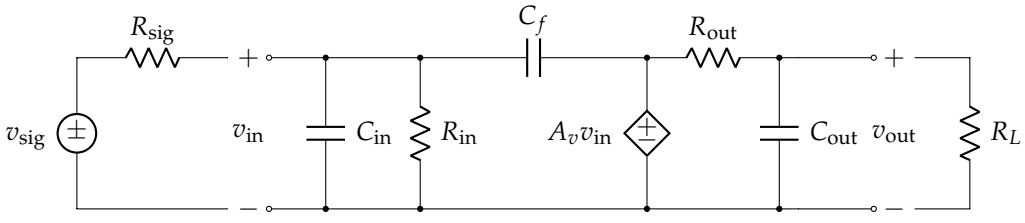


Figure 22: An example amplifier model showing parasitic capacitances.

The general "ZPK" form of the transfer response is

$$H(s) = K\frac{\prod_k (1 - s/\omega_{zk})}{\prod_m (1 - s/\omega_{pm})}$$

where ω_{zk} are the zeros, indexed by k and ω_{pm} are the poles, indexed by m. In this course we will concern ourselves almost exclusively with "simple" poles and zeros in the left half-plane.

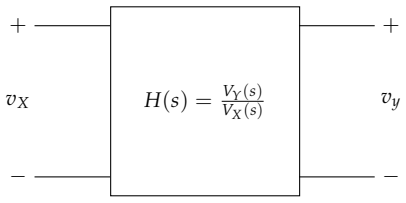


Figure 23: General "black-box" model of a linearized amplifier circuit. Zeros are roots of s in the numerator and poles are roots in the denominator.

In other words, we'll assume that all poles and zeros are real-valued (not complex or imaginary), are all well separated (they do not overlap in value), and are negative valued. If these conditions are satisfied, then we can use a simplified "stick-figure" method to produce approximate magnitude and phase response diagrams, which are called **Bode plots**.

Low-pass systems

For every **pole** ω_{pm}, the magnitude **decreases** by 20 dB per decade at frequencies above the pole. The phase response decreases by 90° between the frequencies $0.1\omega_{pm}$ and $10\omega_{pm}$, and crosses $-45°$ at ω_{pm}.

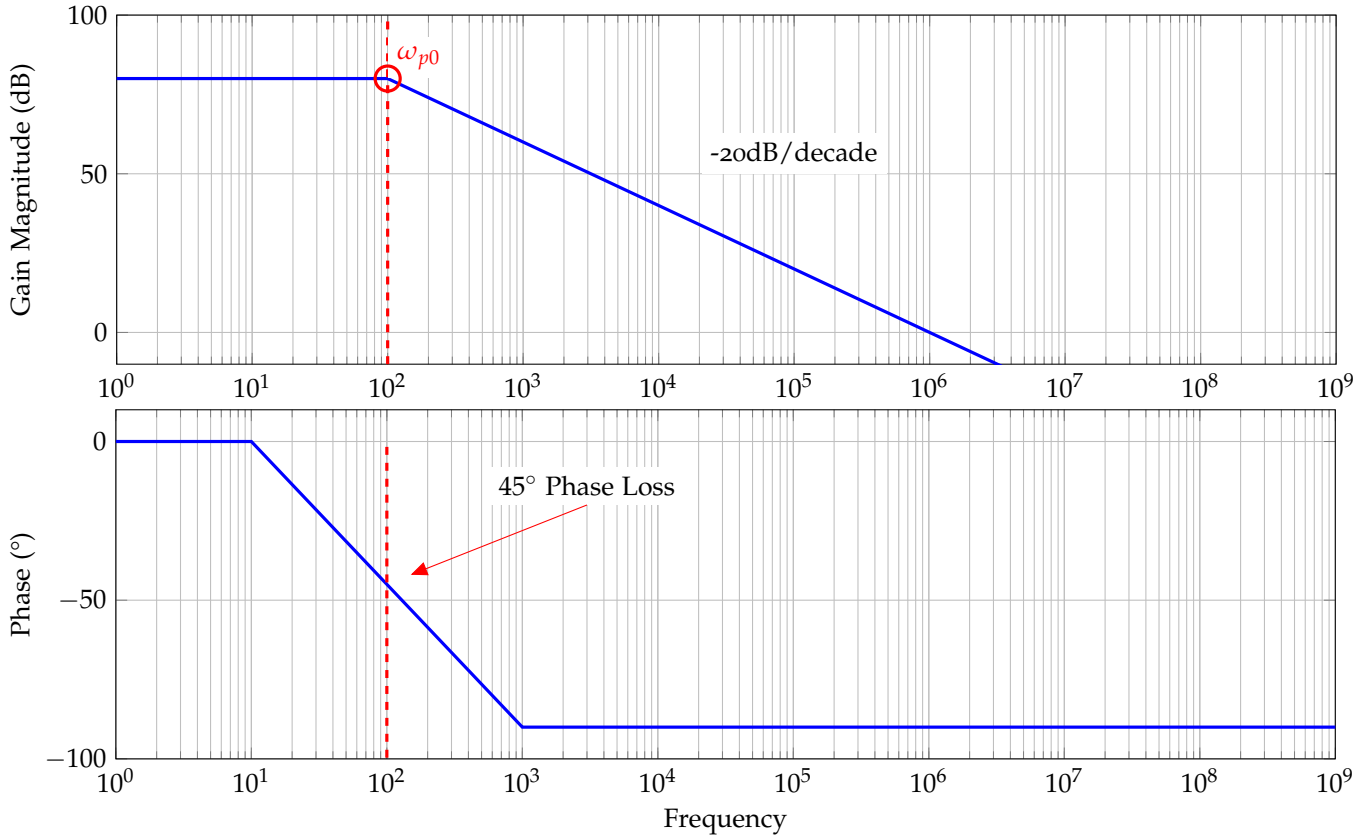


Figure 24: Bode plot of a low-pass system with transfer function $H(s) = K\frac{1}{1+s/\omega_{p0}}$, with a single pole at ω_{p0} and no zeros, and with gain constant $K = 10^4$ corresponding to 80 dB. In this example the pole is at 100 rad/s. The phase response begins to decrease at 10 rad/s, loses 45° per decade, and the phase change concludes at 1×10^3 rad/s.

High-pass systems

For every zero ω_{zk}, the transfer function **increases** by 20 dB per decade at frequencies above the pole. The phase response increases by 90° between the frequencies $0.1\omega_{pm}$ and $10\omega_{pm}$, and crosses $-45°$ at ω_{pm}.

In high-pass systems, there is usually a **zero at the origin**. In that case, there is no phase response associated with the zero (it occurs at infinitely low frequency on the logarithmic scale), and the zero must be canceled by one or more poles at higher frequencies. On the Bode plot, the magnitude response reaches its maximum value and becomes constant after the first pole, ω_{p0}. For frequencies below ω_{p0}, the magnitude response decreases by 20 dB per decade.

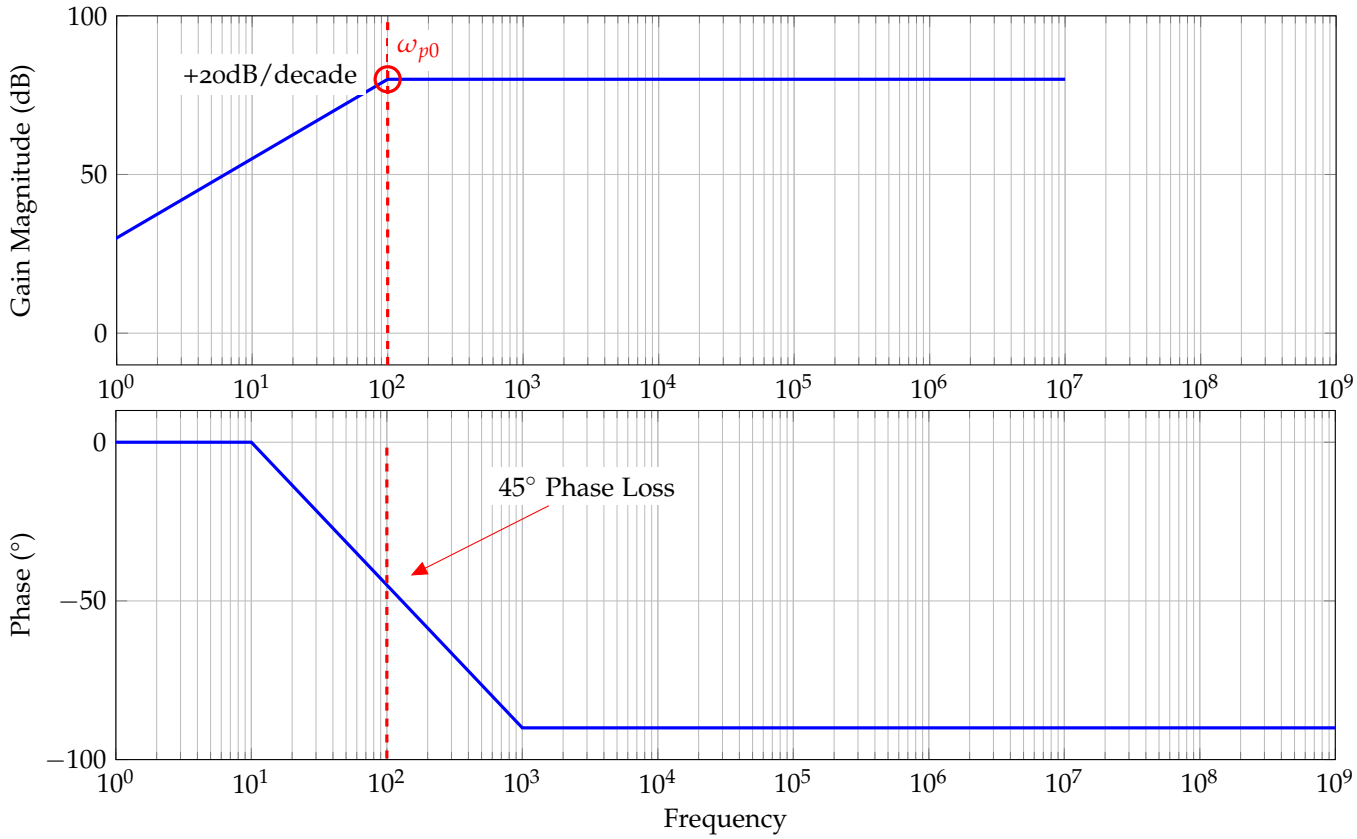


Figure 25: Bode plot of a high-pass system with transfer function $H(s) = K\frac{s}{1+s/\omega_{p0}}$, which has a single zero at the origin, and a single pole at 100 rad/s. The gain constant is $K = 10^4$, corresponding to 80 dB. The phase response is due to the pole; the zero contributes no phase change since it occurs at the origin.

Band-pass systems

Many circuit's exhibit a mix of high-pass and low-pass characteristics. We will especially see this in circuits that use **capacitive coupling** to separate the DC offset from an AC small signal. In a bandpass system, the transfer function's magnitude is highest for middle frequencies between two pole frequencies ω_L and ω_H. This zone is referred to as the circuit's **mid-band** or **pass band**.

A typical band-pass amplifier model is shown below. The signal source has a DC offset voltage V_{SIG}, which is usually undesirable since it will be amplifier along with the signal. In order to amplify just the signal, the offset is rejected by using a **coupling capacitor** C_{C1} to create a high-pass response at the input. The amplifier's output similarly has an undesired DC offset V_{OUT}, which is rejected using the coupling capacitor C_{C2}.

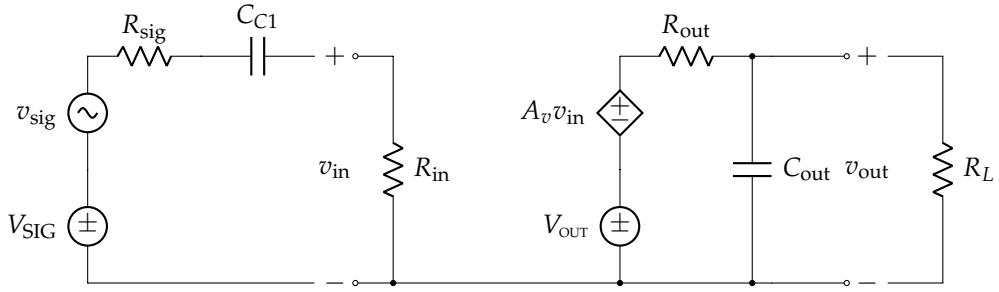


Figure 26: Bandpass amplifier model where coupling capacitors C_{C1} and C_{C2} are used to reject or replace the DC offsets V_{SIG} and V_{OUT}.

To analyze the bandpass circuit, we replace capacitors by their Laplace domain equivalent impedances. We then have

$$
\begin{aligned}
v_{\text{in}} &= v_{\text{sig}} \left(\frac{sC_{C1}R_{\text{in}}}{1 + sC_{C1}\left(R_{\text{in}} + R_{\text{sig}}\right)} \right) \\
v_{\text{out}} &= A_v v_{\text{in}} \left(\frac{R_L}{R_L + R_{\text{out}} + sC_{\text{out}}R_{\text{out}}R_L} \right) \\
&= v_{\text{sig}} A_v \left(\frac{sC_{C1}R_{\text{in}}}{1 + sC_{C1}\left(R_{\text{in}} + R_{\text{sig}}\right)} \right) \left(\frac{R_L}{R_L + R_{\text{out}} + sC_{\text{out}}R_{\text{out}}R_L} \right)
\end{aligned}
$$

The transfer function is

$$
\begin{aligned}
H(s) &= \frac{v_{\text{out}}}{v_{\text{sig}}} \\
&= A_v \left(\frac{R_L}{R_L + R_{\text{out}}} \right) \left(\frac{sC_{C1}R_{\text{in}}}{\left(1 + sC_{C1}\left(R_{\text{in}} + R_{\text{sig}}\right)\right)\left(1 + sC_{\text{out}}\left(R_L \parallel R_{\text{out}}\right)\right)} \right)
\end{aligned}
$$

In this case there is a zero at the origin and two poles located at

two frequencies:

$$\omega_{p0} = \frac{1}{C_{C1}\left(R_{\text{in}} + R_{\text{sig}}\right)}$$
$$\omega_{p1} = \frac{1}{C_{\text{out}}\left(R_L \parallel R_{\text{out}}\right)}$$

Under ideal conditions, the voltage amplifier should have very high R_{out} which places ω_{p0} at a low frequency, and it should have a very low R_{out} which places ω_{p0} at a high frequency. Since there is a zero at the origin, the magnitude rises by 20 dB per decade for frequencies below ω_{p0}, and then becomes flat between ω_{p0} and ω_{p1}. For frequencies higher than ω_{p1} the magnitude falls by 20 dB per decade.

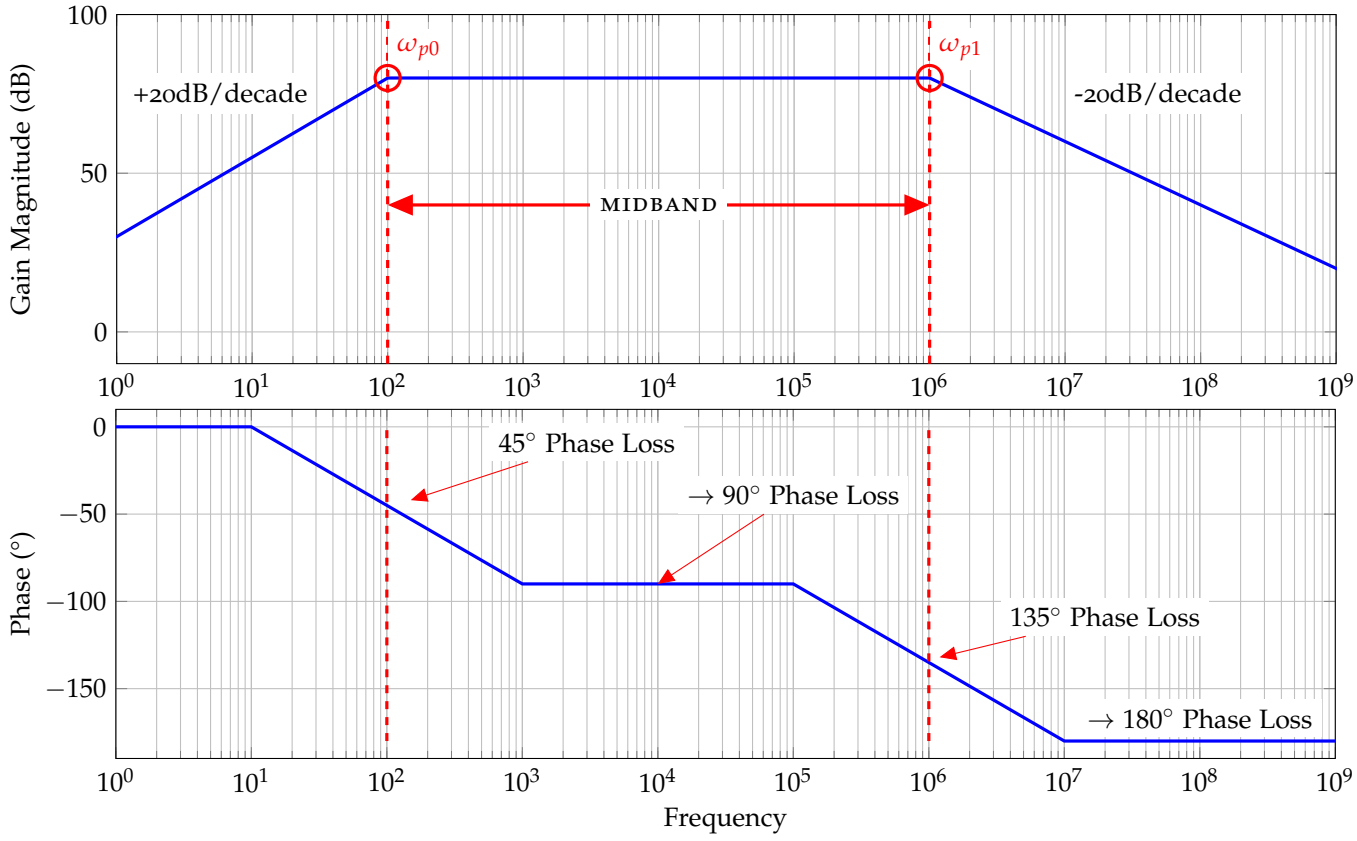


Figure 27: Bode plot of a band-pass system with a single zero at the origin, and two poles at 100 rad/s and 1×10^6 rad/s . The gain constant is $K = 10^4$, corresponding to 80 dB. The phase response is due to the two poles, approaching $-180°$ at higher frequencies.

EveryCircuit Demonstration 4 (Ideal Band-Pass Amplifier Model).

This demonstration shows an implementation of the band-pass model from Figure 26. Examine this circuit and perform both AC simulations (frequency mode) and transient simulations (time mode). Try increasing and decreasing both C_{C1} and C_{out} by $10\times$ (for a total of four different cases), and observe how the pole frequencies change. Verify that the observations match the predictions from our analysis in this section.

EveryCircuit Demonstration 6 (Capacitive Coupling).

This capacitive coupling demonstration shows how we can remove a signal's DC offset and replace it with a different offset. The circuit works through superposition of high-pass and low-pass signal paths. The input AC signal has an offset of 10 V and a zero-to-peak amplitude of 1 V. At the output, the original offset is rejected by the coupling capacitor. A new offset of 1 V is provided by an independent DC voltage source, and is superimposed through the 1 kΩ resistor on the output side.

Harmonic distortion

When a pure sinusoid is input to perfectly linear amplifier, the output is expected to be a pure sinusoid, and its magnitude spectrum should have a single impulse. Real amplifiers are not perfectly linear though, so the output is usually not a perfect sinusoid.

As a result, unexpected features called **harmonics** appear in the output magnitude spectrum. Harmonics are **spurious** impulses that appear at integer multiples of the original **fundamental** signal frequency. So if the input sinusoid has a fundamental frequency component at f_0, the distorted output sinusoid has harmonic components spaced at integer multiples $f_k = kf_0$.

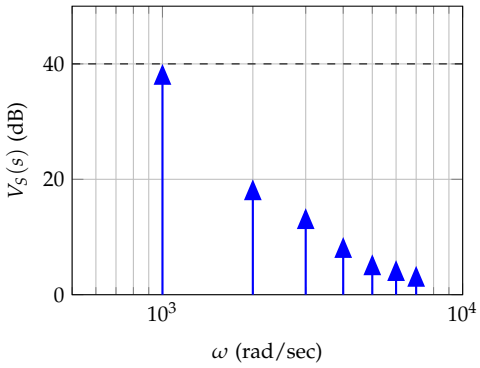


Figure 28: Harmonic "spurs" appear at integer multiples of the fundamental frequency, and represent distortion.

Aliasing

Since the harmonic components can extend to very high frequencies, they may contribute to aliasing effects in a digital oscilliscope's FFT display. Aliasing occurs when a signal violates the Shannon-Nyqvist Sampling Theorem, which states that the sampling rate must be at least twice the highest frequency

present in the signal. On a typical digital oscilloscope, we must be aware of the following considerations:

- The `Sec/Div` knob sets the sampling rate f_S.
- If the signal frequency $f > f_S/2$, then the scope will show an *image* at $f - f_S/2$. So if you increase f beyond $f_S/2$, the signal peak on the FFT will appear to move *backwards*.
- When many high-frequency harmonics are present, their images will overlap again and again over the FFT display, creating an erroneous and confusing plot.
- Higher frequency harmonics can be suppressed by activating an internal bandwidth limit on the oscilloscope's input channel.
- When zooming in to see more detail on the FFT display, do not use the `Sec/Div` knob. Instead, look for a digital zoom setting in the FFT menu.

Operational Amplifier Circuits

Amplifiers with finite open-loop gain

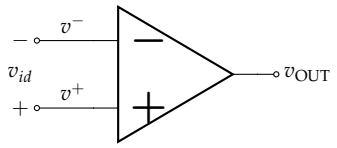


Figure 29: An op amp has two input terminals and one output. The input signal is *differential*, with $v_{id} \triangleq v^+ - v^-$. The output is *single-ended*, with $v_{\text{OUT}} = Av_{id}$.

Operational amplifiers (op amps) are *nearly ideal* differential amplifiers. This means that their output is proportional to the difference of their inputs, and is governed by the characteristic equation

$$v_{\text{OUT}} = A\left(v^+ - v^-\right),$$

where gain is the amplifier's voltage gain. Since op amps are nearly ideal, we expect them to have very high R_{in} and very low R_{out}. Furthermore, there should be **zero current passing into the op amp's input terminals**.

Op amps are almost always used in **negative feedback configurations** where there is some path for current to flow between the amplifier's output and its inverting input terminal. To analyze realistic op amp circuits with feedback, we need to introduce some more refined notation:

$G^\star =$ The **desired** or **ideal** or **nominal** closed-loop gain
$\Rightarrow G_i^\star =$ for an inverting configuration
$\Rightarrow G_{ni}^\star =$ for a non-inverting configuration
$G =$ The **actual** achieved closed-loop gain.
$A =$ The op amp's finite open-loop gain, in volts per volt.
$\epsilon =$ The **error coefficient**
$\Rightarrow G = G^\star\epsilon$

Notice that the concept of "open-loop gain" is distinct from the "open circuit gain," but in this chapter we will consider them to be approximately the same. The open-loop gain refers to the amplifier's gain *without feedback*, whereas the open-circuit gain refers to the gain *without a load*. Since an op amp is expected to have a very low R_{out}, we will assume that loading effects are negligible.

Inverting amplifier

The standard inverting configuration includes an input resistor R_1 and a feedback resistor R_2. Whenever an op amp is connected in a negative feedback configuration, it will exhibit a **virtual short** effect that forces v^- to be approximately equal to v^+. The virtual short occurs because the op amp's open-loop gain tends to be very large. We can **prove** the virtual short effect under the most ideal condition: that the op amp's open loop gain is so large it effectively approaches infinity.

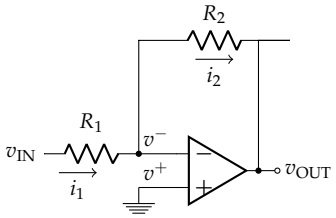


Figure 30: Inverting op amp configuration. No current flows into the op amp's terminals, so $i_2 = i_1$.

Summary: This configuration's characteristics are:

$$G^\star = -\frac{R_2}{R_1}$$
$$G = \epsilon G^\star$$
$$\epsilon = \frac{A}{1 + A + \frac{R_2}{R_1}}$$

The closed-loop gain is the ratio v_{OUT}/v_{IN} when a negative feedback connection is present.

Proof. First suppose $v^- > v^+$. Then we expect to see a large negative voltage at v_{OUT}. By superposition,

$$v^- = v_{IN}\left(\frac{R_2}{R_1 + R_2}\right) + v_{OUT}\left(\frac{R_1}{R_1 + R_2}\right).$$

But if the op amp's gain $A \to \infty$, then $v_{OUT} \to -\infty$ and consequently $v^- \to -\infty$. This creates a contradiction, since we supposed that $v^- > v^+$. On the other hand, if $v^- < v^+$, then $v_{OUT} \to \infty$ and consequently $v^- \to \infty$, which is another contradiction. The only non-contradictory scenario is if $v^- = v^+$. □

Thanks to the virtual short effect, we can say that *ideally* $v^- = 0$, so the current passing through R_1 is

$$i_1 = \frac{v_{IN}}{R_1}.$$

Since there is no current passing into the op amp's input terminals, the entire current i_1 must pass through R_2. Then $i_2 = i_1$ and $v_{OUT} = -i_1 R_2 = -v_{IN}(R_2/R_1)$. This result is based on ideal assumptions, so we can say that the **ideal closed-loop gain** is

$$G_i^\star = -\frac{R_2}{R_1}.$$

The ideal analysis assumes that the op amp's open-loop gain goes to infinity. We can perform a more realistic analysis by accounting for the op amp's **finite open-loop gain**. In this case, the op amp has an **inexact virtual short**, so we should not rely on it in our analysis. Instead, we can solve for the closed-loop gain beginning from the op amp's characteristic equation:

$$
\begin{aligned}
v_{\text{out}} &= A\left(v^+ - v^-\right) \\
\Rightarrow v_{\text{out}} &= A\left(0 - v^-\right) \\
\Rightarrow v^- &= -\frac{v_{\text{out}}}{A} \\
i_2 = i_1 &= \frac{v_{\text{out}} - v^-}{R_2} \\
&= \frac{v^- - v_{\text{in}}}{R_1}
\end{aligned}
$$

Then we have

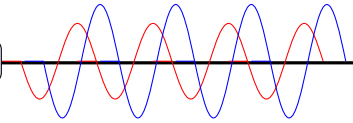
$$
\begin{aligned}
R_1\left(v_{\text{out}}\left(1+\frac{1}{A}\right)\right) &= R_2\left(-\frac{v_{\text{out}}}{A} - v_{\text{in}}\right) \\
\Rightarrow v_{\text{out}}\left(1+\frac{1}{A}+\frac{R_2}{R_1 A}\right) &= -\frac{R_2}{R_1}v_{\text{in}} \\
\Rightarrow G_i = \frac{v_{\text{out}}}{v_{\text{in}}} &= \left(-\frac{R_2}{R_1}\right)\left(\frac{A}{A+1+R_2/R_1}\right)
\end{aligned}
$$

Notice that, in this form, we can express the circuit's actual gain as the product of two terms:

$$
\begin{aligned}
G_i &= G_i^\star \times \epsilon \\
G_i^\star &= -\frac{R_2}{R_1} \\
\epsilon &= \frac{A}{A+1+R_2/R_1}
\end{aligned}
$$

The first term, $G^\star$, is the gain expected if we used an ideal op amp. The second term, ϵ, is an error coefficient that quantifies the effect of using an op amp with finite open-loop gain A.

EveryCircuit Demonstration 8 (Inverting configuration).



This circuit implements an inverting configuration where the op amp's open-loop gain is $A = 10\,\text{V/V}$ (i.e. 20 dB). The resistor values are $R_1 = 1\,\text{k}\Omega$ and $R_1 = 2\,\text{k}\Omega$, so we expect an ideal closed-loop gain of $G^\star = -2\,\text{V/V}$. The input signal has a zero-to-peak amplitude of 1 V, so the output amplitude should be 2 V. Simulate this circuit and observe the output amplitude. It should be 1.54 V. To verify that this matches the prediction from our theory, solve for ϵ and G using the methods described in this section. Then, try increasing the op amp's open-loop gain to $20\,\text{V/V}$ and repeat your calculations to verify that the theory holds up.

Non-inverting amplifier

The non-inverting configuration is similar to the inverting configuration, except the input signal is applied at v^+. Under ideal assumptions, we may appeal to the virtual short so that $v^- = v_{\text{IN}}$, and $i_2 = i_1$. Then

$$\begin{aligned} i_1 &= -\frac{v_{\text{IN}}}{R_1} \\ v_{\text{OUT}} &= v_{\text{IN}} - i_1 R_2 \\ &= v_{\text{IN}} + v_{\text{IN}} \frac{R_1}{R_2} \\ \Rightarrow G_{ni}^\star &= 1 + \frac{R_1}{R_2} \end{aligned}$$

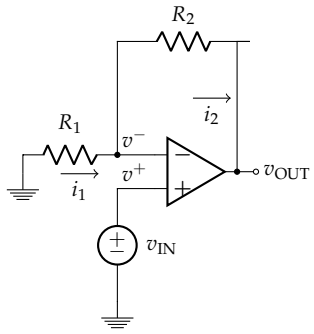


Figure 31: Non-inverting amplifier configuration. The "virtual short" effect causes the op-amp's input terminals to have nearly equal potentials, so $v^- \approx v^+$.

Summary: This configuration's characteristics are:

$$\begin{aligned} G^\star &= 1 + \frac{R_2}{R_1} \\ G &= \epsilon G^\star \\ \epsilon &= \frac{A}{1 + A + \frac{R_2}{R_1}} \end{aligned}$$

To obtain the more realistic gain accounting for finite open-loop gain, we begin from the characteristic equation as before:

$$\begin{aligned} v_{\text{out}} &= A\left(v_{\text{in}} - v^-\right) \\ \Rightarrow v^- &= v_{\text{in}} - \frac{v_{\text{out}}}{A} \\ i_2 = i_1 &= \frac{v_{\text{out}} - v^-}{R_2} \\ &= \frac{v^-}{R_1} \end{aligned}$$

Rearranging we get:

$$\begin{aligned} R_1\left(v_{\text{out}} - v^-\right) &= R_2 v^- \\ \Rightarrow R_1\left(v_{\text{out}} - v_{\text{in}} + \frac{v_{\text{out}}}{A}\right) &= R_2\left(v_{\text{in}} - \frac{v_{\text{out}}}{A}\right) \\ \Rightarrow v_{\text{out}}\left(1 + \frac{1}{A}\left(1 + R_2/R_1\right)\right) &= v_{\text{in}}\left(1 + \frac{R_2}{R_1}\right) \\ \Rightarrow v_{\text{out}}\left(1 + \frac{G_{ni}^\star}{A}\right) &= v_{\text{in}} G_{ni}^\star \\ \Rightarrow v_{\text{out}}\left(\frac{A + G_{ni}^\star}{A}\right) &= v_{\text{in}} G_{ni}^\star \\ \Rightarrow G = \frac{v_{\text{out}}}{v_{\text{in}}} &= G_{ni}^\star\left(\frac{A}{A + G_{ni}^\star}\right) \\ \Rightarrow G &= G_{ni}^\star\left(\frac{A}{A + 1 + R_2/R_1}\right) \end{aligned}$$

Once again we may express the result in two parts, $G^\star$ and ϵ:

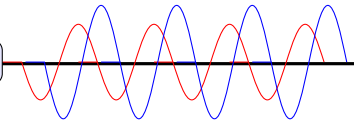
$$\begin{aligned} G_{ni}^\star &= 1 + \frac{R_2}{R_1} \\ \epsilon &= \frac{A}{A + 1 + R_2/R_1} \\ G_{ni} &= G_{ni}^\star \times \epsilon \end{aligned}$$

Notice that the error coefficient, ϵ, is the same for both the inverting and non-inverting configurations.

Generalized Result

Since the error coefficient is the same in both configurations, the closed-loop gain can be generally expressed as

$$\begin{aligned} G &= G^\star \times \epsilon \\ &= G^\star \left(\frac{A}{A + 1 + R_2/R_1} \right) \end{aligned}$$



EveryCircuit Demonstration 10 (Non-inverting configuration).

Make a copy of the inverting configuration circuit and modify it to implement a non-inverting configuration. Keep the parameters from the original exampe, $R_1 = 1\,\text{k}\Omega$, $R_2 = 2\,\text{k}\Omega$ and $A = 10\,\text{V/V}$, and set the input signal amplitude to 1 V. For these parameters, calculate the expected values of $G^\star$, ϵ and G. Simulate the circuit and verify that the output amplitude agrees with your calculations.

Voltage Follower

The voltage follower represents a slightly different case, since there are no resistors.

In this configuration, we have the following device equations:

$$\begin{aligned} v_{\text{OUT}} &= A\left(v^+ - v^-\right) \\ &= A\left(v_{\text{IN}} - v_{\text{OUT}}\right) \\ \Rightarrow G = \frac{v_{\text{OUT}}}{v_{\text{IN}}} &= \frac{A}{A+1} \end{aligned}$$

In this case, the gain can be expressed as

$$\begin{aligned} G_{vf}^{\star} &= 1 \\ \epsilon_{vf} &= \frac{A}{A+1} \\ G_{vf} &= G_{vf}^{\star} \times \epsilon_{vf} \end{aligned}$$

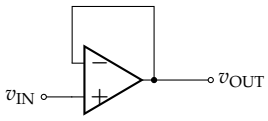
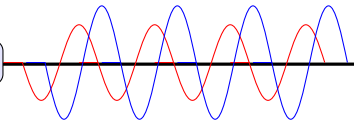


Figure 32: Voltage follower configuration. Due to the "virtual short" effect, $v_{\text{OUT}} \approx v_{\text{IN}}$.

Summary: This configuration's characteristics are:

$$\begin{aligned} G^{\star} &= 1 \\ G &= \epsilon G^{\star} \\ \epsilon &= \frac{A}{1+A} \end{aligned}$$



EveryCircuit Demonstration 12 (Voltage follower configuration).

Make a copy of the inverting configuration circuit and modify it to implement a voltage follower configuration. Keep the same op amp gain from the original example, $A = 10\,\text{V/V}$, and set the input signal amplitude to 1 V. Calculate the expected values of $G^{\star}$, ϵ and G. Simulate the circuit and verify that the output amplitude agrees with your calculations.

Difference amplifiers

To make an amplifier with *fully-differential* input, we can combine inverting and non-inverting configurations. This gives us two gains:

$$G_{ni}^{\star} = 1 + \frac{R_2}{R_1}$$
$$G_{i}^{\star} = -\frac{R_2}{R_1}$$

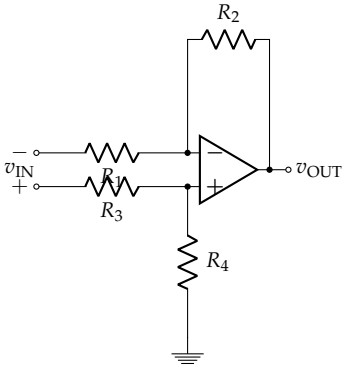


Figure 33: Difference amplifier configuration for amplifying a differential signal. Inverting and non-inverting configurations are superimposed. The resistor-divider $R_3 - -R_4$ is used to ensure the same gain for the inverting and non-inverting signal paths.

To achieve proper differential operation, the inverting and non-inverting gains must be **balanced**, i.e. $G_{ni} = G_i$. In their usual configurations, this is not the case. In order to balance the inverting and non-inverting gains, we insert the voltage divider R_3, R_4, so that:

$$\begin{aligned} G_{ni}^{\star} &\to \left(1 + \frac{R_2}{R_1}\right)\left(\frac{R_4}{R_3 + R_4}\right) \\ &= \frac{R_2}{R_1} \text{ (condition for balance)} \\ \Rightarrow \left(\frac{R_4}{R_3 + R_4}\right)\left(1 + \frac{R_2}{R_1}\right) &= \frac{R_2}{R_1} \end{aligned}$$

Then solving for R_4/R_3 we find that

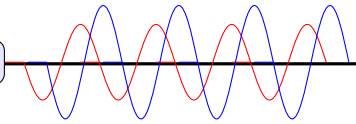
$$\begin{aligned} \frac{R_3}{R_3 + R_4} &= \left(\frac{R_2}{R_1}\right)\left(\frac{R_1}{R_2 + R_1}\right) \\ &= \frac{R_2}{R_1 + R_2} \end{aligned}$$

Then we can invert both sides:

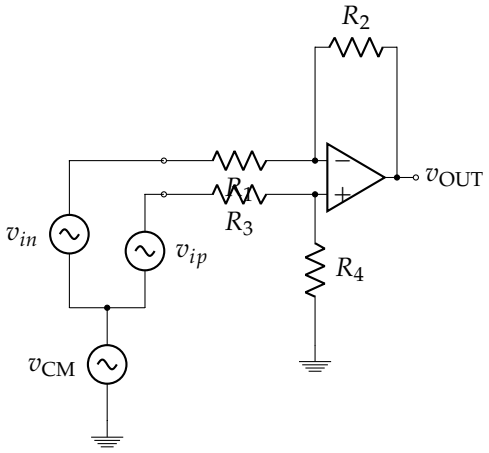
$$\begin{aligned} 1 + \frac{R_3}{R_4} &= 1 + \frac{R_1}{R_2} \\ \Rightarrow \frac{R_3}{R_4} &= \frac{R_1}{R_2} \end{aligned}$$

So **the resistor ratios need to be *matched*.**

EveryCircuit Demonstration 14 (Difference amplifier (differential mode)).



This circuit implements a difference amplifier with both differential and common-mode input circuits. In the initial setup, you should see that the two differential input signals, v_{ip} and v_{in}, have zero-to-peak amplitudes of 1 V and a frequency of 1 kHz. One of the sources, v_{ip}, has a phase of 180° in order to have opposite polarity from v_{in}. The common-mode signal v_{CM} is shared by *both* of the input signals, i.e. they share this signal component; it is *common* to both of them. In the example design, v_{CM} has a small amplitude of 100 mV and a frequency of 300 Hz. **We expect the common-mode signal to be canceled out, so it should not appear at all in the output signal**. To verify this, increase the amplitude of v_{CM} to 5 V, so it will be clearly visible. Notice that the output waveform doesn't change.



Next, modify the value of R_4 by increasing it to 4 kΩ. Keep the amplitude of v_{CM} at 5 V, and let the simulation run for a while. You should observe that a 300 Hz fluctuation is superimposed onto the output signal. **The common-mode is no longer canceled.**

The importance of matching

If the inverting and non-inverting gains are imbalanced, then the **common-mode signal** is not perfectly cancelled. To see this, we now consider the actual gains G_i and G_{ni}, which may differ due to imprecision in actual resistor values:

$$
\begin{aligned}
v_{\text{IN}^+} &= \frac{1}{2} v_{\text{sig}}^{+} + v_{\text{CM}} \\
v_{\text{IN}^-} &= \frac{1}{2} v_{\text{sig}}^{-} + v_{\text{CM}} \\
\Rightarrow v_{\text{OUT}} &= \frac{1}{2} \left(G_{ni} + G_i\right) v_{\text{sig}} + \left(G_{ni} - G_i\right) v_{\text{CM}}
\end{aligned}
$$

The latter part of this result is called the **common-mode gain**, $A_{CM} = (G_{ni} - G_i)$. The **Common Mode Rejection Ratio (CMRR)** is the ratio of the effective differential gain, $A_d = \frac{1}{2}(G_{ni} + G_i)$, to A_{CM}:

$$\text{CMRR} = \frac{A_d}{A_{CM}}$$

This figure is often specified in dB. Ideally it should be infinite.

Input resistance in the difference amplifier

One source of mismatch in the difference amplifier is that the input resistances are unmatched between the two input legs. To evaluate the input resistance, we apply the method described in **??** separately for each leg of the input signal.

At the inverting input, we find that the input resistance is equal to R_1, since $v_{ip} = 0$, so that $v^+ = 0$ and, due to the virtual short, $v^- = 0$. At the non-inverting input, the equivalent resistance is equal to $R_3 + R_4$. If the input signals have a significant series resistance, we will see signal attenuation due to resistive coupling effects, which modifies the gain. Let us assume that both v_{ip} and v_{in} are both connected in series with a resistance equal to R_{sig}. Then this resistance is effectively added in series with R_1 and R_3, so that after accounting for this **loading effect** the gain becomes

$$G_L^\star = \frac{R_2}{R_1 + R_{\text{sig}}}.$$

In other words

$$G_L^\star = G^\star \left(\frac{R_1}{R_1 + R_{\text{sig}}} \right).$$

If we repeat this analysis on the non-inverting signal path, we will find the same ratio. Finally, accounting for finite gain together with the loading effect:

$$G_L = G^\star \left(\frac{R_1}{R_1 + R_{\text{sig}}} \right) \epsilon,$$

where ϵ is now modified due to the presence of R_{sig}:

$$\epsilon = \frac{A}{1 + A + \frac{R_2}{R_1 + R_{\text{sig}}}}.$$

Instrumentation amplifiers

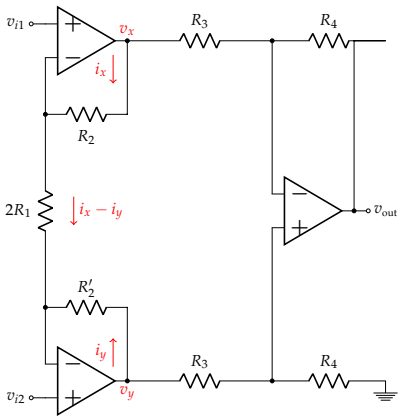


Figure 34: Instrumentation amplifier configuration for amplifying differential signals. Since both inputs are connected to the op amps' non-inverting terminals, they should both have high input resistance and matched electrical characteristics. Compared to difference amplifiers, this configuration is less sensitive to resistor mismatch and has improved CMRR.

Advantages over difference amplifiers:

- Very high input resistance ($R_{\text{in}} \to \infty$).
- Gain controlled by a single resistor ($2R_1$).
- CMRR increased by the gain of the pre-amp stage.

Disadvantages:

- Needs three op amps.
- Higher power consumption.

Instrumentation amplifier A_D analysis.

We have three amplifiers. The first two are non-inverting configurations. Together they are described as a *fully-differential pre-amplifier*. The third op amp is configured as a difference amplifier. The differential gain may be analyzed as a superposition of two non-inverting configurations:

$$v_x = v_{i1}\left(1 + \frac{R_2}{R_1}\right) - v_{i2}\frac{R_2}{R_1}$$
$$v_y = v_{i12}\left(1 + \frac{R_2}{R_1}\right) - v_{i1}\frac{R_2}{R_1}$$

The overall gain of the pre-amplifier stage is then

$$\begin{aligned} A_{D1} &= \frac{v_x - v_y}{v_{i1} - v_{i2}} \\ &= 1 + \frac{2R_2}{2R_1} \\ &= 1 + \frac{R_2}{R_1}. \end{aligned}$$

The difference amplifier contributes a gain of R_4/R_3, so the total differential gain is

$$A_D = \left(1 + \frac{R_2}{R_1}\right)\left(\frac{R_4}{R_3}\right).$$

Instrumentation amplifier A_{CM} analysis.

We expect to obtain a net improvement in CMRR through this configuration, compared to the difference amplifier. In fact, the instrumentation amplifier achieves the following two advantages:

- Eliminates sensitivity to R_1 in the non-inverting configurations by sharing R_1 between the two circuits.
- Eliminates sensitivity to mismatch in R_2.

In the Common-Mode case, set $v_{i1} = v_{i2} = v_{icm}$. Then, due to the *virtual short* effect, the op amp's inverting terminals are also equal to v_{icm}. Therefore the voltage drop across R_1 is zero, so that

$$\begin{aligned} i_x - i_y &= 0 \\ \Rightarrow v_x &= v_y \end{aligned}$$

Note that this result does not depend on the matching between R_2 and R_2'. We may conclude that the pre-amplifier's common-mode gain is

$$A_{CM1} = 0\text{V/V}$$

. This would mean that the instrumentation amplifier has a **theoretically infinite CMRR**. In practice, the op amps themselves will contribute *second order* imperfections ("second order" means they contribute smaller effects than resistor mismatch), resulting in some residual imbalance and a finite CMRR.

Non-Ideal Op Amp Characteristics

We have already discussed **finite open-loop gain**, **finite input resistance** and **common-mode gain** as non-ideal features of op amp circuits. Now we will examine two additional features:

- Input bias current
- Offset voltage

Input Bias Current

Every op amp has a small but non-zero bias current flowing into its input terminals; a typical value might be 10 µA, but this can vary across a wide range for different products. The bias current is typically a **fixed current** that can be modeled as a DC current source.

Example 5 (Bias current in inverting configuration)**.**

In this example we analyze the effect of bias current on an inverting configuration. The theorem of superposition allows us to set $v_{\text{in}} = 0$ to analyze the contribution of I_{bias}. In this case, we see that

$$v^- = 0 \text{ (virtual short)} \Rightarrow v_{\text{out}} = I_{\text{bias}} R_2$$

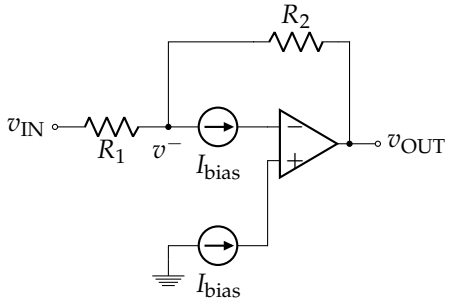


Figure 35: Inverting configuration showing bias-current sources.

By superposition, we can add in the contribution from v_{in}, resulting in

$$v_{\text{out}} = v_{\text{in}} \left(-\frac{R_2}{R_1} \right) + R_2 I_{\text{bias}}.$$

Based on this example, we can see that the effect of I_{bias} is to introduce a DC offset voltage on v_{out}. This places a limitation on the size of R_2 that can be used. Suppose, for instance, that we have

$$I_{\text{bias}} = 10\mu\text{A} \qquad R_2 = 1\text{M}\Omega \qquad V_R = 5\text{V}$$

and the op amp's power rails are at $\pm V_R$. In this case, the bias current induces an output offset voltage equal to

$$I_{\text{bias}} R_2 = 10\text{V},$$

which is greater than the rail of the op amp. As a result, **the op amp will simply saturate**.

Example 6 (Maximum resistance due to I_{bias})**.**

Given I_{bias}, an input signal v_{IN} and a desired closed-loop gain G, how can we determine the maximum allowable value for R_2? Suppose v_{max} is the maximum value of v_{IN}, and v_{min} is the minimum (note that v_{min} can be a negative voltage). Then our circuit must satisfy

$$I_{\text{bias}}R_2 + Gv_{\text{min}} < V_R \;\Rightarrow R_2 < \frac{V_R + Gv_{\text{min}}}{I_{\text{bias}}}.$$

So returning to our example where $I_{\text{bias}} = 10\,\mu\text{A}$ and $V_R = 5\,\text{V}$, and let $G = -10\,\text{V/V}$ and $v_{\text{min}} = -0.1\,\text{V}$, we find

$$R_2 < \frac{(5\,\text{V}) + (1\,\text{V})}{10\,\mu\text{A}} = 600\,\text{k}\Omega.$$

Offset Voltage

Every op amp has a DC offset voltage so that its equation is

$$v_{\text{out}} = A\left(v^+ - v^- + V_{\text{ofs}}\right).$$

When used in high-gain circuits, this offset voltage gets amplified, which may lead to erroneous signal processing in some circuits.

V_{ofs} is **random**, usually varying in the range ± 10mV. V_{ofs} can also **change slowly** over time, making it difficult to zero it out by design.

Example 7 (Inverting configuration with offset voltage).

Suppose an inverting op amp configuration has supply rails equal to $+5$V and -5V, and is configured to have a closed loop gain $G = -R_2/R_1 = -100$V/V. The input signal is a sinusoid with peak-to-peak amplitude 45mV, and the op amp has an offset voltage $V_{\text{ofs}} = 10$mV. Draw the output waveform.

Answer: the output amplitude is 4.5V, but since the offset voltage is also amplified, the output will contain a DC offset equal to 1.01V, hence the output waveform is

$$v_{\text{out}} = 1.01 + 4.5\sin\left(2\pi f t\right).$$

This will result in clipping of the waveform.

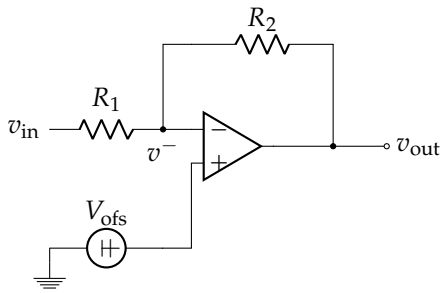


Figure 36: Inverting configuration showing input offset voltage source.

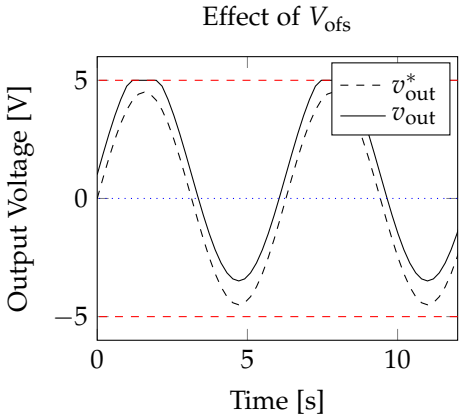
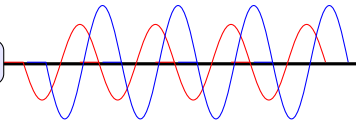


Figure 37: Waveform saturation caused by undesired amplification of the op amp's input offset voltage.

EveryCircuit Demonstration 16 (Non-inverting circuit with bias and offset).



This circuit implements models of both I_{bias} and V_{OFS} in a non-inverting op amp configuration. Since the EveryCircuit op amp model is very ideal, a slight circuit trick is used to model the bias current by steering it into ground instead of into the op amp terminal. This trick doesn't change anything at all about the circuit's behavior. The model uses typical values of $10\,\mu\text{A}$ and $-10\,\text{mV}$ for the bias current and offset voltage, respectively.

We see that the output waveform has a significant DC offset due to the bias and offset effects, and part of the waveform is saturated. To get some experience with these effects, you can experiment with larger and smaller values of each, and with positive and negative values of V_{OFS}. Occasionally the simulation will halt and complain that it can't find a solution. Usually in these cases you can just restart the simulation and will proceed without any problems.

Design question: how can the circuit be modified to minimize the undesirable offset and avoid saturating the output waveform?

Frequency Response of Op Amps

General-purpose op amps are said to be **internally compensated** devices, meaning they are deliberately designed to have a single-pole frequency response with a very low cutoff frequency:

$$A(s) = \frac{A_0}{1 + s/\omega_c}$$

where

ω_c = The low cutoff frequency

A_0 = The DC open-loop gain, in V/V

The frequency response looks like this:

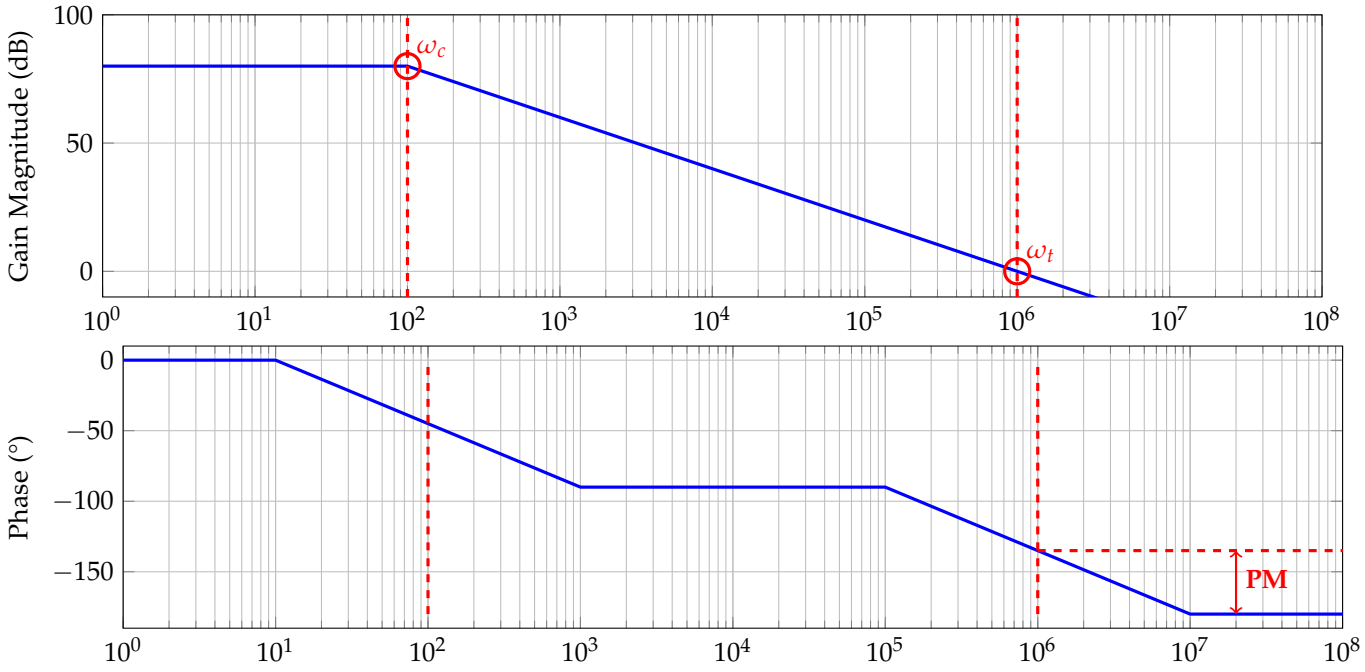


Figure 38: Standard op amp frequency response. In this example, $\omega_c = 100\,\text{rad/s}$, $\omega_t = 1 \times 10^6\,\text{rad/s}$, and $A_{v0} = 80\,\text{dB}$. In real op amp products these parameters can vary significantly for different products. For a given product, ω_t typically shows low part-to-part variation and is a useful figure-of-merit.

The phase response loses 45° about the dominant pole ω_c. There are typically additional poles at frequencies above ω_t, which can cause additional phase loss just prior to ω_t. The **Phase Margin (PM)** measures how much phase is lost at ω_t. Specifically,

$$\text{PM} = 180° + \angle A(j\omega_t)$$

(note that $\angle A(j\omega_t)$ is negative).

For our (introductory) purposes, we will assume that PM = 90°.

Unity-Gain Frequency

A typical op amp has a a very large DC open-loop gain, often greater than 80 dB or 10 000 V/V. Then the magnitude response can be approximated as

$$\begin{aligned} A(\omega) &= \sqrt{\frac{A_0^2}{1+\omega^2/\omega_c^2}} \\ &\approx \frac{A_0\omega_c}{\omega} \end{aligned}$$

The unity-gain frequency ω_t is where the gain magnitude is equal to unity, i.e. 1 V/V:

$$\begin{aligned} 1 &= \omega_c \frac{A_0}{\omega_t} \\ \Rightarrow \omega_t &= \omega_c A_0 \qquad\qquad \text{Note } A_0 \text{ is in V/V} \end{aligned}$$

Because of this result, the unity-gain frequency is often referred to as the **Gain-Bandwidth Product (GBP)**. We can also write the transfer function in terms of ω_t as follows:

$$A(s) = \frac{A_0}{1+sA_0/\omega_t}$$

Closed-Loop Frequency Response

Consider the inverting configuration using an op amp with a one-pole response:

$$\begin{aligned} A_{CL}(s) &= \left(-\frac{R_2}{R_1}\right)\left(\frac{A(s)}{A(s)+1+R_2/R_1}\right) \\ \Rightarrow A_{CL}(s) &= G^\star\left(\frac{\frac{A_0}{1+s/\omega_c}}{\frac{A_0}{1+s/\omega_c}+1-G^\star}\right) \\ \Rightarrow A_{CL}(s) &= G^\star\left(\frac{A_0}{A_0+1-G^\star+s(1-G^\star)/\omega_c}\right) \\ &\approx G^\star\left(\frac{A_0}{A_0+s(1-G^\star)/\omega_c}\right) \qquad \text{since } A_0 \gg 1-G^\star \\ &= G^\star\left(\frac{1}{1+s(1-G^\star)/\omega_t}\right) \end{aligned}$$

We can re-write this as a single-pole transfer function with pole

$$\begin{aligned} \omega_{CL} &= \omega_t/(1-G^\star) \\ &= \omega_t/(1+R_2/R_1)\,. \end{aligned}$$

This result introduces the universal **Gain-Bandwidth Trade-off**. By using feedback, we can convert between bandwidth and

closed-loop gain according to an approximate one-to-one ratio. Note that for a given op amp, **all configurations have the same unity-gain frequency.** Hence we consider ω_t to be the universal parameter of an op amp's frequency response.

The closed-loop frequency response looks like this:

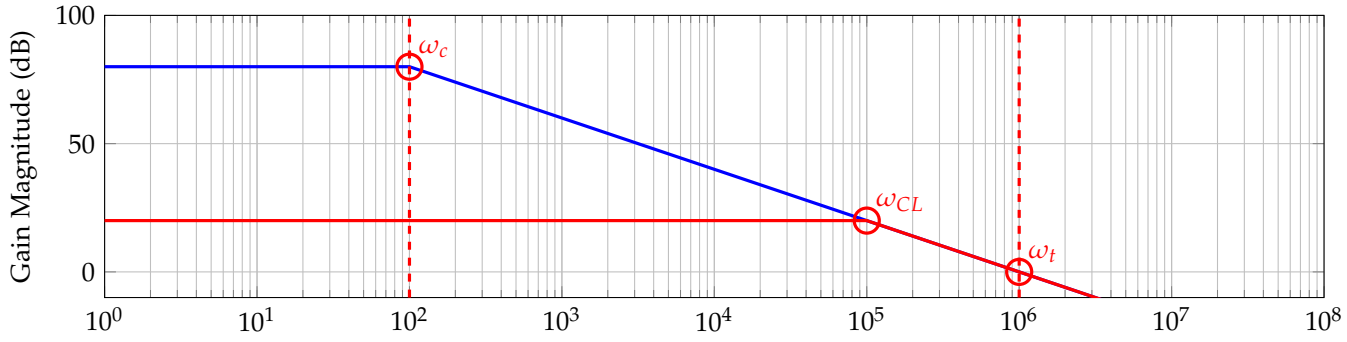


Figure 39: Closed-loop magnitude response (in red) for an op-amp feedback configuration. The open-loop response is also shown (in blue). For frequencies greater than ω_{CL}, the closed-loop response approximately matches the open-loop response.

Example 8 (Closed-loop frequency response, low-gain)**.**

Consider the following parameters:

$$\omega_t = 10\text{MHz}$$
$$R_2/R1 = 10$$

What are the closed-loop DC gain, the 3dB cutoff frequency, and the unity-gain frequency for these parameters?

$$A_{CL} = -10\ \frac{\text{V}}{\text{V}}$$
$$\omega_t = 10\text{MHz}$$
$$\omega_c = 909\text{kHz}$$

Example 9 (Closed-loop frequency response, high-gain).

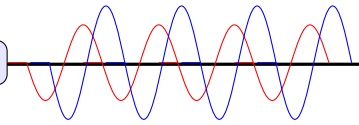
Consider the following parameters:

$$\omega_t = 10\text{MHz}$$
$$R_2/R1 = 100$$

What are the closed-loop DC gain, the 3dB cutoff frequency, and the unity-gain frequency for these parameters?

$$A_{CL} = -100\ \frac{\text{V}}{\text{V}}$$
$$\omega_t = 10\text{MHz}$$
$$\omega_c = 99\text{kHz}$$

Notice that as the desired gain G grows to be very large, the cutoff frequency approximates to ω_t/G.



EveryCircuit Demonstration 18 (Closed-loop frequency response).

This circuit models an op amp with a single-pole transfer function connected in a non-inverting configuration. Since EveryCircuit's built-in op amp model is basically ideal, we have to insert extra components to introduce a pole at the op amp's output node. This is accomplished using an RC low-pass network followed by a **voltage buffer** comprised of a dependent voltage source with a gain of 1 V/V.

Perform a frequency simulation of the closed-loop system and observe the major parameters: the DC gain A_{CL} (in dB), the 3 dB cutoff frequency ω_c, and the unity-gain frequency ω_t. Next, increase the value of R_2 by 10× and then 100×, and observe how it changes A_{CL} and ω_c. Convert A_{CL} to V/V in order to test the predictions from the theory presented in this section. **Verify that ω_t remains constant and is approximately equal to $A_{CL}\omega_c$.**

Slewing

Slewing is very different from ordinary transfer-function based behavior. Slewing can be thought of as **saturation** of v'_{out}, i.e. a second-order saturation effect. As such, it is fundamentally non-linear and introduces harmonic distortion into the signal.

$$SR = \max \frac{d}{dt} v_{\text{out}} \qquad \text{given in V/µs.}$$

Slewing tends to turn the output signal into a **triangle wave**. If the op amp's input signal is a pure sinusoid, then we can determine if the output will be affected by slewing:

$$v^*_{\text{out}} = A_{CL} \mathcal{V}_{\mathcal{A}} \sin(2\pi f t)$$
$$\Rightarrow \frac{d}{dt} v^*_{\text{out}} = 2\pi f A_{CL} \mathcal{V}_{\mathcal{A}} \cos(2\pi f t)$$

where $\mathcal{V}_{\mathcal{A}}$ is the input signal amplitude. This tells us that slewing may occur if $\mathcal{V}_{\mathcal{A}}$ is large, or if the frequency f is large, or if the closed-loop gain A_{CL} is large. The **maximum rate of signal change** must be less than the slew-rate:

$$2\pi f A_{CL} \mathcal{V}_{\mathcal{A}} \leq \text{SR}$$

A signal affected by slewing

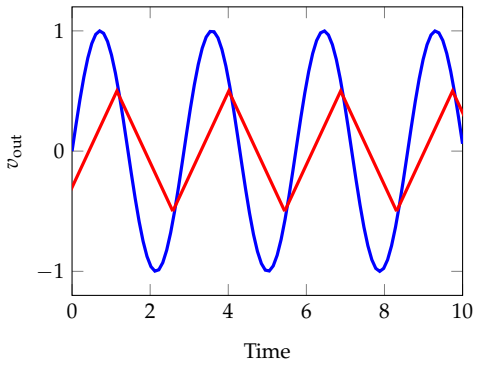


Figure 40: Slew-rate distortion in an op amp circuit.

Example 10.

Slew-rate limiting Suppose an amplifier has the following characteristics:

$$\text{SR} = 1\text{V/µs}$$
$$A_{CL} = 2\ \text{V/V}$$
$$f = 100\text{kHz}$$

What is the maximum input amplitude $\mathcal{V}_{\mathcal{A}}$ which can guarantee no slewing?

$$\mathcal{V}_{\mathcal{A}\max} = \frac{\text{SR}}{2\pi f A_{CL}}$$
$$= 0.796\text{V}.$$

Clearly the slew rate can present real limitations for a circuit.

Although slewing distortion occurs commonly in op amp circuits, there is no easy way to model it in simple simulators like EasyCircuit. To get an accurate prediction of slew-rate

limiting, we need to use a more advanced simulator like SPICE. This is one of the first instances where we can see the need for sophisticated engineering software.

Full Power Bandwidth (FPBW)

The FPBW is the maximum frequency at which an op amp can deliver its full-swing output signal (i.e. rail-to-rail output amplitude). If the op amp has rails at $\pm V_R$, then the maximum output amplitude is $\mathcal{V}_\mathcal{O} = V_R$. Then

$$\text{FPBW} = \frac{\text{SR}}{2\pi V_R}$$

If the op amp is single-supply, then the maximum amplitude is $V_R/2$, so

$$\text{FPBW} = \frac{\text{SR}}{\pi V_R}$$

The circuit will process any frequency less than the FPBW **distortion-free**. For higher frequencies, you may begin to see spurious harmonics in the output spectrum.

Once the FPBW is known, the slewing limit can be predicted as follows:

$$\mathcal{V}_{\mathcal{O}\text{max}} = V_R \frac{\text{FPBW}}{f}$$

Hence if you know the FPBW and the rail voltage, you can estimate the maximum allowable amplitude at a given high frequency.

Op Amp Integrators and Differentiators

If the circuit is analyzed in the **Laplace domain**, we can consider arbitrary impedances to behave as through they were resistors. Then

$$A_{CL} = \frac{v_{\text{out}}(s)}{v_{\text{in}}(s)} = -\frac{Z_2(s)}{Z_1(s)}$$

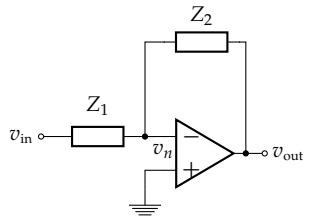


Figure 41: Generalized inverting configuration with complex impedances.

Differentiator: Z_1 is a capacitor

If Z_1 is a capacitor C_1, and Z_2 is a resistor R_2, then

$$Z_1(s) = \frac{1}{sC_1}$$
$$\Rightarrow A_{CL}(s) = -sC_1R_2.$$

In this transfer function, $-C_1R_2$ is just a scale constant. The signal is multiplied by s, resulting in **differentiation** in the time-domain.

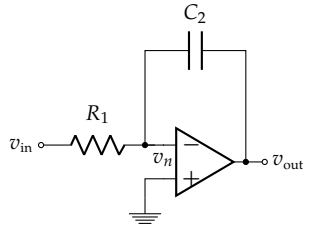


Figure 42: Ideal Miller integrator. $H(s) = \left(\frac{1}{s}\right)\left(\frac{-1}{R_1C_2}\right)$.

Integrator: Z_2 is a capacitor

If Z_2 is a capacitor C_2, and Z_1 is a resistor R_1, then

$$Z_2(s) = \frac{1}{sC_2}$$
$$\Rightarrow A_{CL}(s) = -\frac{1}{sR_1C_2}.$$

In this transfer function, $-1/(R_1C_2)$ is a scale constant. The signal is multiplied by $1/s$, resulting in **integration** in the time-domain.

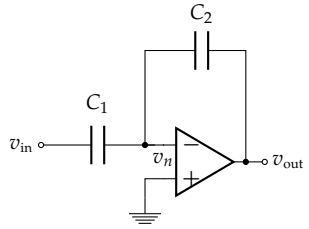


Figure 43: Idealized capacitive inverting configuration.

Z_1 and Z_2 are both capacitors

If both of the impedances are capacitors, then the behavior is similar to an inverting configuration.

$$Z_1(s) = \frac{1}{sC_1}$$
$$Z_2(s) = \frac{1}{sC_2}$$
$$\Rightarrow A_{CL}(s) = -\frac{C_1}{C_2}.$$

In this transfer function, $-C_1/C_2$ is the amplifier's gain. The s terms cancel out, resulting in no integrating or differentiating behavior.

Practical Considerations

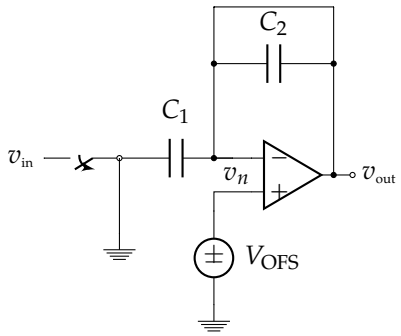
In practice, capacitors cannot simply be left floating at the op amp terminals. Consider the circuit shown in Figure 43. Node v_n is left **floating**, which means **there is nothing to define its potential**. It could literally be anything, which could be disastrous.

Additionally, there is no path for the op amp's DC **bias current** to flow. To address these problems, we have some options:

(a) Include large resistances to passively clear the charge on v^-.

(b) Use ideal switches to periodically *reset* charge on v^-.

Method (b) is commonly used in integrated circuits, where capacitors are easier to make than resistors, and switches are made using MOSFET transistors. One of the key advantages to the circuit in Figure 45 is that it can cancel out the op amp's offset voltage. When used for this purpose it is often called an **auto-zeroing** circuit. We can analyze the auto-zeroing circuit in two phases. In Phase 1, the switches are configured to short across C_2 and to connect the top plate of C_1 to ground. In this phase, the op amp is basically in a voltage-follower configuration. Due to the virtual short effect, v_n should be equal to V_{OFS}, so C_1 gets charged up to a voltage equal to V_{OFS}. Meanwhile C_2 is discharged to a voltage of zero.



Phase 1

In Phase 2, the input is connected, and node v_n is left **floating**, so that no charge can be added or removed from v_n. So any charge added to the outside plate of C_1 has to be balanced by an opposite charge on the outside plate of C_2. The charge $Q_1 = v_{\text{IN}}/C_1$ must be balanced by $Q_2 = v_{\text{OUT}}/C_2 = -Q_1$.

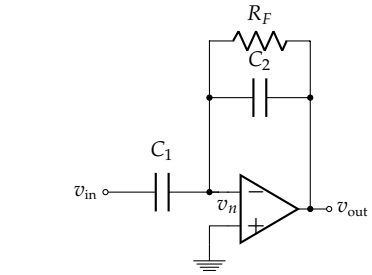


Figure 44: Practical capacitive configuration with DC bypass resistor.

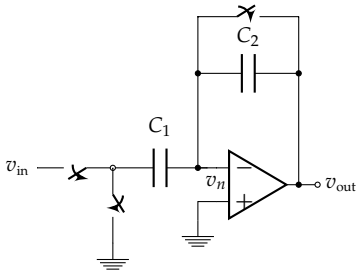
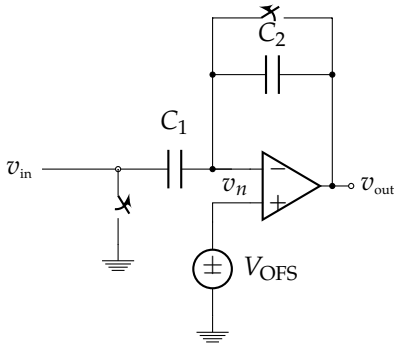


Figure 45: Practical capacitive configuration with switched DC bypass.



Phase 2

Figure 46: The two switching phases of a capacitive inverting configuration.

Therefore

$$\frac{vOUT}{C_2} = -\frac{v_{\text{IN}}}{C_1}$$
$$\Rightarrow \frac{v_{\text{OUT}}}{v_{\text{IN}}} = -\frac{C_2}{C_1}.$$

This interpretation allows us to build practical op amp configurations using only capacitors. There's a good reason for doing this: resistors are big but capacitors are small. When making circuits at the micro or nano scale, it is usually preferred to use capacitors and avoid resistors whenever possible.

Miller Integrator

When Z_2 is a capacitor and Z_1 is a resistor, as in Figure 42, the circuit is called a Miller integrator. The ideal circuit from Figure 42 suffers from a few practical difficulties:

1. What determines the **Initial Condition** of the integrator? Usually we want $v_{out} = 0$ at some starting time $t = 0$.

2. When $v_{in} = 0$, we expect $v_{out} = 0$ for all $t > 0$. However the op amp's **systematic offset voltage** creates a "ghost input" that gets integrated, so the charge on C_2 will go to ∞.

3. There is no **reset mechanism** to zero the charge on C_2. **If v_{in} is a sinusoid centered at 0, the offset will keep charging C_2 without limit.**

To resolve these difficulties, there are two common solutions. The first solution is to use a large **passive bypass resistor** R_F connected in the feedback path, as shown in Figure 47. The bypass resistor serves to zero the DC charge on C_2. If R_F is sufficiently large, then it will have minimal influence on the frequency response above DC, however it may contribute to offset effects due to the op amp's bias current and offset voltage. It will also tend to amplify any DC offset present in the input signal. At very low frequencies, we can treat the capacitor C_2 as an open-circuit, i.e. we simply remove it from the circuit. This reveals the circuit's DC behavior, an inverting integrator with gain $-R_F/R_1$. Similarly at higher frequencies where $(\omega C_2)^{-1} \ll R_F$, we can ignore the presence of R_F and the circuit should behave like an ideal integrator.

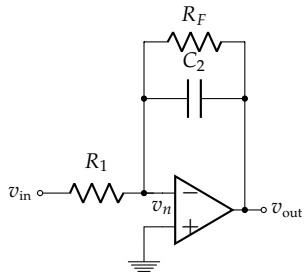


Figure 47: Miller integrator with DC bypass resistor R_F.

The second solution is to use a **switching reset** in the feedback path. We use a switch which is closed periodically to zero the charge on C_2. This method has the advantage of being insensitive to the op amp's bias current, and is only weakly sensitive to the input offset voltage. The main drawback is that the switch also resets the signal integration result, so it is not possible to integrate over a long period of time.

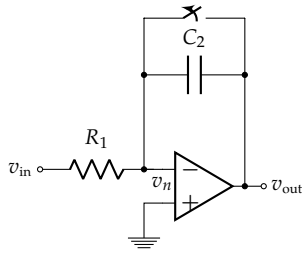


Figure 48: Miller integrator with switched DC bypass.

Frequency Analysis

The Miller Integrator introduces an interesting frequency response. We can solve the **unity-gain frequency** by evaluating

the magnitude:

$$A_{CL}(j\omega) = \frac{1}{\omega R_1 C_2}$$
$$\Rightarrow \omega_t = \frac{1}{R_1 C_2}.$$

We can draw the magnitude response by placing a point at ω_t, then draw backwards adding 20dB per decade at frequencies below ω_t. The response grows toward ∞ as the frequency approaches DC.

Introduction to Diodes

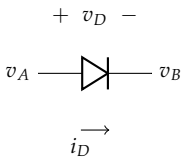


Figure 49: Diode symbol and notation.

A diode is like a **valve** that lets current flow one direction but not the other. It is a **nonlinear device**, which means the traditional linear analysis techniques cannot be directly applied. We begin with some simplified models that are useful for building intuition about diode circuits. After that, we'll build up to more accurate (but difficult) models and techniques. We'll see that accurate simulation using SPICE (or a similar software tool) is essential for designing nonlinear circuits.

Ideal switch model

The simplest way to understand a diode is to consider it as an ideal switch.

- When $v_D > 0$, the switch is closed and i_D can be *any positive value*.
- When $v_D \leq 0$, the switch is open and $i_D = 0$.

Using the switch model, we first make a **hypothesis** as to whether the diode is ON or OFF. Then, we analyze the circuit to **verify it is consistent** with that hypothesis. This approach introduces our first **iterative procedure**: If the circuit contains multiple diodes, we initially assume that **all diodes are OFF** and then analyze the circuit. Any diode with a forward voltage is then turned ON. After changing the diode's state, we must re-analyze the circuit to see if any additional devices need to be turned ON.

Example 11 (Max-Value Circuit).

Analysis steps:

1. Suppose both diodes are OFF. Then $v_C = 0$. But then both D_1 and D_2 have positive potentials across their terminals, so they cannot both be off.
2. Observe that D_2 has the **larger forward potential** across its terminals. Based on this, suppose that D_2 is ON while D_1 remains OFF. In this case, $v_C = v_B = 7\text{V}$, hence the potential across D_1 is $v_C - v_A = -3\text{V}$, which is consistent with the hypothesis.

In this example, we find that D_1 is OFF while D_2 is ON. Based on our analysis, we can generalize the result and describe this circuit by the function

$$v_C = \max(v_A, v_B).$$

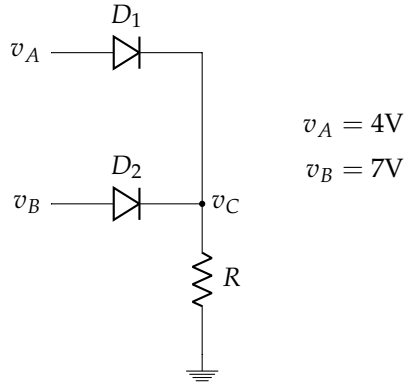


Figure 50: Diode max-value circuit.

Example 12 (Min-Value Circuit).

Analysis steps:

1. Suppose both diodes are OFF. Then $v_C = V_{\text{DD}}$.
2. Observe that D_1 has the **larger forward potential** across its terminals. Based on this, suppose that D_1 is ON while D_2 remains OFF. In this case, $v_C = v_A = 4\text{V}$, hence the potential across D_2 is $v_B - v_C = -3\text{V}$, which is consistent with the hypothesis.

In this example, we find that D_2 is OFF while D_1 is ON. Based on our analysis, we can generalize the result and describe this circuit by the function

$$v_C = \min(v_A, v_B).$$

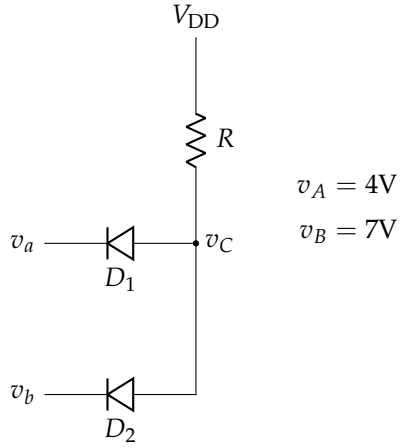


Figure 51: Diode min-value circuit.

Exponential model

A more accurate model of the diode is given by this expression:

$$i_D = I_S \left[\exp\left(\frac{v_D}{nU_T} \right) - 1 \right]$$

Notice that when $v_D = 0$, the current is also zero. When $v_D > 0$, the exponential part rapidly becomes much greater than one. When $v_D < 0$, the exponential part rapidly becomes much smaller than one. This splits the diode into two different modes: called **forward bias** and **reverse bias**, respectively.

Forward bias $i_D \approx I_S \exp\left(\frac{v_D}{nU_T}\right)$.

Reverse bias $i_D \approx -I_S$.

Diode physical device parameters:

I_S = Scale current, typ. $\approx$ 1pA to 1nA
n = Grading coefficient, typ. ≈ 1
U_T = Thermal voltage, $\frac{k_B T}{q} \approx$ 26mV at room temp.
k_B = Boltzmann constant $= 8.6173 \times 10^{-5}$eV/K
T = Temperature (K) $\approx$ 300K at room temp.
q = Elementary charge $=$ 1eV/V

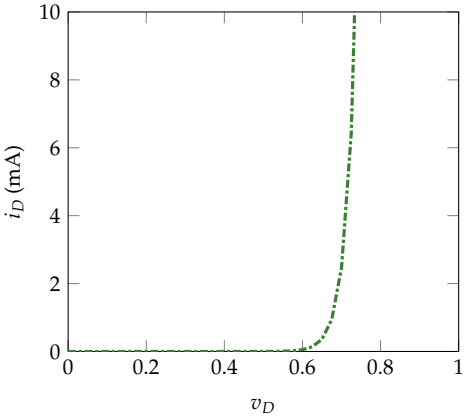


Figure 52: Diode transfer characteristic. The current increases very rapidly when $v_D \approx 0.7$ V.

Constant voltage-drop model

From the physical model, we can see that the i_D curve becomes very steep when $v_D \geq 0.7$V, so we can say this is approximately the **ON voltage** of the diode. When a diode is turned ON, it should have a nearly constant forward voltage drop equal to 0.7V.

Example 13 (Min-value circuit with 0.7V drop model).

Analysis steps:

1. Complete the analysis using the ideal switch model. We find that D1 is ON and D2 is OFF.
2. Estimate a more accurate result by adding a 0.7V forward drop to every diode that is ON, hence

$$v_C = v_A + 0.7\text{V} = 4.7\text{V}$$

Based on our analysis, we can generalize the result and describe this circuit by the function

$$v_C = \min(v_A, v_B) + 0.7\text{V}.$$

Important Note: If $|v_A - v_B| < 0.7$V, then this method does not yield any valid solution. In that case, we must use the full physical model with **iterative analysis** to arrive at the correct solution.

Iterative Analysis

The constant voltage drop model gives us an approximation that is useful for back-of-the-napkin analysis. For a more precise analysis, we must solve the voltages and currents using the full physical model. The resulting equations do not often have closed-form solutions, so we must apply an iterative method based on this procedure:

1. Obtain an initial solution using the switch model.
2. Improve the solution using the constant voltage drop model.
3. Based on that solution, calculate the resulting currents that should flow through linear elements (resistors).
4. From those currents, estimate a more exact voltage drop for each diode.
5. Repeat steps 3 and 4 in a loop until the answers **converge** to a stable answer.

Convergence: How to know when the iterations are finished

Most of the time, we do not carry out iterative analysis by hand; we use SPICE or a similar simulator to perform these calculations for us. "Under the hood," SPICE performs iterative calculations to predict a circuit's behavior. These simulators use two criteria to decide when iterations are complete: **absolute tolerance (abstol)** and **relative tolerance (reltol)**, defined as:

- **abstol** – Simulation continues until all voltages and currents satisfy
$$|\Delta x| < \text{abstol}$$
- **reltol** – Simulation continues until all voltages and currents satisfy
$$\left|\frac{\Delta x}{x}\right| < \text{reltol}$$

Most simulators will allow you to adjust the abstol and reltol parameters. Smaller values result in better accuracy, but will take more time to finish.

Example 14 (Iterative analysis).

In this example, the current i is described by two equations:

$$i = \frac{v_A - v_B}{R}$$
$$i = I_S \exp\left(\frac{v_B}{nV_T}\right)$$

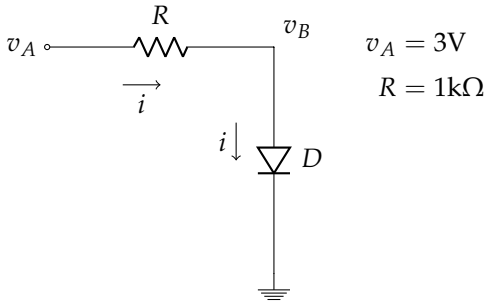


Figure 53: Iterative solution for resistor-diode series configuration.

Because of the exponential term, there is no easy solution. We may solve the circuit by iteration:

1. Using the ideal switch model, we see that D must be ON.

2. Using the 0.7V model, we obtain an initial guess

$$v_B^{(0)} = 0.7\text{V}$$
$$i^{(0)} = 1\text{mA}$$

3. Using the resistor equation, we obtain a new estimate for the current:

$$i^{(1)} = \frac{3 - 0.7}{1\text{k}\Omega} = 2.3\text{mA}$$

4. From the diode equation, we can obtain a new estimate for the voltage:

$$v_D^{(1)} = v_D^{(0)} + nV_T \ln\left(\frac{i^{(1)}}{i^{(0)}}\right) = 0.72166\text{V}.$$

5. We repeat these analyses using the generalized equations

$$i^{(k+1)} = \frac{v_A - v_B^{(k)}}{R}$$
$$v_B^{(k+1)} = v_B^{(k)} + nV_T \ln\left(\frac{i^{(k+1)}}{i^{(k)}}\right)$$

By following this procedure, we obtain the following sequence of results:

k	i [mA]	v_B [V]	Δi [mA]	Δv_B [V]
0	1	0.7	-	-
1	2.3	0.72166	1.3	0.02166
2	2.2783	0.72141	-0.021656	-2.5×10^{-4}
3	2.2786	0.72141	2.4596×10^{-4}	2.8067×10^{-6}

Notice that the changes Δi and Δv_B become smaller with each iteration. This means that the calculations are **converging** onto the correct answer, where all equations find perfect agreement.

Linearized Model

Yet another way of modeling the diode is to use a *linear approximation*.

Recall the definitions of small-signal notation, I_D and V_D are the *operating point* values, i_d and v_d are *small variations*, and i_D and v_D are the actual physical signal values. Hence

$$\begin{aligned} i_D &= I_D + i_d \\ v_D &= V_D + v_d \end{aligned}$$

Looking at this circuit, it is easy to see that $r_d = v_d/i_d$. Since v_d and i_d represent small variations, we can interpret r_d as the derivative:

$$\begin{aligned} r_d &= \frac{dv_D}{di_D} \\ &\approx \left[\frac{di_D}{dv_D}\right]^{-1} \\ &= \left[\frac{I_S}{nV_T}\exp\left(\frac{V_D}{nV_T}\right)\right]^{-1} \\ &= \frac{nV_T}{I_D} \end{aligned}$$

If we use the 0.7V model, and assume room temperature operation with $n = 1$ (so $nV_T = 0.026$V), then the values for this model are

$$\begin{aligned} V_D &= 0.7\text{V} \\ I_D &= 1\text{mA} \\ r_d &= 26\Omega \end{aligned}$$

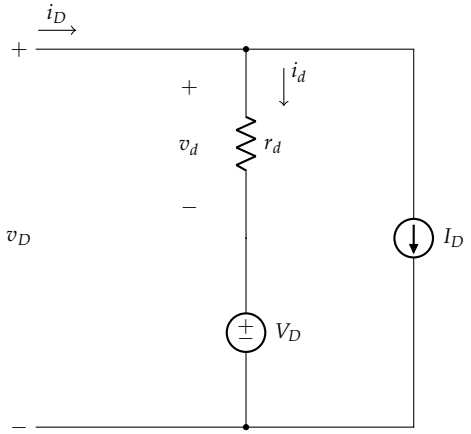


Figure 54: Linearized diode model.

Iteration with the linearized model

Iterative analysis can be combined with small-signal analysis by repeatedly recalculating r_d. In this version, the diode's circuit model looks like this:

The analysis procedure is as follows:

1. Use the constant 0.7V model to obtain an initial guess for all voltages and currents, and for $r_d^{(0)}$.
2. Using linear circuit analysis, find the solution for $v_D^{(k+1)}$.
3. Using the **non-linear** device current equation, calculate the current $i_D^{(k+1)}$ and the small-signal resistance $r_d^{(k+1)}$:

$$i_D^{(k+1)} = i_D^{(k)} \exp\left(\frac{v_D^{(k+1)} - v_D^{(k)}}{nV_T}\right)$$

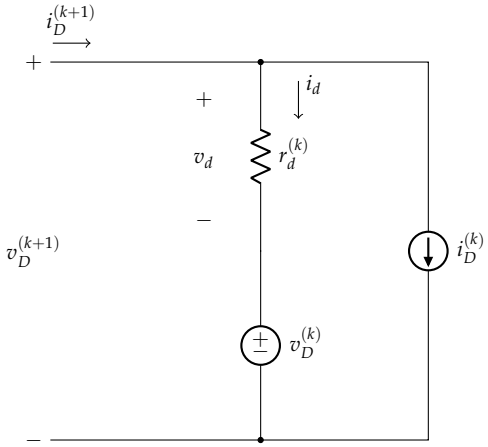


Figure 55: Iterative solution of linearized model parameters.

4. Repeat the calculations until the answer is sufficiently converged.

Beginning from the initial conditions $i_D^{(0)} = 1\text{mA}$, $v_D^{(0)} = 0.7\text{V}$ and $r_d^{(0)} = 26\Omega$, we may solve new values for the voltages and currents. Those new values may then be used to improve the calculation of r_d.

The chief advantage of using small-signal iteration is that it provides **stable** convergence for most circuits, whereas the method of direct iteration can sometimes fail. This method is mathematically equivalent to the **Newton-Raphson method**, and is the most common type of algorithm used in circuit simulators like SPICE.

Example 15 (Iteration with linearized model).

For example, we may reconsider our resistor-diode circuit:

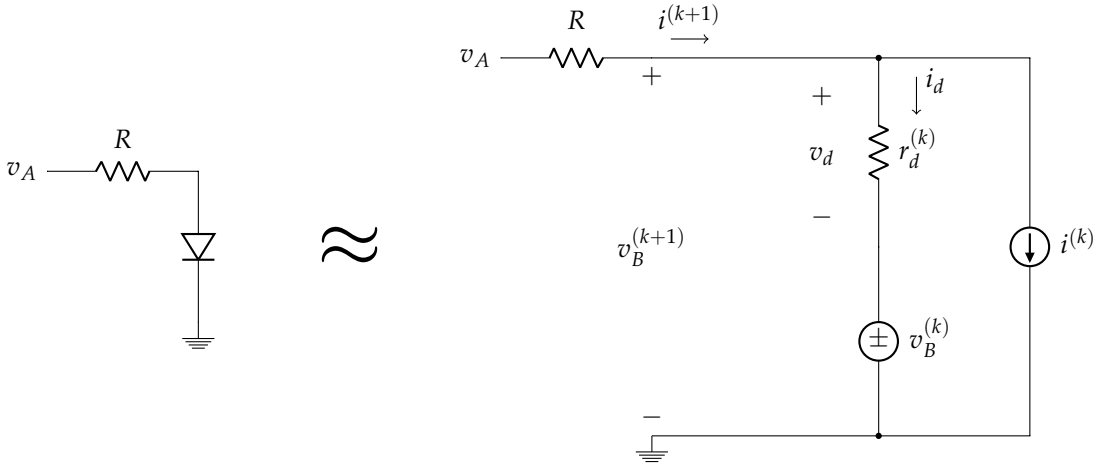


Figure 56: Iterative solution of the resistor-diode series configuration using the linearized model.

In this example, the iterative procedure yields the following table of results:

k	i [mA]	v_B [V]	r_d [Ω]	i_d [mA]	v_d [V]
0	1	0.7	26	2.25504	0.032943
1	3.5504	0.73294	7.3232	−1.0705	−0.009330
2	2.4799	0.72361	10.4843	−0.1934	−0.002112
3	2.2865	0.7215	11.3713	−0.0000079	−0.00009
4	2.2786	0.72141	11.4106	0	0

Diode Circuits

Half-Wave Rectifier

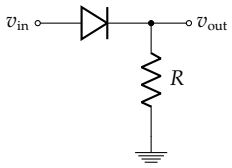


Figure 57: Half-wave rectifier circuit.

The 1/2-wave rectifier circuit passes current only when $v_{out} > 0.7\text{V}$. In this case, the diode's forward voltage drop is close to 0.7V, regardless of the current that flows, so that $v_{out} \approx v_{in} - 0.7\text{V}$. When the diode is OFF, no current flows, so $v_{out} \approx 0\text{V}$. This behavior is approximately described by the expression

$$v_{out} \approx \max\left(0,\, v_{in} - 0.7\text{V}\right).$$

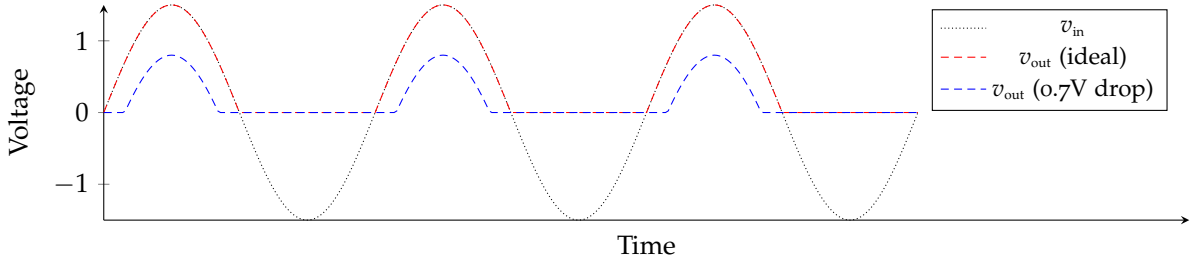


Figure 58: Behavior of the half-wave rectifier. The ideal switch model is compared to the more accurate constant-0.7 V drop model.

Example 16 (Half-wave rectifier with $v_{in} < 0$)**.**

In this example, let $v_{in} = -1\text{V}$ and $R = 1\text{k}\Omega$. We want to solve for v_{out}. First, we assume the diode is OFF and check for consistency. We find that $v_{out} = 0$ and therefore the diode's forward drop is $v_D = -1\text{V}$. Since v_D is negative, the diode must be OFF, so $v_{out} = 0\text{V}$.

Example 17 (Half-wave rectifier with $v_{in} = 1\text{V}$)**.**

In this case the diode is clearly ON. Using the constant voltage drop approximation, we can estimate that $v_{out} \approx 0.3\text{V}$. A more precise estimate may be obtained using the small-signal model:

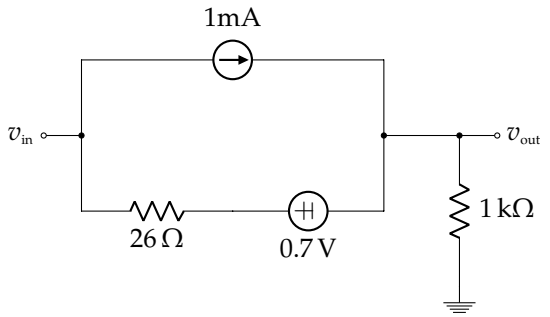


Figure 59: Linearized model of half-wave rectifier.

$$\frac{v_{out} + 0.7\text{V} - v_{in}}{26\Omega} + \frac{v_{out}}{1\text{k}\Omega} - 1\text{mA} = 0$$
$$\Rightarrow v_{out}\left(\frac{1}{26} + \frac{1}{1\text{k}\Omega}\right) = 1\text{mA} + \frac{v_{in} - 0.7}{26}$$

Now solving for v_{out}:

$$\Rightarrow v_{out} = (1\text{mA})\,(26\Omega \parallel 1\text{k}\Omega) + \frac{v_{in} - 0.7}{26}\,(26\Omega \parallel 1\text{k}\Omega)$$

Now notice that $(26 \parallel 1000) \approx 26$ (try it). Then we can simplify the approximation:

$$v_{out} \approx 26\text{mV} + v_{in} - 0.7\text{V}$$
$$\approx \mathbf{v}_{in} - \mathbf{0.684}\text{V}.$$

This provides a more accurate approximation when the resistor R is large. In the case where $v_{out} = 1\text{V}$, we find that

$$v_{out} \approx 0.343\text{V}$$
$$i_D \approx 343\mu\text{A}$$

Resistor-diode regulator

The regulator circuit is similar to the 1/2-wave rectifier, only it interchanges the positions of the diode and resistor. In this circuit, when $v_{in} < 0.7\text{V}$, the diode is either OFF or only weakly ON, so the current is close to zero. In that case, the voltage drop across R is nearly zero, so $v_{out} \approx v_{in}$. When $v_{in} > 0.7\text{V}$, the diode is clearly ON. Using the constant voltage drop model, we find that $v_{out} \approx 0.7\text{V}$, so the waveform is "clipped" at 0.7V.

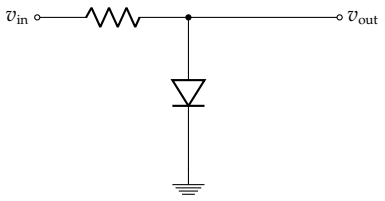


Figure 60: Single-diode regulator circuit.

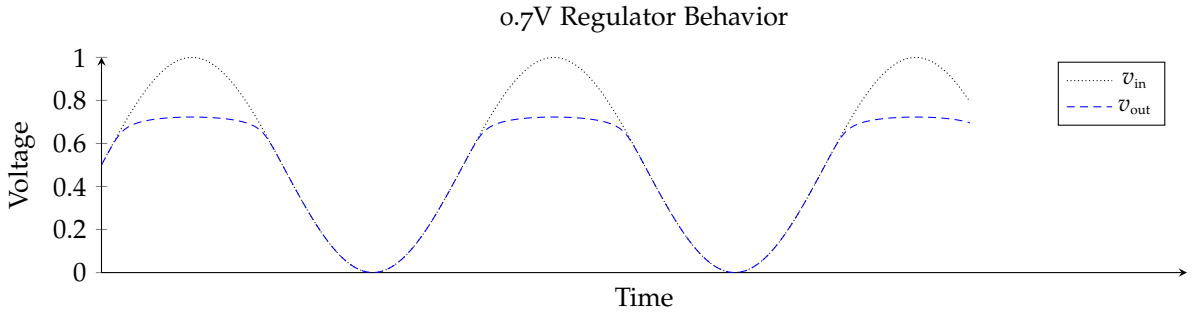


Figure 61: Behavior of the single-diode regulator circuit with $R = 100$. The results from SPICE simulation are more accurate than hand analysis.

A more accurate analysis is obtained using the linearized diode model. By applying the node-voltage method at v_{out}, we find that

$$\begin{aligned} \frac{v_{out} - v_{in}}{R} + \frac{v_{out} - 0.7\text{V}}{26\Omega} + 1\text{mA} = 0 \\ \Rightarrow v_{out}\left(\frac{1}{R} + \frac{1}{26\Omega}\right) = \frac{v_{in}}{R} + \frac{0.7\text{V}}{26\Omega} \\ \Rightarrow v_{out} = v_{in}\frac{26\Omega}{26\Omega + R} + 0.7\text{V}\frac{R}{R + 26\Omega} \end{aligned}$$

As the name implies, regulators are used to produce stable DC voltages. Ideally, a regulator should produce 0.7V regardless of v_{in} (so long as $v_{in} > 0.7\text{V}$). The preceding analysis revealed a slight dependency between v_{in} and v_{out}:

$$\Delta v_{out} = \Delta v_{in}\left(\frac{26\Omega}{R + 26\Omega}\right).$$

In practice, the residual Δv_{out} signal can introduce interference into the circuits that are interfaced with the regulator. According to this analysis, the regulation works best when R is large.

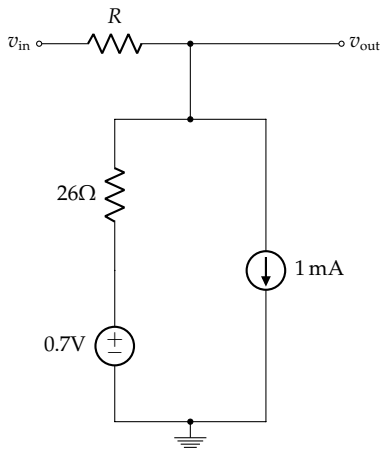
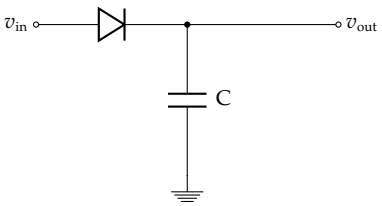


Figure 62: Linearized model of the single-diode regulator.

Peak rectifier

The peak rectifier (or peak detector) circuit is like a rectifier that uses a capacitor in place of the resistor. This circuit can be interpreted as an **integrating rectifier**. Unlike the usual diode circuits, the 0.7V approximation can be misleading when applied to the peak rectifier.

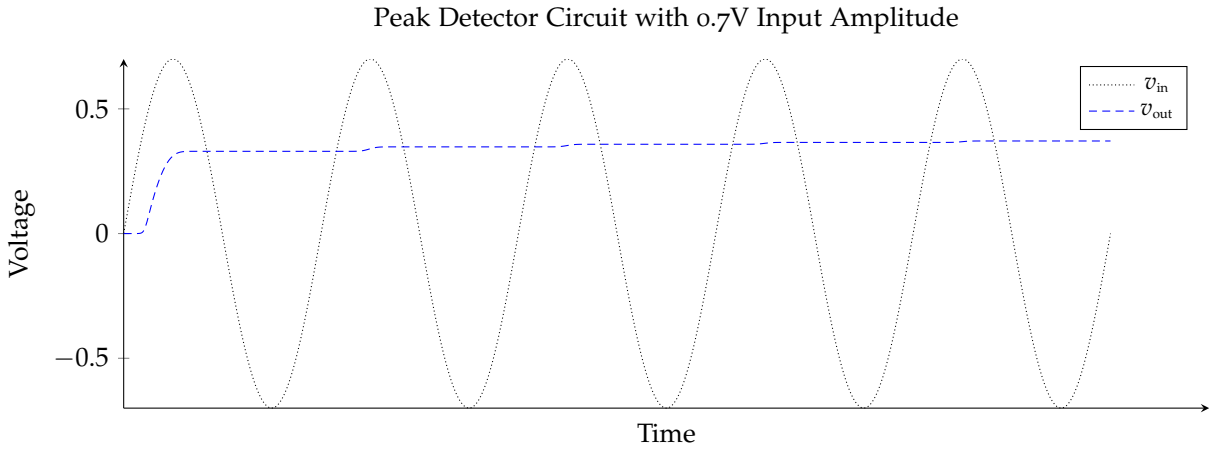


This is because the capacitor integrates all of the current that passes through it:

$$v_{out} = \frac{1}{C} \int_0^{t_F} i_D(t)\, dt.$$

When the diode is OFF, a small current still flows, and that current is steadily accumulated by the capacitor's integrating behavior.

Consider the output from a SPICE simulation where $C = 1\text{nF}$, shown below. In this simulation, we can see that v_{out} rises initially to 0.263V, which is approximately $v_{in} - 0.437\text{V}$. Clearly the 0.7V model is not working.



To understand why the 0.7V model fails, we may examine the diode current, shown in the figure below. Although the current never exceeds 1µA, the small pulses are sufficient to charge C.

In each cycle of the input waveform, the peak current gets smaller, so the output waveform marches in smaller and smaller steps toward the peak value of the input voltage. Given enough time, v_{out} will eventually rise very close to the actual peak. This effect can be used to create AC-to-DC converters.

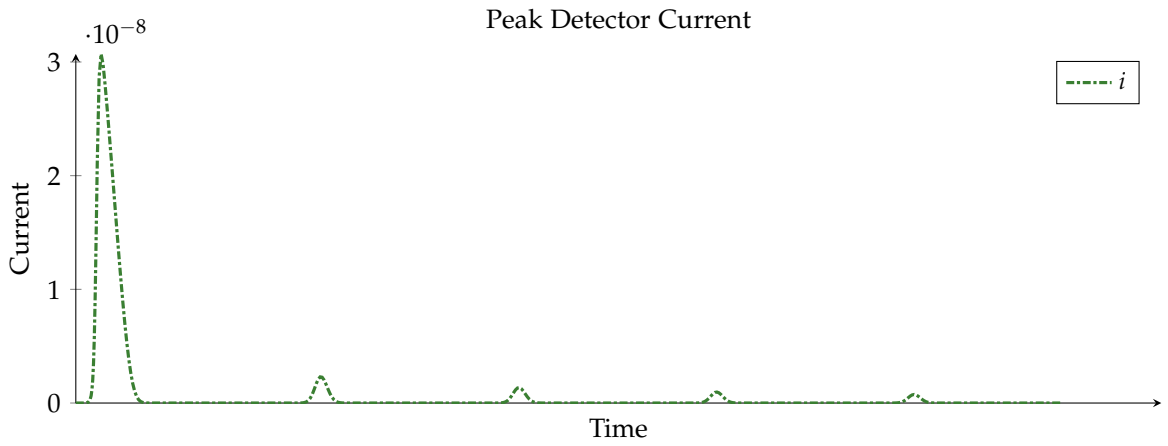


Figure 63: Current delivered into the capacitor in the peak detector circuit.

Envelope detector

The peak detector circuit can also be used in a variety of applications for instrumentation and communication. In these applications, we usually want to detect the **envelope** of some waveform, which requires that v_{out} be allow to drop when v_{in} decreases. This is accomplished by adding a resistor R in parallel with C, resulting in an *envelope detector*:

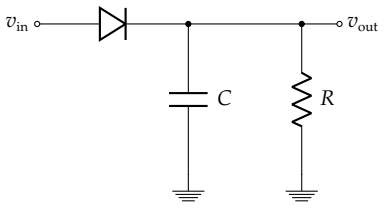


Figure 64: Envelope detector circuit.

An example SPICE simulation result is shown in the plot below. This simulation used the following values:

$$C = 10\mu F$$
$$R = 10k\Omega$$
$$f = 10Hz$$

In this circuit, the diode is able to rapidly charge the capacitor C, which is then slowly discharged by R.

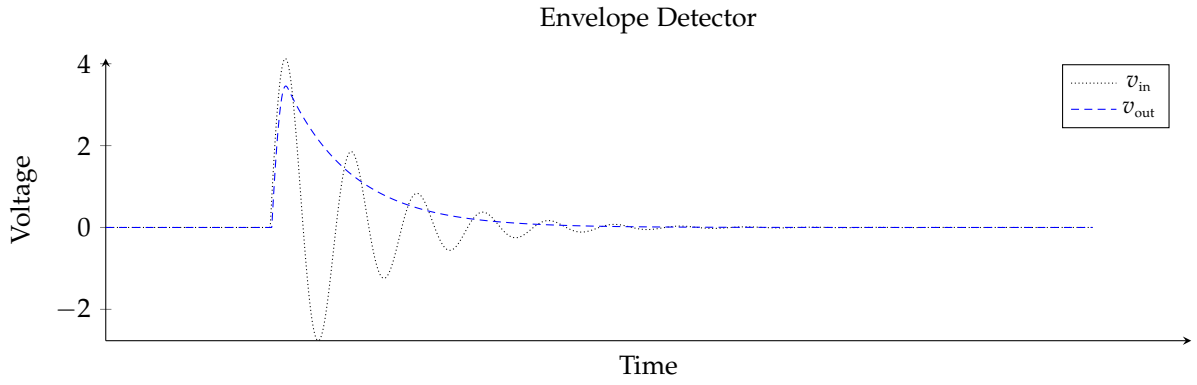


Figure 65: Behavior of the envelope detector circuit.

When the didoe is OFF, the output waveform is described by

the standard RC discharge equation

$$v_{\text{out}}(t) = v_{\text{peak}}\left(1 - \exp\left(-RC\left(t - t_0\right)\right)\right),$$

where t_0 is the time when the diode turns OFF.

Netlist 1: envelope_detector.sp

```
* envelope detector circuit

* Generic diode model:
.model diode d(Is=2.0298e-15, n=1)

* The input is a damped 10Hz sine wave that
* looks like an impulse:
Vin 1 0 SIN(0 5 10 0.25 8)

* Peak detector circuit:
D1 1 2 diode
C1 2 0 10uF
R1 2 0 10k

* Transient simulation:
.tran 1m 1.5

.end
```

Bridge Rectifier

A bridge rectifier circuit, shown below, provides **full-wave** rectification. Node numbers in the figure are indicated in blue, corresponding to the example SPICE description.

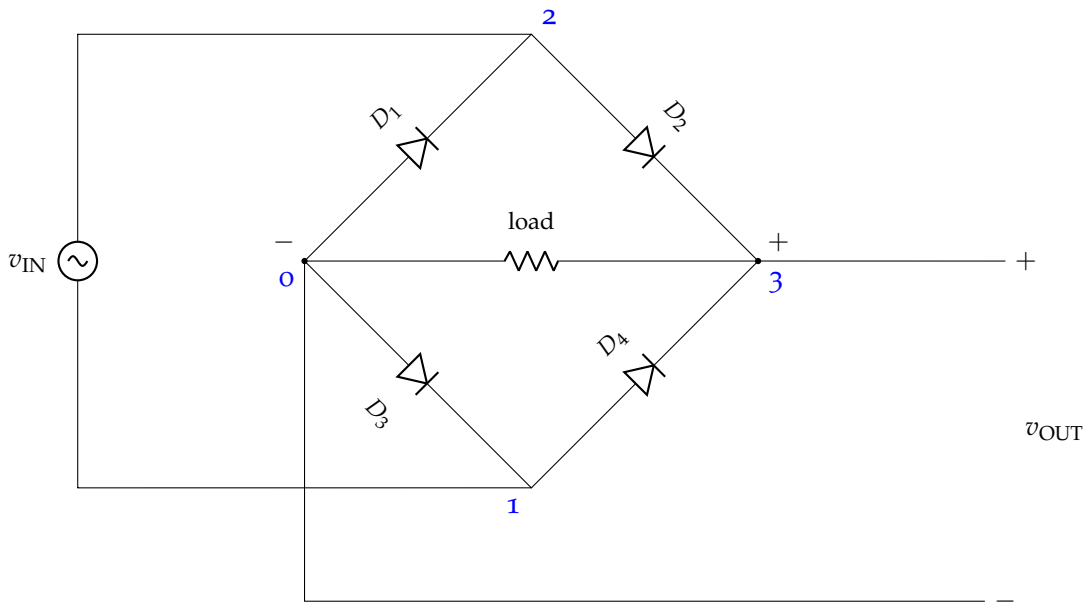


Figure 66: Full wave bridge rectifier circuit.

Analysis:

Case 1: $v_{\text{IN}} > 0$. In this case, we see that **the most positive potential** appears at the anode of D_2. Based on this, we may predict that D_2 is ON while D_1 is OFF. Since the **most negative potential** appears at the anode of D_4, we may conclude that D_4 is OFF.

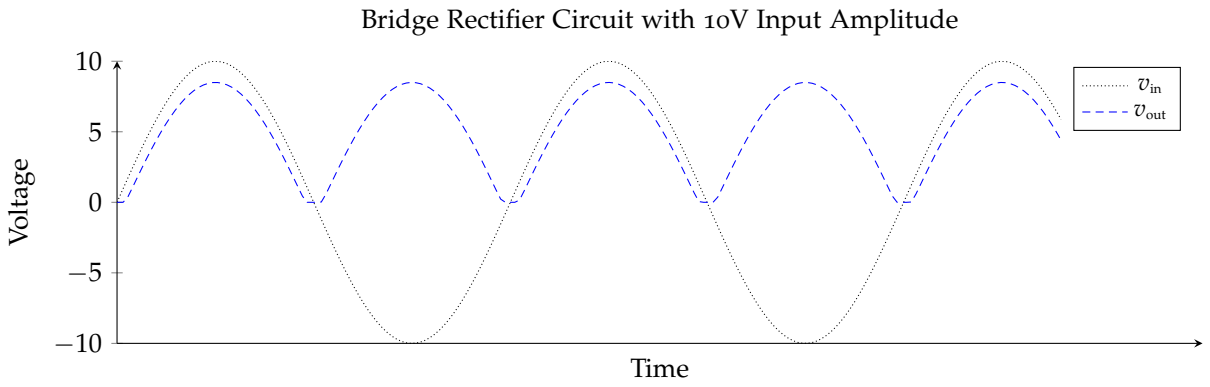
Based on this reasoning, we infer that the current flows in a zig-zag through D_2, then the load, then D_3. The potential appearing across the load is

$$v_{\text{OUT}} \approx v_{\text{IN}} - 1.4\text{V}.$$

Case 2: $v_{\text{IN}} < 0$. In this case the most positive potential appears at the anode of D_4, and the most negative potential appears at the cathode of D_1. We may conclude that the current flows in a zig-zag through D_4, then the load, then D_1. In this case the potential appearing across the load is

$$v_{\text{OUT}} \approx |v_{\text{IN}}| - 1.4\text{V}.$$

The bridge arrangement ensures that the polarity across the load is always oriented right-to-left, regardless of the input polarity.



SPICE simulation example for the bridge rectifier:

Netlist 2: bridge_rectifier.sp

```
* bridge rectifier circuit

* Generic diode model:
.model diode d(Is=2.0298e-15, n=1)

* The input is a 120Hz sine wave:
Vin 2 1 SIN(0 10 120)

* Bridge rectifier:
D1 0 2 diode
D2 2 3 diode
D3 0 1 diode
D4 1 3 diode

* Load resistor:
Rload 3 0 1k

* Transient simulation:
.tran .1m 0.02

.end
```

Voltage Regulators

We previously considered a 0.7V regulator circuit. We can extend this concept to produce other regulated voltages by connecting multiple diodes in series. For example, we may connect four diodes in series to create a 2.8V regulator circuit:

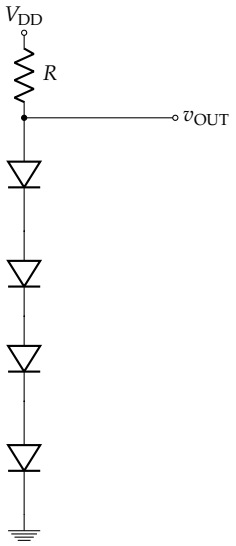


Figure 67: A four-diode voltage regulator.

Ripple Analysis: Line Regulation

The regulator is able to reject ripple waveforms that appear in the supply voltage, however the rejection is not perfect. A close inspection reveals that a small ripple is injected into v_{OUT}:

The regulator's quality is measured by the amount of ripple that appears in v_{out}. More precisely, we want to know the ratio of output ripple amplitude to input ripple amplitude. This quantity is called the **line regulation**, defined as

$$LR = \frac{\Delta v_{OUT}}{\Delta V_{DD}}$$

To predict this, we must calculate the **small-signal** gain of AC signals that are transferred from v_{in} to v_{out}. We previously introduced a **small-signal** model that allows each diode to be replaced by a linear approximation. Now we introduce the concept of an **AC Equivalent Circuit** which we can use to analyze the non-DC behavior.

Deriving the AC Equivalent Circuit

Step 1 To obtain the linear circuit approximation, replace all non-linear devices (e.g. diodes) with their linearized companion models, as in previous examples.

Step 2 To obtain the AC equivalent circuit, set all independent DC sources to zero. This means that independent current sources are replaced by open-circuits, and independent voltage sources are replaced by short-circuits.

After obtaining the AC equivalent circuit, we use all-lower-case notation to indicate the ripple waveforms v_{in} and v_{dd}. Using the AC equivalent circuit, we can solve for the line regulation as the ratio of these small signals:

$$LR \approx \frac{v_{out}}{v_{dd}}$$

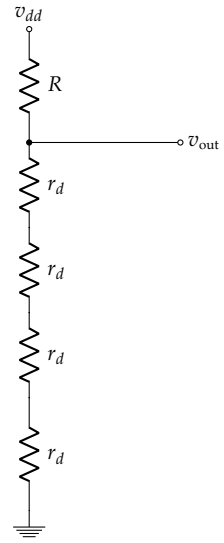


Figure 68: Small-signal equivalent circuit model of the four-diode 2.8 V regulator.

Reminder: The lower-case signal v_{out} represents the small ripple signal appearing in the output. The all-upper-case notation V_{OUT} is used to represent the DC (average) value. The actual physical signal is $v_{OUT}(t) = V_{OUT} + v_{out}(t)$.

Example 18 (Four-diode voltage regulator design).

Let

$$v_{in} = 10\text{V} + (0.5\text{V}) \sin(2\pi f t).$$

Basic analysis

Find R to get an average current of 1mA, resulting in $v_{out} = 2.8\text{V}$.

$$\begin{aligned} 1\text{mA} = I &= \frac{v_{in} - v_{out}}{R} \\ \Rightarrow R &= \frac{v_{in} - v_{out}}{I} \\ &= 7.2\text{k}\Omega \end{aligned}$$

The behavior of this circuit is investigated using SPICE simulation. The results shown below include a supply ripple with zero-to-peak amplitude of 0.5V at 120Hz. From SPICE simulations, we see that the actual output voltage is 2.7863V, which is slightly less than the intended value. The ripple amplitude is also found to be 14.158mV.
By using the small-signal model, we can obtain a reasonable estimate of v_{out}, the small ripple waveform that is superimposed on the regulator's output:

$$\begin{aligned} v_{in} &= (0.5\text{V}) \sin(2\pi f t) \\ v_{out} &= v_{in} \frac{4r_d}{R + 4r_d} \\ &= v_{in} \left(\frac{104}{7200 + 104} \right) \\ &= v_{in} (0.014239) \\ &= (7.12\text{mV}) \sin(2\pi f t) \end{aligned}$$

Then the line regulation is

$$\begin{aligned} LR &= \frac{v_{out}}{v_{dd}} \\ &= \frac{7.12\text{mV}}{500\text{mV}} \\ &= 0.014239 \\ &= 1.4239\% \end{aligned}$$

Netlist 3: basic_regulator.sp

```
* 2.8V regulator circuit

* Generic diode model:
.model diode d(Is=2.0298e-15, n=1)

* The input is a 120Hz sine wave with a 10V offset.
* The supply ripple amplitude is 0.5V
Vin 1 0 SIN(10 0.5 120)

* Regulator circuit
* The output is at node 2
R1 1 2 300
D1 2 3 diode
D2 3 4 diode
D3 4 5 diode
D4 5 0 diode

* Transient simulation:
.tran .1m 0.02

.end
```

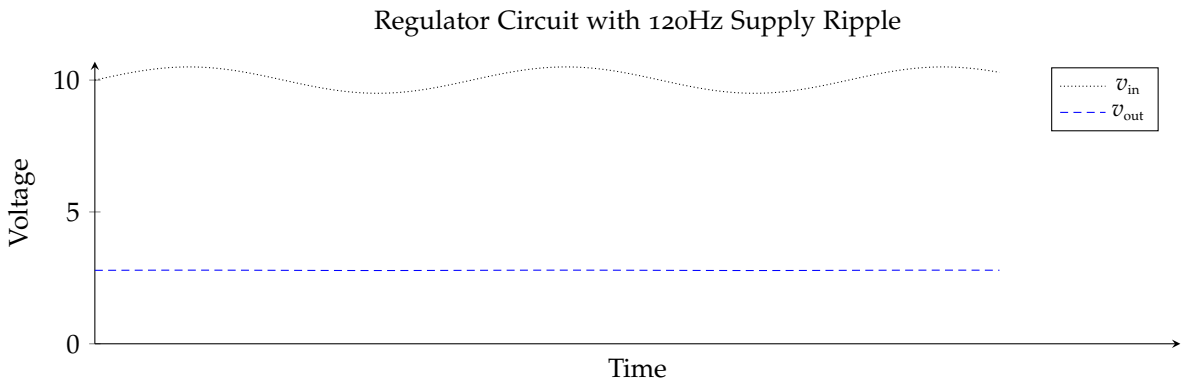


Figure 69: Behavior of the four-diode regulator from SPICE simulation.

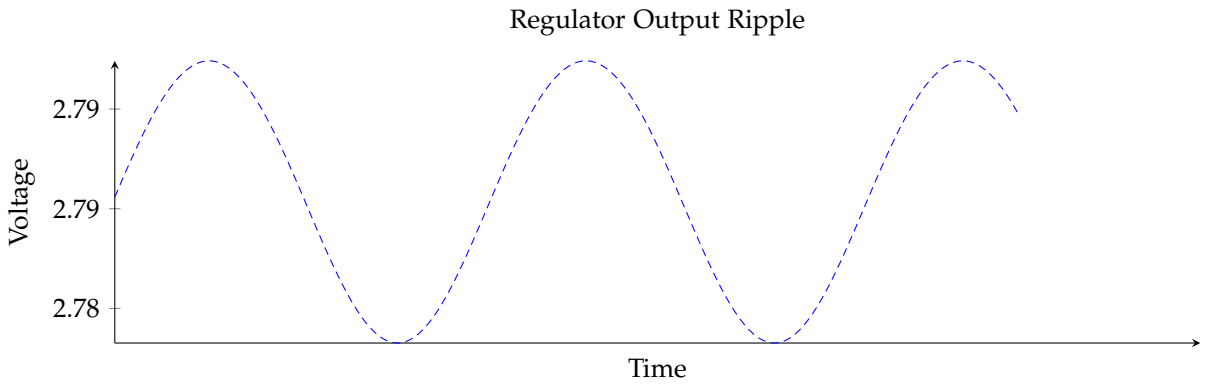


Figure 70: Zoomed view of the ripple voltage on v_{OUT}.

General Analysis

In general, for a diode circuit comprised on N diodes with **bias current** I_D, generated by an input voltage V_{IN} and resistance R, we can produce a regulated voltage $V_{\text{OUT}} = NV_D$, where V_D is the individual diode voltage associate with I_D. Then the line regulation is

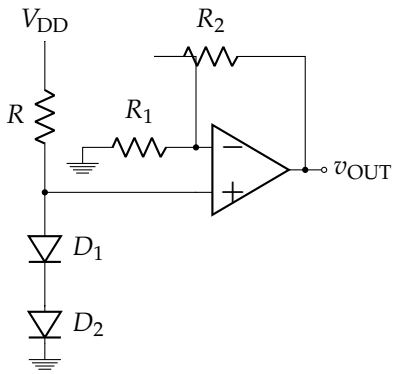
$$\begin{aligned} LR &= \frac{NV_T/I_D}{R + NV_T/I_D} \\ &= \frac{NV_T}{RI_D + NV_T} \\ &= \frac{NV_T}{V_{\text{DD}} - V_{\text{OUT}} + NV_T}. \end{aligned}$$

The smaller we make this value, the better quality we will provide on the regulator output. Things that achieve good quality regulation include:

- A large voltage drop RI_D, i.e. v_{in} should be significantly greater than the regulated v_{out}.

- A small number of diodes N.

Example 19 (Two Diode Regulator with Op Amp Buffer).



Another approach is to use two diodes at 1mA to create a 1.4V reference, which is then multiplied using a non-inverting op amp configuration to yield 2.8V:
For this configuration, the line regulation is:

$$\begin{aligned} LR &= (G)\,\frac{2V_T}{(10\text{V} - 1.4\text{V}) + 2V_T} \\ &= 0.01202\,(\text{V/V})\,. \end{aligned}$$

So it appears that this solution is slightly better, although it may be affected by the tolerances on R_1 and R_2, as well as the op amp's input bias current and finite gain.

Super Diode, Precision Rectifier

This circuit operates in two modes. When the diode is **forward biased**, it is a unity-gain follower. Note that in this configuration v_D can be very near zero, because little current is required to regulate the op amp's inverting terminal.

When the diode is **reverse biased**, the op amp is disconnected from the output node. Therefore it delivers no current to the load, and $v_{out} = 0$. Note that in this configuration, the op amp's loop is open, which will cause v_a to **rail** negative. Because of this issue, this circuit is best used with a single-sided power supply.

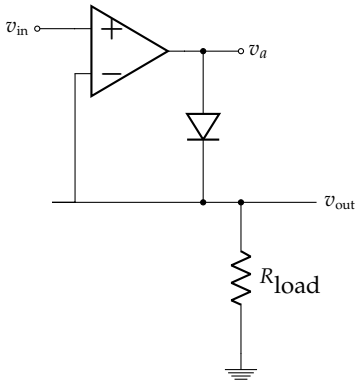
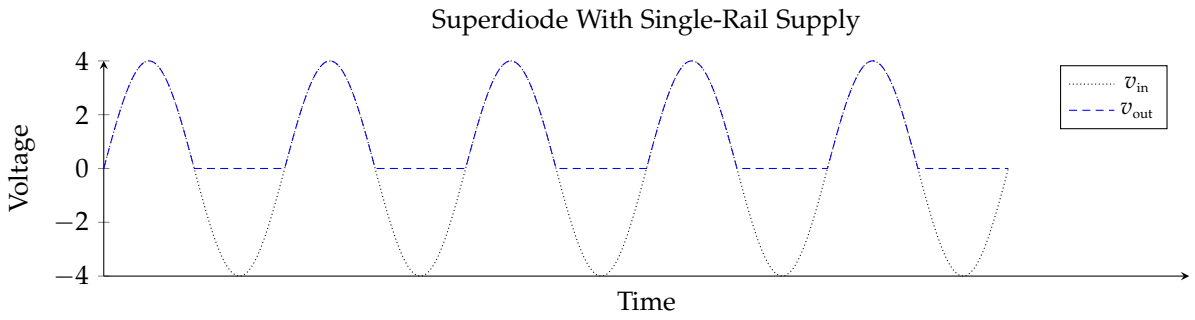


Figure 71: Precision rectifier circuit with op amp feedback.



Netlist 4: superdiode.sp

```
* super-diode precision rectifier simulation

* Include model for 741 op amp:
.include 741.sp

* Generic diode model:
.model diode d(Is=2.0298e-15, n=1)

VIN 1 0 SIN(0 4 10)
VDD 10 0 DC 10V

* 741 instance
* Pin order: v+ v- VR+ VR- vo
X1 1 2 10 0 3 uA741

D1 3 2 diode
RL 2 0 1k

.tran 1m 0.5

.end
```

The superdiode netlist uses a SPICE model for the uA741 op amp. The model is provided by the vendor, and the usage is documented in the model file:

Netlist 5: 741.sp (top lines showing port order)

```
* SPICE model for uA741 op amp
*
* To use a subcircuit, the name must begin with 'X'.  For example:
* X1 1 2 3 4 5 uA741
*
* connections:   non-inverting input
*                |  inverting input
*                |  |  positive power supply
*                |  |  |  negative power supply
*                |  |  |  |  output
*                |  |  |  |  |
.subckt uA741    1  2  3  4  5
```

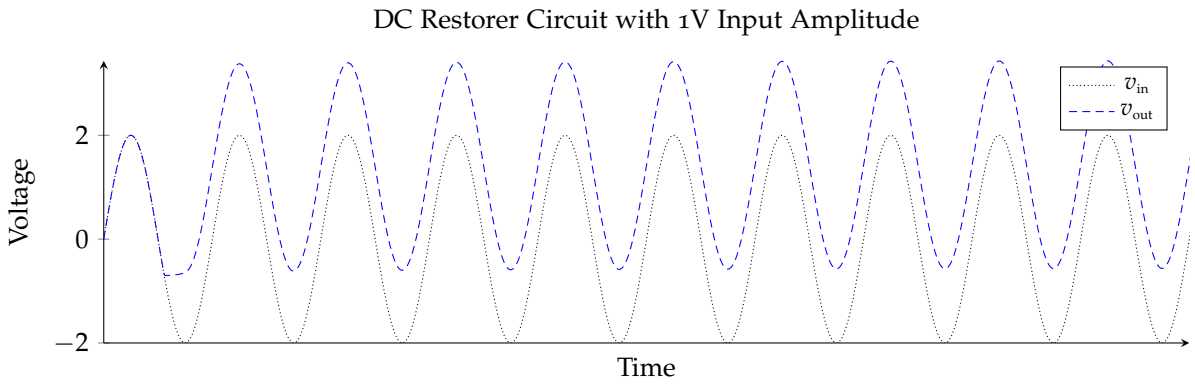


Figure 73: Behavior of DC restorer (clamped capacitor) circuit simulated in SPICE.

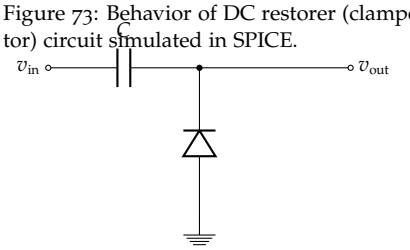


Figure 72: Clamped capacitor circuit.

DC Restoration, Clamped Capacitor

In this circuit the behavior depends on the capacitor's charge q.

$$v_{out} = v_{in} + \frac{q}{C}$$

When the diode is **forward biased**, the capacitor is able to be charged via current flowing through the diode. When the diode is **reverse biased**, no current flows, so that capacitor **holds** its charge.

To analyze the circuit, consider the initial condition $q(t = 0) = 0$, so that initially $v_{out} = v_{in}$. Suppose v_{in} is initially zero, and increases above zero. Then the diode will stay reverse biased, and q doesn't change.

But if v_{in} decreases below zero, then the diode will begin to switch on. The capacitor will accumulate charge equal to

$$\begin{aligned} q(t) &= i_D t \\ &= I_S \exp\left(\frac{-v_{out}}{V_T}\right) t. \end{aligned}$$

This current will be greater than zero as long as $v_{out} < 0$. Consequently, **the capacitor will collect charge until $v_{out} = 0$.** As a result of this process, the capacitor will store a voltage equal to the **minimum value** of v_{in}.

Result: v_{out} is a shifted version of v_{in}, such that its minimum value is equal to zero.

Netlist 6: dc_restorer.sp

```
* DC restoration circuit

* Generic diode model:
.model diode d(Is=2.0298e-15, n=1)

* The input is a 10Hz sine wave:
Vin 1 0 SIN(0 2 10)

* Peak detector circuit:
D1 0 2 diode
C1 1 2 10uF

* Transient simulation:
.tran 1m 2

.end
```

Boost converter

Diodes are frequently used in power conversion circuits. In this appendix we look at one important step-up DC-to-DC converter circuit, known as the *boost converter*. The boost converter consists of an inductor, a diode, a switch and a load capacitance, as in the schematic shown in Figure 74.

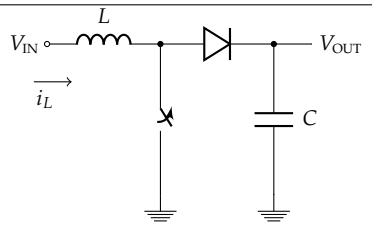


Figure 74: Idealized boost converter circuit.

When the switch closes, the inductor is shorted to ground, resulting in a large current. When the switch opens, the inductor's current cannot change instantly, so the current is forced through the diode into the capacitor. This establishes a large potential across the capacitor.

More precisely, suppose that the switch is initially open and the current i_L is zero. Then, at time $t = 0$, the switch closes abruptly. The current is then

$$i_L(t) = \frac{1}{L}\int_0^t V_{IN} d\tau = \frac{V_{IN}}{L} \times t.$$

Then if the switch opens again at some time t_1, the inductor will possess a stored energy equal to

$$E(t_1) = \frac{1}{2} L I^2.$$

Since the current must continue flowing through the inductor, most of this energy is transferred into the capacitor, and some of the energy is dissipated in the diode. For our purposes, we'll use the ideal switch model and ignore losses in the diode, hence we will assume all the energy is transfered to the capacitor.

If the switch is toggled very rapidly, with period T, then the current i_L will "ripple" up and down, transferring packets of energy in each cycle. Suppose the switch is closed for a time DT, where D is the *duty cycle* of the switching clock. Then the switch is open for a time $(1-D)\,T$. At steady state, the current should grow and shrink by the same amount:

$$\Delta i_L(\text{on}) = V_{\text{IN}} \frac{DT}{L}$$
$$\Delta i_L(\text{off}) = -\left(V_{\text{IN}} - V_{\text{OUT}}\right) \frac{(1-D)\,T}{L}$$

If we set the rise and fall equal to each other (as required for steady-state operation), then we can solve for V_{OUT}:

$$V_{\text{OUT}} = \frac{V_{\text{IN}}}{1-D}$$

Note that this analysis only works if the switching is very fast, so that the inductor current never drains completely to zero.

If the switching clock has a 50% duty cycle, then the circuit acts as a voltage doubler. An example SPICE simulation follows with $V_{\text{IN}} = 20\,\text{V}$ and a switch frequency of 5 MHz with a duty cycle of 62%. The expected output is $V_{\text{OUT}} = 52.6\,\text{V}$. The simulation output, shown in Figure 75, approaches the expected limit after about 1 ms, which corresponds to five thousand switching cycles in this example. Note that the simulation *does* account for energy lost in the diode. The main effect of diode losses is that the circuit takes longer to approach the steady state condition. But just like with the peak detector circuit, the diode's loss gets smaller in each cycle, and some energy is delivered to the capacitor in each cycle, so the circuit asymptotically approaches the ideal limit.

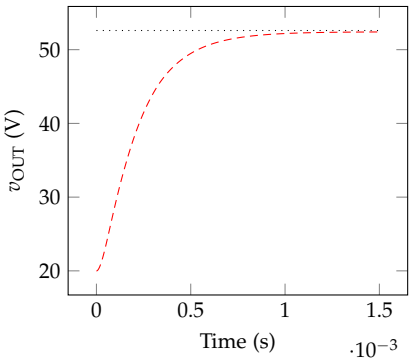


Figure 75: Output voltage from the boost converter when initialized at zero.

Netlist 7: boost_converter.sp

```
* Boost converter simulation

* Generic diode model:
.model diode d(Is=2.0298e-15, n=1)

* switch model:
.model switch sw(Ron=5, Roff=100000, Vt=0.001, Vh=0.0001)


* The input is 10V DC
Vin 1 0 DC 20V
```

```
* The switch control voltage is a high-frequency pulse waveform
Vswitch 4 0 PULSE(0 1 0 1n 1n 125n 200n)

* Boost converter circuit:
* Inductor:
L1 1 2 100u

* Diode, load capacitor and switch:
D1 2 3 diode
CL 3 0 10u
S1 2 0 4 0 switch

.tran 100n 1500u

.end
```

Memristors

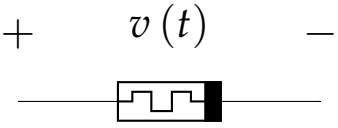


Figure 76: Memristor symbol. The device acts like a time-varying resistor, where the resistance changes in response to $\int v\,(t)\ dt$.

The "memristor" refers to a nonlinear two-terminal device for which the resistance changes in response to an applied voltage or current signal, and in which the resistance stays constant for some period of time when the signal is removed. This behavior is described as a "**mem**ory **r**es**istor**," hence the abbreviated name, "memristor". A device which exhibits resistance memory is said to be a "memristive" device or system.

Memristive devices were first observed more than a century ago, but they lacked a general theory to understand and apply their behavior in the context of circuit engineering. The first general theory was articulated by Leon Chua in two papers, written in 1971 and 1976 [2]. Chua's theory began as a somewhat speculative hypothesis. Interest surged in memristors after a group from HP Labs published a paper titled "The missing memristor found," which showed that nano-scale resistance-switching devices have the characteristics predicted by Chua's theory [3].

[2] ; and

[3]

Axiomatic Circuit Theory

Memristor theory grew out of Chua's axiomatic formulation of nonlinear circuit theory, which accounts for conservation laws (i.e. Kirchoff's Voltage and Current Laws) in networks of nonlinear devices. There are four "fundamental" quantities which must be conserved in any circuit:

- Voltage v (KVL)
- Current i (KCL)
- Charge q (conservation of matter)
- Flux linkage ϕ (conservation of energy)

Among these quantities, flux linkage (we'll call it "flux" for short) tends to cause a lot of confusion, which has led to some

serious debates over the interpretation of memristor theory. We'll briefly examine the theoretical foundations in order to prevent any confusion later.

In a coil inductor, the flux is associated with the magnetic field flux through the plane of a wire loop. In the general theory of circuits, however, flux linkage is a broader concept, and does not always have a magnetic field interpretation. In the theory of circuits, the flux is defined strictly as the integral of voltage over time:

$$\phi \triangleq \int v(t)\, dt.$$

This definition does not need to invoke any concept of a magnetic field. In fact, **the theory of circuits does not reference any electric or magnetic field quantities at all; it is concerned strictly with the relationships between current, voltage, charge and flux in a network of interacting components.** This theory represents physical laws that are as fundamental as electromagnetism, and are valid irrespective of the presence of electric or magnetic fields.

Some recent research hints at a genuine link between magnetic fields and memristor flux within a specific type of resistance-switching memristor . The correctness and applicability of this theory are not yet established. It could help to clear some of the confusion about flux, but such an explanation is not strictly necessary, since the circuit theory definition of flux stands on its own.

We could say that it is a *sufficient theory* for circuit interactions, meaning it is a set of laws or *axioms* which suffice to account for all of the applicable facts within a specified domain. For example, we can state a sufficient theory of addition on the natural numbers $(0, 1, 2, \ldots)$ without considering the concepts of multiplication, division, fractions, rational or irrational numbers, prime numbers, etc. Those additional concepts have no importance to the theory of adding natural numbers. In a similar way, the full theory of electromagnetism is mostly unnecessary for studying circuit networks.

So what is the theory of circuits really about? The theory treats every physical device as a **black box**, defined only by the activity of v, i, q and ϕ on its terminals. It it assumes there are no significant field interactions *between* devices or wires, and addresses two basic questions:

1. What types of device behavior are allowed under the physical laws?

2. How do the devices interact when connected together via wires?

The answers to these questions dictate what kinds of circuits are possible to build, and also provide the foundation for analyzing and simulating those circuits. In effect, this theory forms the basis for SPICE and other circuit simulation methods.

Constitutive Relations

In order to define the behavior of "black box" devices, we begin with a standard format, called the constitutive relation, that can be used to describe any passive two-terminal device. A constitutive relation is an expression that defines a conservative relationship between two of the fundamental quantities. The traditional linear components are defined by familiar relations, expressed in both their usual and differential forms

usual form	differential form
$v = R\,i$	$dv = R\,di$
$q = C\,v$	$dq = C\,dv$
$\phi = L\,i$	$d\phi = L\,di$

In addition to the three standard passive elements, we may define two more constitutive relations based on the definitions of current and flux:

integral form	differential form
$q = \int i\,dt$	$dq = i\,dt$
$\phi = \int v\,dt$	$dv = \phi\,dt$

This brings us to a total of **five relations among the four quantities**, as illustrated in the diagram in Figure 77. Chua noticed that **one relation is missing: the relationship between ϕ and q.** All other relationship pairs are defined, so why is this one absent? He named the hypothetical missing element the memristor and proceeded to analyze what kind of characteristics it should have, according to the physical laws of circuit theory.

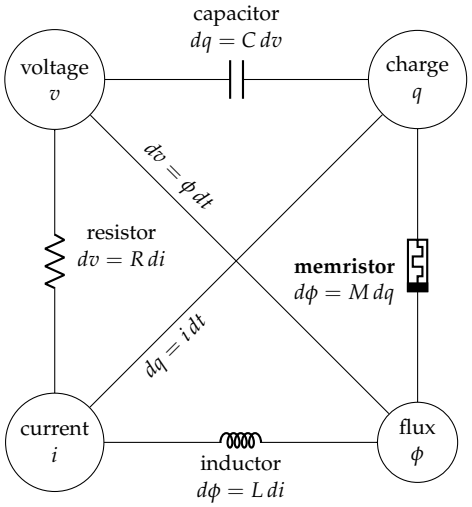


Figure 77: An illustration of the constitutive relationships corresponding to standard components and definitions. One relation is missing in the classical theory: the memristor.

"The memristor is not an invention. Rather it is a description of a basic phenomenon of nature that manifests itself in various dissipative devices, made from different materials, internal structures and architectures."

– Prodromakis, Toumazou & Chua, "Two centuries of memristors" *Nature Materials* vol. 11, 478–481 (2012)

Nonlinear Devices

The memristor is an inherently nonlinear device. In order to understand it, we first need to examine the fundamentals of nonlinear constitutive relations. Nearly all real-world devices exhibit some degree of nonlinearity. The format of constitutive relations allows us to define arbitrary nonlinear behaviors which can then be analyzed using computer simulations.

We begin with nonlinear versions of the standard elements. Since computer simulators will typically work by analyzing **small-signal linearized models** of nonlinear circuits (similar to the methods we used to analyze diodes), nonlinear constitutive

relations are represented in differential form:

$$\begin{aligned} dv &= R(v,i)\,di && \text{non-linear resistor} \\ dq &= C(v,q)\,dv && \text{non-linear capacitor} \\ d\phi &= L(i,\phi)\,di && \text{non-linear inductor} \end{aligned}$$

Notice that in these nonlinear relations, the R, C and L values are not constant; they functions of their electrical state. This theory represents a wide generality of phenomena – for example, a diode can be considered as a nonlinear resistor.

Example 20 (Diode as a nonlinear resistor)**.**

Consider a forward biased diode with the standard equation

$$i = I_S e^{v/nU_T}.$$

We can define the small-signal resistance of this diode as

$$\begin{aligned} R(v,\,i) &= \left[\frac{di}{dv} I_S e^{v/nU_T}\right]^{-1} \\ &= \frac{nU_T}{I_S e^{v/nU_T}} \\ &= \frac{nU_T}{i} \end{aligned}$$

So, for small changes in the current and voltage around the neighborhood around $(v,\,i)$, we can say that the diode's constitutive relation is

$$dv = \frac{nU_T}{i}\,di.$$

In a transient computer simulation for a circuit containing diodes or other nonlinear devices, we begin with an initial state using an iterative technique. Then we slowly advance time in very small steps, each time solving for the small changes dv, di, dq and $d\phi$ at all points in the circuit. We add the changes to the circuit's state and repeat, until reaching the simulation's stop time.

Now the memristor's constitutive relation is given by

$$d\phi = M(\phi,\,q)\,dq.$$

In order to understand this relation, it is helpful to translate it into the domain of voltage and current:

$$\begin{aligned} \frac{d\phi}{dt} &= M(\phi,\,q)\,\frac{dq}{dt} \\ \Rightarrow v &= M(\phi,\,q)\,i. \end{aligned}$$

That looks like a resistor equation, except the resistance is a function of ϕ and q. And since ϕ and q are defined as the time-integral over voltage and current, respectively, we can say that the resistance in M is **a function of the entire history of v and q applied to the memristor**. This long-term memory effect is the reason behind the name "memory resistor."

Memristor Properties

Chua authored several papers exploring the possible attributes of a memristor device. The most important features can be deduced from the constitutive relation itself. First of all, since the memristor behaves like a time-varying resistance, we expect that when zero volts are applied, its current should also go to zero, just as it would for any resistor. Second, if a positive voltage is applied, the flux and charge should begin to change, and we therefore expect the resistance to change over time. Third, if a negative voltage is applied, it should reverse the changes in the flux and charge, and so the resistance should change in the opposite direction.

The above reasoning predicts what is known as **pinched hysteresis**, which is the key attribute of any memristor device. Pinched hysteresis is observed on a type of plot called the **Lissajous figure**, which displays the device's current and voltage when being driven by a period source. For an ordinary resistor, the Lissajous figure should be a straight line. For capacitors and inductors, the Lissajous figure is a circle or ellipse. **For a memristor, we see the "bow tie" pattern shown in Figure 78. This pattern is not observed for any other device, and is considered a "fingerprint" of memristance.**

A second property of memristance is **non-volatility**: when the applied voltage is zero, the current is also zero, therefore the flux and charge should remain constant. In other words, the memristor *remembers* its internal resistance for some period of time when held at zero volts. This property is potentially very useful for memory and storage applications, and is now one of the top priorities in memristor research.

A third property is **lobe narrowing** at higher frequencies. When a high-frequency voltage signal is applied, the positive and negative voltages appear for shorter periods of time, so the flux does not change as much. This should squeeze the hysteresis curves closer together. At progressively higher frequencies, the Lissajous pattern should converge to a single line or curve with no hysteresis at all — like an ordinary resistor. This prop-

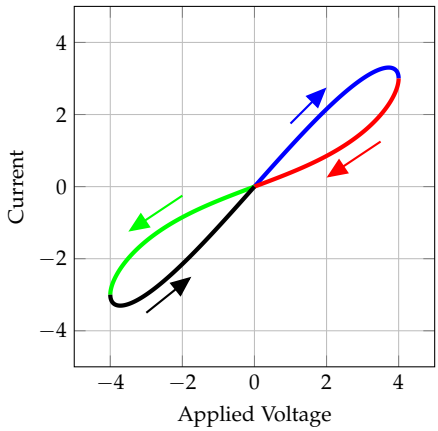


Figure 78: Pinched hysteresis in a memristor, driven by a sinusoidal voltage across its terminals. In this model, we assume resistance increases when a rising positive voltage is applied (blue curve). As the resistance increases, the blue curve's slope becomes flatter. When the positive voltage begins to decrease (red curve), the current is lower due to the increased resistance in this portion of the curve. When the voltage reaches zero, the current also goes to zero. When the applied voltage swings negative (green curve), the resistance decreases so the curve becomes more steep. When the applied negative voltage rises toward zero again (**black curve**), the current is stronger (more negative) due to the decreased resistance. Once the applied voltage returns to zero, the memristor's resistance should return to its original value.

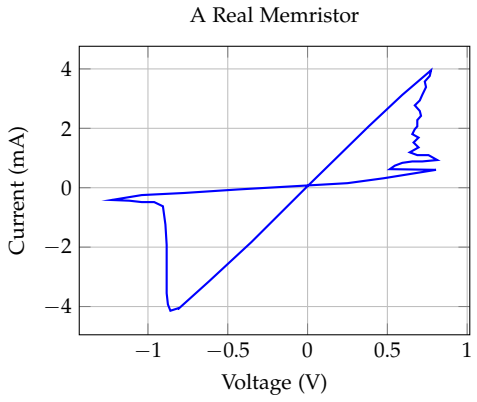


Figure 79: Measured data of a resistance-switching device reported by HP Research. The HP group was the first to identify these devices as memristors.

erty can be useful, for example, in reconfigurable radio circuits, where low-frequency signals can be used to "tune" the device's resistance, whereas radio-frequency signals will see a stable resistance.

A final property is that memristors are purely passive devices. Unlike capacitors and inductors, a memristor does not store any energy. The memory effect in a memristor is usually due to a chemical or structural change, such as a migration of molecules within a solid material. Work must be done to induce those changes, and they cannot be reversed without additional work from an external source.

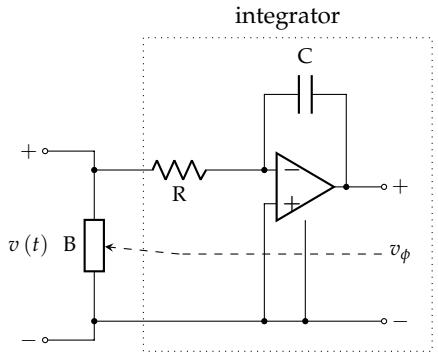


Figure 80: A memristor model based on an op amp integrator and a voltage-controlled nonlinear resistor.

Simulating Memristors

Memristor models are an active area of research, and there is not yet any standard model for simulations, nor is there a memristor device built into SPICE or other simulators. We can nevertheless get some experience with memristive behavior by creating *behavioral models* in SPICE. This is a tricky problem; several behavioral models are available but all of them are sensitive to simulation parameters and may crash the simulation if conditions are not just right. Furthermore memristor models often become invalid outside of a limited range of voltage and frequency, so they are not necessarily general-purpose models.

One of the simplest models uses an **op amp integrator** to model the flux in the memristor, as shown in Figure 80, where the integrator's output voltage v_ϕ should be proportional to the flux ϕ. Then, to model the changing resistance, we have to utilize a special **nonlinear dependent current source** provided by NGSpice. The syntax for a nonlinear source is

```
Bxxx n+ n- i={expr}
```

where *{expr}* is the mathematical expression used to define the device's current. We could use a simple expression like this:

$$i = -K\, v\,(t)\; v_\phi\,(t)$$

where K is a scale constant with units of V/Ω, and the resistance is assumed to decrease with higher values of ϕ. The expression is negative since the Miller integrator circuit is inverting, so we have to invert the sign.

There are a couple of problems with this model: first, we can't allow v_ϕ to ever be positive, since that would turn the memristor into an impossible *energy source*. This problem is fixed by **inserting a diode across the terminals of** C. We add an

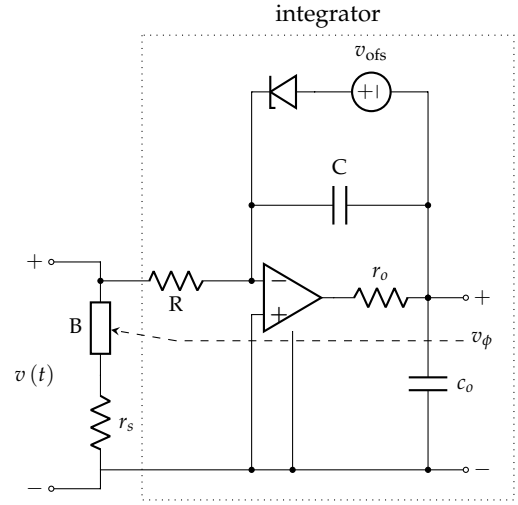


Figure 81: A stabilized and bounded memristor model. The zener diode and offset voltage v_{ofs} are used to constrain the maximum and minimum resistance in the device.

offset voltage source as shown in Figure 81 to compensate for the diode's forward drop. In addition, by using a zener diode, the diode's reverse breakdown voltage limits the maximum flux in the memristor, so **the memristor model should saturate at maximum and minimum resistances**. It is realistic to assume limits on the device's flux, since physical quantities generally don't extend to infinity in real devices.

A second problem is more subtle: this model can sometimes induce numerical instability in SPICE simulations. There are some tricks that improve stability, like inserting small-valued resistors and capacitors around the op amp and the nonlinear current source. These stabilizing devices appear as r_o, c_o and r_s in Figure 81. To verify the qualitative behavior of this model, the simulation below tests the model for a 5 V sinusoidal input at three different frequencies. It should produce six plots representing the Lissajous figure and the integrator state for each case.

memristor_integrator_model.sp

```
* Memristor integrator model

* Generic diode model:
.model diode d(Is=2.0298e-15, n=1, BV=200)

* Memristor subcircuit model:
.subckt memristor nplus nminus
R1 nplus n2 100
C1 n2 n3a 10n ic=0

* Ideal op amp model (dependent v source):
E1 n3 nmin nmin n2 10000

* Stabilizing RC network at op amp's output:
R2 n3a n3 1
C2 n3a nmin  10f

* Zener diode to constrain integrator bounds:
D1 n3a n2a diode
V1 n2a n2  DC -0.8V

* Nonlinear current source:
B1 nplus nmin i={-0.001*(v(n3)-v(nmin))*((v(nplus)-v(nmin)))}

* Stabilizing series resistance:
R3 nmin nminus 1
.ends
```

memristor_integrator_simulation.sp

```
* Memristor simulation based on integrator model

.include memristor_integrator_model.sp

V1 nplus 0 SIN(0 5 100k)
X3 nplus 0 memristor

.control
* Medium-frequency simulation:
tran 10n 50u uic
plot -i(v1) vs v(nplus)
plot v(X3.n3)
wrdata integrator_mid_freq v(nplus) i(v1) v(X3.n3)

* Low-frequency simulation:
alter @V1[sin]=[ 0 5 1k ]
tran .01m .0025 uic
plot -i(v1) vs v(nplus)
plot v(X3.n3)
wrdata integrator_low_freq v(nplus) i(v1) v(X3.n3)

* High-frequency simulation:
alter @V1[sin]= [ 0 5 1MEG ]
tran 1n 5u uic
plot -i(v1) vs v(nplus)
plot v(X3.n3)
wrdata integrator_high_freq v(nplus) i(v1) v(X3.n3)

.endc
.end
```

The simulation results are shown in Figure 84 on the following page. In the mid-frequency case (at 100 kHz) we see the typical pinched hysteresis that represents memristive behavior. The integrator state shows that the device's flux oscillates, corresponding to the integral of the sinusoidal input.

For the low-frequency results (at 1 kHz), the Lissajous figure shows "peaks" on each side of the hysteresis lobes. The peaks represent resistance saturation in the memristor device, which is verified by noting the saturation in the integrator state. If we examine the behavior at even lower frequencies, the hysteresis lobes will almost disappear, and all we will see is a curved line. This is because the resistance saturates early in the sinusoidal period, so the device behaves sort of like a diode, with low resistance for positive voltages and high resistance for negative voltages.

Lastly, in the high-frequency case (at 1 MHz), we see the predicted lobe narrowing. The integrator state shows why: the flux amplitude is reduced at high frequencies. This phenomenon has an easy theoretical explanation: the flux is

$$\begin{aligned}\phi &= \int \mathcal{V}_A \sin(2\pi f t)\ dt \\ &= \left(\frac{\mathcal{V}_A}{2\pi f}\right) \cos(2\pi f t)\end{aligned}$$

so as f increases we expect the amplitude to drop.

The model used in this example is an idealization. Real memristors can exhibit a variety of complex nonlinear behaviors, but all of them possess the essential characteristic of pinched hysteresis and resistor memory. Memristor theory is now understood to encompass many historical and contemporary devices, including 19th century devices like the arc lamp (Figure 82) and coherer (Figure 83). Chua's theory has also been expanded to encompass modern devices like the thermistor, various point contact devices, fluorescent tubes, among others. As a physical theory, memristance is increasingly observed in biological and chemical systems such as synaptic ion channels in neurons, leaves, blood, and even slime molds [4].

Figure 82: Davy's arc lamp exhibits memristive behavior, and is believed to be the first human-made memristor device.

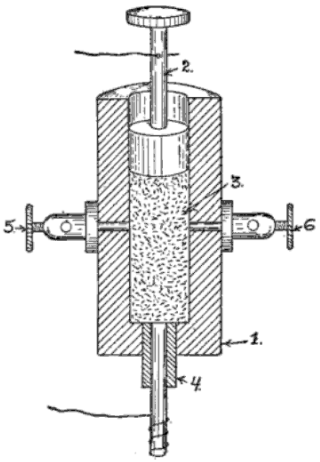
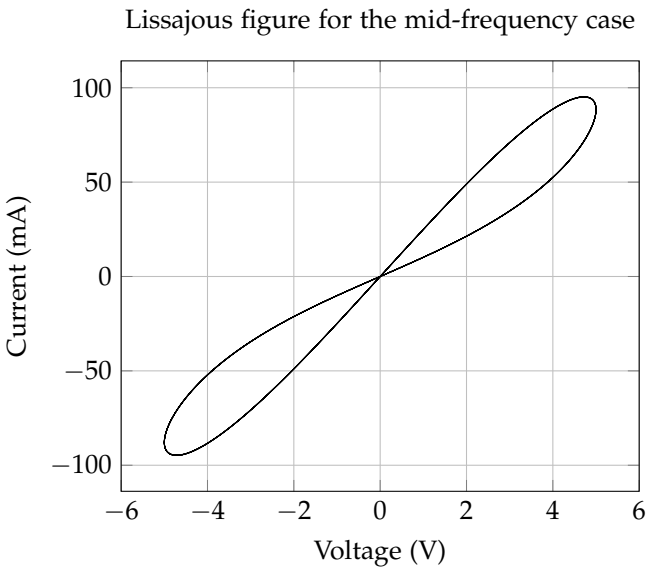
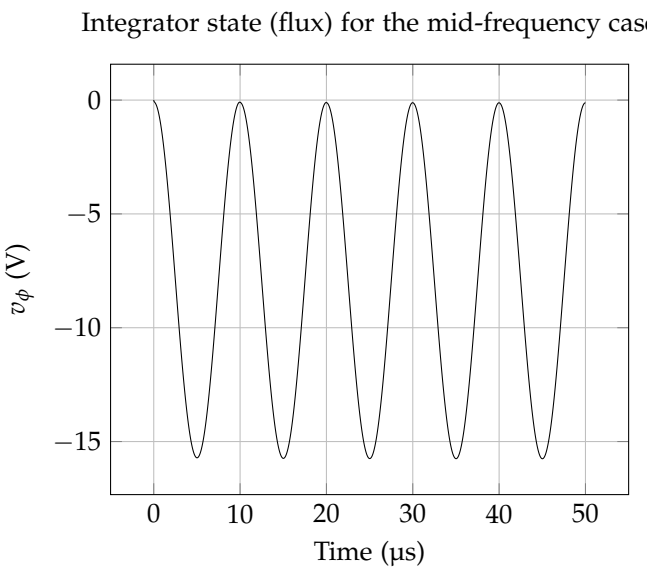
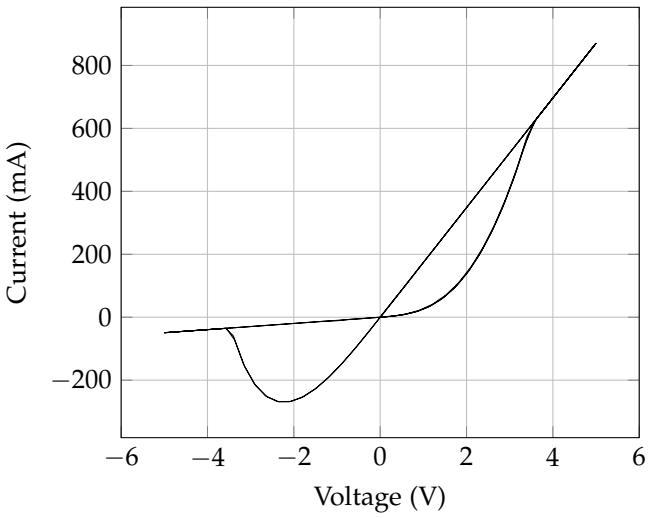


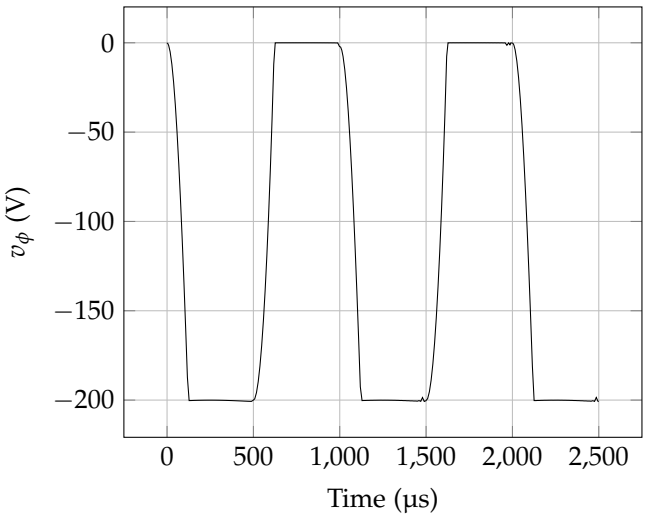
Figure 83: Branly's coherer, basically a tube filled with iron filings, is another early memristor device. The tube's resistance changes in the presence of radio waves. This device was widely used in wireless telegraph receivers from 1890 to 1920. In 1901, Bose reported the first observation of pinched hysteresis in a coherer device.

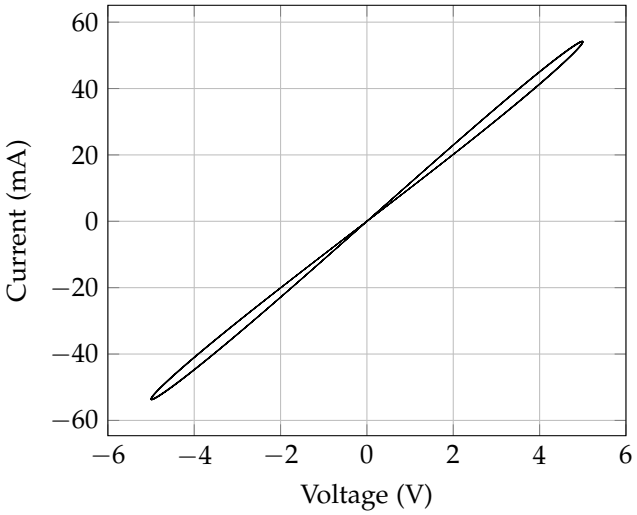
[4]; and











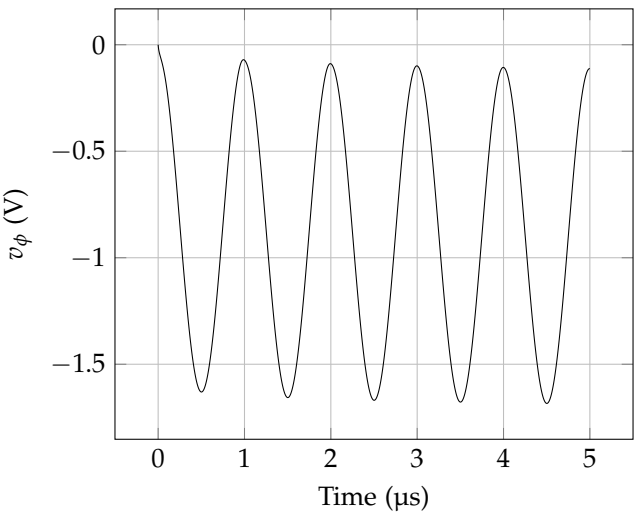


Figure 84: Simulation results for the integrator-based memristor model at three different frequencies.

Memristor Applications

Memristors are now known to have been used in many applications before there was a unified theory to describe them. One of the earliest practical uses was in radio receivers based on the coherer device, that exploited the memristor's diode-like behavior. The coherer was eventually replaced by simpler vacuum tube and solid-state diodes, and for most of the past century memristive behavior was either ignored or expressly avoided.

Thanks to the development of Chua's theory, we can now identify several important applications for memristive behavior. Today, the most promising applications are seen in **Resistive RAM (RRAM)** memories and in **neuromorphic circuits**, which mimic the activity of biological neurons. Applications are also being considered for high-speed computing, and for new types of logic-in-memory architectures which blur the distinction between processors and RAM.

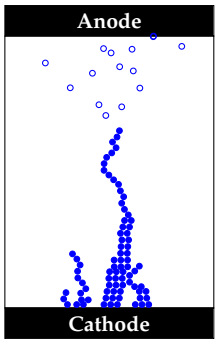


Figure 85: Formation of a metal filament within the resistance-switching memristor.

RRAM

Researchers are currently studying new semiconductor memory technologies that exploit the non-volatility of memristor devices. These are based on a type of memristor called **resistance switching devices**. Resistance-switching memristors are simple structures, often only a few nanometers across. The device has metal plates on the top and bottom, separated by an insulator. The top plate contains a different metal composition from the bottom plate, which allows metal ions to migrate into the insulating material. In this section, we present a simple version of RRAM based on resistance-switching devices. Many other types of memristive RAM are now being researched, so this should be viewed as an introductory example.

Thanks to the migration of metal ions, when voltage is applied in one direction, a **metal filament** tends to grow into the insulator, which lowers its resistance. When the voltage is reversed, the filament breaks up and restores the high-resistance state of the insulator. Due to the small dimensions of the material, this process can occur very rapidly, so **the device tends to switch quickly between maximum and minimum resistance levels, hence the name "resistance switching device."**

In order to read and write data from an RRAM array, we can use a simple diode addressing scheme as shown in Figure 86. Each memristor is connected in series with a diode. At the nanoscale, it's possible to make the diode "for free", as it can be

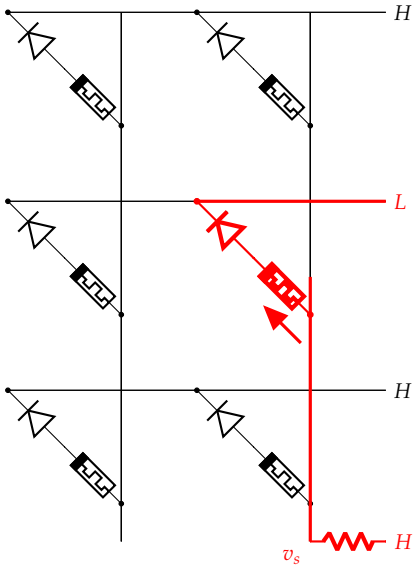


Figure 86: Portion of an RRAM array showing the bias conditions for a SET operation.

made from the wire connections around the memristor. Under normal conditions, all row wires are biased at a "high" voltage level, and all column wires are biased at a "low" voltage, so that every diode is reverse-biased.

To drive a forward voltage across the memristor (called a SET operation), we **activate one single memristor by driving its column voltage high and its row voltage low**. This forward biases the diode in that cell only, and exposes the memristor to non-zero forward voltage, which drives it to its maximum-resistance state.

To drive a reverse voltage across the memristor (called a CLEAR operation), we bias one column connection low, while all other columns are biased high. We then **drive the cell's row voltage to a very high level, sufficient to exceed the diode's reverse breakdown voltage.** This exposes the memristor to a non-zero reverse voltage and drives it to its minimum-resistance state.

To READ the data from a cell, we apply a small forward voltage on the column wire. The memristor then forms a voltage divider with the column's series resistance, allowing us to measure the memristor's state by sampling the divider output at v_s. A high value implies high-resistance, and a low value implies low-resistance.

It is possible that the small voltage applied during a READ operation could alter the resistance state within the cell. Fortunately, many resistance-switching memristors are found to exhibit a **threshold effect**, so that their resistance remains undisturbed as long as the applied voltage remains less than some threshold. With that type of device, we can use a READ voltage less than the threshold to perform a **non-destructive read**: defined as measuring the device's state without disturbing it.

To simulate an RRAM example, we first modify our memristor model by adding a threshold effect. One way to do this is to insert diodes in series with the integrator's input. Then the input voltage will have to exceed the diode's 0.7 V drop in order to influence the integrator's state. The modified memristor model, and an RRAM demonstration, are given in the netlists below. The demonstration considers only a single RRAM cell.

The simulation results are shown in Figure 88, which presents three transient signal traces. The top plot shows the column voltage (red), the row voltage (green), and the voltage seen across the memristor (blue). During a SET operation, the memristor voltage spikes to a high value, but then quickly curves downward as the memristor switches into a

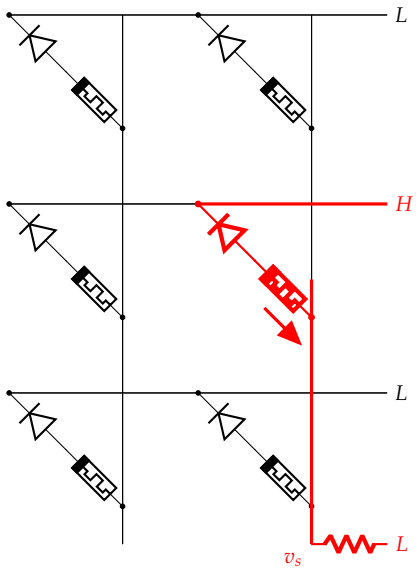


Figure 87: Portion of an RRAM array showing the bias conditions for a CLEAR operation.

high-resistance state. During the CLEAR operation, the memristor shows a small negative voltage at first, but then transitions to a larger negative voltage as the device switches to a low-resistance state.

In between the SET and CLEAR operations, a small, brief pulse is used to READ the memristor's state. The pulse is kept small in order to stay below the memristor's threshold, so that the state will not be disturbed. This type of small pulse is often called a **sub-threshold pulse**. The middle plot in Figure 88 shows a zoomed view of the READ signals. We see that the voltage across the memristor is higher when it is in a low-resistance state. By measuring the voltage output, we can deduce whether the memristor is in a high or low resistance state, hence revealing the stored value of zero or one.

The bottom plot in Figure 88 shows the integrator state within the memristor model, representing the device's resistance state. We see that the device switches quickly between saturated high and low levels, and is not disturbed by the sub-threshold READ signal.

memristor_integrator_model_with_threshold.sp

```
* Memristor integrator model with threshold

* Generic diode model:
.model diode d(Is=2.0298e-15, n=1, BV=200)

* Memristor subcircuit model:
.subckt memristor nplus nminus
D2 nplus nth diode
D3 nth nplus diode
R1 nth n2 100
C1 n2 n3a 10n ic=0

* Ideal op amp model (dependent v source):
E1 n3 nmin nmin n2 10000

* Stabilizing RC network at op amp's output:
R2 n3a n3 1
C2 n3a 0  10f

* Zener diode to constrain integrator bounds:
D1 n3a n2a diode
V1 n2a n2  DC -0.8V

* Nonlinear current source:
B1 nplus nmin i={-0.0001*(v(n3)-v(nmin))*((v(nplus)-v(nmin)))}

* Stabilizing series resistance:
R3 nmin nminus 1
.ends
```

RRAM_threshold.sp

```
* Memristor resistive RAM demo

* Zener diode model:
.model zdiode d(Is=2.0298e-15, n=1, BV=5)

* Load the external memristor model:
.include memristor_integrator_model_with_threshold.sp

* Circuit for one RRAM cell:
R1 1 2 100
X1 2 3 memristor
D1 3 4 zdiode

* VH  is the write-1 pulse
* VL  is the write-0 pulse
* Vrd is the read pulse
VH  1 5 DC 0 PULSE (0 10 00u 10u 10u 400u 2400u)
VL  4 0 DC 0 PULSE (0 10 1200u 10u 10u 400u 2400u)
VRd 5 0 DC 0 PULSE (0 1 660u 10u 10u 30u 1200u)


* Simulation control commands:
.control
tran 100u 20m uic

* Plot write/read/memristor signals:
plot v(1) v(4) 'v(2)-v(3)'

* Plot model integrator state:
plot v(X1.n3)-v(X1.nmin)

* Plot read response:
plot '(v(2)-v(3))*v(5)/3' v(5)

wrdata RRAM_threshold_demo v(1) v(2) v(3) v(4) v(5) X1.n3 X1.nmin

.endc
.end
```

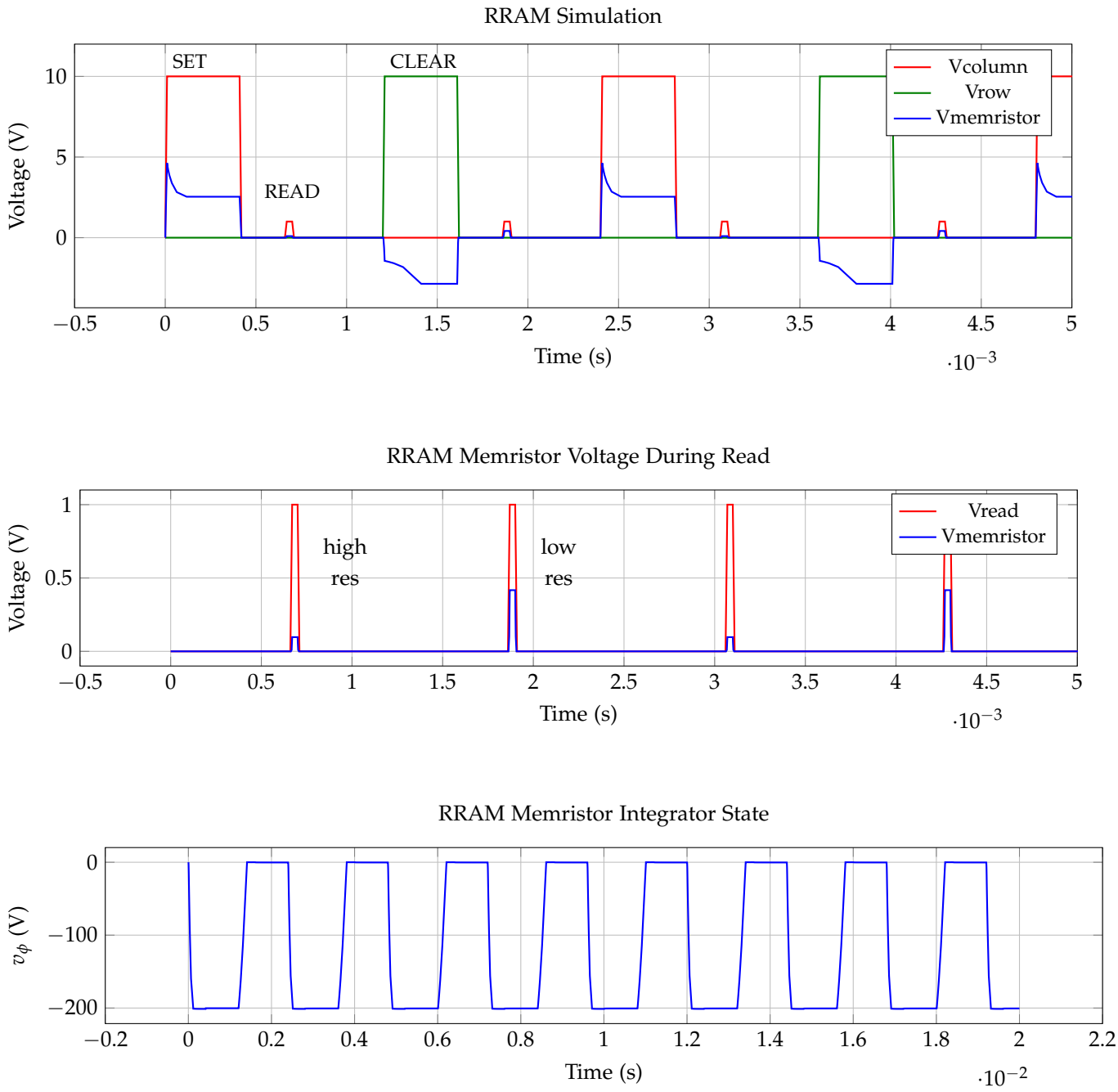


Figure 88: RRAM simulation results using the integrator model with threshold. This is only a hypothetical model, but real RRAM circuits show qualitatively similar behavior.

Exploring Memristor Controversies

Since memristors are a relatively new addition to electronics and circuit theory, there is naturally a fair amount of debate among researchers and practitioners over how the theory should be interpreted. Students might be interested in hearing about some of this controversy. This section provides an overview of some of the major criticisms.

While most researchers have accepted the basic science and historical narrative presented in this chapter, there are a few critics who remain unconvinced. Criticisms appear mainly among practitioners in private industry, writers and commenters in popular magazines and newspapers, and online discussions. For example, critics have been active in editorializing their views on the Wikipedia entry for memristors. A small number of peer-reviewed articles (less than ten) have challenged various aspects of memristor devices and theory, compared to thousands that have adopted, expanded and successfully applied the theory. A number of very critical articles have been published in the notorious pseudoscience server known as "vixra" (don't be fooled by the academic appearance of some vixra articles).

Criticisms can be divided into the following major categories: claims of credit (i.e. who *really* discovered or invented it); pedantic arguments (e.g. real memristors are not "ideal" enough); philosophical disputes; and pessimism about applications. Other arguments may be seen, including outright denial that memristance exists, allegations of fraud, and other crankish fringe complaints. But any worried readers can satisfy such extreme doubts by simply purchasing a memristor array sample chip (currently $199 from Bio-Inspired Technologies) and observing it directly.

"A significant part of the scientific process is to vet descriptions of new ideas or objects, and the bigger the potential impact of a concept, the more rigorous that scrutiny should be. However, intertwined with this process are many human issues of desire for recognition and priority of discovery, as well as an often strong bias to reject anything new without actually understanding it. There are a lot of misconceptions about memristors floating around that are difficult to correct with only a few explanatory pages. Real understanding requires a great deal of hard work, and the resources essential to achieve that understanding already exist in the literature."

Claims of Credit

1. Resistance-switching devices existed before Chua's theory, and before HP Lab's device. They didn't invent anything. Since it's now understood that memristive devices were used in the 19th century, some critics argue that Chua's theory "doesn't count" as a legitimate discovery. Critics have also questioned the novelty of HP Lab's discovery, since various types of thin-film resistance-switching devices were studied going back to the 1960s. Furthermore, 1995 an Indian research group described a device very similar to HP's [5].

[5]

Some criticisms also mention older theories that resemble or

overlap with Chua's theory on memristors. A device called the "memistor" was studied in the 60's, and was similarly given a name based on "memory resistance." It is fairly common for competing theories to emerge in science, and subtle differences can cause one theory to prevail while another is forgotten. The "memistor vs memristor" argument is examined and resolved in an article by Kim and Adhikari[6].

[6]

Answer: So why do Chua and HP deserve so much attention? The reason is that they articulated a unifying theory. Chua's contributions was to express nonlinear circuit theory as a rigorous and *closed* mathematical system. The major contribution from HP Labs was to link Chua's theory with resistance-switching devices. As a result, researchers have been able to refine and improve on Chua's original theory. Critics often claim that the Indian researchers believe that resistance-switching devices *are not memristors*, but their personal beliefs are not especially relevant to the bigger scientific question.

Thanks to contributions from hundreds of other researchers, we now know that "memristance" is a rich concept with broad descriptive power. It covers numerous devices and behaviors that are awkward to describe using traditional circuit concepts, and are absent from electronic simulators like SPICE. In the past, useful circuit techniques may have been missed due to a simple lack of description for these behaviors.

Pedantic Arguments

2. The memristor definition has evolved over time. There is no "ideal memristor" matching the one first proposed by Chua in 1971. All known memristors have some differences that disqualify them as ideal. Real memristors are a closer match to the expanded definition of "memristive systems" developed by Chua in 1976. Researchers now use the term "memristor" when they really mean "memristive system." Critics argue that you can't redefine a term like that.

Answer: Sure you can. This is the process of science. A scientific theory is not like a contract or a piece of legislation; we expect theories and definitions to evolve as they are refined by continuing evidence and reasoning. When critics say "there is no real memristor," they are simply being dishonest. If we allowed their reasoning, we could also claim there is no real capacitor, inductor or resistor. An ideal capacitor, for example, should be able to store an arbitrary amount of charge, and it should retain that charge forever if its terminals are left open.

"Mathematics is an experimental science, and definitions do not come first, but later on."

"I do not refuse my dinner simply because I do not understand the process of digestion."

– Oliver Heaviside, answering criticisms of his Laplace transform method for circuit analysis

But a real capacitor will experience dielectric breakdown if it stores too much charge. A real capacitor will slowly discharge through leakage currents. Ergo, there is no such thing as a capacitor, right?

The concept of memristance continues to evolve, and researchers are not yet settled on the best meaning of "ideal" memristor behavior. What it means to be "ideal" is ultimately decided based on what is *most useful* for circuit analysis and design. We haven't yet figured everything out, and that's just fine. That's what makes memristors interesting as a topic of research.

Philosophical Disputes

3. Argument from radical empiricism: All genuine science must originate from empirical observation, and mathematics is useful only for the subordinate role of description.

"Today's scientists have substituted mathematics for experiments, and they wander off through equation after equation, and eventually build a structure which has no relation to reality... The scientists of today think deeply instead of clearly. One must be sane to think clearly, but one can think deeply and be quite insane."

– Nikola Tesla, arguing against Einstein's theory of general relativity

Answer: As engineers, we may be tempted to embrace this philosophy, since the final proof always appears in the real world. But if we really thought this way, it would invalidate most of the methods used in our profession, and in all the hard sciences. A scientific theory is more than just a handy description of observations. The value of a theory lies in its *predictive utility*: if a theory's axioms correspond to confirmed physical laws, then mathematical predictions from those laws should hold true. Otherwise *we have to reject or modify the theory*. We are therefore *obligated* to accept Chua's prediction of memristors, or else we have to *explain* what's wrong with the theory of circuits. **For a *scientific skeptic*, It's not enough to just reject mathematical predictions; we have to correct the underlying theory.** Memristor critics have not proposed any changes to the theory; in effect, they are simply rejecting mathematics as a legitimate foundation for hard science.

"Who the h— is still believing that memristors might exist in physical reality? By now, it should be clear that the 'memristor' is nothing else but a mathematical curiosity. The above discussed findings are exclusively related to resistance switching materials (ReRAM)."

– "A Physicist" (online comment)

4. Argument from reductionism. Here there are two sub-arguments: (A) that we cannot legitimately separate the definition of *flux* in circuit theory from its magnetic field interpretation; and (B) that we cannot study devices at the circuit level, they can only be understood by studying the specific physics and chemistry that apply within a real device.

Answer: Sub-argument (A) is elaborated in a peer-reviewed article by Vongehr and Meng, published in *Scientific Reports* with the title, "The Missing Memristor has Not been Found." This article postulates an elaborate scenario involving an alternate universe where magnetic fields and inductors don't exist,

and imagines an alternative Chua (perhaps wearing a goatee in this universe) who predicts, through mathematical analysis, that there should be a missing device called an "inductor" that relates current and flux. But since their universe doesn't have magnetic fields, it would be impossible to discover a "real" inductor, even if they discovered some devices deemed *similar*, those devices would not represent true inductors because they would not possess true magnetic fields.

"whatever devices would be discovered without magnetism, none can be the real EM inductor, but the latter is the grounds on which the original real memristor device hypothesis sits."

This "thought experiment" aims toward concluding that there can be no legitimate concept of "flux" except the one associated with magnetic fields, and therefore the discovered memristors are not "real" memristors. The authors furthermore speculate that a "true" memristor is impossible, but offer no concrete evidence for this conclusion, other than to speculate that we may be living the wrong universe.

Sub-argument (B) is a more generalized attack, but similar in spirit to Vongehr and Meng's alternate universe theory. Proponents of this criticism argue that it's meaningless to define devices at the circuit level, because (i) circuit theory was developed to study networks of already-existing devices, (ii) we can't produce a successful device unless we understand the detailed physics internal to that device, and (iii) the original theoretical conception of the memristor gave no indication as to how such a device could be realized.

To answer these arguments, we must first observe the value of circuit theory as an independent and complementary discipline from device physics. Circuit theory is not about making devices, it is about making complex systems. A device is only useful to the extent that it can exist in an electronic system, so the theory of circuits is not somehow subordinate to device physics. As mentioned in the beginning of this chapter, circuit theory rests directly on physical laws; it is not "derived from" Maxwell's electromagnetic theory. It is more correct to say that circuit theory is the subset of electromagnetic theory that is indifferent to fields in space. In the context of circuit theory, we define devices by *what they do at their terminals*, not how they are made or how they work internally.

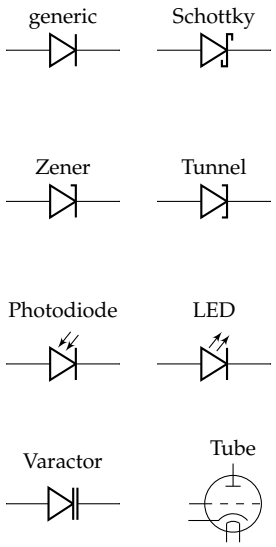


Figure 89: Many types of diodes, with distinct internal physics. All are classified by a single unifying concept.

A perfect example is the **diode**. Historically, the name "diode" referred to a vacuum tube device that worked on the basis of thermionic emission from a heated metal plate, facilitating directional conduction between two electrodes. The term was purportedly coined by William Eccles in 1919 as a portmanteau of "**di**" (two) with "electr**ode**." The same name was soon used to refer to solid-state rectifying semiconductor de-

vices, which operate from different physical principles. It would be counterproductive to argue that we need different names, or to complain that the definition of "diode" was changed to something more broad than its original meaning. To this day, we continue to use the name "diode" for a variety of vacuum tube and the solid-state devices (a short list is shown in Figure 89), even though they have very different constructions and different underlying physics. Their behavior is qualitatively the same, and that's what matters most for the purpose of circuit engineering.

A more subtle example is the phenomenon of **diode reverse breakdown**. When Zener hypothesized the reverse breakdown effect, he proposed that it would be caused by quantum tunneling in the device [7]. When diodes are used in their reverse breakdown mode, they are commonly called "Zener diodes" to honor his *purely mathematical discovery*. We now know when the breakdown voltage is greater than about 5 V, *avalanche ionization* is the primary mechanism, not quantum tunneling. This distinction may be relevant for optimizing a particular design for a particular application, but from the standpoint of circuit theory they both deliver the same qualitative behavior, and for most purposes there's nothing wrong with using the name "Zener diode" to refer indifferently to both physical effects. It would be silly to declare that "Zener was wrong" or to insist that avalanche diodes are not "true" Zener diodes; it's much more *useful* to describe tunnel devices and avalanche devices as subtypes of the general diode class.

[7]

Finally, to answer Vongehr and Meng's alternate universe scenario, imagine that we discovered a device that relates ϕ to i — it stores energy and resists a sudden change in current — but is found to contain no internal magnetic fields of any kind. Perhaps the device is governed by electrochemical reactions or by tiny molecule-sized demons; it doesn't matter. What matters is if the device's behavior is indistinguishable from an inductor, then for all practical purposes *it is an inductor*. It would be meaningless to say that it isn't a "true" inductor, just as it would be meaningless to say that only vacuum tubes are "true" diodes.

Pessimism

5. Non-volatile memristors are impossible. Some researchers have argued that memristors' internal states could be disrupted by noise processes. If that's true, then it will not be possible to store information indefinitely in a memristor memory cell,

and RRAM will not work as a replacement for Flash devices. This argument is put forward by Meuffels and Soni in a draft available at the arxiv server[8] (note that arxiv papers are *not* peer reviewed).

[8]

The article has been cited by a few peer-reviewed articles, and may have some validity. But several critics have also claimed this analysis as evidence that memristors are "impossible" as real devices. Their reasoning is that a memristor is supposed to be non-volatile, meaning it should remember its resistance state indefinitely. If Meuffels and Soni showed that permanent memory is unlikely, then the memristor must be impossible, right?

Answer: We already used an identical argument to prove that capacitors are impossible, since their charge tends to leak away. This argument describes the same kind of problem that we see in DRAM: stored information is temporary. For this reason, DRAM memories require a *refresh* operation to periodically re-write the stored information. Perhaps memristor-based RRAMs will also require a refresh operation. Perhaps that will make RRAMs less competitive compared to other memory technologies. That could all be true, but at most it means that memristors will be disappointing in the RAM market. It doesn't have anything to do with memristors being "impossible" devices. It just means that the non-volatility property is temporary in ressitance-switching devices.

Conclusion

Memristors are a fascinating area of current research in both theory and experiment. This chapter has introduced only the most basic facts about memristors. We examined and answered a few of the major arguments that have generated some controversy on the topic. One can find many more arguments circulating in the wild, but the ones listed here are deemed by the author to be the strongest criticisms. The current scientific consensus is that memristors comprise a legitimate theory bolstered by numerous real devices and phenomena. The reader can gain direct experience by purchasing a resistance-switching sample, or by constructing a replica coherer device, or by studying one of the biochemical substances that are now known to exhibit memristive behavior. More advanced resistors are dangerous to manufacture without proper facilities, so students are urged to exercise caution in their explorations.

Introduction to MOSFETs

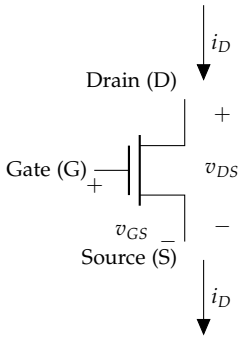


Figure 90: NMOS device symbol, showing the device's three major terminals. Current flows between the drain (D) and source (S) terminals, and is controlled by the gate (G), like a valve. A fourth terminal. known as the *bulk, body, or substrate*, is not shown. The substrate is usually shared by many devices, and for NMOS devices it should be connected to the circuit's *most negative potential* (usually ground or V_{SS}.

Why do we need transistors?

Diodes can perform logic operations, but they cannot perform:

- NOT, NAND or NOR operations – **diode logic gates are not universal!**
- Amplification – signals attenuate as they pass through diode networks, so large-scale systems are impossible with diodes alone.

Amplification is the fundamental characteristic needed for logic circuits – the device must be able to deliver more energy at its output than provided at its input.

MOSFET as switch

The MOSFET has three terminals, source, gate and drain. We may first understand the MOSFET as a logic switch. In this model, the terminal potentials are interpreted as *logic values*, i.e.the logic set $\{0, 1\}$ is mapped to the potential values $\{0, V_{\text{DD}}\}$. Under this model we may consider all signals to be either HIGH or LOW. Then the behavior is as follows:

Device Type	v_G	Device State	v_{DS}
NMOS	HIGH	ON	small
	LOW	OFF	large
PMOS	HIGH	OFF	large
	LOW	ON	small

Notice two things:

- The NMOS and PMOS have *complementary* behavior, i.e. they have opposite states in response to the gate voltage.
- When the device is ON, the drain-source voltage v_{DS} must be quite small. When the device is OFF, the drain-source voltage can be large. This is the behavior we expect from a switch.

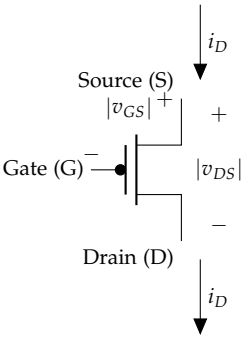


Figure 91: PMOS device symbol, which is *complementary* to the NMOS device. The drain and source terminals are flipped (current is understood to flow vertically downward from the source to the drain). A "bubble" is commonly drawn at the gate to indicate that the device **responds to the logical complement of the gate signal**. For PMOS devices, the substrate is usually connected to the circuit's *most positive potential* (usually V_{DD}.

MOSFET as a current source

The MOSFET is an analog device, meaning it does not merely have "OFF" and "ON" states, but has a continuous range of in-between states. If a MOSFET is balanced in an "almost-on" state, known as its **saturation mode**, it produces an approximately constant current. Therefore **we can use a MOSFET device to implement a DC current source.**

In practice, an NMOS device first begins to switch ON when its V_{GS} crosses a device-specific **threshold voltage,** V_{Th}. When V_{GS} is just slightly above V_{Th}, the device enters its saturation mode. If V_{GS} is increased, the current increases. If V_{GS} is held constant, the current stays constant. If we keep increasing V_{GS}, the device eventually turns fully ON, at which point it no longer acts like a current source, and behaves more like a small resistor.

In the PMOS case, the device first begins to switch ON when its V_{SG} voltage exceeds the device's V_{Th}, which places the device in its saturation mode. The behavior is again complementary to the NMOS: as V_G is **lowered**, V_{SG} increases, and the current increases. Eventually, when V_G becomes low enough, the PMOS device turns fully ON and no longer works like a current source.

If the MOSFET is balanced in its saturation mode, and a time-varying signal is applied to the gate, the device's current will change in response to the gate signal. Hence **the MOSFET in saturation is considered to be a transconductance amplifier: it produces a current output (at the drain) in response to a voltage input (at the gate).**

One of our key design tasks will be to balance the MOSFET in the appropriate mode for our intended application. When making switching circuits, we want the MOSFET to be toggle between ON and OFF states. When making a current source or an amplifier, we want the MOSFET to be suspended in between, in its saturation mode, with a relatively small value of V_{GS} (for NMOS) or V_{SG} (for PMOS).

Electrical Characteristics

The MOSFET is of course more complicated than the switch model implies. To get a more detailed picture of MOSFET behavior, we may consider an alternative invterter circuit known as the resistor-transistor logic (RTL) configuration. The RTL circuit is inferior to the CMOS configuration in that it draws static power when the NMOS device is ON. This is because

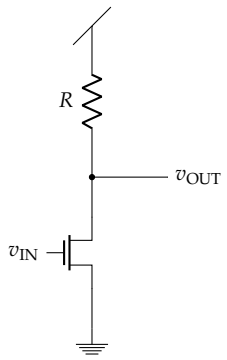


Figure 92: NMOS RTL inverter configuration.

the MOSFET must pull a constant current through the resistor R in order to maintain a low output voltage. It is nevertheless helpful to study the RTL inverter, because its properties are somewhat easier to analyze than the CMOS design.

The RTL inverter's DC transfer characteristic is split into three regions, representing the different modes of the NMOS device:

I **Cutoff**— When v_{IN} is below the devices *threshold*, V_{Th}, it is considered OFF and behaves like an open circuit between the Drain and Source.

II **Saturation** — When v_{IN} is slightly greater than V_{Th}, the device is *partially* turned ON. The output voltage is determined by the current through the MOSFET, which depends strongly on v_{IN}.

III **Triode** — The device is considered *fully* ON when $v_{OUT} < v_{IN} - V_{Th}$. In this mode, the Drain and Source are *almost* short-circuited.

The precise behavior of a MOSFET device is modeled by three different equations corresponding to three operating modes. The equations are qualitatively different, but they should be piecewise continuous (i.e. they should connect at the boundaries between each mode). These equations relate the device's *drain current*, i_{DS}, to the gate-source and drain-source voltages v_{GS} and v_{DS}, respectively. To simplify the equations, we define the **Overdrive Voltage** as

$$v_{OV} \triangleq |v_{GS}| - |V_{Th}|.$$

Then the device equations are:

I **Cutoff** — $v_{OV} \leq 0$:

$$i_{DS} = 0$$

II **Saturation** — $v_{OV} > 0$ and $|v_{DS}| > v_{OV}$:

$$i_{DS} = \frac{1}{2} k v_{OV}^2.$$

III **Triode** — $v_{OV} > 0$ and $|v_{DS}| \leq v_{OV}$:

$$i_{DS} = k \left(v_{OV} |v_{DS}| - \frac{1}{2} |v_{DS}|^2 \right).$$

In these equations, k is a *scale constant* with units $\mu A/V^2$, and is typically on the order of $100\,\mu A/V^2$ to $1\,mA/V^2$. The threshold voltage is a manufacturing parameter that varies widely between different technologies. It is typically between 0.4 V and 2 V.

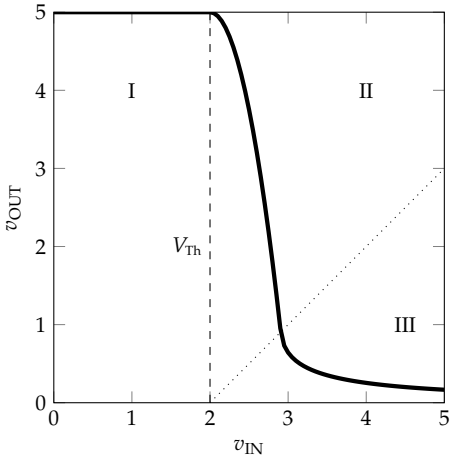


Figure 93: Transfer characteristic of the RTL inverter obtained from a SPICE simulation, showing the operating modes (I) Cutoff, (II) Saturation and (III) Triode.

NMOS RTL Inverter Analysis

Considering the NMOS RTL inverter shown above, suppose $V_{Th} = 2\,V$, $k = 100\mu A/V^2$, $R = 100k\Omega$ and $V_{DD} = 5\,V$. Given the model equations above, solve for the DC transfer characteristic of v_{OUT} as a function of v_{IN}.

Solution: We may divide the analysis into the three regions, and determine the points where these regions meet. Since the MOSFET's source terminal is tied to ground, we observe that $v_{GS} = v_{IN}$ and $v_{DS} = v_{OUT}$. If we imagine that v_{IN} is initially zero, and is slowly increased toward V_{DD}, then we have three subproblems:

I **Cutoff**— When $v_{IN} < V_{Th}$, verify that $v_{OV} < 0$, therefore $i_{DS} = 0$. In that case, there is no current flowing through R, so $v_{OUT} = V_{DD} = 5\,V$.

II **Saturation** — When $v_{IN} > V_{Th}$, then $v_{OV} > 0$ and the device's current is given by the square-law equation. Then v_{out} is determined by the voltage drop across R:

$$\begin{aligned} v_{OUT} &= V_{DD} - \frac{1}{2}Rkv_{OV}^2 \\ &= 5\,V - 5\,(v_{IN} - 2\,V)^2 \end{aligned}$$

As v_{IN} increases, v_{OUT} will decrease until the MOSFET enters the triode mode. That transition happens when $v_{OUT} = v_{OV}$, i.e.

$$\begin{aligned} v_{OV} &= 5\,V - \frac{1}{2}Rkv_{OV}^2 \\ \Rightarrow 0 &= \frac{1}{2}Rkv_{OV}^2 + v_{OV} - 5 \end{aligned}$$

Since the result is a quadratic equation, we can apply the standard formula and solve:

$$\begin{aligned} v_{OV} &= \frac{-1 \pm \sqrt{(1)^2 - 4\left(\frac{1}{2}Rk\right)(-V_{DD})}}{Rk} \\ &= \frac{-1 \pm \sqrt{1 + 2RkV_{DD}}}{Rk} \\ \text{if } 2Rk \gg 1: \; v_{OV} &\approx \sqrt{\frac{2V_{DD}}{Rk}} \end{aligned}$$

Note that we chose the positive result in the quadratic equation, since v_{OV} has to be positive in both saturation and triode, otherwise these equations wouldn't apply. The results are:

$$\begin{aligned} \text{exact: } v_{OV} &= 0.905\,V \\ \text{approx: } v_{OV} &\approx 1\,V \end{aligned}$$

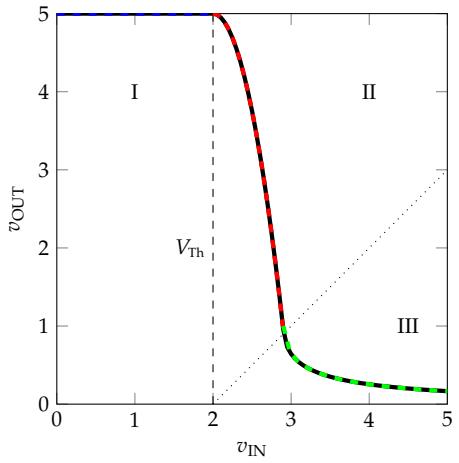


Figure 94: Comparison of analysis and simulation results in the three operating modes, cutoff, saturation, and triode.

And the corresponding values of v_{IN} are:

$$\text{exact: } v_{IN} = 2.905\,\text{V}$$
$$\text{approx: } v_{IN} \approx 3\,\text{V}$$

III **Triode** — When $v_{IN} > 2.5188\,\text{V}$, the device should enter triode, and the new device equation is

$$i_{DS} = k\left(v_{OV}v_{OUT} - \frac{1}{2}v_{OUT}^2\right)$$
$$\Rightarrow v_{OUT} = V_{DD} - Rk\left(v_{OV}v_{OUT} - \frac{1}{2}v_{OUT}^2\right)$$

Once again we can arrange this in the form of a quadratic equation:

$$0 = \frac{1}{2}Rkv_{OUT}^2 - (1 + Rkv_{OV})\,v_{OUT} + V_{DD}$$
$$v_{OUT} = \frac{(1 + Rkv_{OV}) \pm \sqrt{(1 + Rkv_{OV})^2 - 2RkV_{DD}}}{Rk}$$
$$= \frac{(1 + Rkv_{OV}) - \sqrt{1 + (Rkv_{OV})^2}}{Rk}$$

Note that if $Rkv_{OV} \gg 1$ the equation simplifies to $v_{OUT} = 0$.

PMOS RTL inverter analysis

In the PMOS version of the RTL inverter circuit, the circuit is "flipped upside down" and the behavior is transposed. The circuit's logical behavior is the same as the NMOS version, but the fine details are changed. We see that the cutoff, saturation and triode regions now appear in different places:

In this configuration, we start by solving for $|v_{GS}|$, $|v_{DS}|$ and v_{OV} in terms of the terminal signals:

$$\begin{aligned} |v_{GS}| &= V_{DD} - v_{IN} \\ |v_{DS}| &= V_{DD} - v_{OUT} \\ v_{OV} &= |v_{GS}| - |V_{Th}| \end{aligned}$$

We may then proceed with the same analysis steps as before, only this time we imagine that v_{IN} starts at V_{DD} and is slowly decreased down to zero.

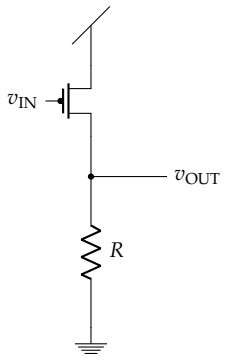


Figure 95: PMOS RTL inverter configuration.

I **Cutoff**— When $v_{OV} < 0$, the MOSFET is OFF so that $i_{DS} = 0$. In that case, there is no current flowing through R, so $v_{OUT} = 0$. This describes the region where $v_{IN} > V_{DD} - |V_{Th}|$.

II **Saturation** — When $v_{OV} > 0$ and $|v_{DS}| > v_{OV}$, the device's current is given by the square-law equation. This corresponds to the case when:

$$\begin{aligned} V_{DD} - v_{OUT} &> V_{DD} - v_{IN} - |V_{Th}| \\ \Rightarrow v_{OUT} &< v_{IN} + |V_{Th}| \end{aligned}$$

In this region, v_{out} is determined by the voltage drop across R:

$$\begin{aligned} v_{OUT} &= \frac{1}{2} R k v_{OV}^2 \\ &= 5\,(V_{DD} - v_{IN} - 2\,\mathrm{V})^2 \end{aligned}$$

As v_{IN} decreases, v_{OUT} will decrease until the MOSFET enters the triode mode. That transition happens when $V_{DD} - v_{OUT} = v_{OV}$, i.e.

$$\begin{aligned} v_{OV} &= 5\,\mathrm{V} - \frac{1}{2} R k v_{OV}^2 \\ \Rightarrow 0 &= \frac{1}{2} R k v_{OV}^2 + v_{OV} - 5 \end{aligned}$$

Notice that **this is the same quadratic equation we obtained for the NMOS circuit,** so we can borrow the results from before:

$$\begin{aligned} \text{exact: } v_{OV} &= 0.905\,\mathrm{V} \\ \text{approx: } v_{OV} &\approx 1\,\mathrm{V} \end{aligned}$$

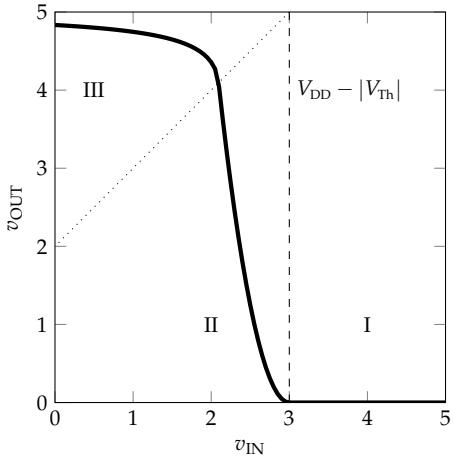


Figure 96: DC transfer characteristic obtained from a SPICE simulation of the PMOS RTL inverter.

And the corresponding values of v_{IN} are:

$$\begin{aligned} v_{IN} &= V_{DD} - v_{OV} - |V_{Th}| \\ \text{exact: } v_{IN} &= 5\,\text{V} - 2.905\,\text{V} = 2.095\,\text{V} \\ \text{approx: } v_{IN} &\approx 5\,\text{V} - 3\,\text{V} = 2\,\text{V} \end{aligned}$$

III **Triode** — When $|v_{DS}| < v_{OV}$, the device enters triode. This corresponds to the case where

$$\begin{aligned} V_{DD} - v_{OUT} &< V_{DD} - v_{IN} - |V_{Th}| \\ \Rightarrow v_{OUT} &> v_{IN} + |V_{Th}| \\ \text{or when } v_{IN} &< v_{OUT} - |V_{Th}| \end{aligned}$$

In this region, the device's current and output voltage change as follows:

$$\begin{aligned} i_{DS} &= k\left(v_{OV}\,|v_{DS}| - \frac{1}{2}v_{DS}^2\right) \\ \Rightarrow v_{OUT} &= Rk\left(v_{OV}\,|v_{DS}| - \frac{1}{2}v_{DS}^2\right) \\ \Rightarrow V_{DD} - |v_{DS}| &= Rk\left(v_{OV}\,|v_{DS}| - \frac{1}{2}v_{DS}^2\right) \end{aligned}$$

Once again we can arrange this in the form of a quadratic equation, but this time we will simplify the equation by leaving it in terms of v_{DS}, so we get:

$$\begin{aligned} 0 &= \frac{1}{2}Rkv_{DS}^2 - (1 + Rkv_{OV})\,|v_{DS}| + V_{DD} \\ |v_{DS}| &= \frac{(1 + Rkv_{OV}) \pm \sqrt{(1 + Rkv_{OV})^2 - 2RkV_{DD}}}{Rk} \\ &= \frac{(1 + Rkv_{OV}) - \sqrt{1 + (Rkv_{OV})^2}}{Rk} \end{aligned}$$

Notice that this is the exact same result as before, only it's "upside down." We can next get the solution for v_{OUT}:

$$\begin{aligned} v_{OUT} &= V_{DD} - |v_{DS}| \\ &= 5\,\text{V} - \frac{(1 + Rkv_{OV}) - \sqrt{1 + (Rkv_{OV})^2}}{Rk} \end{aligned}$$

Finally, if we suppose $Rkv_{OV}1 \gg 1$ then $v_{OUT} \to V_{DD}$.

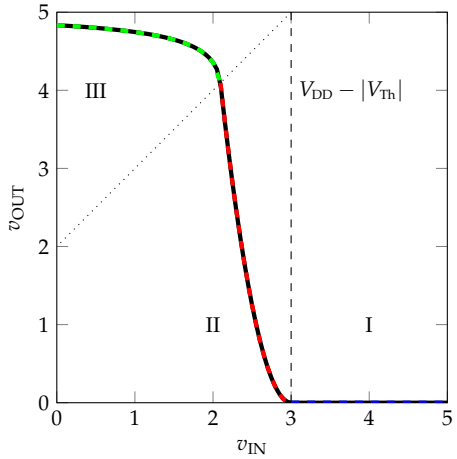


Figure 97: Comparison of analysis and simulation results for the PMOS RTL inverter.

Comparison of NMOS and PMOS versions

Our results show that both the NMOS and PMOS configurations have the same qualitative behavior. They both function as logic inverters. If we were to balance one of these circuits right in the center of its saturation region (II), where the slope is very steep, we could use it as an inverting amplifier. We will soon introduce linearized amplifier models that apply in the saturation region; it will be important to recognize that both NMOS and PMOS devices have the same linearized models in saturation, just as they show the same behavior in the RTL inverter configurations, even though they exhibit complementary logical behavior.

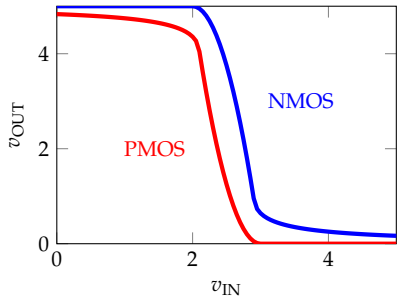


Figure 98: Overlay of the NMOS and PMOS RTL inverter transfer characteristics. Both devices behave as an inverter. They differ slightly in the offset voltage at which they "tip" from high to low.

Behavior in Saturation

We may now go one level deeper and examine the MOSFET's behavior in the saturation mode. First, let's understand why it's called "saturation." Suppose we hold the gate potential fixed so that $v_{GS} = 1\,\text{V}$ and perform a DC sweep on v_{DS} while measuring the current.

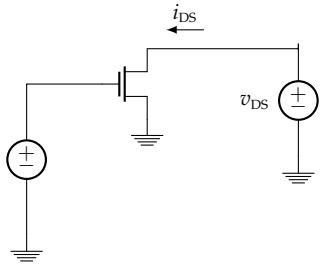


Figure 99: An experiment in which v_{GS} is held constant while sweeping v_{DS}.

According to the triode equation, the current should be a parabola:

$$i_{DS} = k\left((1)v_{DS} - \frac{1}{2}v_{DS}^2\right).$$

But if that were true, the current would begin to decrease when $v_{DS} > 1$ (dashed curve below), and eventually the current would swing negative, creating an impossible free-energy device. Obviously this doesn't happen, instead the device current rises monotonically until it reaches the peak of the parabola, and then flattens out at higher v_{DS} (solid curve). This is why it's called "saturation": when v_{DS} is swept from zero, i_{DS} increases until it saturates at a maximum value.

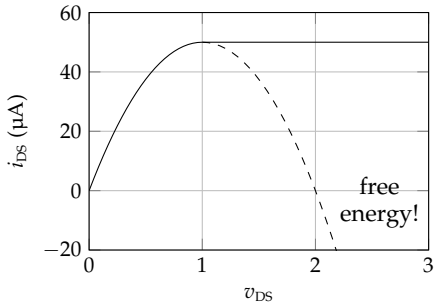


Figure 100: The triode equation predicts decreasing current when $v_{OV} > v_{DS}$ (dashed line). Physical laws dictate that the current should be non-decreasing as v_{DS} is increased, so the current **saturates** at a nearly constant value.

Since the saturation current is approximately constant, we may interpret the MOSFET as a nonlinear **voltage controlled current source** that depends on the gate voltage. For a first-order circuit analysis, we can replace the MOSFET symbol with a dependent current source. When a **small signal** is applied at the gate, we can write the gate voltage as a superposition of the DC signal (V_G) and the small signal (v_g). In that case we can **linearize** the current source by solving the first-order Taylor approximation:

$$i_D \approx V_{GS} + \left(\frac{d\,i_D}{d\,v_{gs}}\bigg|_{V_{GS}}\right) v_{gs}.$$

It will often be useful to analyze the **small-signal equivalent circuit**, which is obtained from the linearized model by **zeroing out DC independent sources**. For this purpose we may simplify the expression:

$$i_d \approx \left(\frac{d\, i_D}{d\, v_{GS}} \bigg|_{V_{GS}} \right) v_{gs}.$$

Since the MOSFET takes a voltage as input and produces a current as output, it is conventionally interpreted as a **transconductance amplifier**. The amplifier's transconductance gain, conventionally denoted as g_m, is defined as the derivative of i_D with respect to v_{GS}:

$$g_m \triangleq \frac{d\, i_D}{d\, v_{gs}} \bigg|_{V_{GS}}$$

Then we can write the device's small-signal behavior as simply

$$i_d = g_m\, v_{gs}.$$

Channel Length Modulation

In a real MOSFET device, the saturation current is not perfectly constant with increasing v_{DS}. Instead, we see an *approximately* linear increase with v_{DS}, which can be partially explained as a variation in the MOSFET's channel dimensions. From the circuit perspective, this behavior is modeled by augmenting the square-law equation with a "fudge factor" λ:

$$i_{DS} = \frac{1}{2} k v_{OV}^2 \left(1 + \lambda\, |v_{DS}|\right).$$

Typically λ is in the range from $0.01\,\mathrm{V}^{-1}$ to $0.1\,\mathrm{V}^{-1}$. When including channel length modulation (CLM), the curves are not completely flat in saturation.

Since we consider the MOSFET to be a transconductance amplifier, we can interpret the slope due to CLM as the **output resistance**:

$$r_o \triangleq \left(\frac{d\, i_D}{d\, v_{DS}} \bigg|_{DC} \right)^{-1}.$$

In this definition, the derivative is evaluated at the DC operating point, which encompasses all the DC values of V_{GS}, V_{DS} and I_D.

Calculating g_m and r_o

The transconductance and output resistance can be calculated in a few different ways. We could measure these parameters experimentally by using an ammeter to observe the changes

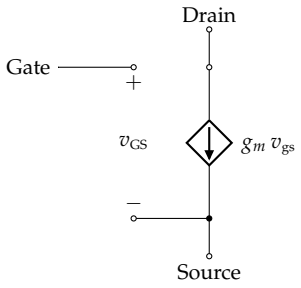


Figure 101: The MOSFET behaves like a dependent current source controlled by the gate-source voltage.

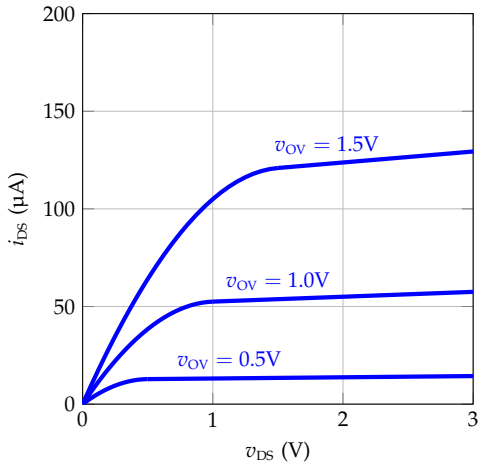


Figure 102: Transfer characteristic of i_D vs v_{DS} for an NMOS device at three different values of v_{OV}. The slope is not completely flat in the saturation region. This means the device should have a **finite differential resistance** when in saturation.

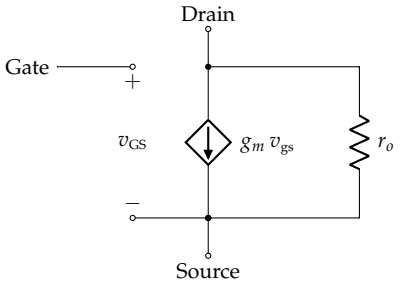


Figure 103: When CLM is included, it appears as an output resistance in the transconductance amplifier model.

in i_D that result from small variations in v_{GS} and v_{DS}. For hand analysis, we can directly integrate the device equations:

$$\begin{aligned}\frac{d\,i_D}{d\,v_{GS}} &= \frac{d}{d\,v_{GS}}\left(\frac{1}{2}k\,(v_{GS}-V_{Th})^2\right)\Big|_{DC}\\ &= k\,(V_{GS}-V_{Th})\\ &= kV_{OV}\end{aligned}$$

This tells us that the transconductance gain is directly proportional to the DC overdrive voltage.

In practice, it is often easier to select a DC bias current I_D, rather than to directly control V_{OV}. In that case, it is useful to express g_m in terms of I_D:

$$\begin{aligned}kV_{OV} &= \sqrt{k\,(k\,V_{OV}^2)}\\ &= \sqrt{2k\,I_D}\end{aligned}$$

This expression tells us that the transconductance gain is proportional to the square root of the DC bias current.

Lastly, to calculate the output resistance we need to consider CLM:

$$\begin{aligned}r_o &= \left(\frac{d}{d\,v_{DS}}\left(\frac{1}{2}kv_{OV}^2\,(1+\lambda\,v_{DS})\right)\Big|_{DC}\right)^{-1}\\ &= \left(\frac{1}{2}kv_{OV}^2\lambda\right)^{-1}\\ &= \frac{1}{\lambda\,I_D}\end{aligned}$$

This expression tells us that the output resistance is inversely proportional to the bias current. Since the transconductance increases with $\sqrt{I_D}$, there is a tradeoff between transconductance and output resistance. This tradeoff will have important consequences for practical circuit design.

Summary, Saturation Mode

Large Signal:

$$\begin{aligned}v_{OV} &= |v_{GS}| - |V_{Th}|\\ i_D &= \frac{1}{2}k\,v_{OV}^2\\ \text{when}\quad &|v_{DS}| > v_{OV}\\ \text{and}\quad &v_{OV} > 0.\end{aligned}$$

Small-Signal:

$$\begin{aligned}g_m &= k\,V_{OV}\\ &= \sqrt{2k\,I_D}\\ r_o &= (\lambda\,I_D)^{-1}\end{aligned}$$

Some important DC configurations

There are a few patterns that appear frequently in MOSFET circuits, and it will be useful to have their solutions available for reference. The two major cases are the diode connection and the passive resistor bias network. The third case is a combination of the first two. Additional configurations can be understood as special cases of these three configurations.

Diode connection: When the MOSFET's gate is directly connected to the drain terminal, it is referred to as a "diode connection." In this configuration, the drain terminal provides

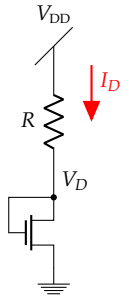


Figure 104: Diode-connected NMOS device.

a negative feedback loop to the gate terminal, so that the circuit settles into a stable DC state. Since v_{GS} is determined by the voltage drop across R, which is in turn determined by the current I_D, the solution is governed by feedback:

$$\begin{aligned} V_{GS} = V_D &= V_{DD} - I_D R \\ I_D &= \frac{1}{2} k V_{OV}^2 \\ &= \frac{1}{2} k \left(V_{DD} - I_D R - V_{Th} \right)^2 \end{aligned}$$

To complete the solution, we define a variable $x = \sqrt{I_D}$, and then arrange the above equation into a quadratic equation:

The diode connection is **guaranteed** to always be in the saturation mode, since $V_{DS} = V_{GS}$ it is always assured that $V_{DS} > V_{OV}$.

$$R\sqrt{\frac{k}{2}}x^2 + x - \sqrt{\frac{k}{2}}\left(V_{DD} - V_{Th}\right) = 0$$

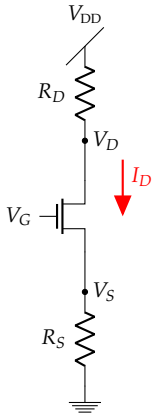
Applying the quadratic formula:

$$\begin{aligned} x &= \frac{-1 \pm \sqrt{1 + 2kR\left(V_{DD} - V_{Th}\right)}}{R\sqrt{2k}} \\ \Rightarrow I_D = x^2 &= \frac{\left(-1 + \sqrt{1 + 2kR\left(V_{DD} - V_{Th}\right)}\right)^2}{2kR^2} \end{aligned}$$

EveryCircuit Demonstration 20 (Diode connected NMOS device).

A diode connected configuration is implemented with $V_{DD} = 5\,\text{V}$, $R = 50\,\text{k}\Omega$, and the MOSFET parameters are $k = 500\,\mu\text{A/V}^2$, $V_{Th} = 0.5\,\text{V}$ and $\lambda = 0.05\,\text{V}^{-1}$. The analysis from this section predicts a bias current of $I_D = 78.8\,\mu\text{A}$, which is quite close to the simulated value. We can also verify that the gate voltage should be $V_G = 1.06\,\text{V}$, which is again quite close to the simulated result.

Passive bias network: In this very general case, resistors are placed adjacent to both the source and drain terminals, and the gate is biased at some constant voltage, such that the device is held in its saturation mode. There are interactions at both the drain and source terminal:



$$V_D = V_{DD} - I_D R_D$$
$$V_S = I_D R_S$$
$$I_D = \frac{1}{2} k \left(V_G - I_D R_S - V_{Th}\right)^2$$

we solve this case in much the same way as the diode-connected circuit. By defining $x = \sqrt{I_D}$, we can obtain a quadratic polynomial:

$$0 = R_S \sqrt{\frac{k}{2}} x^2 + x - \sqrt{\frac{k}{2}} \left(V_G - V_{Th}\right)$$
$$\Rightarrow x = \frac{-1 \pm \sqrt{1 + 2k R_S \left(V_G - V_{Th}\right)}}{R_S \sqrt{2k}}$$

and $I_D = x^2$.

EveryCircuit Demonstration 22 (NMOS bias network).

A NMOS passive bias configuration is implemented with $V_{DD} = 5\,\text{V}$, $R_D = 50\,\text{k}\Omega$, $R_S = 20\,\text{k}\Omega$, $V_G = 1.25\,\text{V}$, and the MOSFET parameters are $k = 500\,\mu\text{A/V}^2$, $V_{Th} = 0.5\,\text{V}$ and $\lambda = 0.05\,\text{V}^{-1}$. The analysis from this section predicts a bias current of $I_D = 22.5\,\mu\text{A}$, which is quite close to the simulated value. We can also verify that the source voltage should be $V_S = 0.45\,\text{V}$, and the drain voltage should be $V_D = 3.875\,\text{V}$. We can then verify that the device is biased in saturation since $V_{DS} = 2.44\,\text{V}$ whereas $V_{OV} = V_G - V_S - V_{Th} = 0.3\,\text{V}$. All of these calculations are very close to the simulation results.

MOSFETs as Switches

Many digital and mixed-signal applications use MOSFET devices as logic switches. Like all devices, MOSFETs make imperfect switches, but with careful design they can be used to realize complex and efficient logic circuits. When operated as a switch, we will primarily use the **triode** and **cutoff** operating modes, corresponding to ON and OFF states, respectively. Cutoff is easy to understand: the device is OFF, no current flows between the source and drain terminals, and we can treat it like an open circuit.

The ON behavior is more complicated. When a MOSFET switches ON, it usually transitions into the saturation mode first before settling into the triode mode. In some configurations, the MOSFET may be prevented from entering triode, making it "stuck" in saturation, where it performs poorly as a switch. As an example, consider the two scenarios shown in Figure 105.

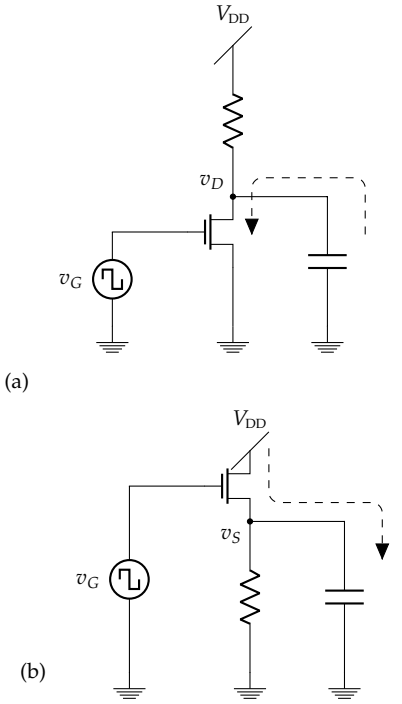


Figure 105: An NMOS device in (a) pull-down configurration, and (b) pull-up configuration.

The pull-down configuration is basically identical to the NMOS RTL inverter circuit. We now include the presence of a parasitic capacitance at the MOSFET's drain node. This capacitor must be charged or discharged in order to change the drain voltage. As a result, the circuit's transient behavior will look somewhat like the DC behavior we saw before. Suppose the device is initially OFF, and the capacitor at v_D is initially charged to V_{DD}. Then, when v_G transitions from zero to V_{DD}, the NMOS device is initially in cutoff. Once v_G crosses the device's threshold voltage, it enters the saturation region and begins to draw current that discharges the capacitor at v_D. After some time, the capacitor is discharged enough so that the device enters triode, and eventually it may be discharged to zero.

The pull-up configuration doesn't work as well. Suppose that the NMOS device in Figure 105(b) is initially OFF, and the capacitor is initially fully discharged so that $v_S = 0\,\text{V}$. Then when v_G switches from zero to V_{DD}, the NMOS device begins to turn on. But since $v_D = V_G = V_{DD}$, **it will never cross into triode** because $v_{DS} > v_{GS} - V_{Th}$, regardless of what happens at v_S. Therefore as the device tries to pull up v_S, its current is governed by the square law, $i_D = k\,v_{OV}^2$. As the capacitor charges up, v_{OV} eventually goes toward zero. In the end, the capacitor cannot be charged any higher than

$$v_{s,\,\text{max}} = V_{DD} - V_{Th},$$

because this is the voltage where v_{OV} becomes zero, and device switches OFF.

The behavior of a PMOS device is similar but complementary. PMOS devices pull-up well, but are not able to pull down. When a PMOS device is used to pull down a signal, the best it can achieve is a source voltage equal to $|V_{Th}|$. The PMOS analysis is identical to the NMOS analysis.

EveryCircuit Demonstration 24 (NMOS RTL inverter).

This example shows the NMOS RTL inverter configuration. THe transient simulation shows that when the NMOS device is ON, the active pull-down is very fast and effective. When the NMOS device is OFF, however, the passive pull-up operation is very slow due to the RC delay. Note that this example uses a manual switch to change the input signal. You have to click it to change the state.

EveryCircuit Demonstration 26 (NMOS pull-up and pull-down).

This circuit shows a combination of the pull-up and pull-down configurations shown in Figure 105. The two configurations are folded together, so there is no resistor current to overcome. The simulation shows that the NMOS device pulls down very well, but the pull-up operation is both slow and incomplete. A PMOS pull-down example is also available, showing the complementary behavior.

Ideal CMOS Inverter

MOSFET switching behavior is exemplified by the CMOS inverter circuit, which is the simplest logic gate. When the input (gate) voltage is high, the NMOS device is ON while the PMOS device is OFF. The NMOS device pulls the output low while the PMOS device does nothing. When the input is low, the NMOS device switches OFF while the PMOS device switches ON, pulling the output high. This results in a very clean logic inverter behavior. More importantly, when the gate is idle (not switching), there is no current through either the NMOS or PMOS device, so we say there is zero static power dissipation. Power is only dissipated during switching, when a small amount of energy must be expended to change the output voltage.

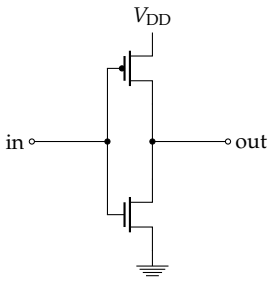


Figure 106: Standard CMOS inverter circuit.

Static CMOS Logic

The most successful and widespread application of MOSFET devices is CMOS logic, which is the foundation of modern digital electronics. The majority of CMOS logic circuits are based on the *static CMOS gate* structure, which uses a **PMOS pull-up network** in parallel with an **NMOS pull-down network**.

Static CMOS gates are based on four basic principles:

- Series (stacked) MOSFETs implement an AND operation
- Parallel MOSFETs implement an OR operation
- NMOS devices invert their outputs (i.e. place a bubble at the network's output port)
- PMOS devices invert their inputs (i.e. place bubbles at the network's input ports).

By using De Morgan's Laws, we can transform any logic function into a PMOS network and an NMOS network, then short their outputs together. According to de Morgan's Laws, a NAND gate is equivalent to an OR gate with inverted inputs, and a NOR gate is equivalent to an AND gate with inverted inputs.

So to produce a gate with some desired function F, we start with a logic gate implementation, and then apply transformations to create two versions: one with a single bubble at the output (for the NMOS network), and another with bubbles on all the top-level inputs (for the PMOS network). In both networks, there should be no bubbles in the connections between gates. We can also insert inverters onto the inputs and outputs in order to complement them as needed.

In the process of obtaining the PDN and PUN circuits, several circuit transformations are allowed. The basic gate conversions from de Morgan's Laws are always permitted, as are the alternative transformations shown in Figure 109. If a bubble is present on a connection, it can be moved to the other end of the connection, where it might be useful for applying a gate transformation. We can also insert "double bubbles" by placing a bubble at the start and end of a connection. Double bubbles can also be removed if needed, since they cancel each other out.

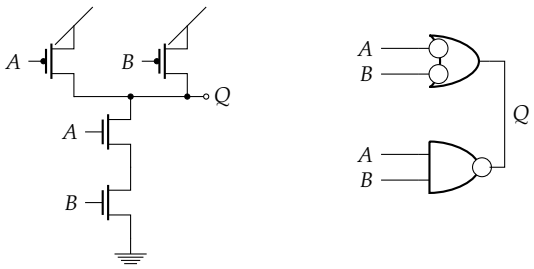


Figure 107: CMOS NAND gate and its logic interpretation. The PMOS pull-up network uses two devices in parallel, representing an OR function with inverted inputs. The NMOS pull-down network uses two devices in series, representing an AND function with inverted output. By de Morgan's Laws, these functions are equal, but complementary.

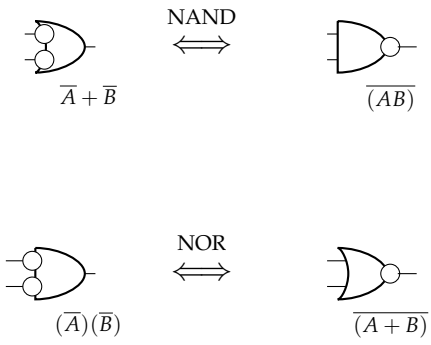


Figure 108: De Morgan's Laws: $\overline{(AB)} = \overline{A} + \overline{B}$ and $\overline{(A+B)} = (\overline{A})(\overline{B})$.

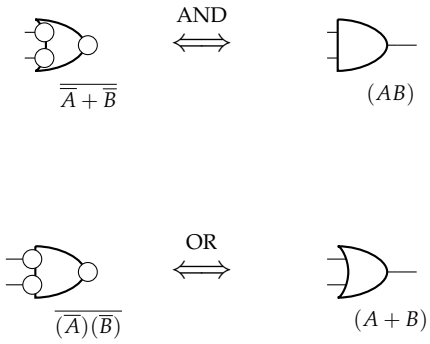
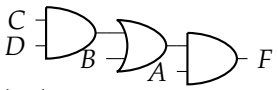


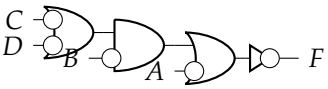
Figure 109: Alternative version of de Morgan's Laws: $AB = \overline{\overline{A} + \overline{B}}$ and $(A+B) = \overline{(\overline{A})(\overline{B})}$.

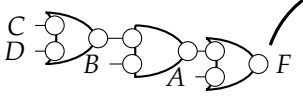
Example 21 (Static CMOS logic design).

Suppose we are given a four-input logic function $F = A(B + CD)$. To synthesize a static CMOS implementation, we begin with a classic logic circuit, and then apply de Morgan's laws to transform it:

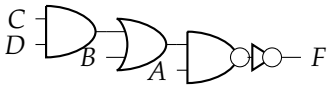


1A. PMOS PUN: Using de Morgan's Laws, make gate transformations to insert bubbles at the signal inputs.



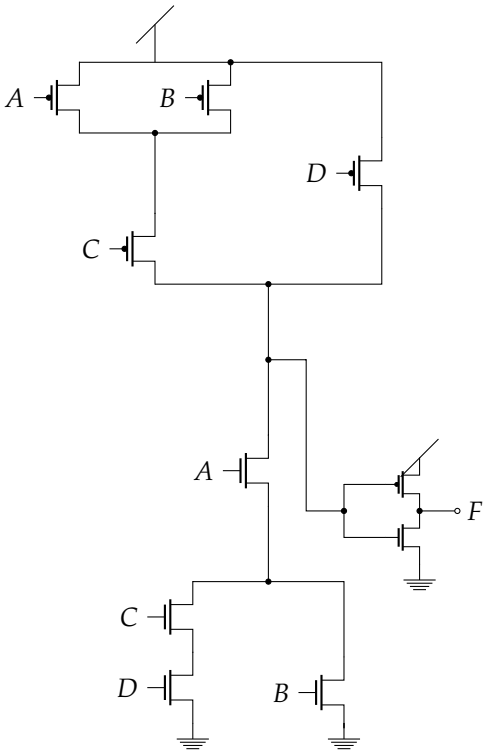


1B. Cancel the double-bubbles and add an inverter to eliminate the output bubble. Now there should be bubbles on all input signals, but no interior bubbles.



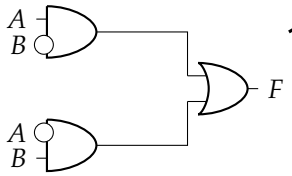
2. NMOS PDN: Since there are already no interior bubbles, simply place a bubble on the output signal and cancel it with an inverter. Now both the PUN and PDN have inverters at the output; for the transistor implementation, we ignore the inverters. An inverter will be attached to the gate's output after we are finished.

4. Transistor implementation: For both networks, begin on the left-most gate and compose a heirarchical structure. OR means a parallel connection, AND means a series (stacked) connection. Once the structures are built, connect the PMOS and NMOS networks together in the middle. Connect V_{DD} to the top of the PMOS network, and ground to the bottom of the NMOS network. Lastly, insert any needed inverters to complete the circuit.

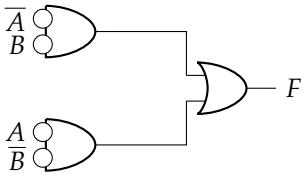


Example 22 (Static CMOS XOR gate).

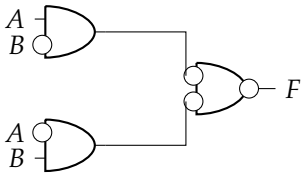
The Exclusive OR (XOR) function is crucial for many logic and arithmetic circuits. The function is $F = A \oplus B = A\overline{B} + \overline{A}B$. To synthesize a static CMOS XOR gate, we begin with a classic logic circuit, and then apply de Morgan's laws to transform it as follows:



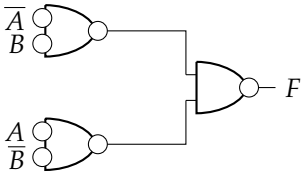
1. PMOS PUN: Just complement two of the input signals to get bubbles on all of them.



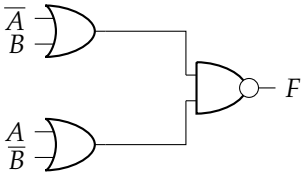
2A. NMOS PDN: Transform one gate to get a bubble on the output.

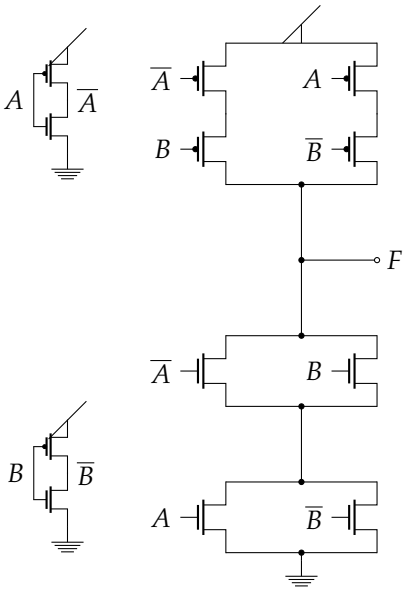


2B. Move the nuissance bubbles, and complement two input signals so the left-side AND gates are fully surrounded with bubbles.



2C. Transform the left-side AND gates to eliminate the bubbles.





3. To complete the transistor implementation, we proceed as before and implement AND gates as stacked connections, while OR gates are parallel connections. The PUN consists of two stacks (from the left-side AND gates) connected in parallel. The PDN consists of two parallel connections, stacked. Inverters are inserted to produce $\overline{A}$ and $\overline{B}$ (all of the signal wire connections are not shown). **In total, 12 transistors are needed.**

Transmission Gates

In order to make a general-purpose switch that can be used for both pull-up and pull-down operations, we can simply connect PMOS and NMOS devices in parallel. This circuit is commonly called a transmission gate. It acts as a passive switch, meaning it cannot directly provide energy to its terminals, it can only transfer energy.

Transmission gates are useful for efficiently realizing several types of logic gates. For example, transmission gates provide one of the simplest realizations for the exclusive-OR (XOR) gate. Transmission gates also provide a natural realization for multiplexor (MUX) gates.

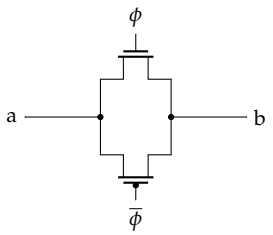


Figure 110: A CMOS transmission gate acting as a switch between nodes a and b. The switch is controlled by a logic signal $\phi \in \{0,\, V_{DD}\}$. The switch is ON when ϕ is high. An inverter is required to produce $\overline{\phi}$, which is needed to switch the PMOS device.

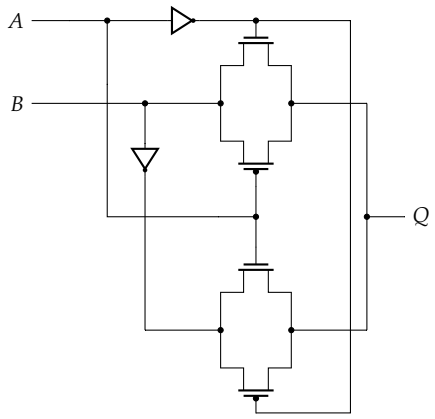


Figure 111: XOR gate based on transmission gates. The logic function is $Q = \overline{A}B + A\overline{B}$. This circuit requires a total of eight MOSFETs.

Analog Switching

In addition to digital applications, transmission gates are useful for switching analog signals. One common application is the **track and hold (T/H)** circuit, which serves as the front-end for many sampling circuits, such as analog-to-digital converters. The most basic T/H circuit contains only a capacitor and switch. When the switch is ON, the capacitor "tracks" the voltage of the input signal. When the switch is OFF, the capacitor is left floating. In this condition, no charge can be added or removed from the capacitor, so it "holds" whatever voltage it contained at the moment when the switch turned OFF.

When a T/H circuit is used to sample a slow-changing analog signal, then there will usually be a small signal difference between each sampling event. As a result, $|v_{DS}|$ will tend to be small each time the transmission gate is switched ON, so the devices start out in triode rather than saturation. This scenario is an example of **small-signal switching**. When the devices are ON, it is useful to think of them as approximate resistors, where the resistance is

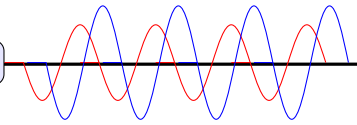
$$\begin{aligned} r_{ON} &\triangleq \left[\frac{d\, i_D}{d\, v_{DS}}\right]^{-1}_{\text{DC, triode}} \\ &= \left[\frac{d}{d\, v_{DS}} k\left(v_{OV} v_{DS} - \frac{1}{2} v_{DS}^2\right)\right]^{-1}_{\text{DC}} \\ &= \left[k\left(V_{OV} - V_{DS}\right)\right]^{-1} \end{aligned}$$

When the device is fully ON, V_{DS} may be very small. In switching applications, V_{OV} is always $V_{DD} - V_{Th}$, so for a fully-ON device we can say that the small-signal equivalent resistance is

$$r_{ON} = \frac{1}{k\left(V_{DD} - V_{Th}\right)}.$$

Lastly, for a transmission gate, the equivalent resistance is the parallel combination of the r_{ON} values for the NMOS and PMOS devices.

EveryCircuit Demonstration 28 (Transmission gate track-and-hold circuit).



This circuit implements a passive T/H circuit based on the transmission gate switch. The circuit has supply voltage $V_{DD} = 5\,\text{V}$ and a 100 nF hold capacitor. The device parameters are as follows:

NMOS		PMOS	
k_n	$= 500\,\mu\text{A/V}^2$	k_p	$= 250\,\mu\text{A/V}^2$
V_{ThN}	$= 0.5\,\text{V}$	V_{ThP}	$= 0.5\,\text{V}$
λ_n	$= 0.05\,\text{V}^{-1}$	λ_p	$= 0.1\,\text{V}^{-1}$

Based on these parameters, the triode ON resistance should be

$$\begin{aligned} r_{ON,n} &= \left[k_n\ (V_{DD} - V_{ThN})\right]^{-1} \\ &= 444\,\Omega \\ r_{ON,p} &= \left[k_p\ (V_{DD} - V_{ThP})\right]^{-1} \\ &= 888.9\,\Omega \end{aligned}$$

In parallel, the total ON resistance for this switch should be $r_{ON} = 296\,\Omega$. When the switch is turned ON, the output signal's rise time is determined by the time constant formed by r_{ON} and C, which works out to be $\tau = 29.63\,\mu\text{s}$. Then the 10–90% rise time is $t_r = 2.2\tau = 65.2\,\mu\text{s}$. This gives an indication as to the minimum switching period for this T/H circuit. In practice, 10–90% is not sufficient to obtain a high-precision sample, so a switching period five to ten times slower may be required.

The simulation example shows an input signal at 100 Hz with a switching period of 500 µs, which is almost 10× higher than the calculated rise time. By zooming in on the waveform, you can see that the tracking accuracy improves gradually with more time in the tracking phase. The slow convergence is a direct consequence of the ON resistance in the transmission gate.

MOSFETs as Amplifiers

We saw previously how the MOSFET device can be interpreted as a transconductance amplifier: the input signal is v_{GS}, and the output signal is i_D. We can build on this concept by configuring the MOSFET in several ways to make different types of amplifiers. In all cases, we will deliberately operate the device in its saturation mode, balanced between its ON/OFF states.

Common-Source Configuration

As a first example, we consider the RTL inverter circuit, only now we will try and balance the circuit at the point where its transfer characteristic is steepest. We refer to this as the **quiescent point, Q point, bias point, or DC operating point**. We then superimpose a small AC signal on top of the DC operating point. **Amplifier design is therefore divided into two tasks: biasing and small-signal analysis.** We'll consider biasing strategies later. In this section we focus on basic small-signal analysis techniques, as they dictate amplifier behavior and potential applications.

To begin with, we consider the common-source configuration and assume it is appropriately biased at a suitable DC operating point. To analyze the small-signal behavior, we replace the MOSFET with its small-signal equivalent model (the transconductance amplifier model). Second, we zero-out any DC independent sources. This means that the V_{DD} node gets shorted to ground, so any devices connected to it are "folded over" onto the ground node.

To analyze the amplifier characteristics, we use the small-signal equivalent circuit to solve for the **gain** and **output resistance**. From the model in Figure 113, we see that the amplifier consists of a current source and two resistors. Since the two resistors appear in parallel, we can merge them as $R_o' = r_o \parallel R$. Then the output voltage is simply the voltage drop across R_o'. Since the current is drawn upward through R_o', the voltage drop is negative. Solving for the gain:

$$v_{out} = -\left(g_m\, v_{in}\right)\, R_o'$$

$$\Rightarrow A_v = \frac{v_{out}}{v_{in}} = -g_m\, R_o'$$

To solve the output resistance, we set the input signal to zero and solve for the equivalent resistance seen looking into the output node. Since there is a literal resistance of R_o' at that node, the output resistance is clearly R_o'.

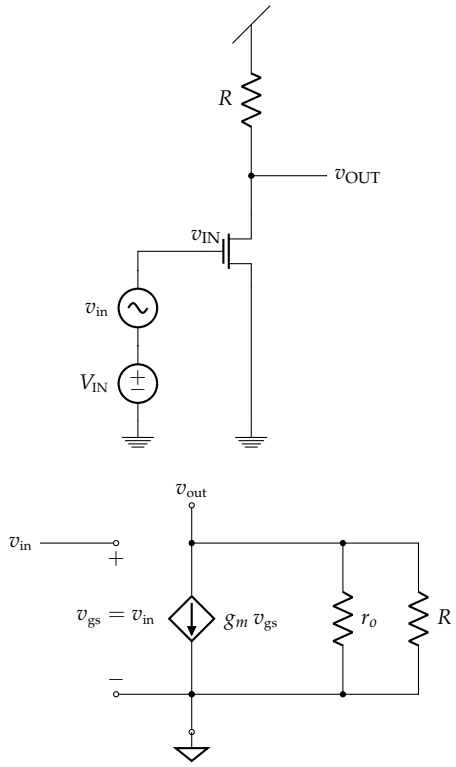
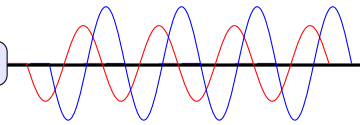


Figure 112: NMOS common-source amplifier configuration and its small-signal equivalent model. Since V_{DD} is shorted out in the small-signal model, the bias resistor R appears in parallel with the device's internal resistance r_o.

Summary: CS amp

- Inverting amplifier
- Output resistance: $R_{OUT} = r_o \parallel R$
- Gain: $A_v = -g_m\, R_{OUT}$

EveryCircuit Demonstration 30 (NMOS Common-Source Amplifier).



This example shows a basic common-source configuration for an NMOS device with $k_n = 500\,\mu\text{A/V}^2$, $\lambda_n = 0.05\,\text{V}^{-1}$ and $V_{\text{ThN}} = 0.5\,\text{V}$. The supply voltage is 5 V. The gate is biased with a DC operating voltage of $V_{\text{IN}} = 0.9\,\text{V}$, and the bias resistor is $R = 50\,\text{k}\Omega$. Capacitive coupling is used at the gate to separate the DC bias voltage from the AC small-signal input. Capacitive coupling is also used at the drain to remove the DC offset from the output signal.

Based on these parameters, we can calculate the device's small-signal characteristics, and then obtain the gain and output resistance as follows:

$$\begin{aligned} V_{\text{OV}} &= 0.9\,\text{V} - 0.5\,\text{V} = 0.4\,\text{V} \\ I_D &= \frac{1}{2}\,k_n\,V_{\text{OV}}^2 = 40\,\mu\text{A} \\ V_{\text{OUT}} &= V_{\text{DD}} - I_D\,R = 3\,\text{V} \\ g_m &= k_n\,V_{\text{OV}} = 200\,\mu\text{A/V} \\ r_o &= (\lambda_n\,I_D)^{-1} = 500\,\text{k}\Omega \\ R_{\text{OUT}} &= r_o \parallel R = 45\,\text{k}\Omega \\ A_v &= -g_m\,R_{\text{OUT}} = -9\,\text{V/V} \end{aligned}$$

Run the transient simulation and verify that the predicted gain and output offset are correct. You will probably notice that the simulated output offset is 2.727 V. Can you explain this discrepancy? (Hint: consider the effect of CLM with $V_{\text{DS}} = 2.727\,\text{V}$, then calculate new values for I_D and V_{OUT}).

PMOS Common-Source Configuration

Now let's consider the complementary PMOS version of the common-source circuit. This circuit is obtained by swapping the vertical positions of the MOSFET and resistor. In the PMOS device, the drain current has an inverse response to the gate voltage: when v_{IN} rises, i_D falls. Since the resistor is positioned between the drain and ground, a smaller current means a smaller output voltage at the drain. The result is that **the small-signal behavior is the same for both the NMOS and PMOS versions**.

To obtain the small-signal equivalent circuit, we zero-out V_{DD} and V_{IN}, so that the PMOS source terminal is connected to small-signal ground. Even though the PMOS device current has an inverse response to the gate voltage, we can flip the device upside down so that the source terminal is folded back onto the ground node. We then obtain **the exact same model as we had for the NMOS version**. What this means is that every NMOS circuit configuration should have a complementary PMOS version with the exact same behavior. The only differences will be in the device's k, V_{Th} and λ parameters.

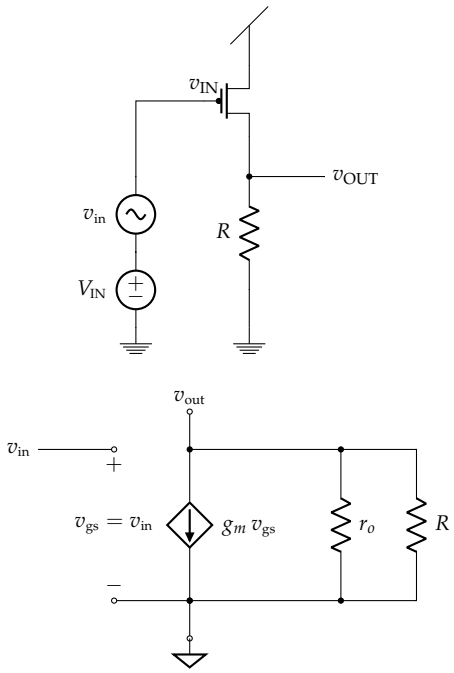


Figure 113: PMOS common-source amplifier configuration. Its small-signal equivalent model is the same as the NMOS version.

EveryCircuit Demonstration 32 (PMOS Common-Source Amplifier).

This example shows a PMOS version of the common-source amplifier. The parameters very similar to the NMOS case: $k_p = 250\,\mu A/V^2$, $\lambda_p = 0.1\,V^{-1}$ and $V_{ThP} = 0.5\,V$. The supply voltage is 5 V. The gate is biased with a DC operating voltage of $V_{IN} = V_{DD} - 0.9\,V$, and the bias resistor is $R = 50\,k\Omega$. Then:

$$\begin{aligned} V_{OV} &= 0.9\,V - 0.5\,V = 0.4\,V \\ I_D &= \frac{1}{2} k_p V_{OV}^2 = 20\,\mu A \\ V_{OUT} &= I_D R = 1\,V \\ g_m &= k_p V_{OV} = 100\,\mu A/V \\ r_o &= \left(\lambda_p I_D\right)^{-1} = 500\,k\Omega \\ R_{OUT} &= r_o \parallel R = 45\,k\Omega \\ A_v &= -g_m R_{OUT} = -4.5\,V/V \end{aligned}$$

Run the transient simulation and verify that the predicted gain and output offset are correct. You will probably notice that the simulated output offset is 1.364 V. Can you explain this discrepancy? (Hint: consider the effect of CLM with $|V_{DS}| = V_{DD} - 1.364\,V$, then calculate new values for I_D and V_{OUT}).

Common-Source with Active Bias

In the previous examples, we considered CS amplifiers where MOSFET is coupled with a resistor. It is often more useful to consider the *active bias* configuration, where the resistor is replaced by an ideal current source. This removes R from the small-signal model. Since the bias current is forced by an ideal DC independent current source, in the small-signal model contains an open-circuit at the MOSFET's drain node. As a result, this configuration achieves the highest possible gain magnitude for a given MOSFET device. The gain and output resistance are

$$A_{vo} = -g_m r_o$$
$$R_{\text{out}} = r_o$$

The gain magnitude of this configuration, $g_m r_o$, is commonly referred to as the **intrinsic gain** of the MOSFET, since it is the highest gain achievable with a single MOSFET device. When the circuit is analyzed with no load attached, it is referred to as the "open-circuit gain" and the subscript letter 'o' is added in A_{vo} to signify this.

In practice, a nearly-ideal current source can be implemented using a MOSFET device with a constant gate voltage. For example, a PMOS device can be substituted in place of the current source. The PMOS gate voltage, V_{GP}, should be chosen so that the device is biased in its saturation mode. In that configuration, the PMOS device is insensitive to the voltage at its drain terminal, so its constant gate voltage maintains a constant bias current I_D.

Since the PMOS device is not perfectly ideal, it contributes a **load effect** due to its intrinsic resistance r_o. In the small-signal model, the NMOS and PMOS r_o's will appear in parallel, so the output resistance and gain are slightly modified:

$$R_{\text{out}} = r_{o,n} \parallel r_{o,p}$$
$$A_v = -g_m R_{\text{out}}$$

By using a PMOS device the circuit's gain is roughly cut in half due to the interaction of r_o's. In general, an amplifier's output node is connected to two branches, one "going up" toward V_{DD} and another "going down" toward ground. The total output resistance is taken as the parallel combination of equivalent resistance looking up with the resistance looking down, i.e. $R_{\text{out}} = R_{\text{up}} \parallel R_{\text{down}}$.

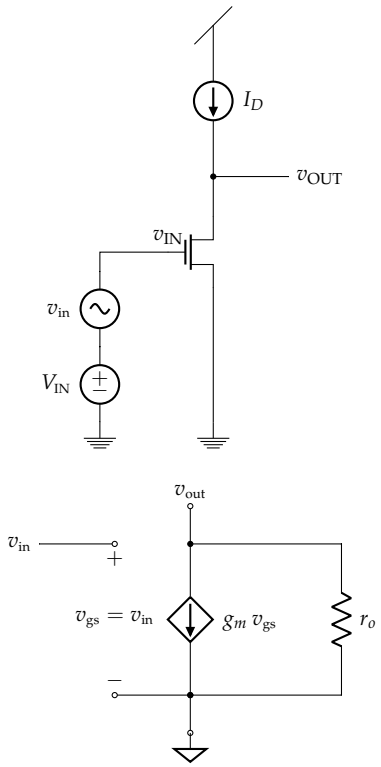


Figure 114: NMOS active-bias common-source amplifier configuration and its small-signal equivalent model. The current source directly forces a DC bias current of I_D in the NMOS device. Since the bias current is forced by a DC independent source, it is **zeroed out** in the small-signal model, leaving an **open-circuit** at the output node.

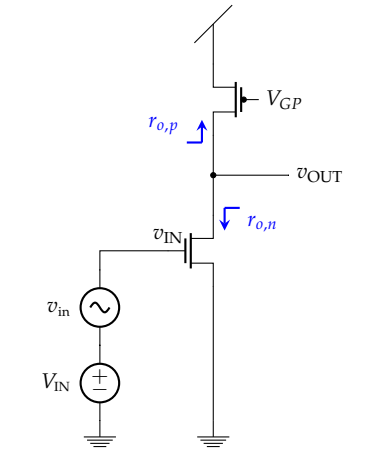


Figure 115: NMOS active-bias common-source amplifier configuration with PMOS bias device. The PMOS device acts as a current source.

Common-Source with Source Degeneration

The active-bias CS amplifier is extremely sensitive to its bias point. If the DC gate voltage is off by a small error, then the circuit is easily driven to its rail voltages and rendered useless. In order to relax the bias sensitivity, we can insert a *degeneration resistor* under the source terminal.

To solve the gain for this configuration, we first observe that **the output node is open-circuited** in the small-signal equivalent circuit model, since DC bias current source was zeroed out. In that case, the current flowing into the output branch must be zero. If a portion of the circuit is enclosed by the dashed box shown in Figure 116, then the total current flowing into the box has to equal the total current flowing out of the box (this is a version of Kirchoff's current law). The MOSFET does not allow any current at its gate terminal, so the gate current is zero. The output terminal is open-circuited, so the drain current is also zero. The only remaining branch is the source terminal, which must be zero since there is no other route for current to flow into the box. Since $i_s = 0$, there is no voltage drop across R_S, so the source voltage is also zero.

As a result of this analysis, the model for solving the gain of this circuit is identical to the model in Figure 114, so the gain must be exactly the same, $A_v = -g_m r_o$. Where the models differ is in the output resistance. To find R_{out} for this circuit, we zero out the input signal and apply a test voltage at v_{out}. Then we solve for the current that flows through the output branch. Since the output node is no longer open-circuited, a non-zero current flows through the drain and source terminals, with $i_d = i_s = i_{\text{out}}$. Also, since $v_{\text{in}} = 0$, the gate-source voltage is $v_{\text{gs}} = -v_s = -i_s R_S$. Based on these considerations, we obtain the circuit shown in Figure 117.

To solve for the output resistance, we consider the voltage drop across r_o. Two downward currents are superimposed on r_o:

$$\begin{aligned} v_{\text{out}} &= v_s + r_o\,(g_m\, v_s + i_s) \\ &= i_s\, R_S + r_o\,(g_m\, i_s\, R_S + i_s) \\ \Rightarrow R_{\text{out}} = \frac{v_{\text{out}}}{i_s} &= R_S + r_o + g_m\, r_o\, R_S \end{aligned}$$

So although the active-bias open-circuit gain is the same when source degeneration is present, the output resistance is much higher. This should result in a more significant coupling effect when a load is connected.

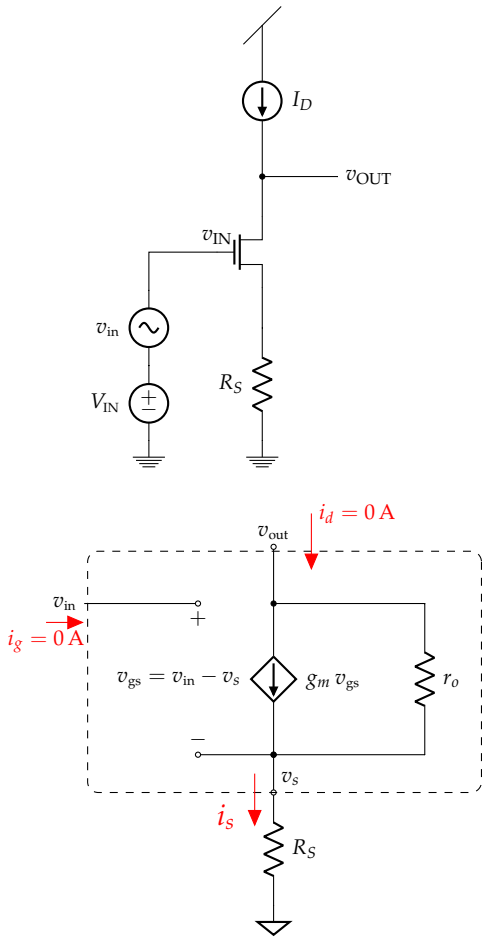


Figure 116: NMOS active-bias common-source amplifier with source degeneration resistor R_S. The effect of R_S is to reduce the amplifier's gain while improving error tolerance in the bias point.

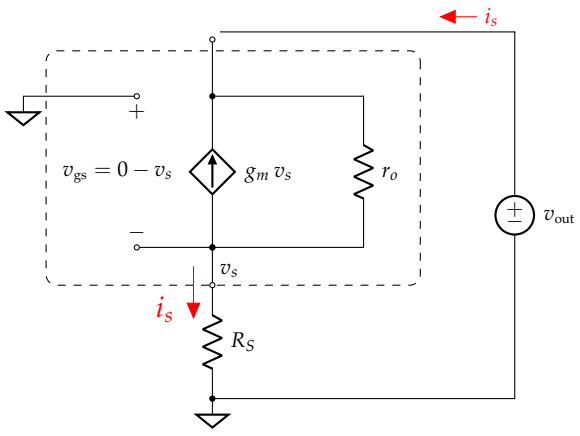


Figure 117: Finding the output resistance for the degenerated amplifier.

Common-Source Amplifier with Passive Bias and Degeneration

In the passive-bias configuration, we can leverage our previous analyses to solve the small-signal behavior without repeating the entire process. This circuit can be viewed as a superposition of the active-bias open-circuit configuration with the bias resistor R_D applied as a load. The gain can be considered as the loaded-gain of the active-bias version:

$$\begin{aligned} A_{vL} &= (-g_m\, r_o)\, \frac{R_D}{R_D + R_{\text{out}}} \\ &= \frac{-g_m r_o R_D}{R_D + R_S + r_o + g_m r_o R_S} \end{aligned}$$

Another way of looking at it is that the resistance R_D summarizes the circuit's "up" branch, and the open-circuit amplifier summarizes the circuit's "down" branch. The two branches can be analyzed separately, and then joined together via a coupling analysis. After coupling, the new overall output resistance is $R'_{\text{out}} = R_{\text{out}} \parallel R_D$.

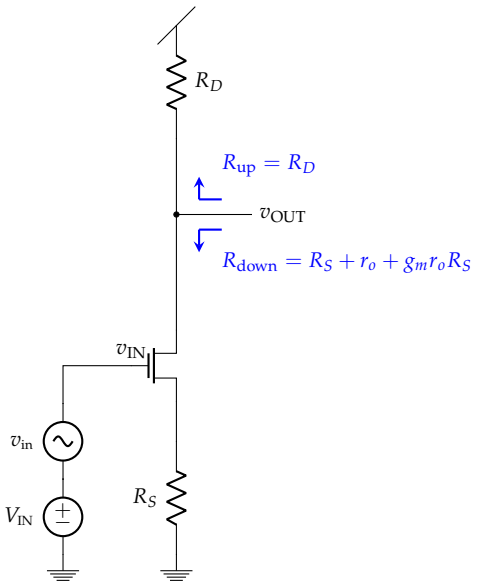


Figure 118: NMOS CS amplifier with passive bias and source degeneration.

Common-Source Amplifier with PMOS Bias and Degeneration

When a PMOS device is used to supply the amplifier's active bias current, we can adopt the same approach as in the passive case. We now consider the amplifier to be loaded by the r_o of the PMOS device:

$$\begin{aligned} A_{vL} &= (-g_m\, r_{o,n})\, \frac{r_{o,p}}{r_{o,p} + R_{\text{out}}} \\ &= \frac{-g_m r_{o,n} r_{o,p}}{R_S + r_{o,n} + r_{o,p} + g_m r_{o,n} R_S} \end{aligned}$$

In both the passive and PMOS biased circuits, we make use of the idealized amplifier model shown below.

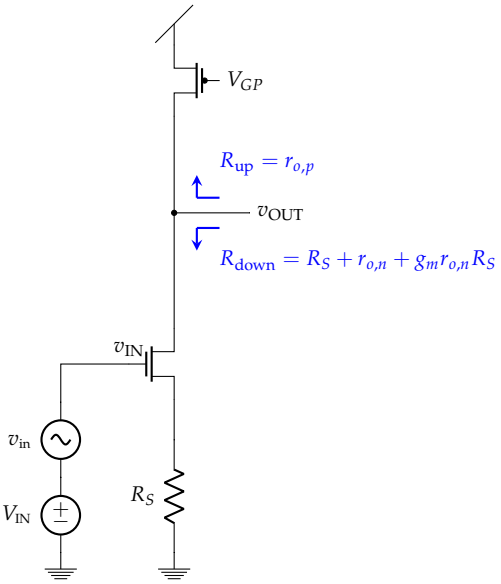


Figure 119: NMOS CS amplifier with passive bias and source degeneration.

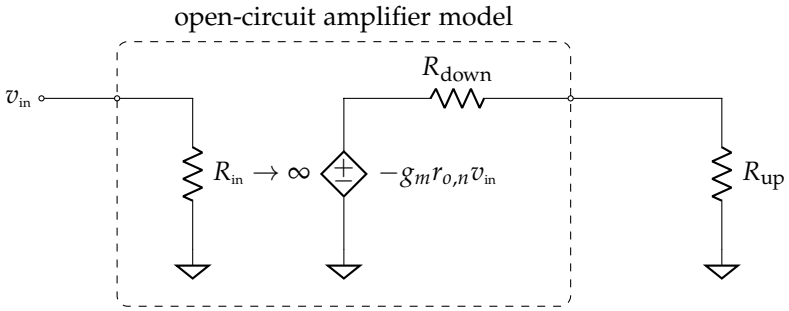


Figure 120: Amplifier model separating the upper bias portion (modeled as a load) from the lower portion (modeled as an open-circuit amplifier configuration).

Example 23 (Passive-biased CS amp with source degeneration).

Consider a passive-biased common-source amplifier like the one shown in Figure 119. The NMOS device has parameters $k = 500\,\mu\text{A/V}^2$, $V_{\text{Th}} = 0.5\,\text{V}$ and $\lambda = 0.05\,\text{V}^{-1}$. The bias resistor is $R_D = 50\,\text{k}\Omega$ and the degeneration resistor is $R_S = 20\,\text{k}\Omega$. If the input offset voltage is $V_{\text{IN}} = 1.25\,\text{V}$, what is the circuit's gain?

To solve this problem, we first solve the DC operating point and then calculate the small-signal parameters. Referring back to the bias configurations studied earlier in the chapter, we see that this circuit is already covered by the passive bias network analysis. In a previous example, we found that the bias current should be $I_D = 22.5\,\mu\text{A}$. Then $g_m = \sqrt{2kI_D} = 150\,\mu\text{A/V}^2$, $r_o = (\lambda I_D)^{-1} = 889\,\text{k}\Omega$, $g_m r_o = 133\,\text{V/V}$, and $R_{\text{down}} = 3.58\,\text{M}\Omega$. The high intrinsic gain looks pretty promising, but the resistive coupling effect is going to ruin it. Putting all this together, the amplifier's loaded gain is

$$\begin{aligned} A_{vL} &= (-g_m r_o)\left(\frac{R_D}{R_D + R_{\text{down}}}\right) \\ &= -1.83\,\text{V/V}. \end{aligned}$$

With the large source degeneration resistance, the circuit does not make a very good amplifier.

If the degeneration resistance could be removed (while keeping the bias current the same), then the output resistance would be a much smaller value of r_o, and in that case the loaded gain would be much better:

$$A_{vL} = (-g_m r_o)\left(\frac{R_D}{R_D + r_o}\right) = -7\,\text{V/V}.$$

Benefits of source degeneration

From the preceding analysis, it sounds like source degeneration is purely harmful, since it significantly reduces the gain. There are three good reasons for understanding source degeneration:

1. It is sometimes an unavoidable feature of some circuits.
2. It models the coupling behavior when multiple MOSFET amplifiers are folded together into a complex circuit.
3. It provides a looser error tolerance for biasing the common-source amplifier.

Of these reasons, the third point is the most practical consideration at this stage in our study of MOSFET amplifiers. A high-gain CS amplifier can be difficult to successfully bias in practice. Since the transfer characteristic is very steep, a slight error can cause the amplifier to rail, making it useless. **By inserting a degeneration resistor, we can flatten the transfer characteristic and make it more tolerant to bias error.**

A collection of simulated DC transfer characteristics is shown in Figure 122. The degeneration resistance is varied from zero up to 20 kΩ. With increasing values of R_S, two drawbacks are visible. First, the gain is diminished, which is evident from the flatter slope in curves with higher R_S. Second, the output signal range is diminished, since the transfer characteristic flattens out at a higher voltage. This limits the minimum output voltage that can appear, so we can't produce a full 5 V rail-to-rail signal in this example.

Using a bypass capacitor

For applications where only high-frequency signals need to be amplified, a win-win solution is possible by inserting a bypass capacitor across the degeneration resistor. The bypass resistor has the effect of shorting out R_S when processing high-frequency signals. But at DC, the capacitor has no effect on the circuit.

At a given frequency f, the bypass capacitor C_B behaves approximately like a resistance of magnitude $(2\pi f C_B)^{-1}$. At higher frequencies this resistance tends toward zero, hence "bypassing" R_S. The amplifier's gain will then tend toward the loaded gain without degeneration:

$$A_{vL} \to -g_m r_o \left(\frac{R_D}{R_D + r_o} \right).$$

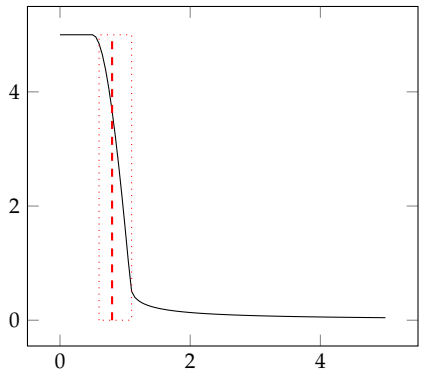


Figure 121: DC transfer characteristic of a CS amplifier with no source degeneration. The amplifier will not function if the input offset strays outside the red box, leaving little tolerance for error.

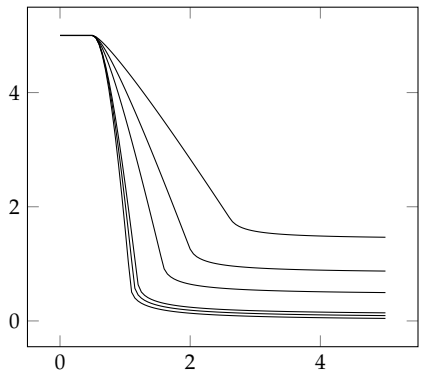


Figure 122: DC transfer characteristic of a CS amplifier with several values of source degeneration. The degenerated amplifier is more forgiving of bias errors, but has a flatter slope and therefore lower gain.

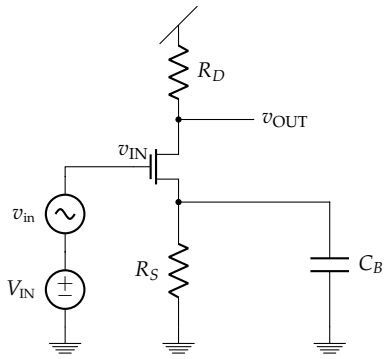
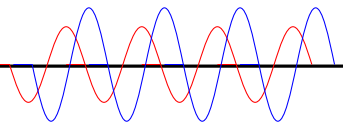


Figure 123: NMOS CS amplifier with passive bias, source degeneration and bypass capacitor.

Netlist 8: DC sweep of common-source degeneration resistances

```
* Common-Source amplifier with source degeneration
.model ntype NMOS(KP=100e-6,VTo=0.5,LAMBDA=0.05)

VDD  ndd  0 DC 5V
VIN  ndc  0 DC 1V
vsig nsig 0 SIN(0 0.1 1k)

RDC  ng   ndc 1Meg
Cin  nsig ng  10uF

RL   nout 0    1Meg
Cout nd   nout 10uF

M1 nd  ng ns 0 ntype W=1u L=200n
RD ndd nd 50k
RS ns  0  1

.control
* Foreach loop to scan through RS values:
foreach RSval 0 500 1000 5000 10000 20000

alter RS = $RSval
dc VIN 0 5 0.05

end

plot dc1.nd dc2.nd dc3.nd dc4.nd dc5.nd dc6.nd
wrdata cs_degenerated_dc dc1.nd dc2.nd dc3.nd dc4.nd dc5.nd dc6.nd
.endc
.end
```

EveryCircuit Demonstration 34 (CS amplifier with bypass capacitor).



This example implements the NMOS common-source amplifier described in the SPICE netlist above, with a passive bias resistor $R_D = 50\,\text{k}\Omega$, a degeneration resistor $R_S = 20\,\text{k}\Omega$, and a bypass capacitor $C_B = 10\,\mu\text{F}$. With the bypass capacitor in place, the gain is close to $-7\,\text{V/V}$ as we predicted in example 23. If you remove the bypass capacitor, you should notice that the gain drops to about $-1.8\,\text{V/V}$ as predicted.

The circuit is highly tolerant of different DC input bias voltages. Try adjusting the DC gate voltage source between 1 V and 1.8 V. The circuit continues to function throughout this range, while maintaining a gain close to the $-7\,\text{V/V}$ target. Try repeating the simulations with R_S and C_B removed (i.e. shorted out). The results will not be as robust for different gate offset voltages.

Amplifier analysis: general principles

We've now seen several different ways to configure the common-source amplifier. All of these configurations can be unified into a general-purpose small-signal analysis procedure. To analyze any configuration, we only need the following information:

1. The **ideal amplifier model** is obtained by analyzing the **open-circuit gain** of an active-bias configuration.

2. The ideal output resistance is equal to the equivalent resistance looking into the corresponding terminal of the ideal active-bias configuration.

3. To account for the circuit's real bias source (whether passive, PMOS, or something else), we consider the bias device to be a load resistance which forms a voltage divider at the amplifier's output.

This general framework is suitable for analyzing all MOSFET amplifier configurations. To solve the terminal resistances, we only need two general-purpose theorems that reveal the resistance looking into the drain and source terminals.

Resistance into the drain: In any configuration, we can quickly solve R_{drain}, the equivalent small-signal resistance looking into the drain terminal of a MOSFET device. To do this, we first summarize any circuitry present under the source terminal, and treat it as a single equivalent resistance R_S. Then the circuit is reduced to the exact same model as the CS amplifier with source degeneration. We found that

$$R_{\text{drain}} = R_S + r_o + g_m r_o R_S$$

This result covers all possible cases. When the source terminal is connected directly to ground, $R_S = 0$, then $R_{\text{drain}} = r_o$. If there is an ideal current source under the source terminal, then $R_S \to \infty$, in which case $R_{\text{drain}} \to \infty$.

Resistance into the source: To find the resistance looking into the source terminal, we summarize any circuitry present above the drain as an equivalent resistance, R_D. We then apply a test voltage at the source and solve for the current that flows

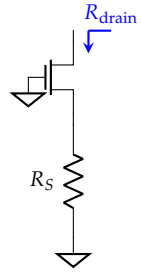


Figure 124: General resistance into the drain terminal: $R_{\text{drain}} = R_S + r_o + g_m r_o R_S$.

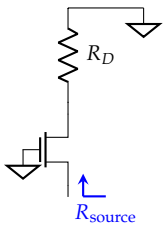


Figure 125: General resistance into the source terminal: $R_{\text{source}} = \frac{R_D + r_o}{1 + g_m r_o}$.

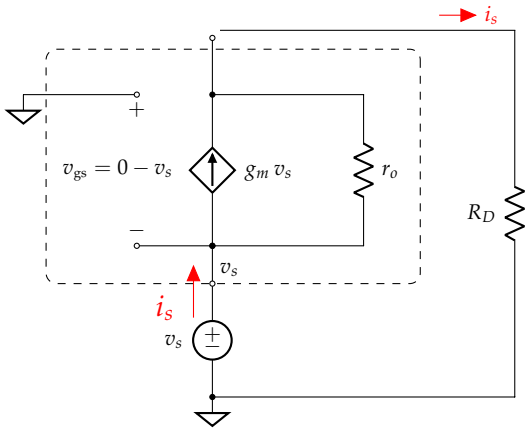


Figure 126: Model for solving the resistance looking into the source terminal.

into the source terminal:

$$\begin{aligned} i_s &= \frac{v_d}{R_D} \\ v_d &= v_s + r_o\left(g_m v_s - i_s\right) \\ \Rightarrow i_s\left(R_D + r_o\right) &= v_s\left(1 + g_m r_o\right) \\ \Rightarrow R_{\text{source}} = \frac{v_s}{i_s} &= \frac{R_D + r_o}{1 + g_m r_o}. \end{aligned}$$

In the coming sections we will apply these general principles to an expanding array of configurations.

Common-Gate amplifier configuration

In the common-gate (CG) configuration, the input signal is applied to the source terminal, the output is sampled from the drain terminal, while the gate terminal is held at a constant bias voltage. In the small-signal equivalent model, the gate voltage is zeroed-out to small-signal ground, and the bias current source is zeroed-out so that it becomes an open-circuit. Since no current flows out of the open-circuited drain terminal, there must also be no current flowing through R_S, i.e. $i_s = 0$. Therefore $v_s = v_{\text{in}}$. Then v_{out} is determined by the voltage drop across r_o:

$$\begin{aligned} v_{\text{out}} &= v_{\text{in}} + g_m r_o v_{\text{in}} \\ \Rightarrow A_{vo} = \frac{v_{\text{out}}}{v_{\text{in}}} &= 1 + g_m r_o. \end{aligned}$$

The output resistance for this configuration is the resistance looking into the drain, which we already know is:

$$R_{\text{out}} = R_{\text{drain}} = R_S + r_o + g_m r_o R_S.$$

The input resistance is the resistance seen looking into the source terminal. Since there is an ideal current source connected above the drain, the effective resistance above the drain is $R_d \to \infty$, so for this configuration,

$$R_{\text{in}} = R_{\text{source}} = \infty.$$

Passive-bias configuration: If the I_D current source is replaced by a resistor R_D, we can consider R_D as a load resistance. Then the amplifier's gain is revised by considering the coupling ratio:

$$A_{vL} = \left(1 + g_m r_o\right)\frac{R_D}{R_D + R_S + r_o + g_m r_o R_S}.$$

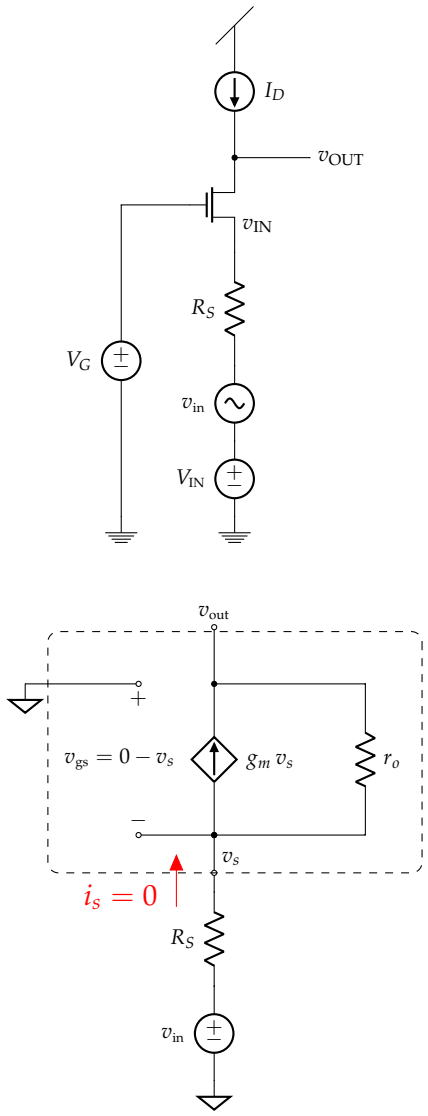


Figure 127: NMOS Common-Gate amplifier configuration with ideal active bias, and its small-signal equivalent circuit model. The signal source is assumed to have a series resistance of R_S.

Capacitive coupling: When the CG amplifier is used to amplify AC signals, we can use a procedure similar to the bypass method that we applied in the CS amplifier with source degeneration. In this configuration we can similarly leverage R_S to provide a more tolerant bias point at DC, while bypassing R_S to mask its effect at higher frequencies.

An ensemble of DC transfer characteristics are shown in Figure 129. The steepest curve corresponds to an R_S near zero. The steepest curve offers the best gain, but the flattest curve offers the most tolerant bias point. By using capacitive coupling, the circuit will "see" the flatter high-R_S curve at DC, but will "see" the steeper low-R_S curve at high frequencies.

Input resistance: In some applications, we are specifically interested in the input resistance coupling for the passive-bias configuration. The CG input resistance is defined as the resistance looking into the source terminal, R_{source}, which depends on the value of R_D together with any load resistance that might be present. This creates a tricky situation: we can either account for R_D via the input coupling OR account for it via the output coupling. **If we model input and output coupling effects at the same time, R_D will be double-counted.**

In our previous output-side analysis, we considered the open-circuit analysis and later inserted R_D as a load. In the input-side analysis, we consider the short-circuit configuration with R_S removed, then insert it as a coupling resistance on the front side. In that case, the solution changes a little:

$$
\begin{aligned}
R_{\text{in}} &= \frac{R_D + r_o}{1 + g_m r_o} \\
R_{\text{out}} &= R_D \parallel r_o \\
A_{vs} &= \frac{R_D}{R_D + r_o} + g_m \left(R_D \parallel r_o\right) \\
A_{vL} &= \left(\frac{R_D}{R_D + r_o} + g_m \left(R_D \parallel r_o\right)\right) \frac{R_{\text{in}}}{R_{\text{in}} + R_S} \\
&= \frac{\left(\frac{R_D}{R_D + r_o} + g_m \frac{R_D r_o}{R_D + r_o}\right)\left(R_D + r_o\right)}{\left(1 + g_m r_o\right)\left(\frac{R_D + r_o}{1 + g_m r_o} + R_S\right)} \\
&= \frac{R_D + g_m R_D r_o}{R_D + r_o + R_S + g_m r_o R_S} \\
&= \left(1 + g_m r_o\right) \frac{R_D}{R_D + r_o + R_S + g_m r_o R_S}
\end{aligned}
$$

The same result we obtained before.

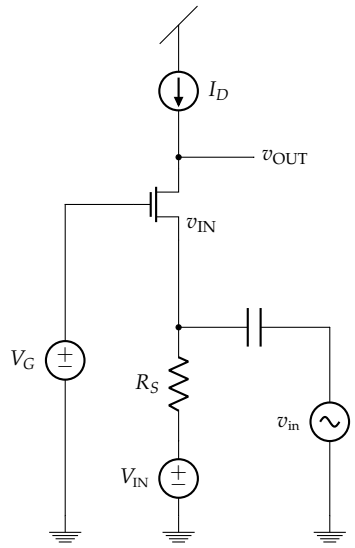


Figure 128: NMOS Common-Gate amplifier with capacitive input coupling to bypass R_S.

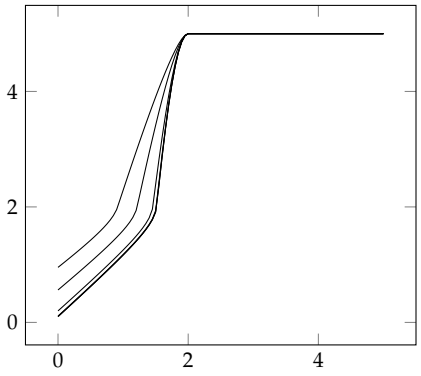


Figure 129: DC transfer characteristic of a CG amplifier with different values of R_S. A higher R_S provides a flatter characteristic, and is therefore more tolerant to bias error.

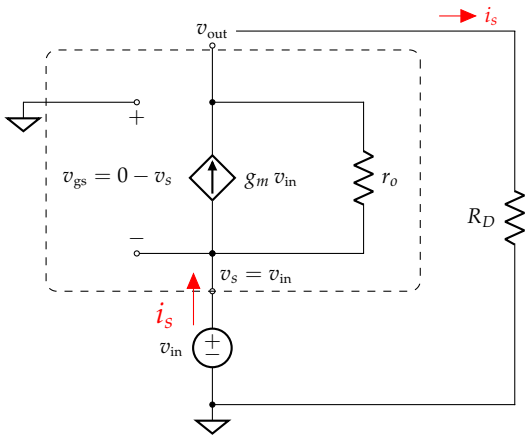


Figure 130: Short-circuit model for input-side coupling analysis, with R_S removed while R_D remains.

Example 24 (Common-gate configurations).

Consider a CG amplifier with passive-bias where $R_D = 50\,\text{k}\Omega$, $R_S = 10\,\text{k}\Omega$, $V_G = 2.5\,\text{V}$ and $V_{\text{IN}} = 1.6\,\text{V}$. The NMOS device parameters are $k = 500\,\mu\text{A/V}^2$, $V_{\text{Th}} = 0.5\,\text{V}$ and $\lambda = 0.05\,\text{V}^{-1}$, and the supply voltage is $V_{\text{DD}} = 5\,\text{V}$. What gain and output resistance will be achieved in the series configuration (**??**) and the bypass configuration (**??**)?

To begin with, we solve the DC operating point, then the small-signal parameters, and the open-circuit characteristics of this amplifier:

$$\begin{aligned} V_{\text{OV}} &= V_G - V_S - V_{\text{Th}} \\ V_S &= V_{\text{IN}} + I_D R_S \\ I_D &= \frac{1}{2} k V_{\text{OV}}^2 = \frac{1}{2} k \left(V_G - V_{\text{Th}} - V_{\text{IN}} - I_D R_S \right) \end{aligned}$$

By defining $x = \sqrt{I_D}$, we can express a quadratic equation:

$$R\sqrt{\frac{k}{2}} x^2 + x - \sqrt{\frac{k}{2}} \left(V_G - V_{\text{Th}} - V_{\text{IN}} \right) = 0$$

then solve using the quadratic formula:

$$I_D = \left(\frac{-1 + \sqrt{1 + 2kR\left(V_G - V_{\text{Th}} - V_{\text{IN}} \right)}}{R\sqrt{2k}} \right)^2 = 15.3\,\mu\text{A}$$

Then the small-signal parameters are:

$$\begin{aligned} g_m &= \sqrt{2kI_D} = 123.6\,\mu\text{A/V} \\ r_o &= (\lambda I_D)^{-1} = 1.3\,\text{M}\Omega \\ 1 + g_m r_o &= 162.6\,\text{V/V} \end{aligned}$$

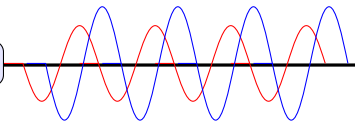
From this point, the two circuits will diverge in the value of R_{drain} at higher frequencies. In the bypass configuration, R_S is masked for AC signals, so $R_{\text{drain}} = r_o$. For the series configuration, however, R_S has a big effect:

$$R_{\text{drain}}\ (\text{series}) = R_S + r_o + g_m r_o R_S = 2.9\,\text{M}\Omega$$

Then the gain for the two configurations is

$$\begin{aligned} A_{vL} &= (1 + g_m r_o) \left(\frac{R_D}{R_D + R_{\text{drain}}} \right) \\ &= 2.73\,\text{V/V}\ (\text{series version}) \\ &= 6.0\,\text{V/V}\ (\text{bypass version}) \end{aligned}$$

EveryCircuit Demonstration 36 (Common-Gate configuration).



This demonstration shows a basic common-gate amplifier with a passive bias resistor $R_D = 50\,\text{k}\Omega$. The NMOS device has the familiar characteristics: $k = 500\,\mu\text{A/V}^2$, $V_{\text{Th}} = 0.5\,\text{V}$ and $\lambda = 0.05\,\text{V}^{-1}$. The supply voltage is $V_{\text{DD}} = 5\,\text{V}$, the NMOS gate is biased at a constant $V_G = 2\,\text{V}$, and the input signal has a DC offset of $V_{\text{IN}} = 1.1\,\text{V}$, and the input AC small signal v_{in} has an amplitude of 100 mV. The input signal has zero series resistance.

To determine the DC operating point and small-signal characteristics, we can start by directly calculating V_{OV} since there is no resistor below the source terminal:

$$\begin{aligned} V_{\text{OV}} &= V_G - V_S - V_{\text{Th}} \\ &= V_G - V_{\text{IN}} - V_{\text{Th}} && = 0.4\,\text{V} \end{aligned}$$

Then the DC bias current, output offset, and small-signal parameters are:

$$\begin{aligned} I_D &= \frac{1}{2} k V_{\text{OV}}^2 = 40\,\mu\text{A} \\ V_D &= V_{\text{DD}} - I_D R_D = 3\,\text{V} \\ V_{\text{DS}} &= V_D - V_{\text{IN}} = 1.9\,\text{V} > V_{\text{OV}} \checkmark \\ g_m &= k\,V_{\text{OV}} = 200\,\mu\text{A/V} \\ r_o &= (\lambda I_D)^{-1} = 500\,\text{k}\Omega \end{aligned}$$

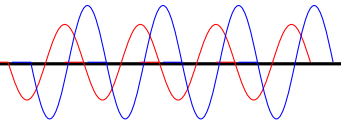
Finally the gain and output resistance are

$$\begin{aligned} A_{vo} &= (1 + g_m r_o)\left(\frac{R_D}{R_D + r_o}\right) \\ &= (101\,\text{V/V})\left(\frac{50}{550}\right) = 9.18\,\text{V/V} \\ R_{\text{out}} &= R_D \parallel r_o = 45.45\,\text{k}\Omega \end{aligned}$$

Now measure the amplitude of the output signal in EveryCircuit, and verify that the gain is a little over 9 V/V, as predicted by our analysis.

As an exercise, try removing R_D and replace it with an ideal current source (pointing down) supplying 40 µA. Predict the effect this will have on the circuit, and verify your prediction in the EveryCircuit simulation (note: for this exercise you will need to reduce the amplitude of v_{in} to 1 mV, and carefully increase the DC offset to $V_{\text{IN}} = 1.11\,\text{V}$).

EveryCircuit Demonstration 38 (Common-Gate bypass configuration).



This circuit implements the passive-bias CG configuration described in example 24. The bypass version is shown. Run the transient simulation and verify that the results align with the predictions from example 24. As an exercise, remove the bypass capacitor and reposition the AC input in series between the DC offset and R_S. Repeat the simulation, and verify that the gain decreases to a value close to what was predicted.

Source Follower configuration

If the input signal is applied to the gate while the output is sampled from the source terminal, the circuit is called a common-drain configuration, more popularly known as a *source follower* since the source terminal "follows" the gate signal with a small-signal gain close to one.

For the ideal active-biased open-circuit configuration, the small-signal model is quite simple. We see immediately that $v_{\text{out}} = g_m r_o (v_{\text{in}} - v_{\text{out}})$, so the open-circuit gain is simply

$$A_{v0} = \frac{g_m r_o}{1 + g_m r_o}.$$

The output resistance is a little more subtle. Since the resistance is looking into the source terminal, we should have $R_{\text{out}} = R_{\text{source}}$ with $R_D = 0$. Then

$$R_{\text{out}} = \frac{r_o}{1 + g_m r_o} \approx \frac{1}{g_m}.$$

The $1/g_m$ approximation is accurate when $g_m r_o \gg 1$, which is usually true.

Passive-loaded configuration: If a resistor R_S is used instead of the ideal current source, we can treat it as a load applied to the ideal open-circuit configuration. Then the loaded gain is

$$\begin{aligned} A_{vL} &= \frac{g_m r_o}{1 + g_m r_o} \frac{g_m R_S}{1 + g_m R_S} \\ &\approx \frac{g_m R_S}{1 + g_m R_S} \end{aligned}$$

This approximation applies when $g_m r_o \gg g_m R_S$, which is often the case when using a passive bias. Finally the output resistance is

$$R_{\text{out}} = \frac{1}{g_m} \parallel R_S.$$

This tends to be much lower than the output resistance of the CG and CS configurations. For that reason, the SF configuration can be useful as an output buffer to drive small-resistance loads without suffering signal attenuation due to output resistance coupling.

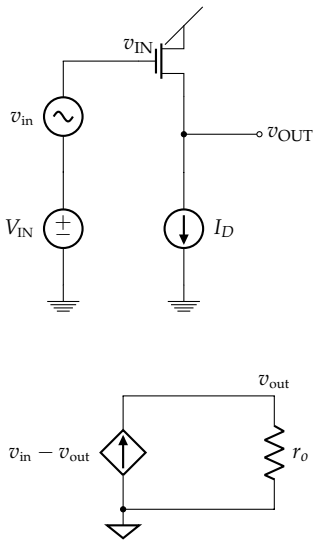


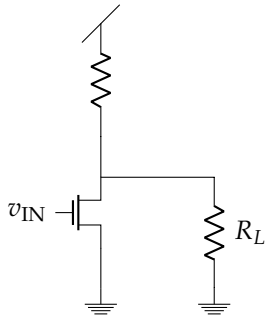
Figure 131: Source follower configuration with ideal active bias, and its small-signal equivalent model.

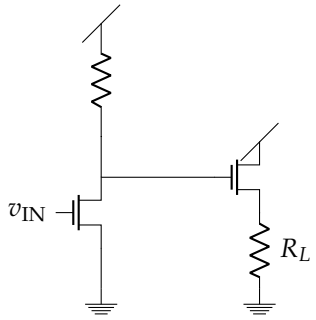
Example 25 (Source Follower output resistance).

Suppose a CS configuration has an open-circuit gain of $A_{vo} = 20\,\mathrm{V/V}$ an output resistance $R_{\text{out}} = 500\,\mathrm{k\Omega}$, and needs to drive a load $R_L = 10\,\mathrm{k\Omega}$. If the CS amplifier is connected directly to the load, the gain will be attenuated so that $A_{vL} = 20 \times 10/(10 + 500) = 0.392\,\mathrm{V/V}$. Now suppose a SF configuration is inserted in between the CS output and the load resistor, and the SF circuit has $g_m = 200\,\mu\mathrm{A/V}$, $r_o = 500\,\mathrm{k\Omega}$, and the load itself acts as a passive bias resistance of $R_S = 10\,\mathrm{k\Omega}$. Then the overall gain of this two-stage circuit will be

$$A_{vL} = 20\,\mathrm{V/V}\left(\frac{g_m r_o}{1 + g_m r_o}\right)\left(\frac{g_m R_L}{1 + g_m R_L}\right) = 19.99\,\mathrm{V/V},$$

so there is almost no attenuation at all.





Biasing MOSFET amplifiers

In the previous examples, we considered the DC solution for passive resistor-biased amplifier configurations. Passive bias designs are convenient in that they can be fully analyzed, and can be made tolerant to bias errors and parametric variation. But there are several drawbacks to passive-bias designs:

- Bias resistors load the amplifier and significantly lower the gain.
- Due to the DC voltage-drop across each bias resistor, the dynamic range is limited; for example, with $V_{DD} - 5\,\text{V}$, a passive-biased CS amplifier may deliver an output amplitude no greater than 1 V.
- For integrated circuit applications, resistors are physically large and expensive to fabricate on-chip.

To address these limitations, there are two major alternatives: **current-mode biasing** and **feedback biasing**. There are other, more advanced, bias solutions, but for now we will focus our attention on these two methods.

Current-mode biasing

One of the most important bias strategies is based on the **current mirror** configuration shown in Figure 132. The current mirror consists of two devices, both biased in saturation, connected with the same gate and source voltages. Since the saturation current depends only on v_{GS}, and both devices have the same v_{GS}, they should both have the same current. On the input side, a current is forced into the drain terminal of a diode-connected device. The diode connection regulates v_{GS} to support the forced current I_D. On the output side, the device can be used as a current source that delivers $i_{out} = I_D$ to a load connected at its drain terminal.

Current mirrors can be used to generate bias currents, and to source multiple copies of a reference current. This can be quite useful for amplifier biasing. In the structure shown in Figure 133, the reference current (i.e. the input) is initially determined by a resistor R. The current in this configuration was previously found to be

$$I_D = \frac{\left(-1 + \sqrt{1 + 2kR\left(V_{DD} - V_{Th}\right)}\right)^2}{2kR^2}.$$

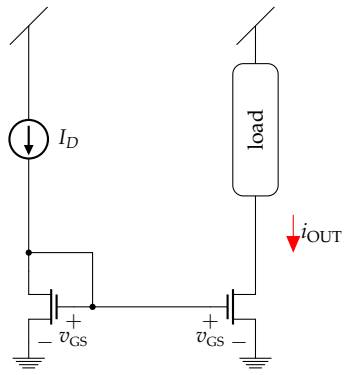


Figure 132: MOSFET current mirror. The input current (on the diode connected side) is copied at the output branch.

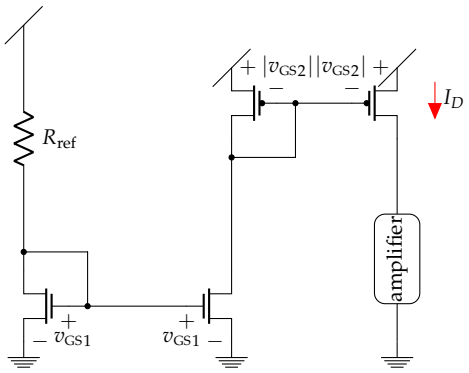
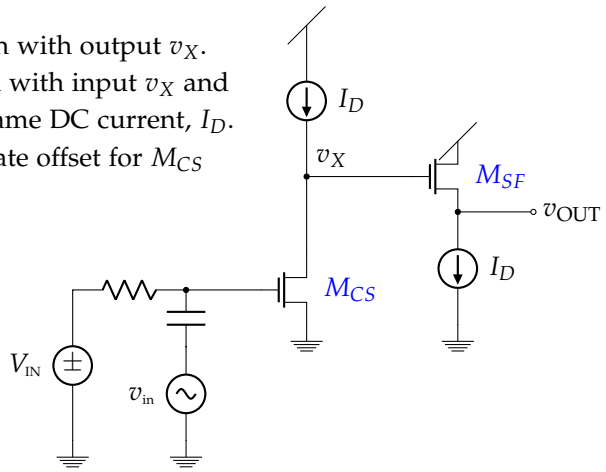


Figure 133: Example showing two current mirrors. In the NMOS mirror, the reference current is setup by the interaction between R and the diode-connected NMOS device. The PMOS mirror then generates a new copy of the reference current, which is used to bias an amplifier.

Example 26 (Current-mirror active bias).

Consider the CS-SF configuration from the previous example. By using current mirrors, we can convert this to an active-bias configuration, replacing the resistors with MOSFETs as shown below.

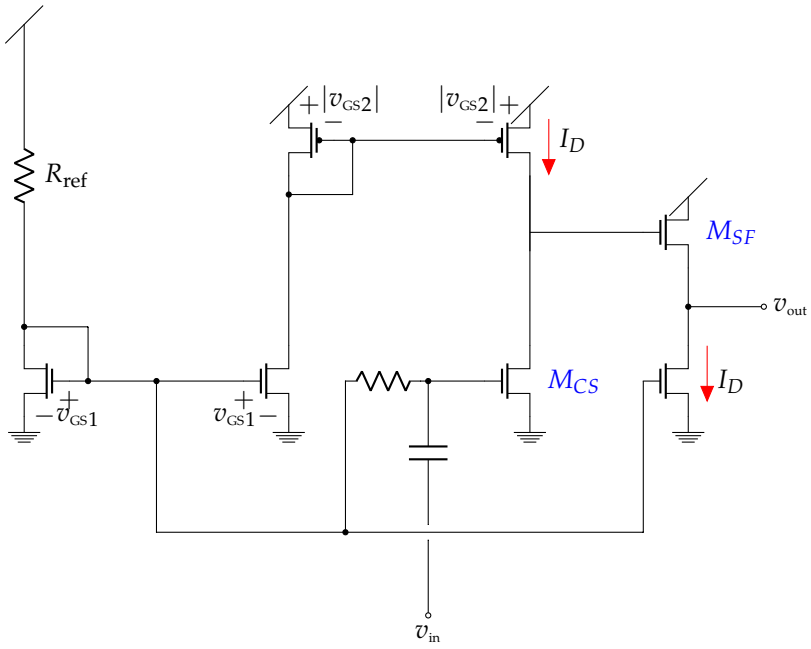
Device M_{CS} is in a common-source configuration with output v_X. Device M_{SF} is in a source-follower configuration with input v_X and output v_{OUT}. Both devices are biased with the same DC current, I_D. Capacitive coupling is used to separate the DC gate offset for M_{CS}



Design problems:

- Generate the correct offset voltage at V_{IN}.
- Implement the I_D current sources.

In the expanded circuit below, current mirroring is used to generate both the V_{IN} offset voltage and the current sources for I_D. There are two current mirrors, an NMOS mirror producing three copies of I_D via v_{GS1}, and a PMOS mirror producing one copy of I_D via $|v_{GS2}|$.



To keep M_{CS} in saturation, it's DC current should be equal to I_D. If all the NMOS devices are *matched*, and if they all have the same V_{GS}, then they should all have the same current. Hence by setting $V_{IN} = v_{GS1}$, the current mirror provides both the gate offset voltage and the current sources.

Symmetry in current mirrors

When using a current-mirror bias network, the output signal's DC offset is not obvious. We usually want an amplifier's DC offset to be balanced in the center of its operating range, but how can we control this? The answer lies with symmetry.

In the two-mirror bias structure shown in Figure 134, the NMOS mirror is used to set the gate offset voltage, and the PMOS mirror is used to source the active bias current. For this setup to work, all devices must be perfectly *matched*, i.e. they must have the same physical parameters (k, V_{Th}, λ, etc) and the same geometry (W and L). They must furthermore all operate in saturation, so that the sensitivity to their drain voltages is minimized. Then they should all have the same device current, I_D.

Using the square-law device equation, we can solve for all the voltages in this circuit except for one: v_X. To obtain a solution for v_X, we note that the PMOS devices have the exact same gate and source voltages, and the same device current. In that case, we may make an argument from physical symmetry: **if two devices are known to have exactly the same electrical state in all variables except one, then they must also be matched in the remaining unknown variable.** In other words, $v_X = v_Y = V_{DD} - |v_{GS2}|$. In the next example, we find that this isn't always the best bias point.

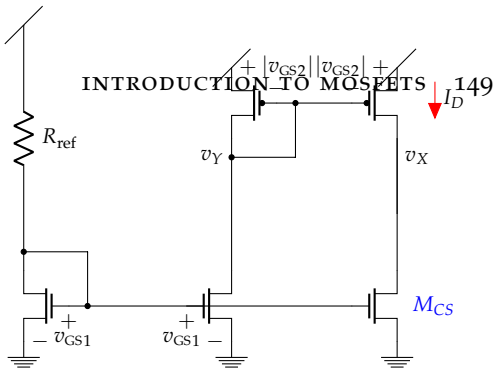


Figure 134: Two current mirrors used to bias a common-source amplifier. By symmetry, we can infer that $v_X = v_Y$.

Example 27 (Output offset with current-mirror bias).

Suppose the circuit of Figure 134 is constructed with the following parameters: $k_n = 5\,\text{mA/V}^2$, $k_p = 3\,\text{mA/V}^2$, $V_{ThN} = V_{ThP} = 0.5\,\text{V}$, $\lambda_n = 0.01\,\text{V}^{-1}$, $\lambda_p = 0.05\,\text{V}^{-1}$, $R_{ref} = 50\,\text{k}\Omega$ and $V_{DD} = 5\,\text{V}$. What is the circuit's complete DC operating point?

Using the previous analysis of the diode-connected MOSFET, we calculate I_D and then v_{GS1} and v_{GS2}

$$I_D = \frac{\left(-1 + \sqrt{1 + 2kR\left(V_{DD} - V_{Th}\right)}\right)^2}{2kR^2} = 86.3\,\mu\text{A}$$

as

$$V_{OV1} = \sqrt{\frac{2I_D}{k_n}} = 0.186\,\text{V} \qquad V_{OV2} = \sqrt{\frac{2I_D}{k_p}} = 0.24\,\text{V}$$

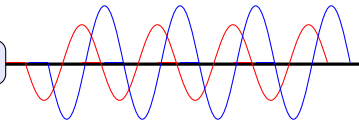
$$V_{GS1} = V_{OV1} + V_{ThN} = 0.686\,\text{V} \qquad |V_{GS2}| = V_{OV2} + V_{ThP} = 0.74\,\text{V}$$

Then the value of v_Y and v_X is

$$v_y = v_X = V_{DD} - |v_{GS2}| = 4.26\,\text{V}$$

In the result from Example 27, notice that **the output offset is very close to the maximum output voltage. It is near the top of its range.** That means the positive leg of an output signal will be clipped. To resolve this problem, we need to break the symmetry by a small amount. One option is to slightly increase the k of device M_{CS} (by increasing its width), so that its device current is slightly greater than I_D. That will pull down the output offset. A second option is to slightly decrease the gate bias voltage at M_{CS}. **Both of these methods are risky,** since it can be challenging to calculate the exact variation required. Real MOSFETs may deviate slightly from our model equations, and manufacturing variations can result in physical parameters that are slightly different from the ones on the data sheet.

EveryCircuit Demonstration 40 (Two stage amp with current mirror bias).



This demonstration implements the current-mirror bias network from Example 26. In order to correct for the output bias problem discussed in Example 27, the width of device M_{CS} is slightly increased. Based on the parameters from Example 27, we found that the bias current is $I_D = 86.3\,\mu\text{A}$. Continuing this analysis, we find the small-signal parameters are

$$g_m = \sqrt{2k_n I_D} = 929\,\mu\text{A/V}$$
$$r_{o,n} = (\lambda_n I_D)^{-1} = 231.7\,\text{k}\Omega$$
$$r_{o,p} = (\lambda_p I_D)^{-1} = 115.9\,\text{k}\Omega$$

Then the expected gain is $A_{vL} = -g_m(r_{o,n} \parallel r_{o,p}) = -71.77\,\text{V/V}$.

In the demonstration, since M_{CS} is slightly wider, its k increases to $5.85\,\mu\text{A/V}^2$. Using an ammeter in the simulation, we can see that this increases the bias current to about $110\,\mu\text{A}$, hence g_m becomes $1.13\,\text{mA/volt}$, $r_{o,n} = 181.8\,\text{k}\Omega$ and $r_{o,p} = 90.9\,\text{k}\Omega$. Then the loaded gain should be $-68.5\,\text{V/V}$. In the simulation, the gain is observed to be $-83.8\,\text{V/V}$, somewhat higher than the prediction.

As an exercise, try setting the width of M_{CS} to 10, so that it matches the widths of all other devices in the circuit. You should see that **the output waveform saturates** due to the output offset being near the upper edge of saturation for the PMOS device.

Feedback biasing

In order to achieve a more reliable bias solution that is highly tolerant to both manufacturing variation and model inaccuracy, we can exploit the power of negative feedback. Since our goal is to achieve an output bias near the center of its dynamic range, we can directly enforce this condition by using an **error amplifier loop** like the one shown in Figure 135. The idea is that if V_X deviates from the desired value, $V_X^\star$, then the amplifier responds by pushing V_X strongly in the opposite direction.

At this point you may ask, "if I have an op amp, why don't I just use it as the amplifier instead of using it to bias a MOSFET circuit?" In practice, we don't need to use a full fledged op amp for this bias configuration. A simpler MOSFET-based differential amplifier is adequate. There are several configurations that can be used for error amplification. One example is the circuit shown in Figure 136, in which a CS configuration is superimposed onto a SF configuration. When two amplifier stages are superimposed in this manner, it is referred to as a **folded configuration**.

The new circuit in Figure 136 now has three devices that will act as CS amplifiers. The primary amplifier is NMOS M_{CS1}, and the error amplifier bias network has PMOS amplifiers M_{CS2} and M_{CS3}. In the small-signal domain, M_{CS2} has an open-circuit gain of $-g_m r_o$, but is loaded by the $1/g_m$ source resistance of M_{SF2}. So the loaded gain of M_{CS2} is

$$A_{CS2} = -g_m r_o \left(\frac{1/g_m}{1/g_m + r_o} \right) \approx -1,$$

where the approximation is due to the fact that $r_o \gg 1/g_m$, so the $1/g_m$ term is removed from the denominator, allowing both g_m and r_o to be canceled. Next, since M_{SF2} is a source follower, its gain is approximately one. So the small-signal voltage arriving at the gate of M_{CS3} is $v_Y \approx (v^+) - (v^-)$. Finally the error signal v_y is amplified by M_{CS3} with a gain of $-g_m r_o$, which supplies the amplification in this feedback loop.

The purpose of the error amplifier is to make small adjustments in the current of M_{CS3} in order to precisely control the offset voltage at v_X. The device current in M_{CS3} should still be very close to I_D, and its gate voltage should therefore be very close to

$$v_Z \approx V_{DD} - V_{ThP} - \sqrt{\frac{2I_D}{k_p}},$$

where I_D is the DC bias current in M_{CS1} and M_{CS3}. Furthermore, the DC currents in M_{CS2} and M_{SF2} should be equal to

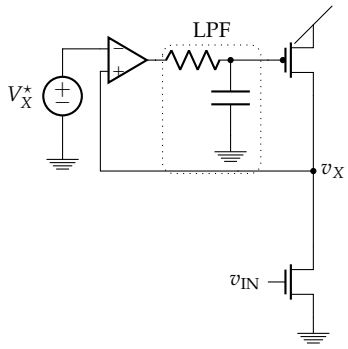


Figure 135: An error amplifier loop is used to regulate the DC offset at V_X so that it stays close to the desired value $V_X^\star$. This configuration adapts to changes in the gate offset at V_{IN}. Note that the feedback signal is connected to the op amp's non-inverting input. This is because the PMOS device acts as an inverting amplifier, so there is a net negative sign around the feedback loop.

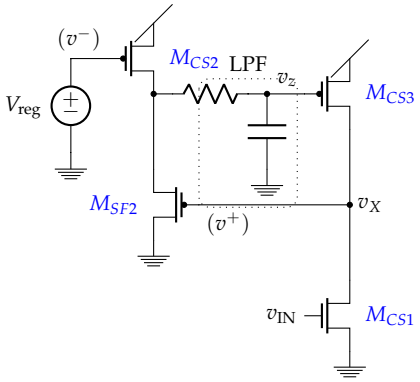


Figure 136: A differential error amplifier made by **folding together** a PMOS common-source amplifier with a PMOS source-follower. In the small-signal model, we see that the SF device loads the CS configuration, so that the overall gain is close to one.

each other, so

$$\frac{1}{2}k_p\left(V_{DD}-V_{reg}-V_{ThP}\right)^2=\frac{1}{2}k_p\left(v_Z-v_X-V_{ThP}\right)^2$$
$$\Rightarrow \left(V_{DD}-V_{reg}-V_{ThP}\right)=\left(V_{DD}-V_{ThP}-\sqrt{\frac{2I_D}{k_p}}-v_X-V_{ThP}\right).$$

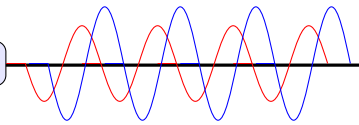
So, in order to achieve $v_X = V_X^\star$, we should set the control voltage at

$$V_{reg}=V_X^\star+V_{ThP}+\sqrt{\frac{2I_D}{k_p}}.$$

If $k_p \gg I_D$, then

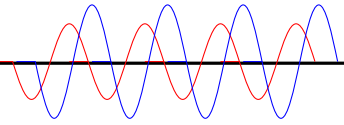
$$V_{reg}\approx V_X^\star+V_{ThP}.$$

The feedback bias methods discussed here are introductory. A variety of more sophisticated bias techniques can also be used, but are beyond the scope of this chapter.



EveryCircuit Demonstration 42 (Two-stage amplifier with ideal feedback bias).

This example modifies the design from Example 27 to use feedback bias with an ideal op amp as the error amplifier, like the solution shown in Figure 135. In this configuration, all devices are matched, so there is no need to manipulate the width of M_{CS}.



EveryCircuit Demonstration 44 (Two-stage amplifier with PMOS feedback bias).

This example modifies the design from Example 27 to use feedback bias with an error amplifier like the one from Figure 136. In this design, the control voltage is $V_{reg} = 3.0\,\text{V}$, corresponding to $V_X^\star + V_{ThP}$. All NMOS devices are matched, and the error amplifier feedback is achieved without using an op amp.

Frequency response of CMOS amplifiers

In this section we will introduce the basic concepts of CMOS amplifier frequency response and bandwidth. All circuits have a maximum operating frequency, beyond which they exhibit rapid signal attenuation. Frequency limitations arise from the **parasitic capacitances** that exist on every node in a physical circuit. Capacitance arises from the wire connections at a node, from the substrate and insulation materials of a wire, a printed circuit board or chip, and from the internal junction physics of the MOSFET itself. In this section, we will focus on the analysis of capacitive effects and not concern ourselves with calculating specific capacitance values around MOSFET devices (we will assume the capacitance values are known or given).

In an amplifier circuit, it is usually sufficient to assume that a "lump" capacitance is connected to each node that summarizes all the neighboring parasitic effects. We therefore insert lump capacitors connected from the input and output nodes to ground, called C_{in} and C_{out}, respectively. By representing these capacitors in the Laplace domain, we can treat them as part of the input and output impedances, and their effect is captured by the resistor-divider coupling ratios at the input and output nodes. Then the amplifier's frequency response is

$$\begin{aligned} A_v(s) &= A_{vo}\left(\frac{R_{\text{in}} \parallel \frac{1}{sC_{\text{in}}}}{R_{\text{sig}} + R_{\text{in}} \parallel \frac{1}{sC_{\text{in}}}}\right)\left(\frac{R_L \parallel \frac{1}{sC_{\text{out}}}}{R_{\text{out}} + R_L \parallel \frac{1}{sC_{\text{out}}}}\right) \\ &= A_{vo}\left(\frac{R_{\text{in}}}{R_{\text{sig}} + R_{\text{in}}}\right)\left(\frac{R_L}{R_L + R_{\text{out}}}\right) \\ &\times \left(\frac{1}{1 + s(R_{\text{sig}} \parallel R_{\text{in}})C_{\text{in}}}\right)\left(\frac{1}{1 + s(R_{\text{out}} \parallel R_L)C_{\text{out}}}\right) \end{aligned}$$

This implies there are two poles:

$$\begin{aligned} \omega_{p1} &= \left[C_{\text{in}}\left(R_{\text{in}} \parallel R_{\text{sig}}\right)\right]^{-1} \\ \omega_{p2} &= \left[C_{\text{out}}\left(R_{\text{out}} \parallel R_L\right)\right]^{-1} \end{aligned}$$

It will often be the case that one pole is much larger than the other. In that case may **consider the smaller pole to be dominant**, and treat the circuit as a one-pole system governed by the dominant pole.

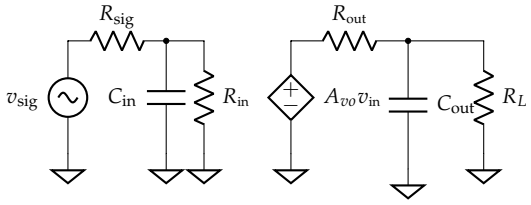


Figure 137: High-frequency amplifier model showing input and output capacitances.

Frequency response of the common-source configuration

The common source configuration benefits from an infinite input resistance, so the input pole is determined solely by R_{sig} and C_{in}. The output pole is quite sensitive to the amplifier's output and load resistances.

Input-dominant pole. If R_{sig} is significant, then the input node may dominate the frequency response. In that case, the circuit's cutoff frequency is approximately $\omega_{in} \approx (R_{sig}C_{in})^{-1}$.

Miller effect. If the input pole is dominant, then it may be necessary to account for any feedback capacitance, C_{fb}, that may bridge between the input and output terminals. In practice feedback capacitances are usually much smaller than the lump capacitances at the input and output terminals. Due to the negative feedback path in the common-source configuration, the effect of C_{fb} is amplified, and may alter the frequency response.

The Miller effect appears in the small-signal equivalent impedance seen looking into the input terminal. Using the simplified circuit model in Figure 140, we find that

$$\begin{aligned} v_{out} &= -A_v v_{in} \\ i_{in} &= (v_{in} - v_{out})\, sC_{fb} = (v_{in} + A_v v_{in})\, sC_{fb} \\ \Rightarrow Z_{in} &= \frac{v_{in}}{i_{in}} = [(1 + A_v)\, sC_{fb}]^{-1} \end{aligned}$$

This result shows that the *effective* capacitance is $(1 + A_v)\, C_{fb}$, i.e. the feedback capacitor is "amplified" by the gain of the inverting configuration. A similar analysis shows that there is no such effect on the output side.

Thanks to the Miller effect, in a high-gain amplifier a small feedback capacitance could prove to dominate the bandwidth. Sometimes this is a nuisance, but it can also be desirable for special applications.

Output dominant pole. If $R_{out} \parallel R_L$ is very large, and R_{sig} is comparatively small, then the output pole will tend to dominate the frequency response. In that case, it is sometimes desirable to load the amplifier with a smaller R_L, in order to increase the bandwidth. Since this also decreases the amplifier's gain, there will be a strong tradeoff between gain and bandwidth.

Gain/Bandwidth Tradeoff. For the output-dominant case, the tradeoff between gain and bandwidth manifests in multiple ways. When using a small load resistance, $R_L \ll R_{out}$, the

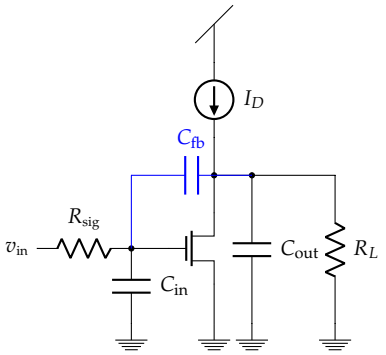


Figure 138: Common-source configuration showing input and output capacitances. In some cases there may be a feedback capacitor C_{fb}, which can affect the bandwidth if the input pole is dominant.

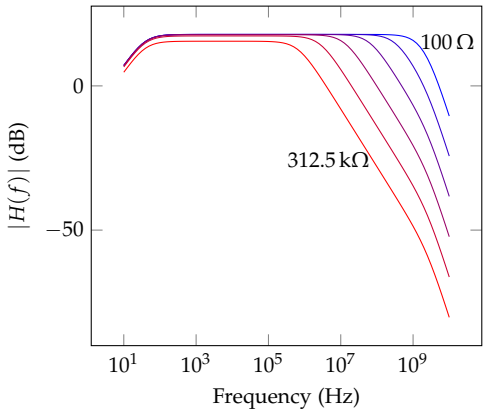


Figure 139: Simulated transfer functions for a CS configuration with input-dominant pole for $C_{in} = 1\,\text{pF}$. Each curve represents a different R_{sig}, starting from 100 Ω and increasing by 5× up to 312.5 kΩ. The cutoff frequency varies from the order of GHz down to a few hundred kHz.

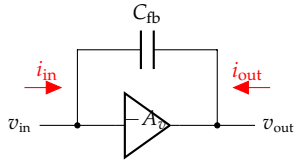


Figure 140: Simplified circuit model for analyzing the Miller effect.

tradeoff is clear:

$$\begin{aligned} A_{vL} &= A_{vo}\left(\frac{R_L}{R_L + R_{\text{out}}}\right) \\ &\approx A_{vo}\left(\frac{R_L}{R_{\text{out}}}\right) \\ f_c &= [2\pi(R_{\text{out}} \parallel R_L)C_{\text{out}}]^{-1} \\ &\approx [2\pi R_L C_{\text{out}}]^{-1} \end{aligned}$$

We see that there is a one-to-one exchange between gain and bandwidth. In other words, the **gain-bandwidth product** is constant: GBW $= A_{vo}/(2\pi R_{\text{out}}C_{\text{out}})$.

For a larger load, or when no resistive load is present, we also see the tradeoff as a property of the bias current. The amplifier's open-circuit gain is

$$\begin{aligned} A_{vo} &= -g_m r_o \\ &= -\left(\sqrt{2kI_D}\right)\left(\frac{1}{\lambda I_D}\right) \\ &= -\sqrt{\frac{2k}{\lambda^2 I_D}} \end{aligned}$$

Meanwhile the cutoff frequency is

$$\begin{aligned} f_c &= (2\pi R_{\text{out}}C_{\text{out}})^{-1} \\ &= \frac{\lambda I_D}{2\pi C_{\text{out}}} \end{aligned}$$

As a result, a **higher I_D means higher bandwidth but lower gain.** To put it another way, a high-bandwidth amplifier consumes more power and has less gain than a low-bandwidth amplifier.

In the case of a passive-biased CS amplifier, we again see the same tradeoff. Suppose the amplifier has a source degeneration resistor such that $R_S = R_D$ with a source bypass capacitor, and the gate offset is at $V_{\text{DD}}/2$. Then, using the quadratic formula, we find that the DC solution, open-circuit gain, and bandwidth are

$$\begin{aligned} I_D &= \left(\frac{-1 + \sqrt{1 + 2kR\,(V_{\text{DD}}/2 - V_{\text{Th}})}}{R\sqrt{2k}}\right)^2 \\ &\approx \frac{V_{\text{DD}}/2 - V_{\text{Th}}}{R} \\ A_{vo} &\approx -g_m R = -\sqrt{2kI_D}R \\ &\approx -\sqrt{2kR\,(V_{\text{DD}}/2 - V_{\text{Th}})} \\ f_c &= (2\pi R C_{\text{out}})^{-1} \end{aligned}$$

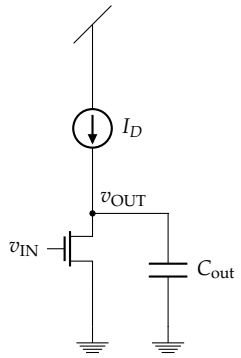


Figure 141: Active-bias CS configuration, when the output pole is dominant, shows a strong gain/BW tradeoff: $A_V \propto \sqrt{I_D}$, whereas $f_c \propto I_D^{-1}$.

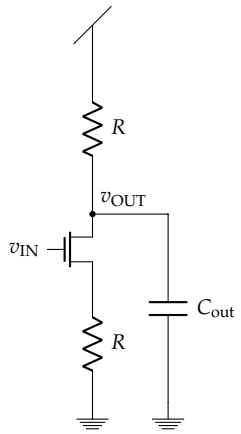


Figure 142: Passive-bias CS configuration, when the output pole is dominant, shows a similar gain/BW tradeoff: $A_V \propto \sqrt{R}$, whereas $f_c \propto R^{-1}$.

So once again we see the same tradeoff between gain and bandwidth. This scenario was simulated using Listing 10 with results shown in Figure 143. The observed gain and -3 dB bandwidth are close to what is predicted by our analysis.

The table below summarizes the results. From the data, we see that the prediction is more accurate for larger R. The netlists used to simulate the input-dominant frequency response and the gain-bandwidth tradeoff are shown after the data table.

	Predicted			Simulated		
R (kΩ)	I_D (µA)	A_v (dB)	f_c (MHz)	I_D (µA)	A_v (dB)	f_c (MHz)
5	400	7	31.8	223	8	32
10	200	10	15.9	131	11.6	15.7
25	80	14	6.4	61	16	12.7
50	40	17	3.18	33	19.5	3.7
100	20	20	1.59	17.5	22.7	1.86

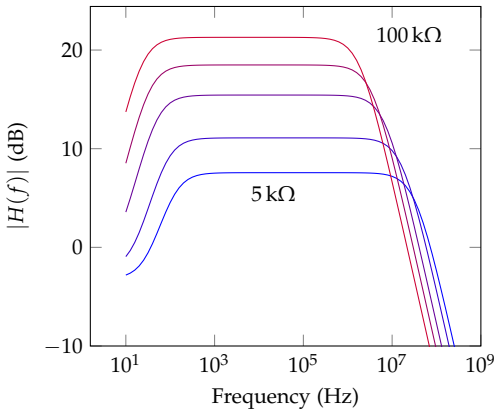


Figure 143: Simulated transfer functions for the CS configuration from Figure 142. The NMOS device has $k = 500\,\mu A/V^2$, $V_{Th} = 0.5\,V$, $V_{DD} = 5\,V$ and $V_{IN} = 2.5\,V$. R is varied from $5\,k\Omega$ up to $100\,k\Omega$.

Netlist 9: AC simulation of CS configuration

```
* Common-Source amplifier with source degeneration
.model ntype NMOS(KP=100e-6,VTo=0.5,LAMBDA=0.05)

VDD  ndd  0 DC 5V
VIN  ndc  0 DC 2.5V
vsig nsig 0 DC=0 AC=1 SIN(0 0.01 1k)

Rsig nsig  nin   10
RDC  ng    ndc   1Meg
CC1  nin   ng    10uF
Cin  ng    0     1pF

RL   nout 0    1Meg
CC2  nd   nout 10uF
Cout nd   0    1fF

M1 nd  ng ns 0 ntype W=1u L=200n
RD ndd nd 50k
RS ns  0  50k
CS ns  0  1uF
.control
* INPUT DOMINANT CASES:
* Foreach loop to scan through RSig values:
foreach RSigval 100 500 2500 12500 62500 312500
alter Rsig = $RSigval
AC dec 10 10 10G
end

* OUTPUT DOMINANT CASES:
```

```
* Foreach loop to scan through Cout values:
alter Rsig=10
foreach Coutval 10f 50f 250f 1.25p 6.25p 31.25p 10p
alter Cout = $Coutval
AC dec 10 10 10G
end

plot db(ac1.nd) db(ac2.nd) db(ac3.nd) db(ac4.nd) db(ac5.nd) db(ac6.nd)
plot db(ac7.nd) db(ac8.nd) db(ac9.nd) db(ac10.nd) db(ac11.nd) db(ac12.nd)
.endc
.end
```

Netlist 10: Gain/BW tradeoff in CS configuration

```
* Common-Source amplifier gain/bw tradeoff
.model ntype NMOS(KP=100e-6,VTo=0.5,LAMBDA=0.05)

VDD  ndd  0 DC 5V
VIN  ndc  0 DC 2.5V
vsig nsig 0 DC=0 AC=1 SIN(0 0.01 1k)

Rsig nsig  nin   10
RDC  ng    ndc   1Meg
CC1  nin   ng    10uF
Cin  ng    0     1pF

RL   nout 0    1Meg
CC2  nd   nout 10uF
Cout nd   0    1pF

M1 nd  ng ns 0 ntype W=1u L=200n
RD ndd nd 50k
RS ns  0  50k
CS ns  0  1uF
.control

* Sweep R values to see Gain/BW tradeoff:
foreach Rval 5k 10k 25k 50k 100k
alter RD=$Rval
alter RS=$Rval
DC VIN 0 5 0.1
AC dec 10 10 10G
set gm=@m1[gm]
set rv=$Rval
let av=$gm*$rv
let avp=sqrt(2*5e-4*$rv)
let avdb=20*log10(av)
let fc=1.0/(2*3.1415*$rv*1e-12)
echo $rv
print @m1[gm] @m1[id] av avp avdb fc
end
```

```
plot db(ac1.nd) db(ac2.nd) db(ac3.nd) db(ac4.nd) db(ac5.nd)
wrdata cs_gbw db(ac1.nd) db(ac2.nd) db(ac3.nd) db(ac4.nd) db(ac5.nd)
.endc
.end
```

Introduction to BJTs

The *bipolar junction transistor* was the first practical transistor device for mass production, and defined the semiconductor industry from the 1950s into the 1980s. Today, BJTs are not as widespread as MOSFETs, but are still very important for niche applications. Some areas where BJTs excel include high-voltage applications, and radio-frequency power amplifiers, where BJTs are able to drive antennas and transmission lines with very good linearity and, hence, low distortion. BJTs also have a high transconductance, and it is usually easier to make a good discrete BJT amplifier, whereas MOSFETs may need several devices in order to achieve a good bias configuration, making them more suited for integrated circuit designs.

The chief drawbacks to BJT devices are high power consumption (BJT bias currents must usually be on the order of 1 mA to 100 mA), and comparatively high voltage overhead (unlike MOSFETs, the BJT's overdrive voltage is not adjustable and cannot be reduced for low-voltage applications). In addition, BJT devices cannot be miniaturized to nano-scale dimensions, so they cannot achieve the same performance enhancements or cost improvements that come with MOSFET scaling. Lastly, BJT devices pass current through their base terminals (comparable to the MOSFET's gate), which makes them inefficient for logic circuits, and complicates amplifier analysis and design.

BJTs are built out of PN junctions (diodes), and normally have three operating modes corresponding to the diode states:

- **cuttoff**: both junctions are *not forward* biased, i.e. $v_{BE} < 0.4\,\text{V}$ and $v_{CB} > -0.4$. Note that the junctions do not have to be reverse biased; a very weak forward bias is sufficient to shutoff the junction for most applications. This mode roughly corresponds to the MOSFET's cutoff mode, and is the appropriate OFF mode for switching circuits.
- **active**: emitter-base junction forward biased, $v_{BE} \approx 0.7\,\text{V}$, collector-base junction *not forward biased*, $v_{CB} > -0.4\,\text{V}$.

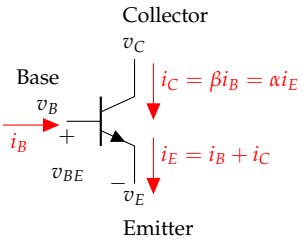


Figure 144: The NPN BJT device, showing its behavior in the **active mode** (when $v_{BE} \approx 0.7\,\text{V}$ and $v_{CE} > 0.3\,\text{V}$). In this mode, the device can be described as a current amplifier with constant gain β.

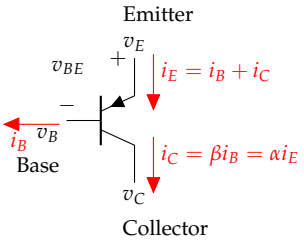


Figure 145: The PNP BJT device, showing its behavior in the **active mode** (when $v_{EB} \approx 0.7\,\text{V}$ and $v_{EC} > 0.3\,\text{V}$). The PNP device's behavior is complementary to the NPN.

Active Mode Summary:

$$
\begin{aligned}
i_B &= I_S \exp\left(\frac{v_{BE}}{nU_T}\right) \\
i_C &= \beta i_B \\
&= \alpha i_E \\
\alpha &= \frac{\beta}{\beta + 1}
\end{aligned}
$$

Note that these conditions imply that $v_{CE} > 0.3\,\text{V}$. In this mode, the base-current is defined by the forward-bias diode equation, $i_B = I_S \exp\left(v_{BE}/nU_T\right)$, and the collector current is $i_C = \beta i_B$, where β is the device's **current gain**, n is the **forward emission coefficient** (usually close to 1.0), U_T is the **thermal voltage** (26 mV at room temperature), and I_S is a **scale current** on the order of pA. This mode roughly corresponds to the MOSFET's *saturation* mode, and is the appropriate DC bias mode for amplifier circuits.

- **saturation**: both junctions are forward biased, $v_{BE} \approx v_{BC} \approx 0.7\,\text{V}$. This mode roughly corresponds to the MOSFET's *triode* mode, and is the appropriate ON mode for use in switching circuits.

In most of our designs, **the BJT will used as an amplifier, and will be operated in its active mode.** In atypical situations, there are two additional modes that may arise in BJT circuits:

- **Reverse active:** when the collector and emitter terminals are swapped, the device can be used with the base-collector junction in forward bias while the base-emitter junction is not forward biased, i.e. $v_{BC} \approx 0.7\,\text{V}$ and $v_{EC} > 0.3\,\text{V}$. In this configuration, the device functions similarly to the forward active mode, but with a much smaller current gain, β. This can easily happen by accident when connecting discrete devices in lab experiments, and highlights a key difference between BJTs and MOSFETs: **whereas MOSFETs are often symmetric devices, BJTs are not. You cannot interchange the collector and emitter terminals.**

- **Avalanche breakdown:** since the BJT is built from diode junctions, reverse-breakdown can occur in one or both of the junctions. The breakdown voltages are usually large enough that they are not encountered in ordinary BJT circuits. In some applications, a BJT can be deliberately forced into avalanche breakdown, which can be useful for high-speed switching of large currents as may be needed in pulse-based instrumentation or radio frequency transmitters. When a BJT is operated in its avalanche breakdown mode, it is referred to as an *avalanche transistor*. Some specialty devices may be built specifically for avalanche operation, but ordinary BJTs can also be operated in the avalanche mode.

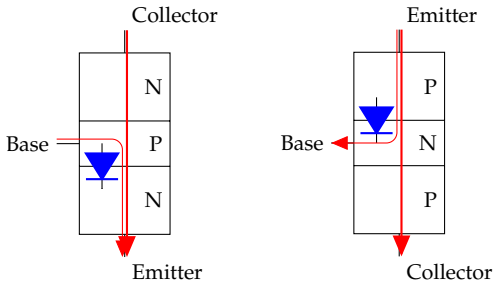


Figure 146: Physical concept of the BJT device. By applying a forward bias across the base-emitter diode, a proportionally larger current is induced between the collector and emitter.

DC passive bias configurations

BJTs are primarily used as amplifiers, so we will consider active-mode bias configurations. Since the device is extremely sensitive to the base-emitter voltage, it is usually necessary to place resistors in series with the base and emitter terminals. These resistors provide "elastic" voltage drops that help maintain an appropriate v_{BE}.

Two common passive bias configurations are shown in Figure 147 and Figure 148. In both cases, we will begin by choosing a desired bias current for I_E (usually on the order of 1 mA), assume that $v_{BE} \approx 0.7\,\text{V}$, and follow a Kirchoff voltage loop across the emitter-base junction. The values of R_C and R_E can be chosen based on gain analysis (which will be addressed later; they will typically be on the order to 10 kΩ), so our bias task will be to calculate the R_B values.

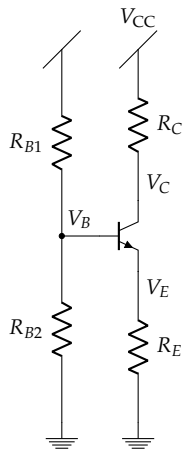


Figure 147: Voltage-divider bias configuration.

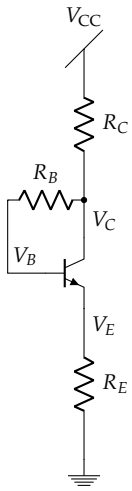


Figure 148: Feedback bias configuration.

Voltage-divider bias: for the configuration in Figure 147, Ohm's Law indicates that $V_E = I_E R_E$, and the Kirchoof voltage loop indicates that $V_B = I_E R_E + v_{BE} = V_E + 0.7\,\text{V}$. The resistors R_{B1} and R_{B2} should be chosen to achieve this voltage. If we assume that I_B is small enough that we can ignore its contribution to the voltage divider, then

$$V_{CC}\frac{R_{B2}}{R_{B1}+R_{B2}} = I_E R_E + 0.7\,\text{V}$$
$$R_{B2} = R_{B1}\frac{V_B}{V_{CC}-V_B}$$

On the collector side, $V_C = V_{CC} - I_C R_C$. The circuit should be biased so that the minimum expected voltage at v_C is at least 0.3 V greater than V_E.

Feedback bias: for the configuration in Figure 148, Ohm's Law again indicates that $V_E = I_E R_E$. In this case, we take the Kirchoof voltage loop from the emitter, to the base, then across the collector and up to V_{CC}. Note that the feedback connection merges the base current together with the collector current, so the total current in R_C is I_E. Then

$$V_{CC} = I_E R_E + v_{BE} + I_B R_B + I_E R_C$$
$$\Rightarrow V_{CC} = I_E R_E + 0.7\,\text{V} + \frac{I_E R_B}{\beta+1} + I_E R_C$$
$$\Rightarrow R_B = \frac{\beta+1}{I_E}\left(V_{CC} - 0.7\,\text{V} - I_E\left(R_E + R_C\right)\right)$$

Example 28 (Voltage divider bias).

A voltage-divider network like the one in Figure 147 has $V_{CC} = 5\,\text{V}$, $R_C = 2\,\text{k}\Omega$, $R_E = 0.5\,\text{k}\Omega$, $R_{B1} = 10\,\text{k}\Omega$, and the device has $\beta = 100\,\text{A/A}$. If the desired emitter current is 1 mA, what is the correct value for R_{B2}?

Based on the given parameters, we can directly calculate $V_E = I_E R_E = 0.5\,\text{V}$, and $V_B = V_E + 0.7\,\text{V} = 1.2\,\text{V}$. Then

$$\begin{aligned} V_{CC}\frac{R_{B2}}{R_{B1}+R_{B2}} &= 1.2\,\text{V} \\ \Rightarrow V_{CC}R_{B2} &= 1.2\,\text{V}\,(R_{B1}+R_{B2}) \\ \Rightarrow R_{B2} &= R_{B1}\frac{1.2\,\text{V}}{V_{CC}-1.2\,\text{V}} \\ &= 3.16\,\text{k}\Omega \end{aligned}$$

Example 29 (Feedback bias).

A voltage-divider network like the one in Figure 148 has $V_{CC} = 5\,\text{V}$, $R_C = 3\,\text{k}\Omega$, $R_E = 0.5\,\text{k}\Omega$, and the device has $\beta = 100\,\text{A/A}$. If the desired emitter current is 1 mA, what is the correct value for R_B?

As in Example 28, we can directly calculate $V_E = I_E R_E = 0.5\,\text{V}$, and $V_B = V_E + 0.7\,\text{V} = 1.2\,\text{V}$. Then

$$\begin{aligned} R_B &= \frac{\beta+1}{I_E}\left(V_{CC} - 0.7\,\text{V} - I_E\left(R_E + R_C\right)\right) \\ &= 80.8\,\text{k}\Omega. \end{aligned}$$

On the collector side, we may estimate the collector voltage as $V_{CC} - I_E R_C = 5\,\text{V} - 1\,\text{mA} \times 3\,\text{k}\Omega = 2\,\text{V}$.

BJT small-signal characteristics

The small-signal characteristics are obtained using the same differential techniques that we employed for MOSFET devices. As with MOSFETs, the most important characteristics are the **transconductance gain** g_m and the **intrinsic resistance** r_o. Unlike MOSFETs, the BJT also has a resistance associated with the base emitter junction, called r_π.

We first obtain the transconductance by applying the differential definition:

$$\begin{aligned} g_m &\triangleq \left.\frac{d\, i_C}{d\, v_{BE}}\right|_{\text{DC}} \\ &= \frac{d}{d\, v_{BE}} \beta I_S \exp\left(\frac{v_{BE}}{nU_T}\right) \\ &= \left.\frac{1}{nU_T} \beta I_S \exp\left(\frac{v_{BE}}{nU_T}\right)\right|_{\text{DC}} \\ &= \frac{I_C}{nU_T}. \end{aligned}$$

Early effect resistance: the BJT's r_o parameter is due to a phenomenon called the Early effect, which is very similar to Channel Length Modulation in MOSFET devices. The Early effect accounts for a slight sensitivity between the collector current and the collector-emitter voltage:

$$i_C = \beta I_S \exp\left(\frac{v_{BE}}{nU_T}\right)\left(1 + \frac{v_{CE}}{V_A}\right)$$

where V_A is called the **Early voltage**, with units of V. The Early effect is algebraically identical to CLM if we recognize that $V_A = \lambda^{-1}$. Then the r_o resistance is

$$\begin{aligned} r_o &\triangleq \left[\left.\frac{d\, i_C}{d\, v_{CE}}\right|_{\text{DC}}\right]^{-1} \\ &= \left[\left.\frac{d}{d\, v_{CE}} \beta I_S \exp\left(\frac{v_{BE}}{nU_T}\right)\left(1 + \frac{v_{CE}}{V_A}\right)\right|_{\text{DC}}\right]^{-1} \\ &= \frac{V_A}{I_C}. \end{aligned}$$

Lastly, the differential base-emitter resistance is defined as

$$r_\pi \triangleq \left[\left.\frac{d\, i_B}{d\, v_{BE}}\right|_{\text{DC}}\right]^{-1}.$$

Recall that $i_B = i_C/\beta$. In that case this derivate is the same as the one that defines g_m, except for the constant β factor. Therefore $r_\pi = \beta/g_m$.

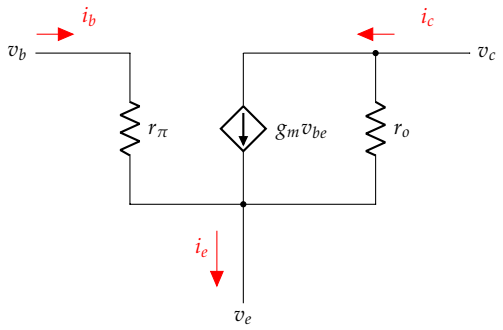


Figure 149: Standard 'TT' model of the BJT device. The base-emitter diode induces a differential resistance named r_π.

Small-Signal Summary:

$$\begin{aligned} g_m &= \frac{I_C}{nU_T} \\ r_o &= \frac{V_A}{I_C} \\ r_\pi &= \frac{\beta}{g_m} \end{aligned}$$

BJT amplifiers with passive bias

The BJT configurations are very similar to their MOSFET counterparts. One crucial difference is that **the BJT allows some current to flow through the base terminal.** This often means that BJT amplifiers have finite input resistance, which can create resistive coupling effects at the input terminals. We'll begin our study with a simplified analysis and then consider the terminal resistances afterward.

In order to simplify our analyses, we'll assume that the resistance seen looking into the BJT's collector terminal is very large, much larger than R_C. We'll furthermore assume that the input resistance looking into the base is very large compared to the equivalent series resistance of the signal source, and that β is very large so $i_e \approx i_c$, and that g_m is very large so that $g_m R_E \gg 1$. With all of these assumptions, the gain is found by the following analysis

$$
\begin{aligned}
v_{\text{out}} &= -g_m v_{be} R_C \\
v_{be} &\approx v_{\text{in}} - i_e R_E = v_{\text{in}} + \frac{v_{\text{out}} R_E}{R_C} \\
\Rightarrow v_{\text{out}} &\approx -g_m R_C \left(v_{\text{in}} + \frac{v_{\text{out}} R_E}{R_C} \right) \\
\Rightarrow \frac{v_{\text{out}}}{v_{\text{in}}} &\approx \frac{-g_m R_C}{1 + g_m R_E} \\
&\approx -\frac{R_C}{R_E}.
\end{aligned}
$$

In the feedback-biased case, we may reach a similar conclusion if the value of R_B is much larger than R_C and also much larger than R_{sig}. In that case, the negative feedback loop created by R_B will have little impact on the small-signal analysis, and we arrive at the same result for the gain.

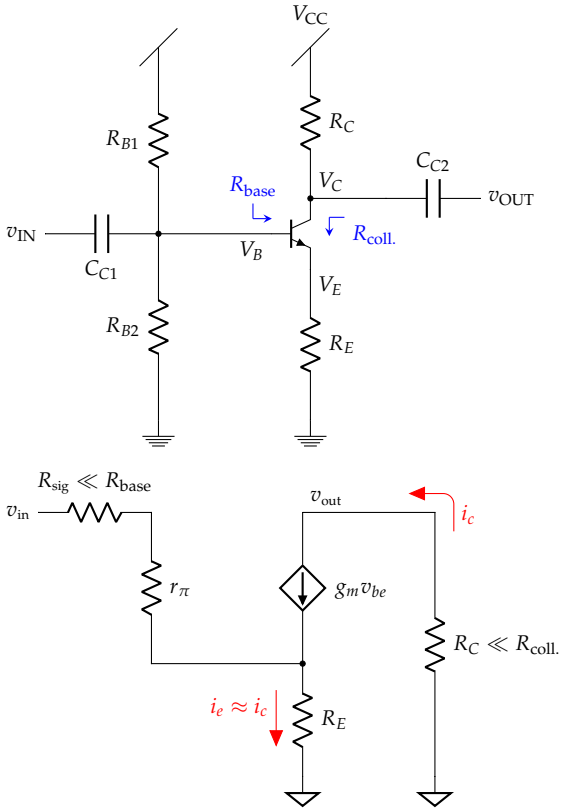


Figure 150: Common-Emitter configuration based on the voltage-divider bias network, and its simplified small-signal model.

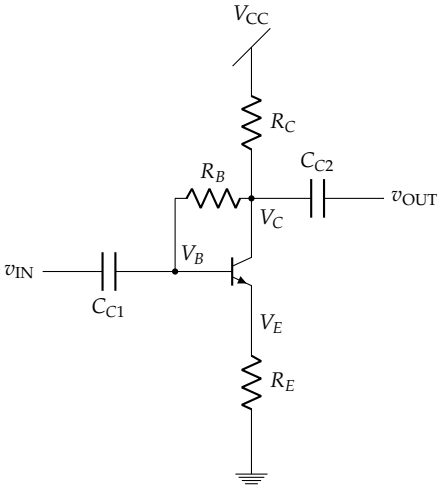
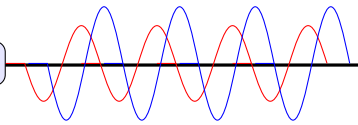


Figure 151: Common-emitter configuration based on the feedback bias network.

EveryCircuit Demonstration 46 (Common-Emitter with feedback bias).



Using the bias network from Example 29, we introduce input and output signals using the capacitive-coupled connections shown in Figure 151. A bypass capacitor is used to eliminate the AC influence of R_E, so the AC gain should be $-g_m R_E$. Since $I_C \approx 1\,\text{mA}$, the transconductance and gain should be

$$g_m = \frac{1\,\text{mA}}{26\,\text{mV}} = 38.46\,\text{mA/V}$$
$$\Rightarrow -g_m R_C = -115\,\text{V/V}$$

The simulation verifies a gain of 104.5 V/V, which is close to our prediction. The small discrepancy is due to the assumptions and approximations made in our analysis.