COCOA: Collaborative Compendium on Analog Integrated Circuits

Table of contents

Preface

This open-source book is meant to be a collaborative effort, bringing together insights from students, professionals, and the broader community of analog integrated circuit designers. It will leverage the new possibilities associated with open-source process design kits (PDKs) and open-source chip design software to build up a knowledge base with reproducible examples in a "live and dynamic" online format.

As of its initial creation in August 2024, it is merely a skeleton with the following structure.

- Part I: Learn to crawl — Square-law transistors, biasing and small-signal analysis
- Part II: Learn to walk — Real transistors, noise, mismatch, and distortion
- Part III: Dare to run — Knowledge base for state-of-the-art circuits

Part I is dedicated to learning about powerful abstractions that are necessary to analyze and design analog integrated circuits. This initial exposure is driven with the simplest possible transistor models so that we can focus on fully understanding these abstractions and don't get distracted by second-order effects that will only become meaningful later on. An added and important side benefit is that we can perfectly match hand calculations and simulations, which not only validates our methods, but also abandons the widespread misconception that circuit simulators rely on some form of "magic" to produce their outputs.

Part II turns expands our horizon to analog circuits with modern transistors, which usually do not obey square-law equations. The good news is that all major abstractions, analysis and design approaches still apply. We just need to abandon the idea of predicting the transistor characteristics with simple equations, and instead rely on lookup tables or advanced model expressions (e.g., EKV-based). Additionally, this is a good time in the overall learning process for dealing with the major pain points of analog design: noise, mismatch and distortion. We review these impairments for the most important circuit primitives, forming a scalable basis for understanding larger circuits.

Part III is meant to capture "deep dives" on commonly used circuits. It will be a forum where you can learn, for example, about the "best" way of going about the design of a bandgap reference, an LC voltage-controlled oscillator, or a successive approximation ADC. This part will undoubtedly be in constant flux and ideally thrive on pointers to reproducible open-source design repositories, simulation data, layouts, etc.

Acknowledgements

This project was inspired by Vijay Janapa Reddi's CS249R Book and borrows from its implementation. Special thanks to James T. Meech for getting this work off the ground (by porting Chapter 1 of Part I to Quarto).

Contributors

We extend our sincere thanks to the diverse group of individuals who have generously contributed their expertise, insights, and time to improve both the content and codebase of this project.

Boris Murmann

James T. Meech

Claudio Talarico

Copyright

Conventions

Please follow these conventions as you contribute to this online book:

1. **Clear Structure and Organization:**
    - **Chapter Outlines:** Begin each chapter with an outline that provides an overview of the topics covered.
    - **Sequential Numbering:** Utilize sequential numbering for chapters, sections, and subsections to facilitate easy reference.
2. **Accessible Language:**
    - **Glossary:** Include a glossary that defines technical terms and jargon.
    - **Consistent Terminology:** Maintain consistent use of terminology throughout the book to avoid confusion.
3. **Learning Aids:**
    - **Diagrams and Figures:** Employ diagrams, figures, and tables to visually convey complex concepts.
    - **Sidebars:** Use sidebars for additional information, anecdotes, or to provide real-world context to the theoretical content.
4. **Interactive Elements:**
    - **Colabs and Projects:** Integrate exercises and projects at the end of each chapter to encourage active learning and practical application of concepts.
    - **Case Studies:** Incorporate case studies to provide a deeper understanding of how principles are applied in real-world situations.
5. **References and Further Reading:**
    - **Bibliography:** Include a bibliography at the end of each chapter for readers who wish to delve deeper into specific topics.
    - **Citations:** Maintain a consistent style for citations, adhering to recognized academic standards like APA, MLA, or Chicago.
6. **Supporting Materials:**
    - **Supplementary Online Resources:** Provide links to supplementary online resources, such as video lectures, webinars, or interactive modules.
    - **Datasets and Code Repositories:** Share datasets and code repositories for hands-on practice, particularly for sections dealing with algorithms and applications.
7. **Feedback and Community Engagement:**

- **Forums and Discussion Groups:** Establish forums or discussion groups where readers can interact, ask questions, and share knowledge.
- **Open Review Process:** Implement an open review process, inviting feedback from the community to continuously improve the content.

8. **Inclusivity and Accessibility:**

   - **Inclusive Language:** Utilize inclusive language that respects diversity and promotes equality.
   - **Accessible Formats:** Ensure the textbook is available in accessible formats, including audio and Braille, to cater to readers with disabilities.

9. **Index:**

   - **Comprehensive Index:** Include a comprehensive index at the end of the book to help readers quickly locate specific information.

Implementing these conventions can contribute to creating a textbook that is comprehensive, accessible, and conducive to effective learning.

Part I.

Learn to Crawl — Square-Law Transistors, Biasing and Small-Signal Analysis

1. Introduction

With the development of the integrated circuit, the semiconductor industry is undoubtedly the most influential industry to appear in our society. Its impact on almost every person in the world exceeds that of any other industry since the beginning of the Industrial Revolution. The reasons for its success are as follows:

- Exponential growth of the number of functions on a single integrated circuit.
- Exponential reduction in the cost per function.
- Exponential growth in sales (economic importance) for approximately forty years.

This growth has led to ever-increasing performance at lower prices for consumer electronics such as cellular phones, personal computers, audio players, etc. The computational power available to the individual has increased to the point that it has changed the way we think about problem solving. Communication technology including wired and wireless networks have fundamentally changed the way we live and communicate.

The innovation responsible for these impressive results is the integration of electronic circuit components fabricated in silicon **integrated circuit** (IC) technology. Today, many of the ICs shaping new applications contain both analog and digital circuitry, and are therefore called mixed-signal integrated circuits. In mixed-signal ICs, the analog circuitry is typically responsible for interfacing with physical signals, and concerned for example with the amplification of a weak signal from an antenna, or driving a sound signal into a loudspeaker. On the other hand, digital circuitry is primarily used for computing, enabling powerful functions such as Fast Fourier Transforms or floating point multiplication.

This module was written as an introduction to the analysis and design of analog integrated circuits in **complementary metal-oxide-semiconductor** (CMOS) technology. In this first chapter, we will motivate this subject by looking at an example of a mixed-signal IC and by highlighting the need for a systematic study of the fundamental principles and proper engineering approximations in analog design.

Chapter Objectives

- Provide a motivation for the study of elementary analog integrated circuits.
- Provide a roadmap for the subjects that will be covered throughout this module.
- Review fundamental concepts for the construction of two-port circuit models.

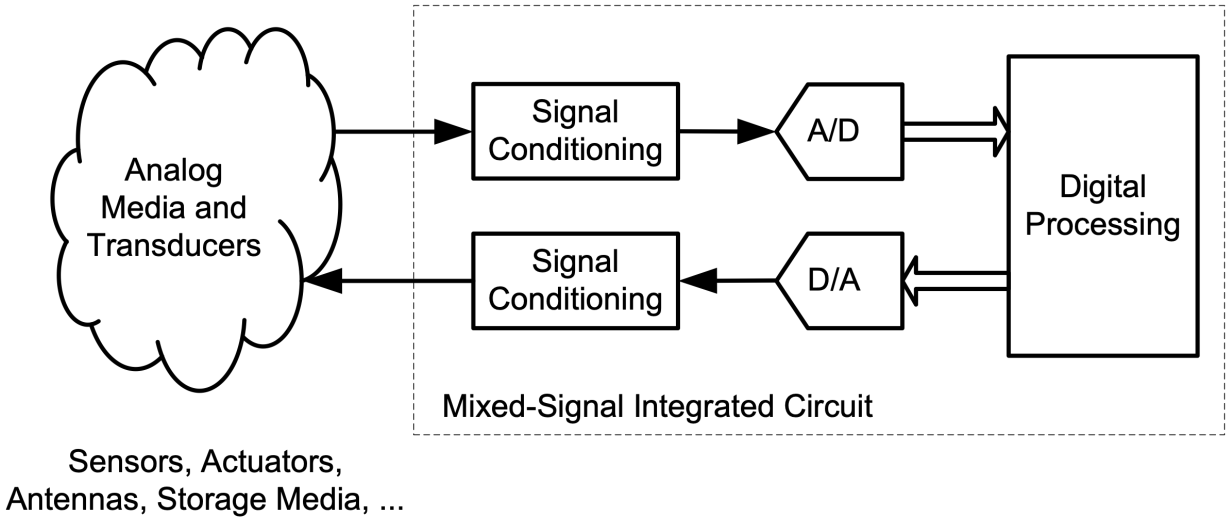


Figure 1.1.: Block diagram of a mixed-signal system.

1.1. Mixed-Signal Integrated Circuits

Figure 1.1 shows a generic diagram of a mixed-signal system, incorporating a mixed-signal integrated circuit. To the left of this diagram are the transducers and media that represent the sources and sinks of the information processed by the system. Examples of input transducers include microphones or photodiodes used to receive communication signals from an optical fiber. Likewise, the output of the system may drive an antenna for radio-frequency communication or a mechanical actuator that controls the zoom of a digital camera. At the boundary between the media and transducers are typically signal conditioning circuits that translate the incoming and outgoing signals to the proper signal strength and physical format. For instance, an amplifier is usually needed to increase the strength of the receive signal from a radio antenna, so that it can be more easily processed by the subsequent system components. In most systems, the signal conditioning circuitry interfaces to analog-to-digital and digital-to-analog converters, which provide the link between analog quantities and their digital representation in the computing back-end of the system.

1.1.1. Example: Single-Chip Radio

The block diagram of a modern mixed-signal integrated circuit is shown in Figure 1.2. This design incorporates most of the circuitry needed to realize a modern cellular phone. For instance, it contains a front-end low-noise amplifier (LNA) to condition the incoming antenna signal. The amplified signal is subsequently frequency shifted, converted into digital format and fed into a digital processor. Even though the block diagram in Figure 1.2 looks quite complex, all of its elements can be mapped into one of the blocks of the generic diagram of Figure 1.1.

Figure 1.3 shows the chip photo of the single-chip radio, with some of the system's key building blocks annotated. As evident from this diagram, the digital logic dominates the area of this particular IC. This situation is not uncommon in modern mixed-signal ICs, not least because

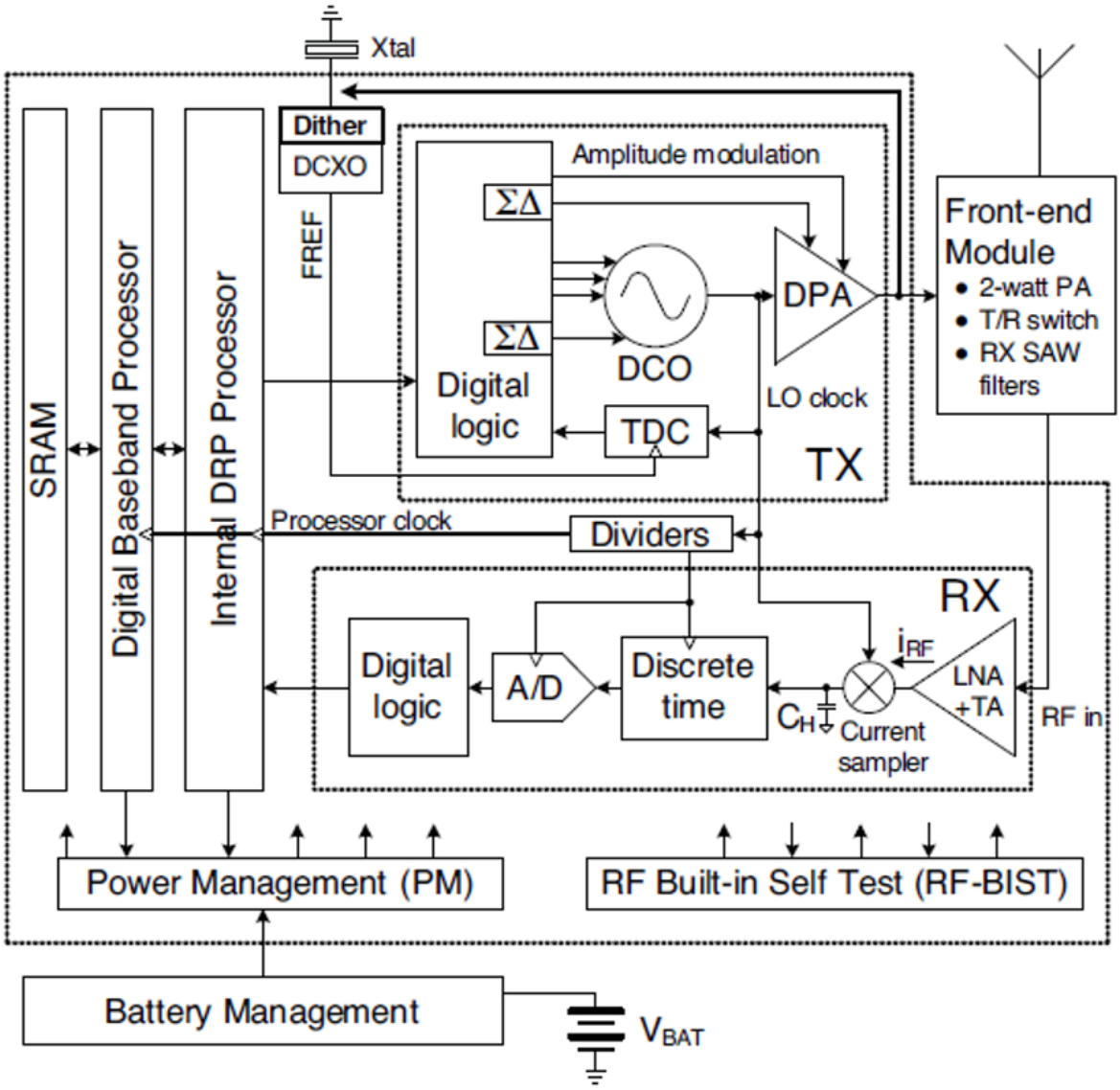


Figure 1.2.: Block diagram of a single-chip radio (Staszewski et al. 2008).

the utilized digital algorithms have reached an enormous complexity, requiring millions or tens of millions of logic gates.

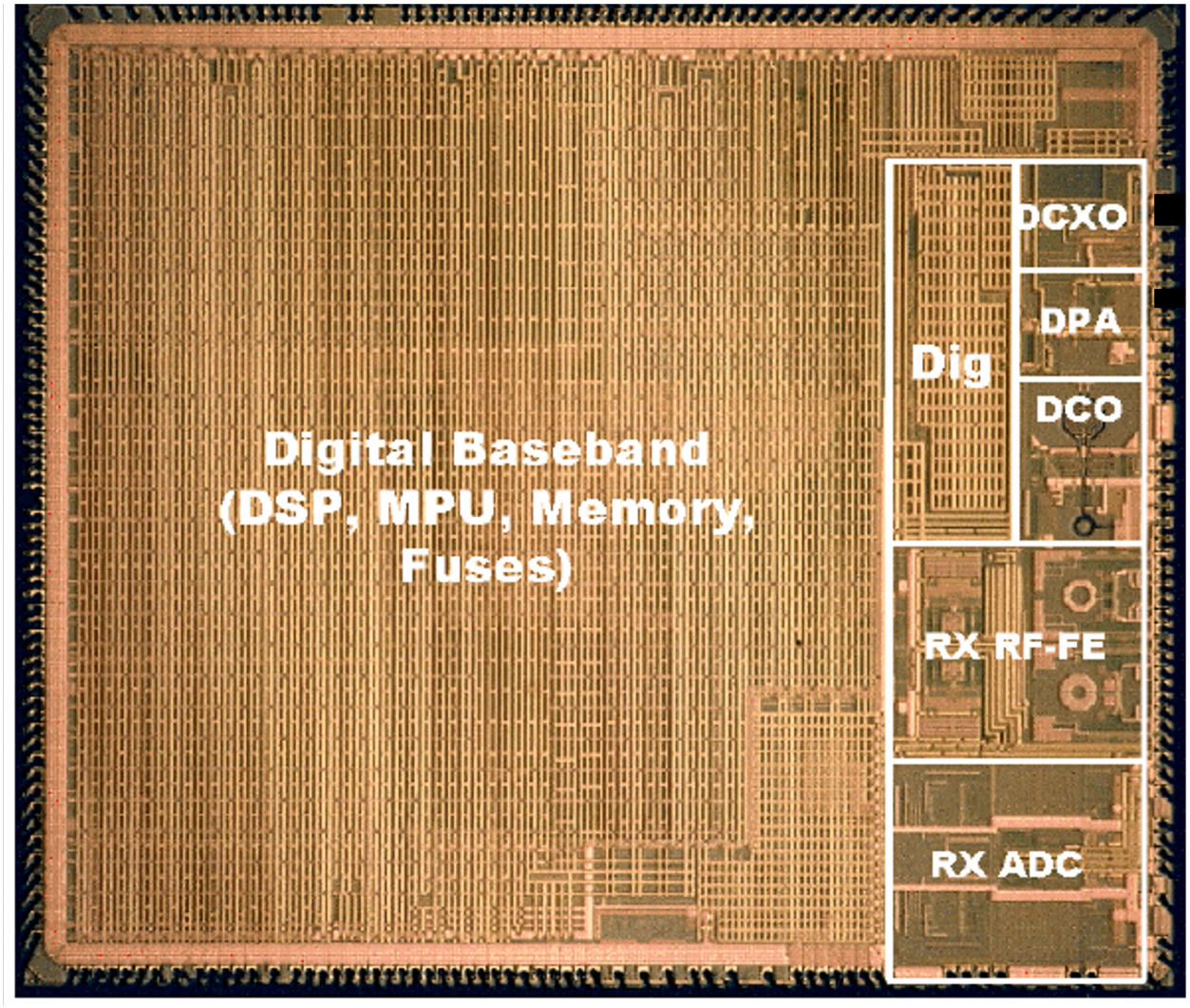


Figure 1.3.: Chip photo of a single-chip radio (Staszewski et al. 2008).

Despite the dominance of digital logic within most systems, the analog interface components are equally important, as they determine how and how much information can be communicated between the physical world and the digital processing backbone. In many cases, the performance of the signal conditioning and data conversion circuitry ultimately determines the performance of the overall system.

1.1.2. Example: Photodiode Interface Circuit

Figure 1.4 shows an example of a signal conditioning circuit that plays a critical role in fiber-optic communication systems. In such a system, a photodiode is used to convert light intensity to electrical current (i_{IN}). In order to condition the signal for further processing, the diode current is converted into a voltage (v_{OUT}) by a so-called **transimpedance amplifier**. This amplifier must

be fast enough to process the incoming light pulses, which often occur at frequencies of multiple gigahertz. In addition, the amplifier must obey certain limits on power dissipation, or the system may become impractical in terms of heat management or power supply requirements.

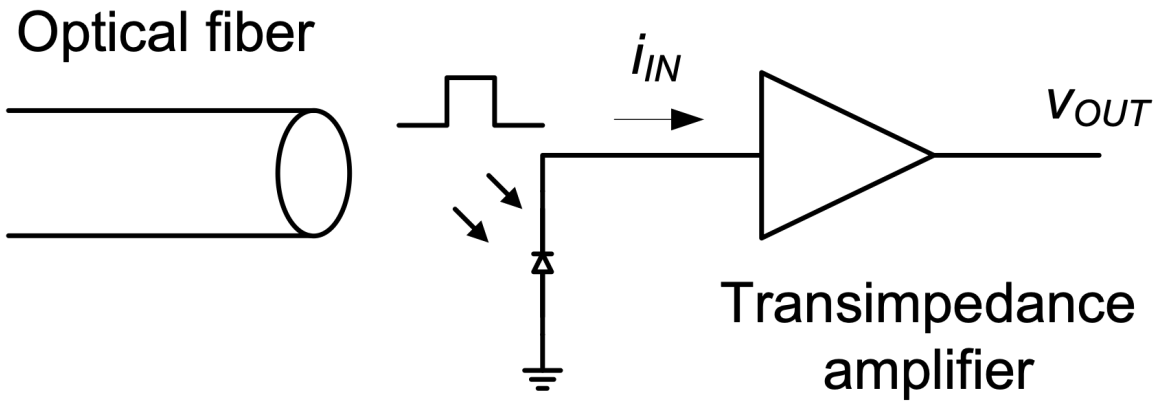


Figure 1.4.: Photodetector circuit for fiber-optic communication.

Limitations in speed and power dissipation are, in general, among the main concerns in the interface circuitry of mixed-signal systems. Since new products tend to demand higher performance, the analog designer is constantly concerned with the design and optimization of system-critical building blocks, aiming for the best possible performance that can be achieved within the framework of the target application and process technology.

A specific example for the circuit realization of a transimpedance amplifier is shown in Figure 1.5. It consists of three transistor stages, each of which serves a specific purpose and design intent. This is true for most amplifier circuits; even though the full schematic of a particular realization may be com- plex, it can usually be broken up into smaller sub-blocks that are more easily understood. Specifi- cally, for the amplifier of Figure 1.5, the experienced designer will recognize that the circuit consists of a cascade connection of a **common-gate**, **common-source**, and **common-drain** stage. These sub-blocks form the basis for a large number of analog circuits, and can be viewed as the "atoms" or fundamental building blocks of analog design. In this module, you will learn to analyze these blocks from first principles, and to reuse the gained knowledge for the design of more complex circuits. The circuit of Figure 1.5 will be analyzed in detail in Chapter 6 of this module, building upon the principles covered in Chapters 2–5.

1.2. Managing Complexity

As evident from the example of Section 1-1-1, modern integrated circuits are highly complex and require a hierarchical approach in design and analysis. That is, a modern integrated circuit is far too complex to be fully understood and analyzed in a single sheet schematic at the transistor level. Typically, a mixed-signal IC is represented by a block diagram as the one shown in Figure 1.2. At the level of this description, suitable specifications are derived for each block, which may itself contain several sub-blocks. The blocks and sub-blocks are then designed and optimized until they meet the desired target specifications.

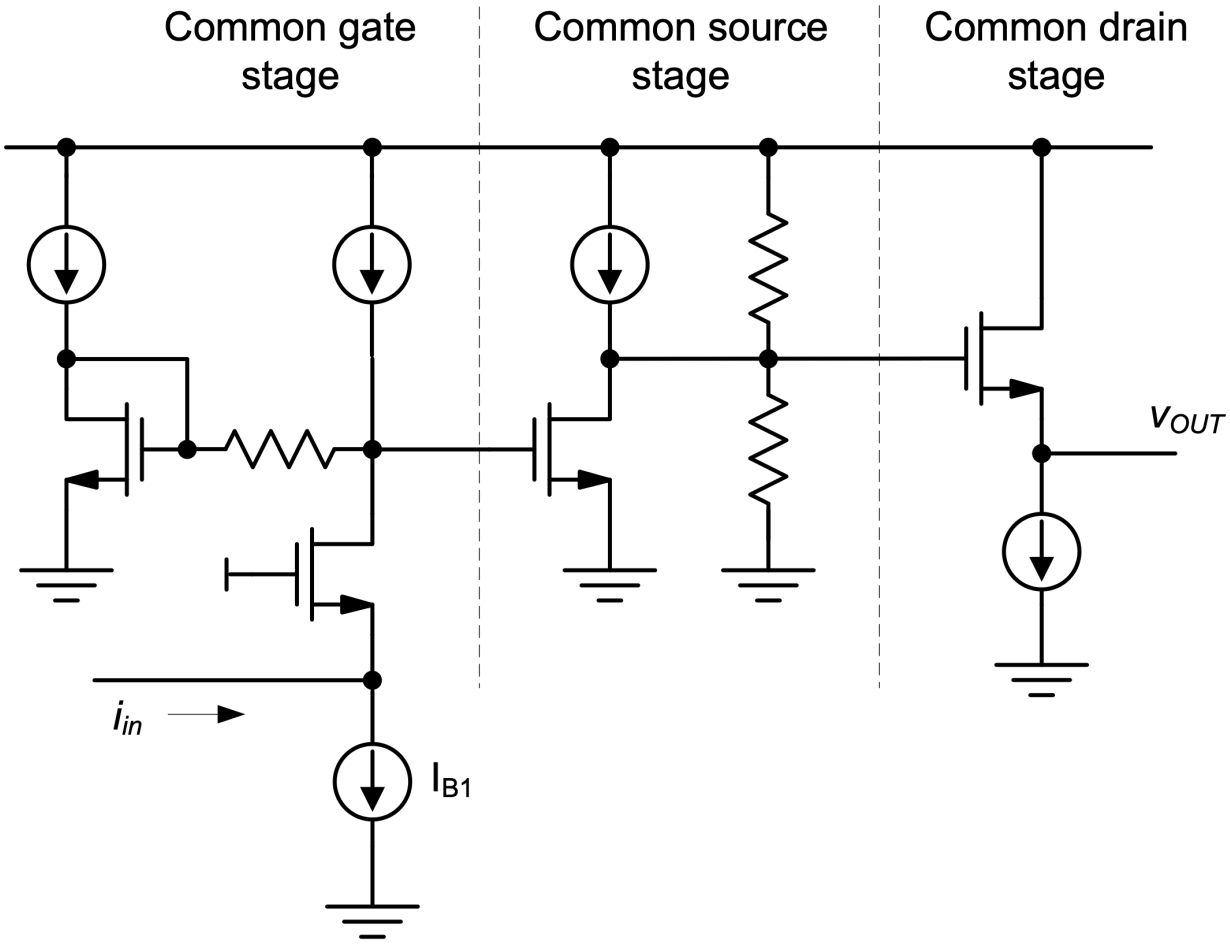


Figure 1.5.: Example realization of a transimpedance amplifier.

Figure 1.6 illustrates examples of the various levels of abstraction that come into play in the design of a modern integrated circuit. At the highest level, the constituent elements can be partitioned into analog and digital blocks. An example of a high-level analog block is an analog-to-digital converter, whereas a microprocessor is an example of a large digital block. These blocks themselves contain smaller functional units, as, for example, operational amplifiers in the case of an analog-to-digital converter. The operational amplifiers themselves contain the aforementioned elementary transistor stages, which are the main subject of this module.

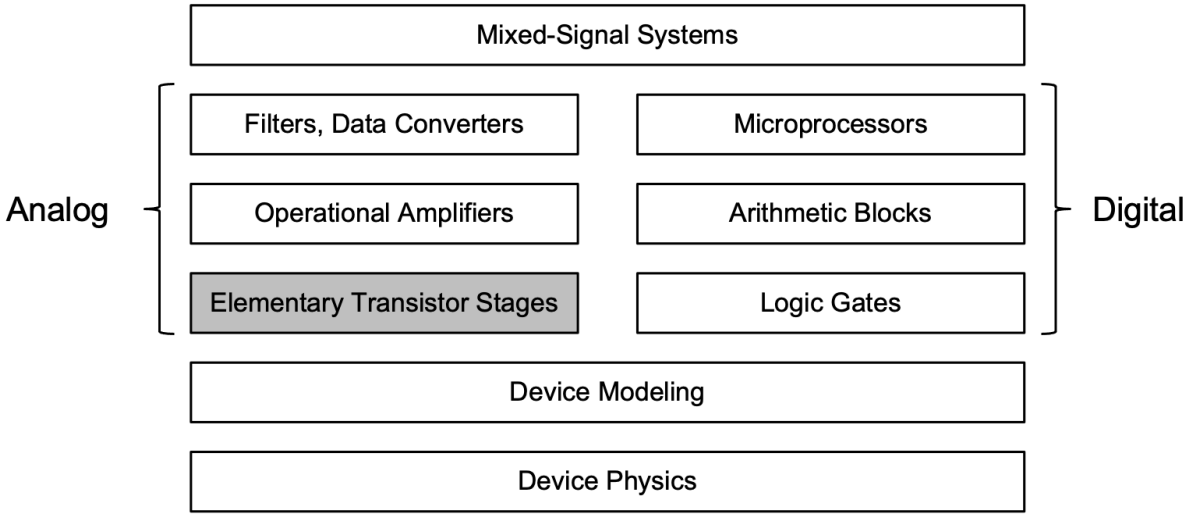


Figure 1.6.: Levels of abstraction in integrated circuit design.

Interestingly, even at the level of elementary transistor stages, is often not possible to work with a perfect model or description of the circuit. This is particularly so because the physical effects in the constituent transistors are highly complex and often impossible to capture perfectly with a tractable set of equations for hand analysis. Therefore, making proper engineering approximations in transistor modeling is an important aspect in maintaining a systematic design methodology. For this particular reason, the presentation in this module follows a "just-in-time" approach for the modeling of transistor behavior. Rather than deriving a complete transistor model in an isolated chapter (as done in most texts), we begin with only the basic device properties and increase complexity throughout the module upon demand, and where needed to gain further insight and accuracy. With this approach, the reader learns to appreciate the complexity of a refined model, and will be able to assess and track potential limitations of working with simplified models.

1.3. Two-Port Abstraction for Amplifiers

High-level system block diagrams, such as Figure 1.2 , are typically drawn as unidirectional flowcharts and do not capture details about the electrical behavior of each connection port and how certain blocks may interact once they are connected. Unfortunately, electrical signals are not unidirectional, and connecting two blocks always means that there is some level of interaction through the voltages and currents at the connection points.

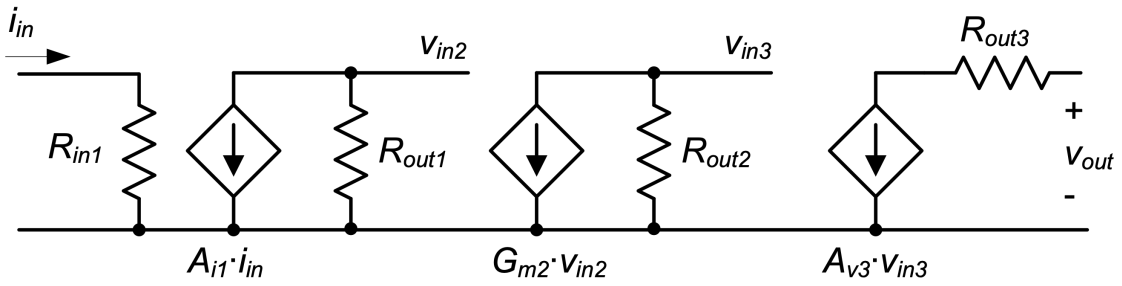


Figure 1.7.: Two-port model of the transimpedance amplifier circuit in Figure 1.5.

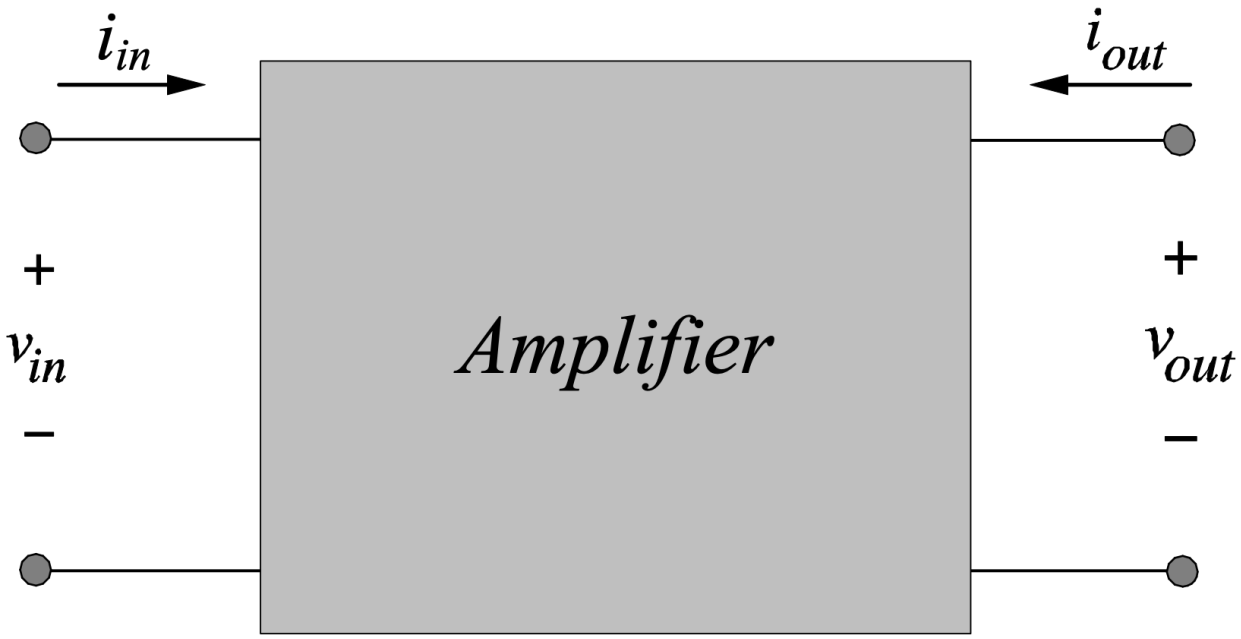


Figure 1.8.: General amplifier two-port.

The commonly used linear two-port modeling abstraction for amplifiers and amplifier stages allows the designer to take these effects into account while maintaining a high level of abstraction. For instance, the circuit of Figure 1.5 can be approximately modeled as shown in Figure 1.7 (the details on obtaining this model are discussed later in this module). Each stage of the overall amplifier is represented via a simplified circuit model that captures its essential features. Once this model is created, the interaction among stages can be analyzed at this high level of abstraction, without requiring detailed insight on how each stage is implemented. The two-port modeling approach is particularly useful in the design of amplifiers, as it can help shape the thought process on how the various stages should be configured to optimize performance. In the following subsections, we will review some of the basic concepts of amplifier two-port modeling used in this module.

1.3.1. Amplifier Types

In this module, we model amplifier circuits as blocks that have an input and output port, where the term "port" refers to a pair of terminals. For each port, we can define input and output currents and voltages as shown in Figure 1.8. Depending on the intended function, we distinguish between the four possible amplifier types listed in Table 1-1. For example, an amplifier that takes an input current and amplifies this current to produce a proportional output voltage is called a transresistance amplifier. In this context, it is important to emphasize that in a general practical amplifier circuit, the input and output ports will always carry both nonzero voltages and currents, and there exist transfer functions between all possible combinations of input/output variables. What truly defines the type of an amplifier is what the circuit designer deems as the main quantities of interest in the amplifier's application.

Table 1.1.: Amplifier types.

Amplfier Type	**Input Quantity**	**Output Quantity**
Voltage Amplifier	Voltage	Voltage
Current Amplifier	Current	Current
Transconductance Amplifier	Voltage	Current
Transresistance Amplifier	Current	Voltage

Now, in order to model the inner workings of each amplifier type, we can invoke the four corresponding two-port amplifier models shown in Figure 1.9. Each amplifier model has an input and output resistance (or more generally, a frequency dependent impedance) and a controlled source to model the amplification.

- In the **voltage amplifier model**, the controlled source is a voltage-controlled voltage source. Ideally, the input resistance is infinite (open circuit, no current flow). The ideal output resistance is zero (ideal voltage source).
- The **current amplifier model** has a current-controlled current source. Ideally, the input resistance is zero (short circuit, no voltage across the input port) and the output resistance is infinite (ideal current source).
- The **transconductance amplifier model** has a voltage-controlled current source. Ideally, the input resistance is infinite (open circuit, no current flow). The ideal output resistance is also infinite (ideal current source).

- The **transresistance or transimpedance[1] amplifier model** has a current-controlled voltage source. Ideally, the input resistance is zero (short circuit, no voltage across the input port). The ideal output resistance is also zero (ideal voltage source).

From these four models and their ideal behavior, we note that the two-ports containing a voltage-controlled source should ideally have large input resistance (R_{in}). This minimizes the signal loss due to resistive voltage division between the source voltage (v_s) and the control voltage (v_{in}). In contrast, the two-ports that use a current-controlled source should have small input resistance to minimize the signal loss due to current division between the source current (i_s) and the control current (i_{in}). In this context, "large" and "small" refer to the value of R_{in} relative to the source resistance (R_s).

On the output side, if the variable of interest is a voltage, the output resistance (R_{out}) should be small so that only a small amount of the amplified voltage is lost through the division with the load (R_L). Conversely, for a current output, the output resistance should be large to minimize current division losses. Again, "large" and "small" are taken as relative measures comparing Rout to R_L.

Consider for example the voltage amplifier of Figure 1.9(a). To calculate the transfer function of the overall circuit (v_{out}/v_s), the input voltage, including its source resistance, is connected to the input of the two-port model and the load resistance is connected to the output. The full circuit is shown in Figure 1.10.

Applying the voltage divider rule at the input and output of the circuit gives

$$\frac{v_{out}}{v_s} = \left(\frac{v_{in}}{v_s}\right) \cdot A_v \cdot \left(\frac{v_{out}}{v_x}\right) = \left(\frac{R_{in}}{R_{in} + R_s}\right) \cdot A_v \cdot \left(\frac{R_L}{R_L + R_{out}}\right) \tag{1.1}$$

As we can see from this expression, the overall voltage gain is maximized when the amplifier has a large input resistance (relative to R_s) and a small output resistance (relative to R_L). For the ideal case of infinite input resistance and zero output resistance, v_{out}/v_s becomes equal to A_v.

For the sake of compact notation, we will often want to use a symbol for the overall circuit gain. The notation used in this module uses primed variables to distinguish between the gain of the controlled source and the gain of the overall amplifier circuit. For example, for the above-discussed voltage amplifier we define $A'_v = v_{out} \ / \ v_s$. This notation is meant to emphasize the connection between the two symbols. A'_v is usually smaller than A_v, but can approach A_v for ideal source and load configurations.

Example 1-1: Transfer Function Transconductance Amplifier.

For the transconductance amplifier circuit in Figure 1.11, calculate the overall transconductance $G'_m = i_{out} \ / \ v_s$.

SOLUTION

Applying the voltage divider rule at the input and the current divider rule at the output yields the following result:

$$G'_m = \frac{i_{out}}{v_s} = \left(\frac{v_{in}}{v_s}\right) \cdot G_m \cdot \left(\frac{i_{out}}{i_x}\right) = \left(\frac{R_{in}}{R_{in} + R_s}\right) \cdot G_m \cdot \left(\frac{R_{out}}{R_L + R_{out}}\right)$$

[1]The term transimpedance is sometimes used to refer to an amplifier that is primarily meant to realize a transresistance. Referring to "impedance" highlights the fact that the transfer function will usually be frequency-dependent.

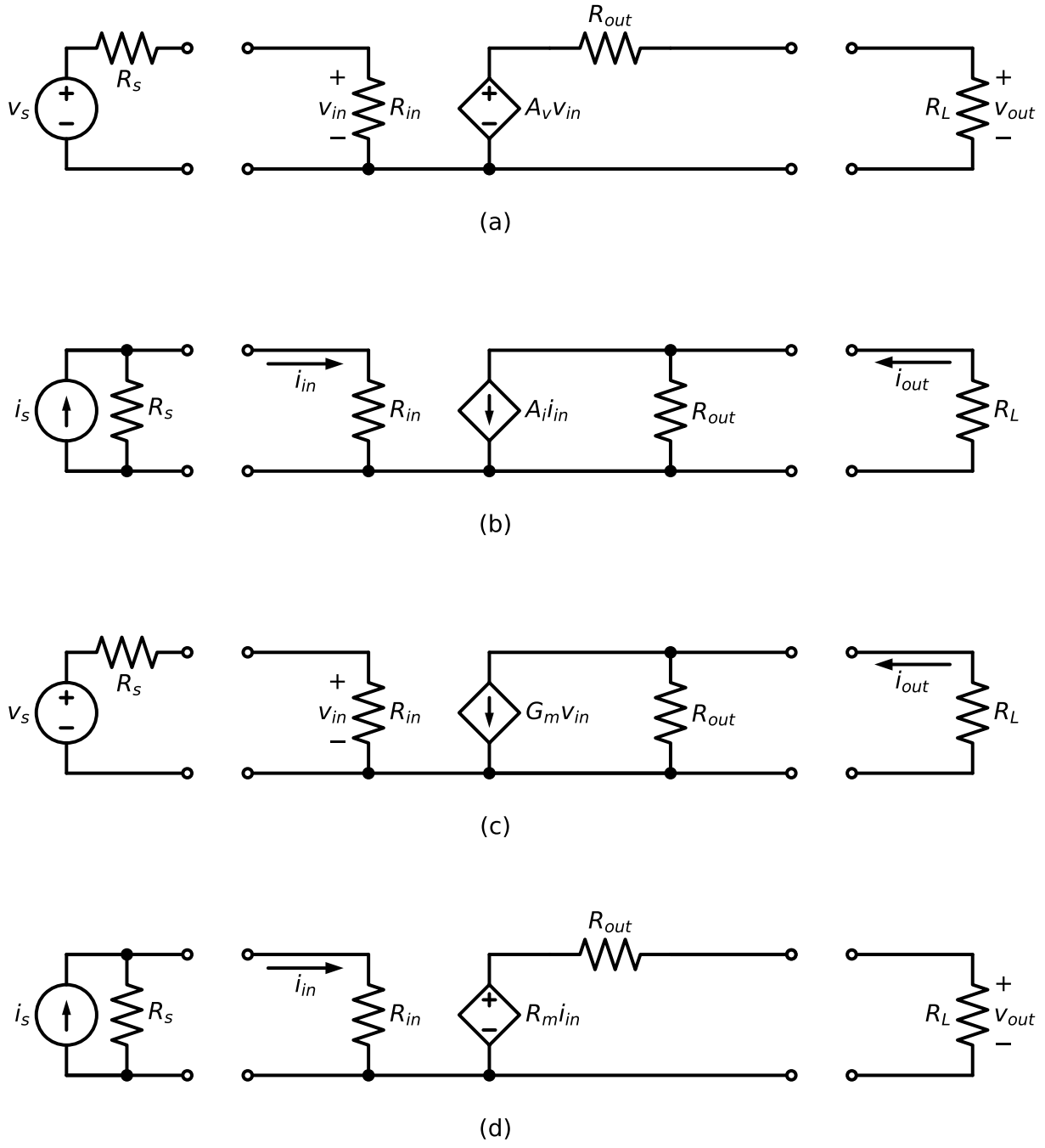


Figure 1.9.: Two-port amplifier models with input source and load: (a) voltage amplifier, (b) current amplifier, (c) transconductance amplifier, and (d) transresistance amplifier

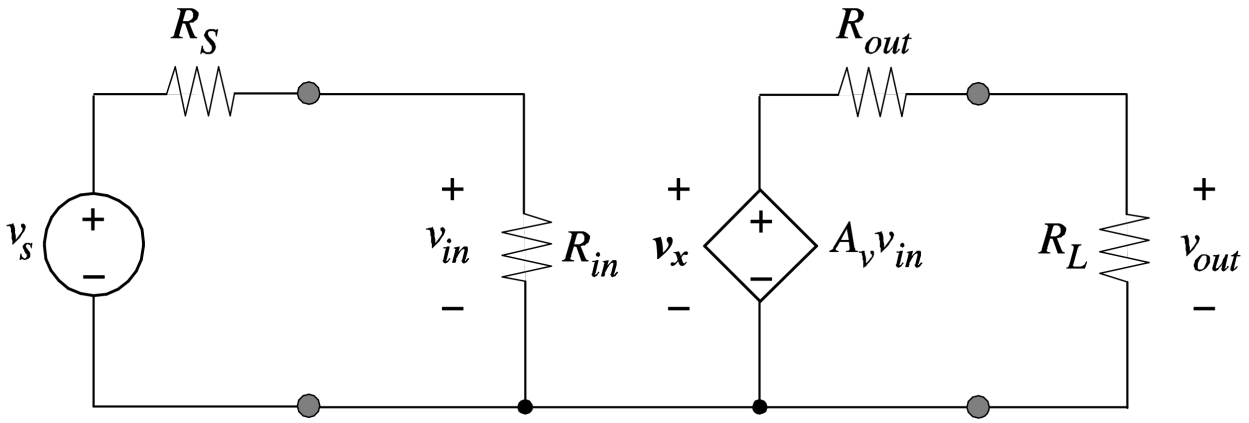


Figure 1.10.: Voltage amplifier with connected source and load resistances.

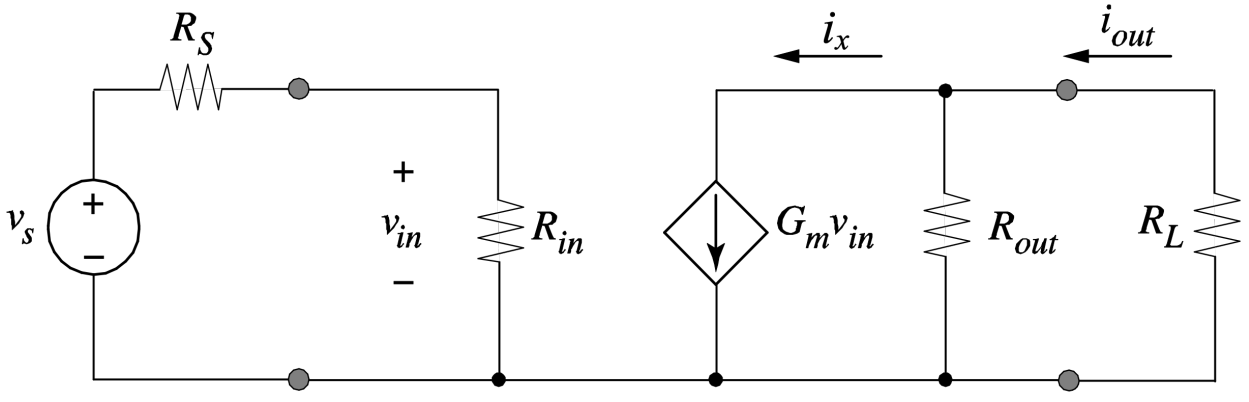


Figure 1.11.: **Figure Ex 1-1**

Thus, the overall transconductance gain is maximized when the amplifier has a large input resistance (relative to R_s) and a large output resistance (relative to R_L). For the ideal case of infinite input and output resistances, G'_m becomes equal to G_m.

As a final remark for this sub-section, it is important to recognize that all of the models in Figure 1.9 can be used interchangeability to describe the exact same electrical behavior (see Problem P1-1). For instance, a voltage amplifier model can be converted into a transconductance amplifier model by applying a Thevénin to Norton transformation for the controlled source.

A corollary to this equivalence is that we can for example use a transconductance amplifier model to describe a voltage amplifier circuit. This is illustrated through the circuit of Figure 1.12, which is electrically equivalent to that of Figure 1.10 (see Problem P1-2). Note that the output is taken as the voltage across the output port instead of the output current; this indicates that the circuit is viewed as a voltage amplifier. Just as in the original circuit of Figure 1.10, we require a large input resistance and small output resistance for this circuit to maximize the overall voltage gain.

The choice of amplifier model depends on several factors. At first glance, it seems natural to model each amplifier type using its "native" model that directly corresponds to the intended function. For example, we could always describe a voltage amplifier using the corresponding voltage amplifier model that contains a voltage controlled voltage source. However, as we shall see throughout this module, it is sometimes more convenient to align the amplifier model with the physical amplification mechanism or a structural feature of the underlying transistor circuit. For instance, the common-source voltage amplifier discussed in Chapter 2 naturally invokes a transconductance-based model due to the physical model of the employed transistor.

1.3.2. Unilateral versus Bilateral Two-Ports

All of the two-port models shown in Figure 1.9 are called **unilateral**, because they can only propagate a signal from the input port to the output port and not the other way around. For instance, injecting a current into the output port of the current amplifier of Figure 1.9(b) will not induce a current at the input port. Unfortunately, many practical transistor circuits are not unilateral, and exhibit **bilateral** behavior when analyzed in detail, and especially at high frequencies.

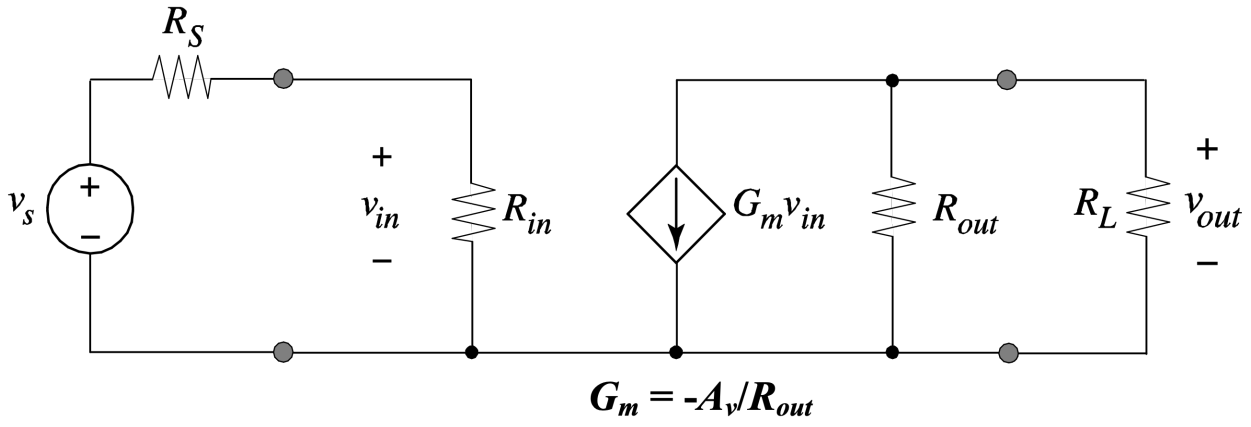


Figure 1.12.: Voltage amplifier with an underlying transconductance amplifier model

Figure 1-12

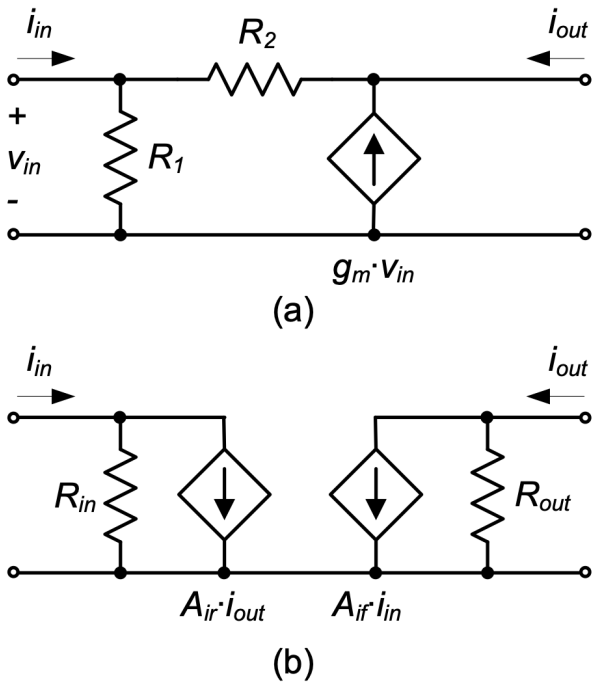


Figure 1.13.: (a) Example of a bilateral current amplifier. (b) Corresponding bilateral current amplifier two-port model.

An example of a bilateral current amplifier is shown Figure 1.13(a). Note that in this circuit, resistor R_2 couples the input and output networks and it can therefore transfer currents in both directions. Consequently, the unilateral model of Figure 1.9(b) cannot perfectly represent this circuit. When it is desired to capture the bilateral behavior, the two-port model in Figure 1.13(b) could be employed in principle. Here, the controlled source Air models the reverse current transfer from the output back to the input. Alternatively, one could employ other bilateral and more general two-port models based on admittance parameters (Y), impedance parameters (Z), and hybrid or inverse-hybrid parameters (H or G); see advanced circuit design texts such as (Gray et al. 2009a). These models are particularly useful when reverse transmission (i.e., feedback from the output to the input) is incorporated in the circuit as part of the intended design.

There are two reasons why we will work exclusively with unilateral two-port approximations in this module. First, the circuits considered are designed primarily to implement forward gain rather than reverse gain; feedback circuits are not treated in this module. For example, referring to the model of Figure 1.13(b), the reverse gain Air will be negligibly small in any current amplifier circuit that we will consider. Second, a clear drawback of working with bilateral two-port models would be a significant increase in analysis complexity. As we have seen in Example 1-1, the overall transfer function of a unilateral two-port circuit can be written by applying simple voltage and current divider rules. This also extends to cascade connections of multiple two-ports. For example, the overall transfer function of the circuit in Figure 1.7 is easily written by inspection, without requiring extensive algebra. With reverse transmission included, the transfer function analysis will generally require solving a linear system of equations. In light of the fact that we do not intend to design circuits in this module that have significant reverse transmission, this increase in complexity is not welcome, and would also hinder us from developing intuition from inspection-driven analysis.

1.3.3. Construction of Unilateral Two-Port Models

We will now describe the general procedures to calculate the controlled sources, as well as the input and output resistances, for the unilateral two-port models of Figure 1.9. The approach is based on applying test voltages and currents to find the desired model parameters.

The most important parameter of any amplifier circuit is its gain. To identify the gain parameters for the models of Figure 1.9(a)-(d), we apply the tests shown in Figure 1.14(a)-(d), respectively.

- To calculate the gain term A_v of a voltage amplifier model, we apply a test voltage at the input with zero source resistance and measure the open-circuit output voltage. $A_v = v_{oc}/v_t$ is therefore also called the **open-circuit voltage gain**.
- To calculate the gain term A_i of a current amplifier model, we apply a test current at the input with infinite source resistance and measure the short-circuit output current. $A_i = i_{sc}/i_t$ is therefore also called the **short-circuit current gain**.
- To calculate the gain term G_m of a transconductance amplifier model, we apply a test voltage at the input with zero source resistance and measure the short-circuit output current to find $G_m = i_{sc}/v_t$.
- To calculate the gain term R_m of a transresistance amplifier model, we apply a test current at the input with infinite source resistance and measure the open-circuit output voltage to find $R_m = v_{oc}/v_{it}$.

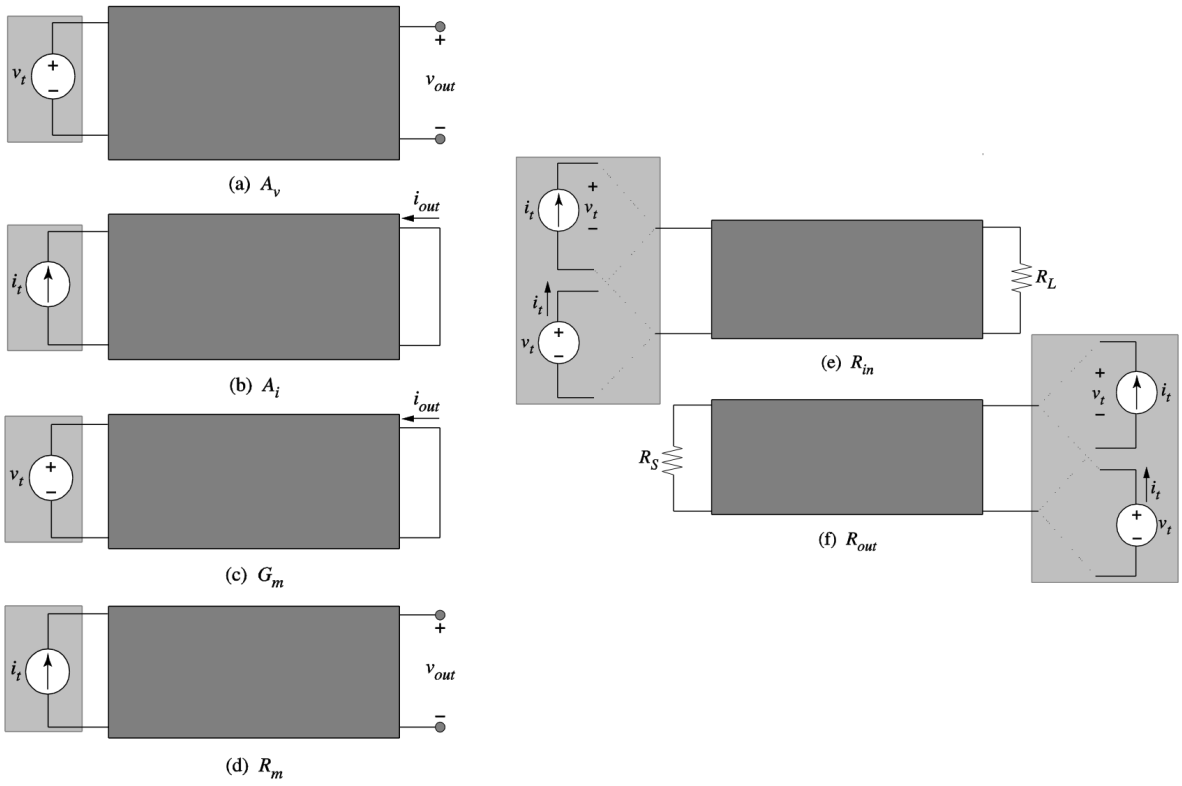


Figure 1.14.: Method to calculate two-port amplifier model parameters: (a) voltage gain A_v, (b) current gain A_i, (c) transconductance G_m, (d) transresistance R_m, (e) input resistance R_in, and (f) output resistance R_out.

The rationale behind these tests can be understood by considering, for example, the case of the voltage amplifier model of Figure 1.9(a). When driven with an ideal voltage source, the effect of any resistance at the input port is eliminated, and the controlled source is directly stimulated by the applied test source (without any voltage division). Likewise, by measuring the resulting output voltage open-circuited, any resistance in series with the controlled source has no effect and the measurement therefore accurately extracts the parameter A_v. Similar explanations apply to the test cases for the remaining amplifier models.

The test setup for extracting the input and output resistances for all amplifier models is shown in Figure 1.14(e), and (f), respectively.

- To calculate the **input resistance** R_{in} we apply a test voltage and measure the current coming from the test source, or apply a test current and mea- sure the voltage across the test source. In this test, the load resistance (R_L) must be connected to the output port as shown in Figure 1.14(e).
- To calculate the **output resistance** R_out, we apply either a test voltage or a test current source at the output port and measure the respective current or voltage from the source. Here, the input source must be set equal to zero. This means that input voltage sources are shorted and input current sources are open-circuited. Only the source resistance (R_S) is left across the input terminals as shown in Figure 1.14(f).

The above procedures extract the input and output resistances perfectly and without any approximations, even if the circuit is bilateral. As we shall see through the examples below, R_{in} and R_{out} do not depend on R_L and R_S, respectively, in a perfectly unilateral amplifier. However, this is not the case in a bilateral amplifier, and therefore the general procedure includes R_L and R_S in the test setup.

In summary, the above procedures for measuring unilateral two-port model parameters aim at finding the best possible unilateral representation of an arbitrary amplifier circuit, which itself may or may not be unilateral. The obtained models are approximate when the amplifier is bilateral, since they do not include a controlled source that captures reverse transmission from the output back to the input. In most cases considered in this module, the reverse transmission term is negligible. Exceptions will be highlighted and treated as appropriate.

Example 1-2: Two-Port Model Calculations for a Unilateral Amplifier

For the transconductance amplifier in Figure 1.15, calculate the following two-port model parameters: the transconductance G_m, the input resistance R_{in}, and the output resistance R_{out}. Also, compute the overall transfer function $G'_m = i_{out} \,/\, v_s$.

SOLUTION

To find the transconductance, we short the output port and apply an ideal test voltage source (v_t) at the input (see Figure 1.15). From this circuit, we see that

$$G_m = \frac{i_{sc}}{v_t} = \frac{g_m v_x \cdot \frac{R_3}{R_3+R_4}}{v_t} = \frac{g_m v_x \cdot \frac{R_2}{R_1+R_2} \cdot \frac{R_3}{R_3+R_4}}{v_t} = g_m \cdot \frac{R_2}{R_1+R_2} \cdot \frac{R_3}{R_3+R_4}$$

Next, to find R_in, we apply a test voltage at the input and connect the load resistance R_L at the output (Figure 1.16). From this circuit, we find that the input resistance is simply the series

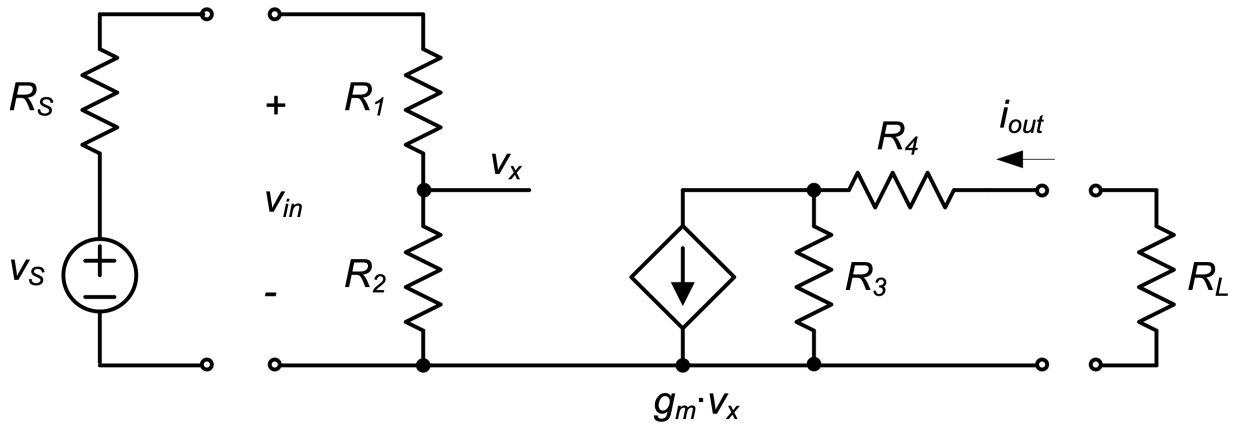


Figure 1.15.: Figure Ex1-2A

connection of R_1 and R_2, i.e., $R_{in} = R_1 + R_2$. Note that the output network does not influence this result.

Finally, to find R_{out}, we apply a test voltage at the output and connect the source resistance R_S across the input port (the source v_s is replaced by a short), (see Figure 1.16(c)). In the resulting circuit, v_x must be zero, because no current is flowing in the input network. Thus, the controlled source carries no current and we conclude that $R_{out} = R_3 + R_4$.

In order to compute the transfer function of the complete circuit, we can reuse the result obtained in Example 1-1.

$$G'_m = \frac{i_{out}}{v_s} = \left(\frac{R_{in}}{R_{in} + R_s}\right) \cdot G_m \cdot \left(\frac{R_L}{R_L + R_{out}}\right)$$

Substituting G_m, R_{in}, and R_{out} from the above calculation yields the final result.

$$G'_m = \frac{i_{out}}{v_s} = \frac{g_m R_2 R_3}{(R_1 + R_2 + R_S)(R_L + R_3 + R_4)}$$

In the preceding example, we have seen that the source and load resistances have no effect on the extracted two-port parameters. In the following example, we will investigate a bilateral circuit to show that in general, the input and output resistances depend on R_S and R_L, which must therefore always be included in the general two-port modeling calculations.

Example 1-3: Two-Port Model Calculations for a Bilateral Amplifier

For the current amplifier in Figure 1.13(a), calculate the following unilateral two-port model parameters: the current gain A_i, the input resistance R_{in}, and the output resistance R_{out}. Also, compute the overall transfer function i_{out}/v_s using the obtained unilateral two-port model. Compare the result to a direct KCL-based analysis of the transfer function. Assume that the circuit is driven by a current source with resistance R_S and loaded by a resistance R_L. For algebraic simplicity, assume $R_1 = 1/g_m$ (this case corresponds to the common-gate amplifier circuit covered in Chapter 4).

SOLUTION

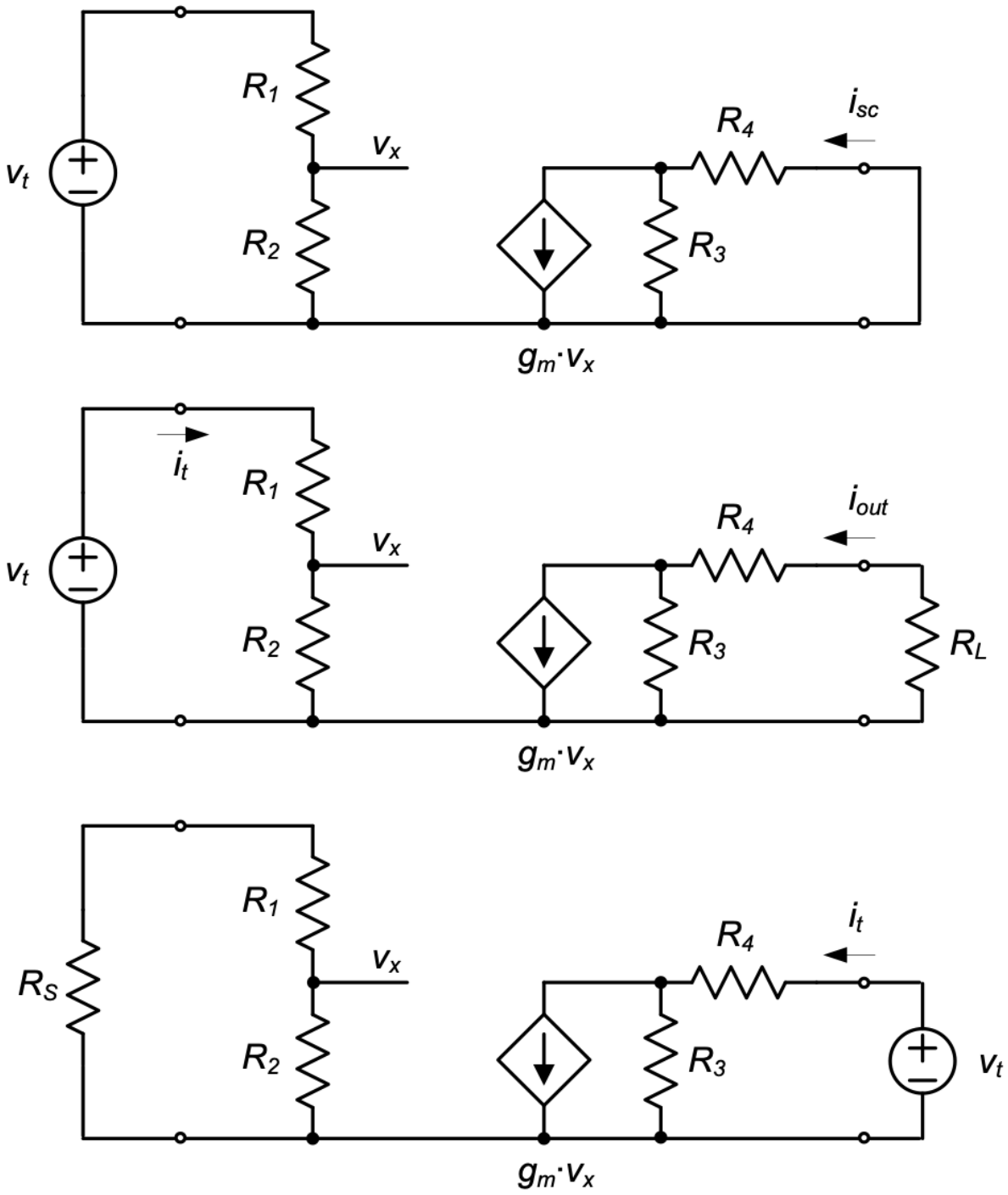


Figure 1.16.: Figure Ex1-2B

To find the current gain Ai, we short the output port and apply an ideal test current source (i_t) at the input (see Figure 1.17(a)). From this circuit, we see that

$$v_{in} = \left(g_m + \frac{1}{R_2}\right)^{-1} \cdot i_t$$

and

$$i_{sc} = -\left(g_m + \frac{1}{R_2}\right) \cdot v_{in}$$

Thus, Ai = isc/it = –1.

Next, to find R_{in}, we apply a test voltage at the input and connect the load resistance R_L at the output (Figure 1.17(b)). From this circuit, we note that the input resistance is not easily identified by inspection. Hence we write KCL for the two nodes of the circuit (v_t and $v_o ut$).

$$0 = -i_t + g_m v_t + \frac{v_t - v_o ut}{R_2}$$

$$0 = -g_m v_t + \frac{v_{out}}{R_L} + \frac{v_t - v_o ut}{R_2}$$

Solving this system of equations yields

$$R_{in} = \frac{v_t}{i_t} = \frac{R_2 + R_L}{1 + g_m R_2} = \frac{1 + \frac{R_L}{R_2}}{g_m + \frac{1}{R_2}}$$

Note from this result that $R_i n$ depends on R_L, as mentioned previously; this dependency stems from the bilateral structure of the circuit. Also note that Rin approaches $1/g_m$ when R_2 is large compared to R_L and $ 1 / g_m$. We will revisit this important point in Chapter 4, in the context of a common-gate amplifier circuit.

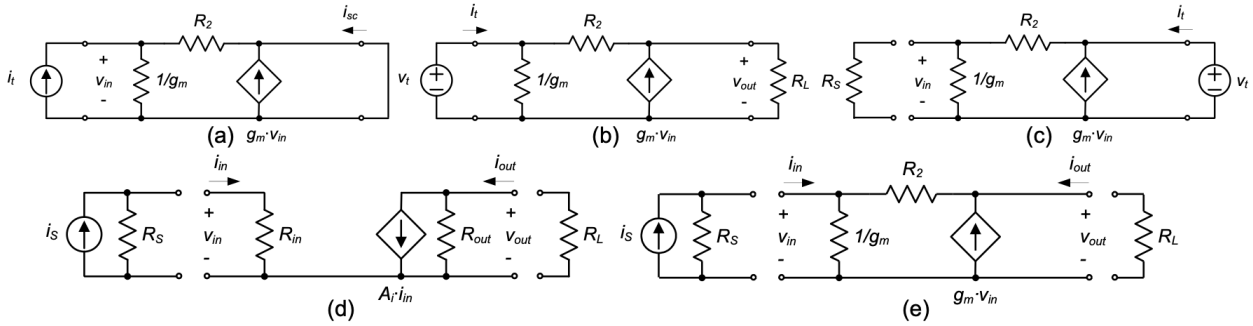


Figure 1.17.: Figure Ex1-3

Now, to find R_{out}, we apply a test voltage at the output and connect the source resistance R_S across the input port Figure 1.17. Again, we must write KCL at the two circuit nodes and solve the resulting system of equations. This yields

$$R_{out} = \frac{v_t}{i_t} = R_2 R_S + g_m R_2 R_S$$

Again, note that R_out is a function of R_S; this is the case for any bilateral circuit.

Finally, to compute the transfer function based on the obtained unilateral model, we consider the circuit shown in Figure 1.17(d). By inspection, we see that

$$A_i' \frac{i_{out}}{i_s} = \Big(\frac{R_S}{R_{in} + R_S}\Big) A_i \frac{R_{out}}{R_L + R_{out}}$$

Substituting A_i, R_{in}, and R_{out} from the above calculation into this expression yields

$$A'_{i,Two-Port} = \frac{i_{out}}{i_s} = -\frac{R_S(1 + g_m R_2)(R_2 + R_S + g_m R_2 R_S)}{(R_L + R_2 + R_S + g_m R_2 R_S)^2}$$

We now wish to compare this result to the accurate transfer function of the circuit, obtained by direct calculation and without approximating the circuit as a unilateral two-port. For this purpose, we consider the full circuit shown in Figure 1.17(e) and write KCL for its two nodes.

$$0 = -i_s + g_m v_i n + \frac{v_{in}}{R_S} + \frac{v_{in} - v_{out}}{R_2}$$

$$0 = -g_m v_{in} + \frac{v_{out}}{R_L} + \frac{v_{out} - v_{in}}{R_2}$$

Solving this system of equations for v_{out} and substituting $i_{out} = -v_{out}/R_L$ yields

$$A'_{i,Exact} = \frac{i_{out}}{i_s} = -\frac{R_S(1 + g_m R_2)}{R_L + R_2 + R_S + g_m R_2 R_S}$$

The discrepancy factor between the two results is given by

$$\frac{A'_{i,Two-Port}}{A'_{i,Exact}} = \frac{R_2 + R_S + g_m R_2 R_S}{R_L + R_2 + R_S + g_m R_2 R_S} = \frac{1 + R_S\Big(\frac{1}{R_2} + g_m\Big)}{1 + \frac{R_L}{R_2} + R_S\Big(\frac{1}{R_2} + g_m\Big)}$$

From this result, we see that the discrepancy factor approaches unity (no error) when R_2 is much larger than R_L, a condition that is often satisfied in practice (the ideal load for a current amplifier is a short circuit). In this case, the unilateral two-port model will accurately describe the behavior of the circuit.

The outcome of the above example captures the main spirit in which we justify relying on unilateral two-port models in this module. Even though the considered amplifier is strictly speaking bilateral, a unilateral model describes its behavior to within the desired engineering accuracy, provided that reason- able boundary conditions hold.

1.4. Integrated Circuit Design versus Printed Circuit Board Design

In the design of analog circuits, the underlying technology has a significant impact on the choice of architecture, because it tends to restrict the availability and specification range of the underlying active and passive components. For instance, a designer working with discrete components on a printed circuit board may be subjected to the following constraints:

- Limit the component count below 100 elements to achieve a small board area.
- Resistors can be chosen in the range of 1Ω–10MΩ.
- Capacitors can be chosen in the range of 1pF–10,000 F.
- The resistor and capacitor values match to within 1–10%.
- The available (discrete) bipolar junction transistors match to within 20% in their critical parameters.

In contrast, the designer of a CMOS system-on-chip may face the constraints summarized below:

- Avoid using resistors; use as many MOSFET transistors as needed (within reasonable limits, on the order of hundreds to several thousands) to realize the best possible circuit implementation.
- Capacitors can be chosen in the range of 10 fF–100 pF.
- The critical parameters in the MOSFET transistors be made to match to within 1%, but vary by more than 30% for different fabrication runs.
- Capacitors of similar size can match to within 0.1%, but vary by more than 10% for different fabrication runs.

As a consequence of the vastly different constraints that apply to the design of analog circuits in CMOS technology, the resulting practical and preferred circuit architectures differ substantially from the ones that would be used in a printed circuit board design. For example, a discrete voltage amplifier may utilize large AC coupling capacitors to simplify and decouple the biasing of the individual gain stages (see example in Figure 1.18). In contrast, it is typically not possible to use AC coupling techniques (except for very high-frequency designs) in integrated circuits, primarily due to the restriction on maximum capacitor size.

The material covered in this module is primarily concerned with analog integrated circuit design. While this choice does not affect many of the key principles used in the analysis and the discussed circuits, it does affect the architectural choices made in arriving at a practical design. For instance, large AC coupling capacitors are not used throughout the discussion. Also, where appropriate, we will invoke certain assumptions about the typical matching of component parameters in CMOS to eliminate impractical design choices.

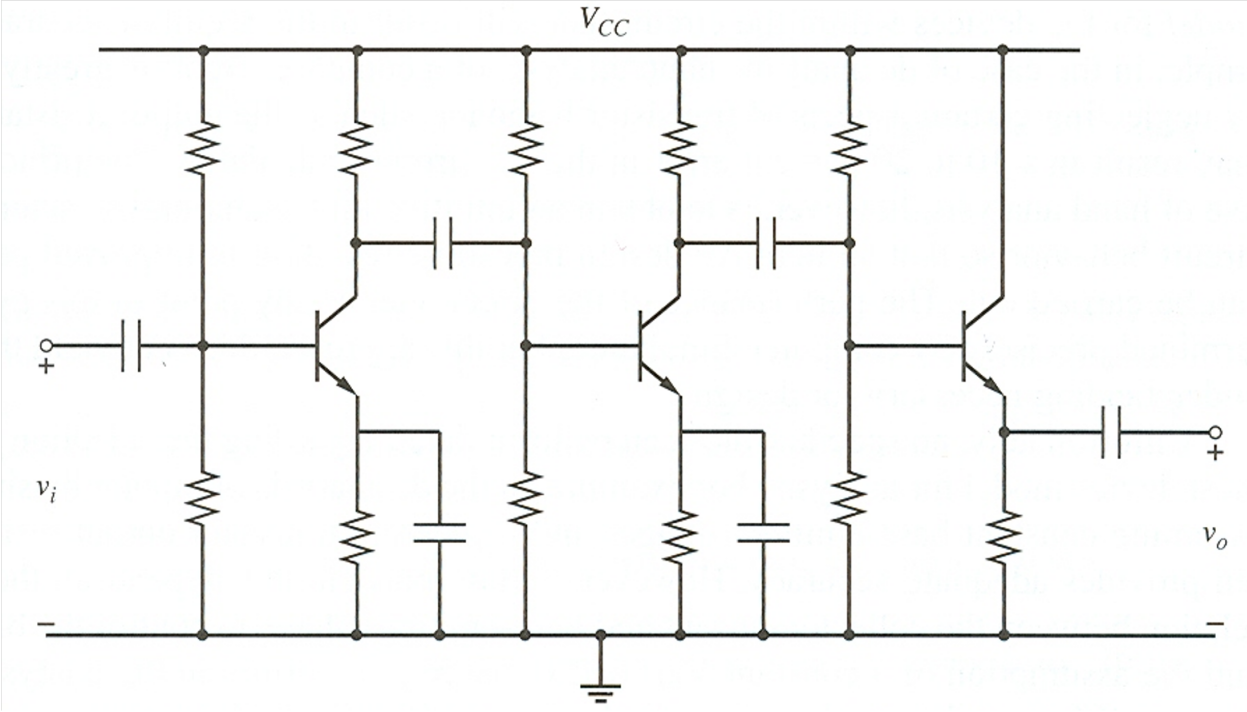


Figure 1.18.: Example of a discrete amplifier circuit using bipolar junction transistors (BJTs).

1.5. Prerequisites and Advanced Material

The reader of this module is expected to be familiar with the basis concepts of linear circuit analysis (see Ulaby and Maharbiz 2013), including

- Passive components (resistors, capacitors)
- Kirchhoff's voltage and current laws (KVL and KCL)
- Independent and dependent voltage and current sources; Thevénin and Norton representation of controlled sources
- Two-port representation of circuits; calculation of port resistances and frequency dependent impedances
- Manipulation of complex variables and numbers
- Phasor analysis and Laplace domain representation of passive circuit elements
- Bode plots

The derivations of device models in this module assume familiarity with basic solid-state physics and electrostatics as treated in introductory texts on solid-state device physics (see Pierret 1996). A few sections of this module are marked with an asterisk (*) to indicate advanced material that may in some cases go beyond the learning goals of an introductory course. These sections can be skipped at the instructor's discretion without affecting the overall flow and context.

1.6. Notation

This module follows the notation for signal variables as standardized by the IEEE. Total signals are composed of the sum of DC quantities and small signals. For example, a total input voltage v_{IN} is the sum of a DC input voltage V_{IN} and a small-signal voltage v_{in}. The notation is summarized below.

- Total quantity has a lowercase variable name and uppercase subscript
- DC quantity has an uppercase variable name and uppercase subscript
- Small-signal quantity has a lowercase variable name and lowercase subscript

1.7. Summary

This chapter offered a brief motivation for the topics covered in this module, which focuses on the analysis and design of elementary amplifier stages in CMOS technology. These elementary stages can be viewed as the "atoms" of analog circuit design and a thorough understanding of the blocks is a necessary prerequisite for the design of advanced analog circuits design, as for instance in the context of large systems-on-chip. At all levels of circuit design, complexity is managed using hierarchical abstraction and model simplification using proper engineering approximations. The unilateral two-port models reviewed in Section 1-3 and used throughout this module, are an example of such abstractions.

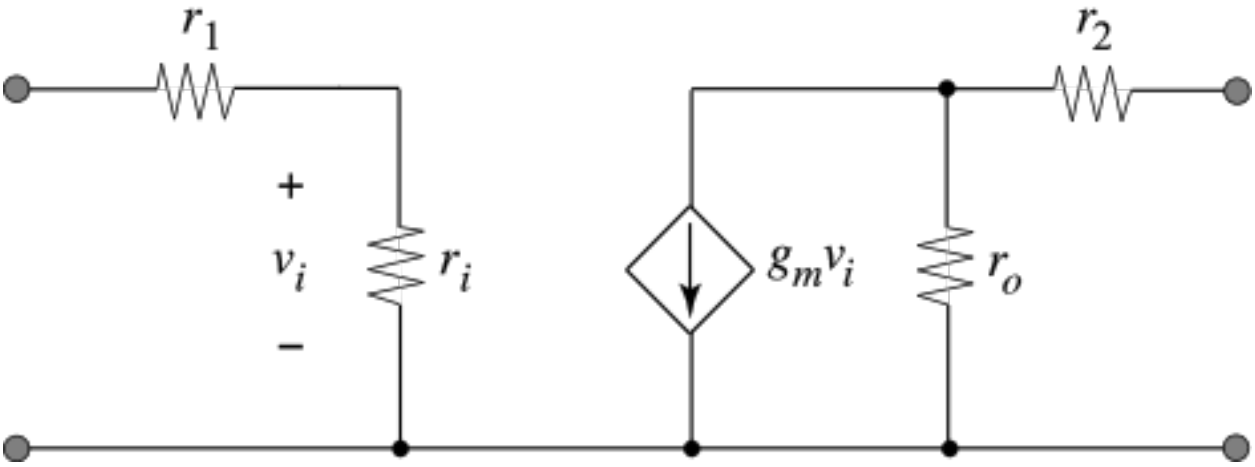


Figure 1.19.: Figure P1-1

1.8. Problems

P1.1 Given the amplifier circuit in Figure 1.19 (a) Find the input and output resistance. (b) Construct an equivalent circuit using a voltage amplifier two-port model and determine all model parameters symbolically. (c) Repeat part (b) for a current amplifier model. (d) Repeat part (b) for a transconductance amplifier model. (e) Repeat part (b) for a transresistance amplifier model.

P1.2 Convince yourself that the circuits of Figure 1.10 and Figure 1.12 are equivalent by showing sym- bolically that both circuits have the same overall voltage gain $A'_v = v_{out} / v_s$.

P1.3 You are given an input voltage source with a source resistance, R_S. (a) Use the unilateral voltage amplifier two-port model found in P1.1 to find the overall voltage gain when the amplifier is driving a load resistor R_L. (b) Specify whether the resistances r_1, r_i, r_o, r_2 in the small signal model should be increased, be decreased, or remain the same to improve the overall voltage gain.

P1.4 You are given an input current source with a source resistance, R_S. (a) Use the unilateral current amplifier two-port model found in P1.1 to find the overall current gain when the amplifier is driving a load resistor R_L. (b) Specify whether the resistances r_1, r_i, r_o, r_2 in the small-signal model should be increased, be decreased, or remain the same to improve the overall current gain.

P1.5 Given the circuit model in Figure 1.20 for an amplifier circuit (with r_i and r_o removed and r_1=r_2=0)

(a) Find the input and output resistance. Note that since this circuit is bilateral, R_S must be considered when computing R_{out}, and R_L must be considered when computing R_{in}.
(b) Construct a two-port model for a unilateral voltage amplifier.
(c) Construct a two-port model for a unilateral current amplifier.
(d) Construct a two-port model for a unilateral transconductance amplifier.
(e) Construct a two-port model for a unilateral transresistance amplifier.

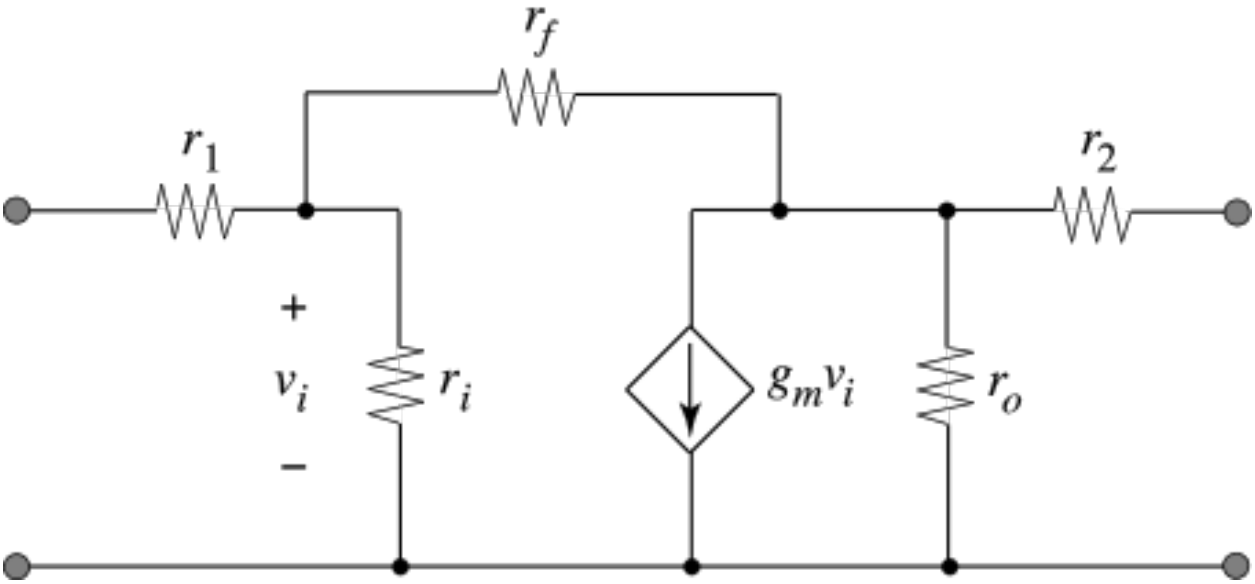


Figure 1.20.: Figure P1-5

P1.6 Consider the two-port model of a voltage amplifier as shown in Figure 1.9(a) with the following parameters: $A_v = 10$, $R_{in} = 5$ kΩ, and $R_{out} = 100$ Ω.

(a) Draw the two-port model for a transresistance amplifier by conversion from the voltage amplifier model.
(b) Draw the two-port model for a transconductance amplifier by conversion from the voltage amplifier model.

(c) Draw the two-port model for a current amplifier by conversion from the voltage amplifier model.

P1.7 Compute the transresistance v_{out}/i_{in} for the circuit of Figure 1.7 using the following parameters: $A_{i1} = 1$, $G_{m2} = 10$ mS, $A_{v3} = 0.8$, $R_{in1} = 50\ \Omega$, $R_{out1} = 500\ \Omega$, $R_{out2} = 1$ kΩ, and $R_{out3} = 100\ \Omega$. Using this result, lump the entire circuit into a single transresistance amplifier as shown in Figure 1.9(d). Draw the resulting model, including R_{in} and R_{out}.

P1.8 Consider the amplifier circuit of Figure 1.13

(a) with $R_1 = 1/g_m = 1$ kΩ and $R_2 = 100$ kΩ and $R_S = R_L = 10$ kΩ. Compute all component values for the bilateral two-port current amplifier model of Figure 1.13

(b) Note that A_{if}, R_{in},and R_{out} can be described as explained in Section 1-3. Similar to A_{if}, A_{ir} is found by short-circuiting the input port and by injecting a test current into the output port. Compare the relative magnitude of A_{if} and A_{ir}.

2. Transfer Characteristic of the Common-Source Voltage Amplifier

Chapter Objectives

- Review the MOSFET device structure and basic operation as described by the square-law model.
- Introduce large- and small-signal analysis techniques using the common-source voltage amplifier as a motivating example.
- Derive a small-signal model for the MOSFET device, consisting of a transconductance and output resistance element.
- Provide a feel for potential inaccuracies and range limitations of simple modeling expressions.

2.1. First-Order MOSFET Model

The device-level derivations of this section assume familiarity with basic solid-state physics and electrostatics. For a ground-up treatment from first principles, the reader is referred to introductory solid-state device material (see Reference 1).

2.1.1. Derivation of I-V Characteristics

The basic structure of an **enhancement mode n-channel** MOSFET is shown in Figure 2.1(a). It consists of a lightly doped p-substrate (**bulk**), two heavily doped n-type regions (**source** and **drain**) and a conductive gate electrode that is isolated from the substrate using a thin silicon dioxide layer of thickness t_{ox}. Other important geometry parameters of this device include the **channel length** L (distance between the source and drain) and the **channel width** W.

As we shall see, the name "n-channel" stems from the fact that this device conducts current by forming an n-type layer underneath the gate. A **p-channel** device can be constructed similarly using an n-type bulk and p-type source/drain regions. The differentiating details between n- and p-channel devices are summarized in Section 2-1-2. For the time being, we will use the n-channel device to discuss the basic principles.

In order to study the electrical behavior of a MOSFET, it is useful to define a schematic symbol and conventions for electrical variables as shown in Figure 2.1(b) . The variables V_{GS}, V_{DS}, and V_{BS} describe the voltages between the respective terminals using the commonly used ordered subscript convention $V_{XY} = V_X - V_Y$. The current flowing into the drain node is labeled I_D.

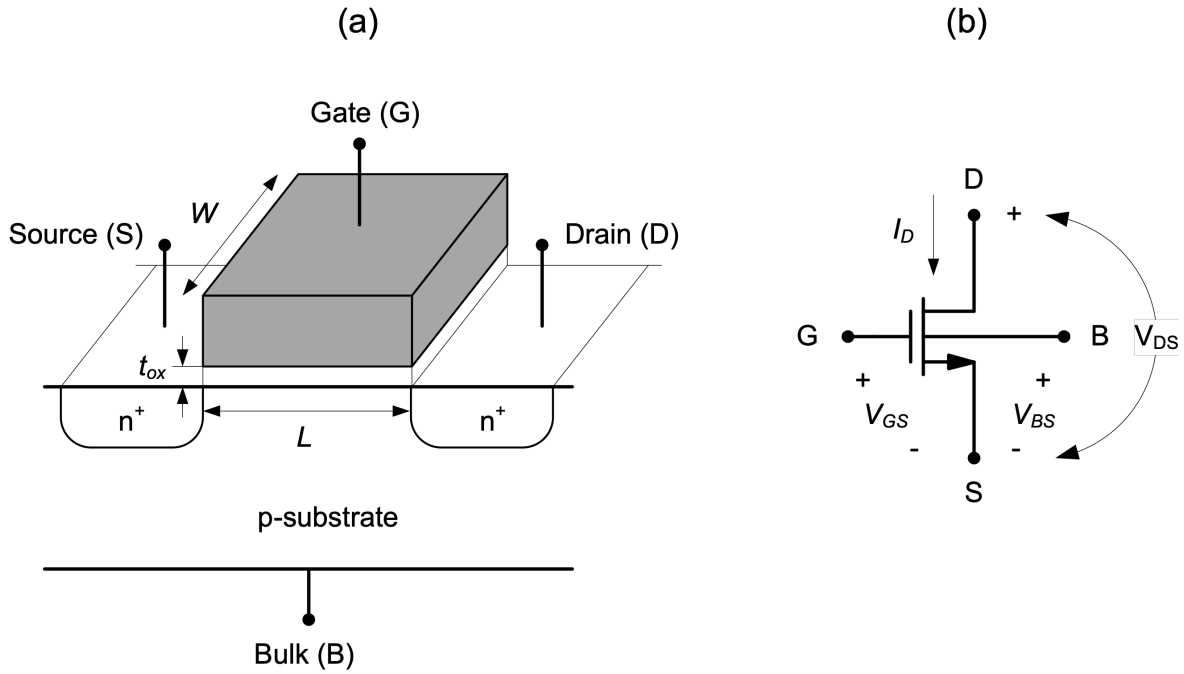


Figure 2.1.: (a) Cross-section of an n-channel MOSFET.

It is important to note that the MOSFET device considered here is perfectly symmetric; i.e., the drain and source terminal labels can be interchanged. It is a common convention to assign the source to the lower potential of these two terminals, since this terminal is the source of electrons that enable the flow of current. We will see later that this convention, together with the arrow that marks the source (and the direction of current flow), provides useful intuition when reading a larger circuit schematic.

We now begin our analysis of the MOSFET device by considering the condition shown in Figure 2.2(a), where the bulk and source are connected to a reference potential (GND), $V_{GS} = 0\ V$ and $V_{DS} = 0\ V$. Under this condition, the drain and source terminals are isolated by two reverse-biased pn-junctions and their **depletion regions**, which prevent any significant flow of current. Applying a positive voltage at the drain ($V_{DS} > 0\ V$) increases the reverse-bias at the drain-bulk junction and will only increase the width of the depletion region at the drain, while $I_D = 0$ is still maintained (to first-order).

Consider now $V_{GS} = 0$ as shown in Figure 2.2(b). This positive voltage at the gate attracts electrons from the source. With increasing V_{GS}, a larger amount of electrons is supplied by the source, and ultimately, a so-called **inversion layer** forms underneath the gate. The voltage V_{GS} at which a significant number of mobile electrons underneath the gate become available is called the **threshold voltage** of the transistor, or V_t. In order to differentiate the threshold voltages and other device parameters of n-and p-channel devices, we will utilize the subscripts n and p throughout this module. E.g., we denote the threshold voltage for n-channels and p-channels as V_{Tn} and V_{Tp}., respectively.

With the inversion layer under the gate, the drain and source regions are now "connected" through a conductive path and any voltage between these terminals ($V_{GS} > 0$) will result in a flow

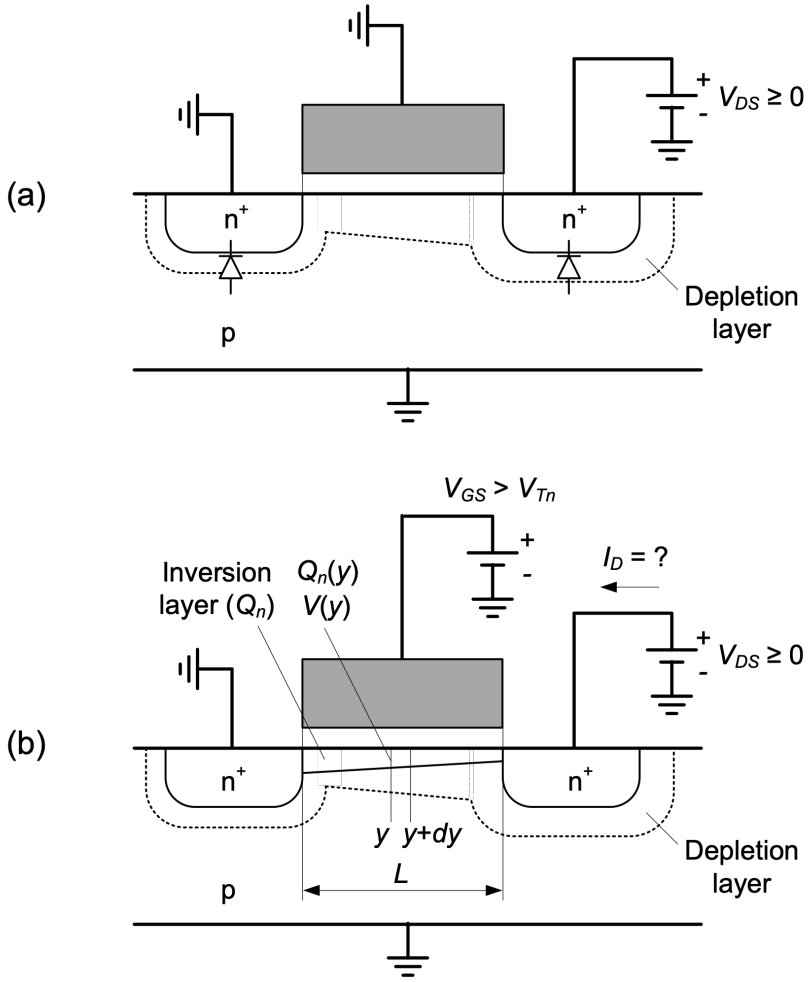


Figure 2.2.: (a) n-channel MOSFET with $V_{GS} = 0$, (b) $V_{GS} > V_{Tn}$,

of drain current. How can we calculate this current? In order to answer this question, the following approximations are useful:

1. The current primarily depends on the number of mobile electrons in the channel times their velocity.

2. The number of mobile electrons in the channel is set by the vertical electric field from the gate to the conductive channel (**gradual channel approximation**).

3. The threshold voltage is constant along the channel; this assumption neglects the so-called body effect.

4. The velocity of the electrons traveling from the source to the drain is proportional to the lateral electric field in the channel.

Figure 2.2(b) establishes relevant variables for further analysis. The auxiliary variable y ranges from 0 to L and is used to express electrical quantities as a function of the distance from the source. The inversion layer charge density (per unit area) and voltage at position y in the channel are denoted as $Q_n(y)$ and V_y,respectively. With these conventions in place, we can translate the above-listed assumptions into the following equations:

$$I_D = -W \cdot Q_n \cdot v(y) \tag{2.1}$$

$$Q_n(y) = -C_{ox} \cdot (V_{GS} - V(y) - V_{Tn}) \tag{2.2}$$

$$v(y) = -\mu_n \cdot E_y \tag{2.3}$$

In these expressions, v is the velocity of the carriers, C_{ox} is the **gate capacitance** per unit area (between the gate electrode and the conductive channel). The term μ_n is called **mobility**, and it relates the drift velocity of the carriers to the local electric field.

As indicated in Equation 2.2, the mobile charge density at coordinate y depends on the local potential, since the voltage across the oxide is given by $V_{GS} - V(y)$. An inversion layer is present at any location under the gate where this voltage difference is larger than the threshold (V_{Tn}). Assuming that the inversion layer extends from source to drain as drawn in Figure 2.2(b), we have $V(L) = V_{DS}$ and $Q_n(L) = -C_{ox}$ $(V_{GS} - V_{Tn} - V_{DS})$. This implies that V_{DS} cannotexceed $V_{GS} - V_{Tn}$ for an inversion layer that extends across the entire channel. For the time being, we will solve for the drain current for this condition and later extend the obtained result for the case of $V_{DS} = V_{GS} - V_{Tn}$.

Now, by combining Equation 2.1 through Equation 2.3 and noting that the electric field is given by $E(y) = -dV(y)/dy$, we can write

$$I_D = \mu_n C_{ox} W \cdot (V_{GS} - V_y - V_{Tn}) \frac{dV_y}{dy} \tag{2.4}$$

This result describes the current density profile along the channel. The terminal current, I_D, can be found by separating the variables and integrating along the direction of y

$$\int_0^L I_D \, dy = \mu_n C_{ox} W \int_0^{V_{DS}} (V_{GS} - V(y) - V_{Tn}) \, dV \tag{2.5}$$

which yields a closed-form solution for the drain current

$$\boxed{I_D = \mu_n C_{ox} \frac{W}{L} \left(V_{GS} - V_{Tn} - \frac{V_{DS}}{2} \right) V_{DS}} \tag{2.6}$$

Note that this expression is valid for $V_{DS} < V_{GS} - V_{Tn}$, as assumed above. In order to extend the obtained result for $V_{DS} = V_{GS} - V_{Tn}$, we continue by inspecting the shape of the inversion layer for various V_{DS} (see Figure 2.3). For $V_{DS} = 0$ V [case (a)], no current flows and $V(y) = 0$ for all y. Provided that $V_{GS} > V_{Tn}$, a uniform inversion layer exists underneath the gate. For small $V_{DS} > 0$, a current flows in the inversion layer, which causes increasing $V(y)$ and decreasing inversion layer charge along the channel. As V_{DS} approaches V_{GS} - V_{Tn}, $Q_n(L)$ approaches zero with a point of diminishing charge at the drain. This effect is called **pinch-off**.

What happens when we increase V_{DS} beyond the point of pinch-off? Further analysis based on solving the two-dimensional **Poisson Equation** at the drain predicts that the pinch-off point will move from L to L–ΔL, where ΔL is small relative to L. Even though no inversion layer exists in the region from L–ΔL to L, the device still conducts current. The charges arriving at $y = L$–ΔL are being swept to the drain by the electric field present in the depletion region of the surrounding pn junction.

To first-order, and neglecting the small change in channel length ΔL, the current becomes independent of V_{DS} and is approximately given by the current at the onset of pinch-off, i.e., at $V_{DS} = V_{GS} - V_{Tn}$. Substituting this condition into Equation 2.6, we obtain

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

for $V_{DS} \quad V_{GS} - V_{Tn}$.

Equation 2.6 and Equation 2.7 are plotted in Figure 2.4 as a function of V_{DS} and some fixed $V_{GS} > V_{Tn}$. The operating region for $V_{DS} < V_{GS} - V_{Tn}$ is commonly called the **triode region**. This name stems from the direct dependence of the drain current on the drain-source voltage, which is qualitatively similar to the behavior of vacuum tube "triodes." The region $V_{DS} \quad V_{GS} - V_{Tn}$ is called the **saturation region** due to the saturation in current at large V_{DS}. In this region, the device operates essentially like a current source; the current is (to first-order) independent of the applied V_{DS} and $I_D = I_{Dsat} =$ constant. The quantity $V_{GS} - V_{Tn}$ is often called **gate overdrive**.

The drain-source voltage at which the drain current saturates is called V_{DSat}. From the above first-order analysis, it is clear that $V_{DSat} = V_{GS} - V_{Tn}$. Nonetheless, it is useful to distinguish between these two quantities, because they may differ significantly when a more elaborate device model is used. V_{DSat} is generally not exactly equal to $V_{GS} - V_{Tn}$ when second-order effects, for example related to small geometries and modern device structures, are considered.

From a circuit perspective, the device's behavior in the triode region is similar to a resistor: the current increases monotonically with increasing terminal voltage. Even though the dependence of I_D on V_{DS} is nonlinear (as seen from Equation 2.6), it is sometimes useful to approximate the

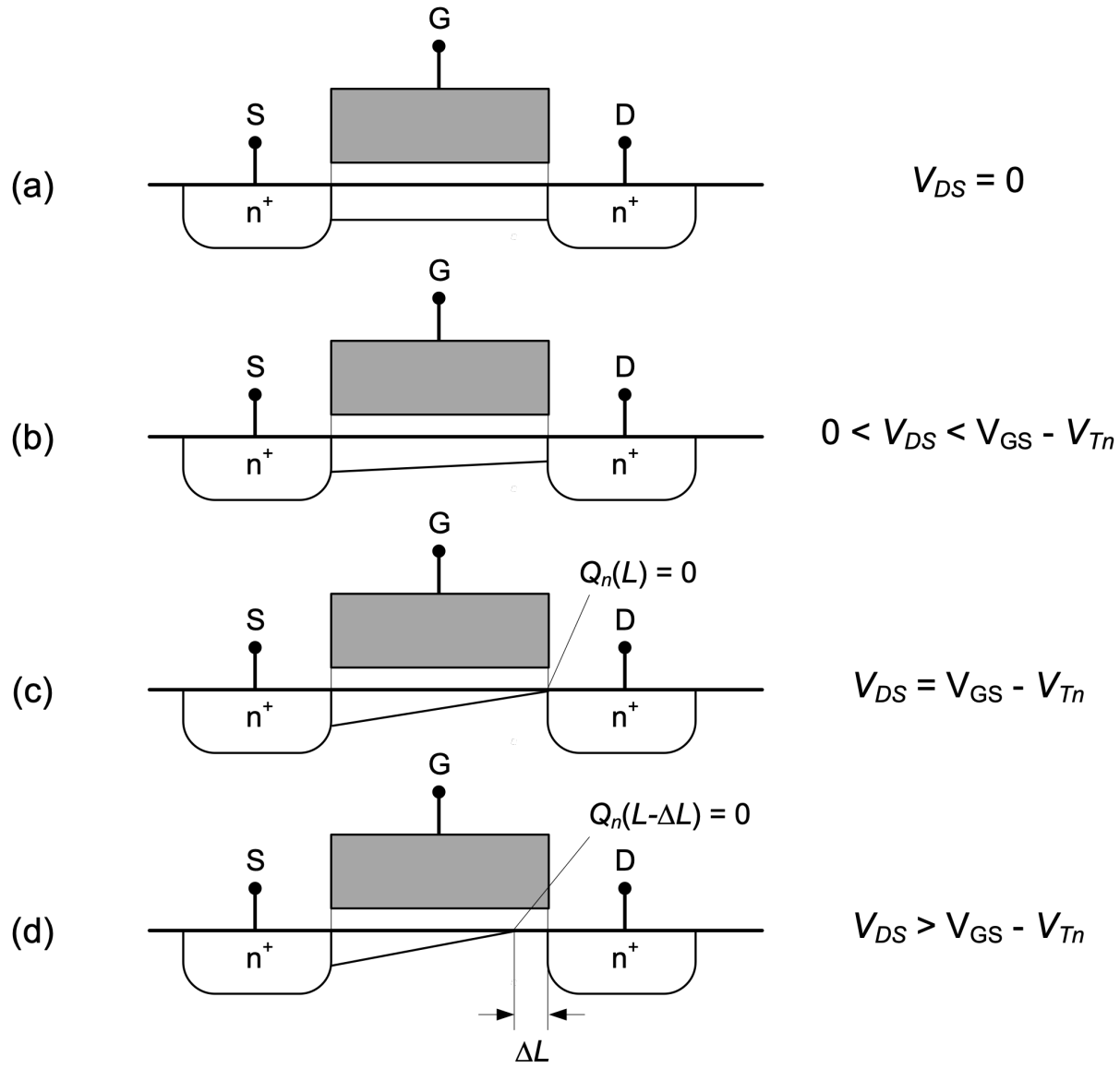


Figure 2.3.: Channel profile for varying V_{DS}.

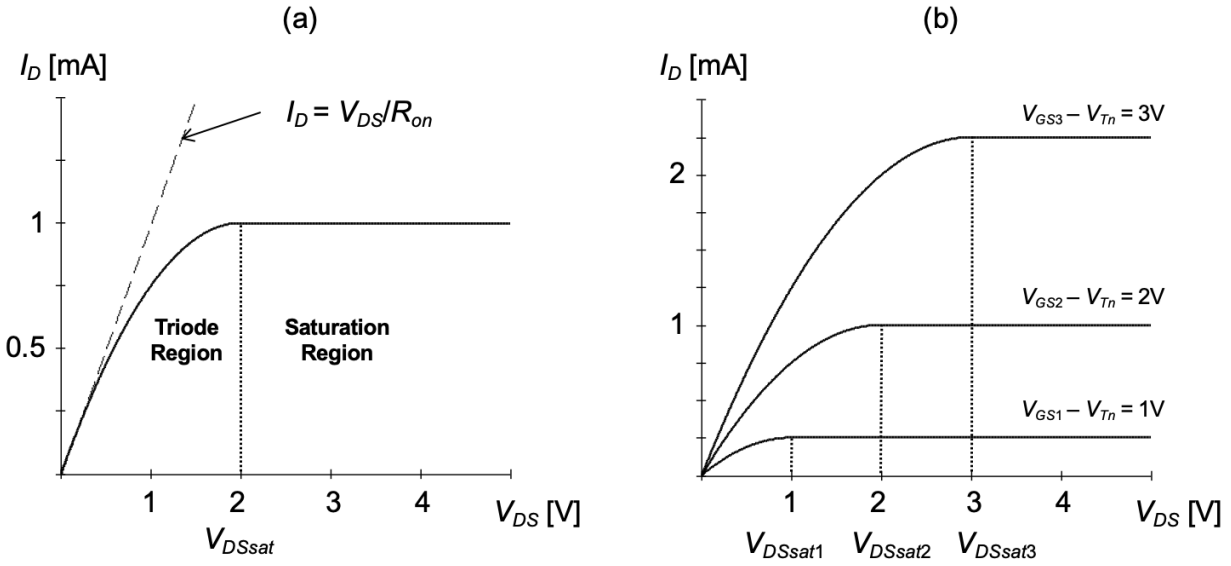


Figure 2.4.: (a) n-channel I-V characteristic for a fixed value of V_{GS} – V_{Tn}= 2V. (b) I-V plots with varying V_{GS} – V_{Tn} (drain characteristic). Parameters: $\mu Cox = 50\ \mu A/V^2$ and $W/L = 10$.

characteristic using a linear I-V law, shown as a dashed line in Figure 2.4(a). For $V_{DS} \ll V_{GS}$ – V_{Tn} , we can approximate Equation 2.6 as

$$I_D \equiv \mu_n C_{ox} \frac{W}{L}(V_{GS} - V_{Tn})V_{DS} \tag{2.8}$$

Under this approximation, I_D depends linearly on V_{DS}, and we can define the so-called **on-resistance** of the device as

$$R_{on} = \frac{V_{DS}}{I_D} \equiv \frac{1}{\mu_n C_{ox} \frac{W}{L}(V_{GS} - V_{Tn})} \tag{2.9}$$

It is interesting to interpret the dependencies in this expression using basic intuition. Increasing the aspect ratio W/L decreases R_{on} since the conductive path becomes shorter and/or wider; this is a basic property of any conductor. The on-resistance also decreases with increasing C_{ox} and V_{GS} – V_{Tn}; this is because the inversion charge increases with these quantities ($Q = CV$). Larger mobility (μ_n) means that the carriers travel faster for the same applied voltages (electric field). This increases the current (charge per unit of time) and therefore also results in smaller R_{on}.

As seen from Equation 2.7, the magnitude of the drain current in saturation depends on the square of the gate overdrive V_{GS} – V_{Tn}. This is further illustrated in Figure 2.4(b), which shows I-V plots for increasing multiples of V_{GS1} – V_{Tn} = 1V. Doubling and tripling the gate overdrive increases the saturation current by factors of four and nine, respectively. Note that R_{on} is reduced only by factors of two and three in these cases, respectively.

The plot in Figure 2.4(b) is often called the **drain characteristic**, because the drain-source voltage (as opposed to the gate-source voltage, which is included as a parameter on the curves), is swept

along the x-axis. Alternatively, the term **output characteristic** is sometimes used, primarily because V_{DS} can often be viewed as the output port voltage of the device; we will see this in the example discussed in Section 2-2.

Another commonly used characterization plot for MOSFETs is the so-called **transfer characteristic**, which shows the drain current as a function of V_{GS} for a fixed value of V_{DS}. If V_{DS} is chosen large enough such that the device operates in the saturation region for all applied V_{GS}, I_D follows from Equation 2.7 and the plot is shaped like a parabola as drawn in Figure 2.5.

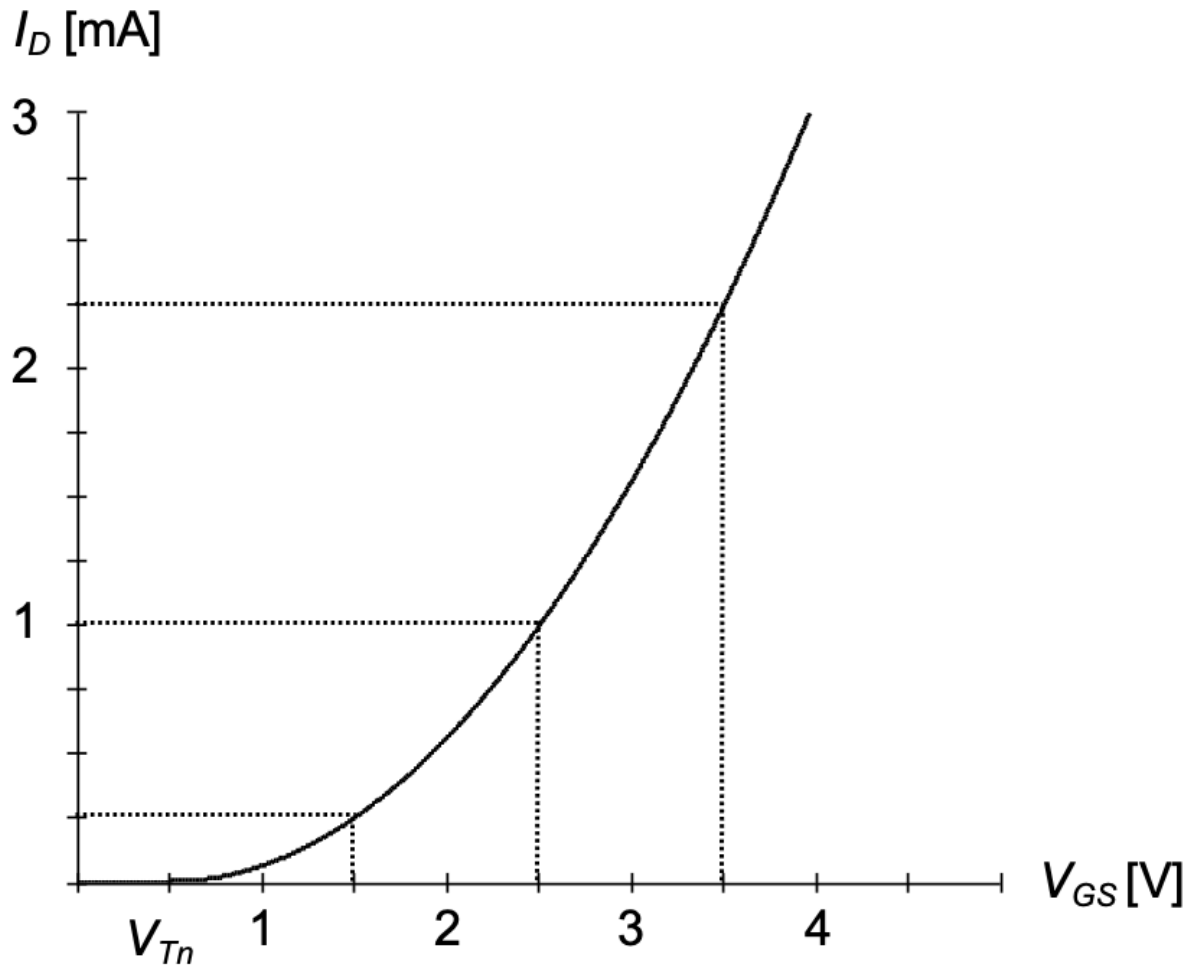


Figure 2.5.: Plot of n-channel drain current as a function of V_{GS} (transfer characteristic). Parameters: $V_{DS} = 5$ V, $\mu C_{ox} = 50$ A/V^2,$V_{Tn} = 0.5$ V, and $W/L = 10$.

Table 2.1 summarizes the first-order MOSFET I-V relationships that were discussed in this section. This set of equations (and extended versions thereof) is often called the **square-law model** since one of its primary features is the quadratic dependence of the saturation current on $V_{GS} - V_{Tn}$. When working with this device model, it is important to remember that it predicts the behavior of real MOSFETs only with limited accuracy. This is primarily so because we have made several simplifications in the model's derivation. The most significant shortcom ings that result from these assumptions can be summarized as follows:

1. In reality, the saturation current has a weak dependence on V_{DS}. This is primarily due to a shortening of the channel length (ΔL) with increasing V_{DS} and also due to the drain voltage dependence of the mobile charge in the channel. We will address this issue in Section 2-2.
2. For transistors built in modern technologies, several second-order effects related to small geometries and large electric fields become significant. This typically results in a saturation current law exponent that is less than two, and $V_{DSat} < V_{GS} - V_{Tn}$. In addition, the drain current does not scale strictly proportional to $1/L$ and the threshold voltage is not constant, but a function of the drain voltage.
3. For $V_{GS} < V_{Tn}$ the device is not completely off, but carries a small current that exponentially depends on V_{GS}. This operating region is called the **sub-threshold region**.
4. For small values of $V_{GS} < V_{Tn}$, on the order of a few tens of millivolts, the region underneath the gate is only moderately inverted, and the square law model tends to predict the drain current with poor accuracy.

Table 2.1.: First-order MOSFET model summary

	"ON" $V_{GS} \geq V_{Tn}$	"OFF" $V_{GS} < V_{Tn}$
$V_{DS} < V_{DSat}$	$I_D = \mu_n C_{ox} \frac{W}{L}(V_{GS} - V_{Tn} - \frac{V_{DS}}{2})V_{DS}$	$I_D = 0$
$V_{DS} \geq V_{DSat}$	$I_D = \frac{1}{2}\mu_n C_{ox} \frac{W}{L}(V_{GS} - V_{Tn})^2$	$I_D = 0$

Despite these shortcomings, the first-order MOSFET model possesses many of the critical features needed to study the fundamentals of analog circuit design. Many of the second-order effects not featured in the basic model can be treated using advanced device physics and often result in a high-complexity model that is unsuitable for hand-calculations and intuition building.

Within the range of circuits treated in this mod ule, we typically begin by applying the first-order model. Then, only when the circuit appears to be sensitive to second-order dependencies not covered by this model, we will look for extensions. A treatment in this fashion has the advantage that the reader can develop a feel for where and when modeling extensions and parameter accuracy are critical.

In general, the tradeoff between modeling accuracy and complexity is a recurring theme at all levels of analog circuit design; the issue is not limited to the introductory material covered in this module. More accurate models can always be generated at the expense of complexity and time. An experienced analog designer will often use the simplest possible model that will predict the behavior of his or her circuit with sufficient (but not perfect) accuracy. This also implies that analog circuit designers must always be on the lookout for model inadequacies. We will encounter and discuss situations where either model expansions or critical insight on modeling accuracy are needed throughout this module.

2.1.2. P-Channel MOSFET

The n-channel MOSFET discussed so far conducts current through an electron inversion layer in a p-type bulk. Similarly, we can construct a **p-channel** device that operates based on forming an

inversion layer of holes in an n-type bulk. The structure of such a MOSFET, which consists of $p+$ source and drain regions in an n-type bulk, is shown in Figure 2.6. In many process technologies, the n-type bulk region is formed by creating an n-type well (**n-well**) in the p-type substrate that is used to form n-channels. Such a technology is called an **n-well technology**. In general, a technology that offers both n- and p-channel devices is called **CMOS** technology, where CMOS stands for Complementary Metal-Oxide-Semiconductor.

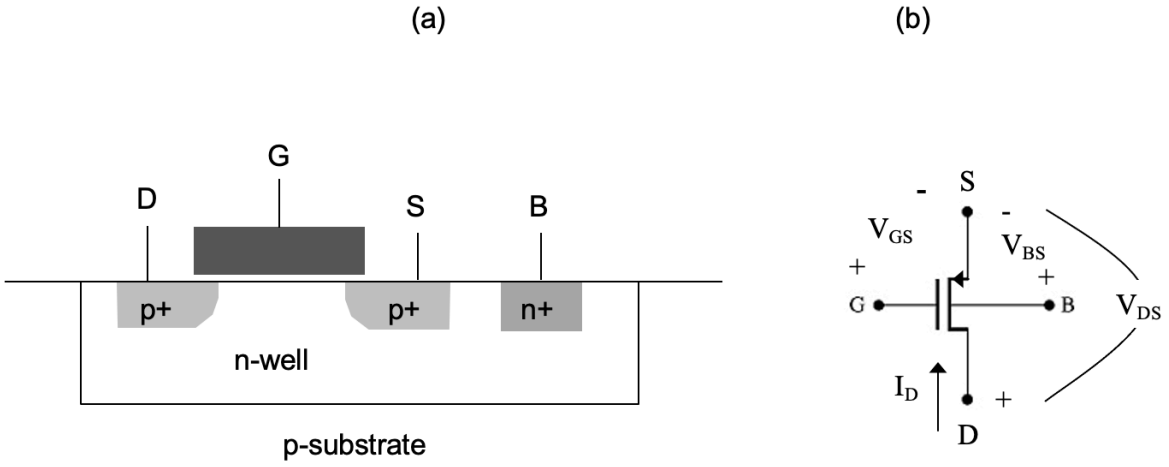


Figure 2.6.: (a) Cross-section of a p-channel MOSFET. (b) Schematic symbol.)

The drain current equations for a p-channel MOSFET can be derived using exactly the same approach as used for the n-channel device since the basic physics are the same. For a p-channel device, the gate must be made negative with respect to the p-type source in order to form an inversion layer of holes; the threshold voltage V_{Tp} is therefore typically negative. Since holes drift across the channel from the source to the drain in the p-channel MOSFET, the drain voltage must be negative with respect to the source, and the drain current (defined as flowing into the drain terminal) is negative. Therefore, in the on-state of the transistor, V_{GS} and V_{DS} are negative quantities, and the source lies at the highest potential among the four terminals. The drain current for a p-channel in saturation, i.e., $V_{GS} < V_{DS}$ and V_{DS} $V_{GS} - V_{Tp}$, is given by

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

A practical problem for the circuit designer is to keep track of the minus signs and negative quantities in the p-channel equations. A solution to this issue is to "think positive," and work with the physically intuitive positive quantities V_{SG} (instead of V_{GS}), V_{SD} (instead of V_{DS}) in all hand calculations. Following this approach, we can rewrite Equation 2.10 as

$$-I_D = \frac{1}{2}\mu_p C_{ox}\frac{W}{L}(V_{GS} - V_{Tp})^2 \tag{2.11}$$

Note that the right-hand side of this equation yields a positive number. The minus sign included on the left-hand side remains necessary because I_D, as defined in Figure 2.6(b), is a negative quantity.

2.1.3. Standard Technology Parameters

For use throughout this module, it is convenient to define standard MOSFET parameter values as given in Table 2.2. The chosen values are representative of a CMOS technology with a minimum channel length, or **feature size** of 1 μm. As we learn more about the behavior of MOSFETs in later sections, this list of parameters will grow and we will augment it as needed.

In the context of defining these parameters, it is important to make a clear distinction between **technology parameters** and **design parameters**. Technology parameters are typically fixed in the sense that a circuit designer cannot alter their values. For instance, the mobility in a MOSFET depends on how the transistor is made, and the underlying recipe remains unchanged and will be reused for an extended time to manufacture a large variety and quantity of integrated circuits. In most modern CMOS technologies, the width and length of a MOSFET remain as the only parameters that the circuit designer can choose (within appropriate limits) to alter the device's electrical behavior.

In determining the transistor geometries, the designer will usually work with electrical variables and parameters that describe the circuit and its functionality, for example in terms of currents and voltages. From these electrical descriptions and specifications, the widths and lengths of the transistors are then calculated, and sometimes adjusted via an iterative process. In this task, intermediate electrical parameters, as for instance the gate overdrive of a MOSFET, are also legitimately viewed as parameters that are under the control of the circuit designer.

Example 2-1: P-Channel Drain Current Caculation

A p-channel transistor is operated with the following terminal voltages relative to ground: $V_G = 2.5\ V$, $V_S = V_B = 5\ V$, $V_D = 1\ V$. Calculate the drain current I_D using the standard technology parameters given in Table 2.2 and assuming $W/L = 5$.

SOLUTION

From the given terminal voltages, we find $V_{SG} = 5\ V – 2.5\ V = 2.5\ V$ and $V_{SD} = 5\ V – 1\ V = 4\ V$. Since $V_{SG} > V_{Tp}$, the transistor is on, and since $V_{SD} > V_{SG} + V_{Tp} = 2.5\ V – 0.5\ V = 2\ V$, it operates in saturation. Therefore, using Equation 2.11 we find

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

$$-I_D = \frac{1}{2}\cdot 25\frac{\mu A}{V^2}\cdot 5\cdot (2.5V - 0.5V)^2 = 250\mu A$$

$$-I_D = -250\mu A$$

Table 2.2.: Standard technology parameters for the first-order MOSFET model

Parameter	**n-channel MOSFET**	**p-channel MOSFET**
Threshold voltage	$V_{Tn} = 0.5$V	V_{Tp} = -0.5V
Transconductance parameter	$\mu_n C_{ox} = 50\mu A/V^2$	$\mu_p C_{ox} = 25\mu A/V^2$

2.2. Building a Common-Source Voltage Amplifier

We will now utilize our first-order understanding of MOSFETs to construct a basic voltage amplifier. We begin by noting that the drain current in all regions of operation can be controlled by varying the gate-source voltage. One way to utilize this effect to build a voltage amplifier is to apply the input such that it controls V_{GS}. An output voltage can then be generated by letting the drain current flow through a resistor, as shown in Figure 2.7(a). The top terminal of the resistor is connected to a **supply voltage**, V_{DD}. In this scheme, a larger V_{IN} causes the drain current to increase and V_{OUT} to decrease, since a larger V_{IN} makes the transistor a "better conductor" (more inversion charge), which forces the voltage at the output port closer to ground. This type of circuit is therefore categorized as an **inverting amplifier**. Furthermore, this transistor stage is called a **common-source amplifier**, since the source terminal of the MOSFET is common to the input and output ports of the circuit.

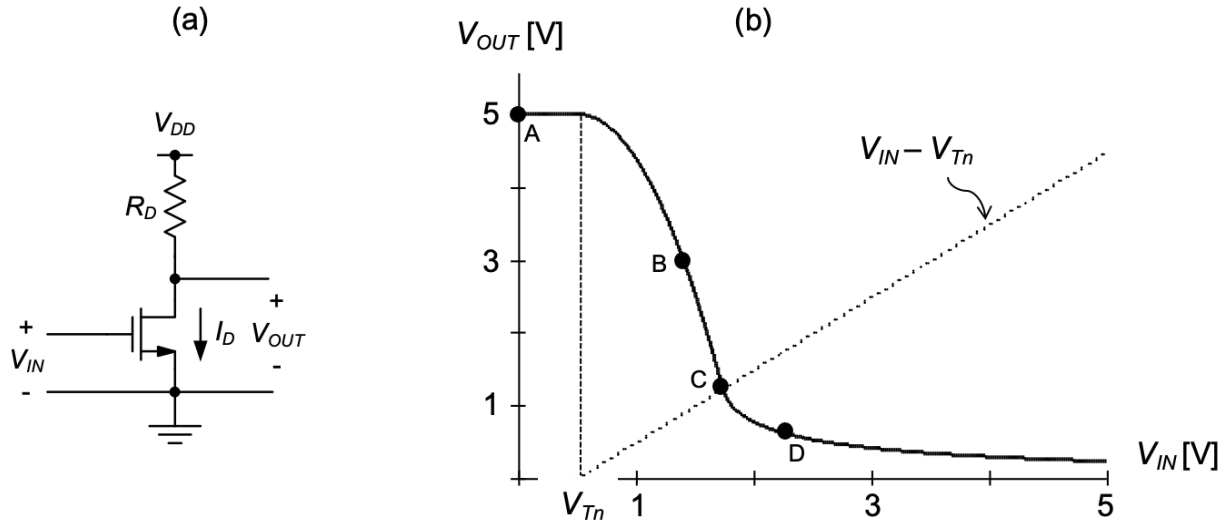


Figure 2.7.: (a) Basic common-source amplifier schematic. (b) Voltage transfer characteristic for $V_{DD} = 5\ V$, $R_D = 5\ k\Omega$, and $W/L = 20$.

2.2.1. Voltage Transfer Characteristic

In order to derive the voltage transfer characteristic of the circuit (V_{OUT} as a function of V_{IN}) we begin by applying Kirchhoff's laws at the output node. This yields

$$V_{OUT} = V_{DD} - I_D R_D \tag{2.12}$$

The drain current ID in this expression depends on V_{GS} and V_{DS} of the transistor, as described in Table 2.1. Given the structure of the circuit in Figure 2.7(a), we note that $V_{GS} = V_{IN}$ and $V_{DS} = V_{OUT}$. Using this information, we can construct a piecewise function that relates the input and output voltages of the circuit. For this derivation, imagine that we sweep V_{IN} from 0 V to the supply voltage, V_{DD}.

First, we note that for $V_{IN} = V_{GS} < V_{Tn}$, no current flows in the transistor; this implies $V_{OUT} = V_{DD}$. This behavior is shown in Figure 2.7(b) as a horizontal line for the input voltage range 0 $V_{IN} < V_{Tn}$, between points A and the vertical line at V_{Tn}. As V_{IN} increases to values greater than or equal to V_{Tn}, the transistor conducts current, and V_{OUT} must be less than V_{DD}. In order to calculate how V_{OUT} changes as a function of V_{IN}, we must first determine the transistor's region of operation. As we increase V_{IN} above V_{Tn}, does the MOSFET operate in saturation or in the triode region?

To answer this question, we must determine if V_{DS} is smaller or larger than V_{GS} - V_{Tn}. For V_{IN} just above V_{Tn}, $V_{GS} - V_{Tn}$ is smaller than V_{DS}, which is still close to V_{DD} at the onset of current conduction. Therefore, the device must initially operate in saturation as we transition from the "OFF" state of the transistor into the region where $I_D > 0$. Under this condition, the output voltage is given by

$$V_{OUT} = V_{DD} - R_D \cdot \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{IN} - V_{Tn})^2 \tag{2.13}$$

and the voltage transfer characteristic shows a drop that is quadratic in V_{IN} as seen in Figure 2.7(b).

As we continue to increase V_{IN}, V_{GS} - V_{Tn} also increases while V_{OUT} continues to decrease. At a sufficiently large V_{IN}, V_{DS} can approach V_{GS} - V_{Tn} and the condition for current saturation may no longer hold; the device then transitions into the triode region. The input voltage at which this transition occurs (point C in Figure 2.7(b)) can be computed by setting the right-hand side of Equation 2.13 equal to $V_{IN} - V_{Tn}$, and solving for V_{IN}. It is interesting to note that graphically, point C can be found through the intersection of the voltage transfer characteristic with the line $V_{IN} - V_{Tn}$. The intersect corresponds to the point where $V_{OUT} = V_{DS} = V_{IN} - V_{Tn} = V_{GS}$ - V_{Tn}, i.e., the transition point between saturation and triode for the MOSFET.

For the region where the MOSFET operates in the triode region, we have

$$V_{OUT} = V_{DD} - R_D \cdot \mu_n C_{ox}\frac{W}{L}(V_{IN} - V_{Tn} - \frac{V_{OUT}}{2})V_{OUT} \tag{2.14}$$

Unfortunately, solving this expression for V_{OUT} yields an unwieldy square-root expression that is best analyzed graphically. As we can see from the plot in Figure 2.7(b), the most important feature here is that the slope of the voltage transfer characteristic diminishes for large V_{IN}; i.e., the slope of the curve at point D is smaller than the slope at point C. Qualitatively, this can be explained by viewing the MOSFET as a resistor, whose value continues to decrease with V_{IN}. For very large V_{IN}, the output voltage must asymptotically approach 0 V. This can be shown by approximating the MOSFET by its on-resistance for small V_{DS} as given by Equation 2.9. The output voltage in the vicinity of point D can then be expressed by considering the resistive voltage divider formed by R_D and R_{on}.

$$V_{OUT} \cong \frac{V_{DD} \cdot R_{on}}{R_D + R_{on}} = \frac{V_{DD}}{1 + R_D \cdot \mu_n C_{ox}\frac{W}{L}(V_{IN} - V_{Tn})} \tag{2.15}$$

This result confirms that for large input voltages, V_{OUT} will asymptotically approach zero.

Example 2-2: Voltage Transfer Calculations for a Common-Source Amplifier

2. Transfer Characteristic of the Common-Source Voltage Amplifier

Consider the circuit of Figure 2.7(a) with the following parameters: $V_{DD} = 5\ V$, $R_D = 10\ k\Omega$.

a. Using the standard technology parameters of Table 2.2, calculate the required aspect ratio W/L such that $V_{OUT} = 2.5\ V$ for $V_{IN} = 1\ V$.

b. Assuming W/L =10, calculate the input voltage V_{IN} that yields $V_{OUT} = 2.5\ V$.

SOLUTION

a. As a first step, we can calculate the drain current that results in $V_{OUT} = 2.5\ V$ using Equation 2.12.

$$V_{OUT} = V_{DD} - I_D R_D$$
$$2.5V = 5V - (I_D \cdot 10k\Omega)$$
$$I_D = 250A$$

Since V_{GS} – V_{Tn} = 0.5 V < V_{DS} = 2.5 V, we know that the device must operate in the saturation region. Therefore,

$$I_D = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{GS} - V_{Tn})^2$$
$$250A = \frac{1}{2}\cdot 50\frac{A}{V^2}\cdot\frac{W}{L}\cdot(1V - 0.5V)^2$$

and solving for the aspect ratio yields $W/L = 40$. Note that this answer can also be found by direct evaluation of Equation 2.13, without computing I_D initially.

b. since V_{IN} is unknown in this part of the problem, we cannot immediately determine the operating region of the MOSFET. In such a situation, it is necessary to guess the operating region, and later test whether the guess was correct. Let us begin by assuming that the device operates in saturation. We can then write

$$I_D = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{GS} - V_{Tn})^2$$
$$250A = \frac{1}{2}\cdot 50\frac{A}{V^2}\cdot 10\cdot(V_{GS} - 0.5V)^2$$

The two solutions to this equation are $V_{GS1} = 1.5\ V$, and V_{GS2} = –0.5 V. Since we know that the device is off for $V_{GS} < V_{Tn}$, it is clear that V_{GS2} is a non-physical solution that must be discarded. For the obtained V_{GS1}, we must now verify that the device operates in saturation, as initially assumed. It is straightforward to see that this is indeed the case since V_{GS1} – $V_{Tn} = 1\ V < V_{DS} =$ 2.5 V. Therefore, the final answer to this problem is V_{IN} = 1.5 V.

If we had initially guessed that the device operates in the triode region, we would write

$$I_D = \mu_n C_{ox}\frac{W}{L}(V_{GS} - V_{Tn} - \frac{V_{DS}}{2})V_{DS}$$
$$250A = 50\frac{A}{V^2}\cdot 10\cdot(V_{GS} - 0.5V - \frac{2.5V}{2})\cdot 2.5V$$

- The two solution to this equation is $V_{GS} = 1.95\ V$. Since $V_{GS} - V_{Tn} = 1.45\ V < V_{DS} = 2.5$, we see that the obtained result contradicts the assumed operation in triode. Therefore, the next logical step would be to evaluate the saturation equation, as already done above.

2.2.2. Load Line Analysis

A generally useful tool for graphical analysis in electronic circuits is the so-called **load line analysis**. The basis for such an analysis in the context of our circuit is the fact that the current flowing through the transistor (I_D) is equal to the current flowing through the resistor (which is viewed in this context as the load of the circuit). Therefore, if we draw the I-V characteristics of the MOSFET and R_D in one diagram, valid output voltages lie at the intersection of the two curves (equal current). This is further illustrated in Figure 2.8. The load line equation in this plot follows from solving Equation 2.12 for I_D and is given by

$$I_D = \frac{V_{DD} - V_{OUT}}{R_D} \tag{2.16}$$

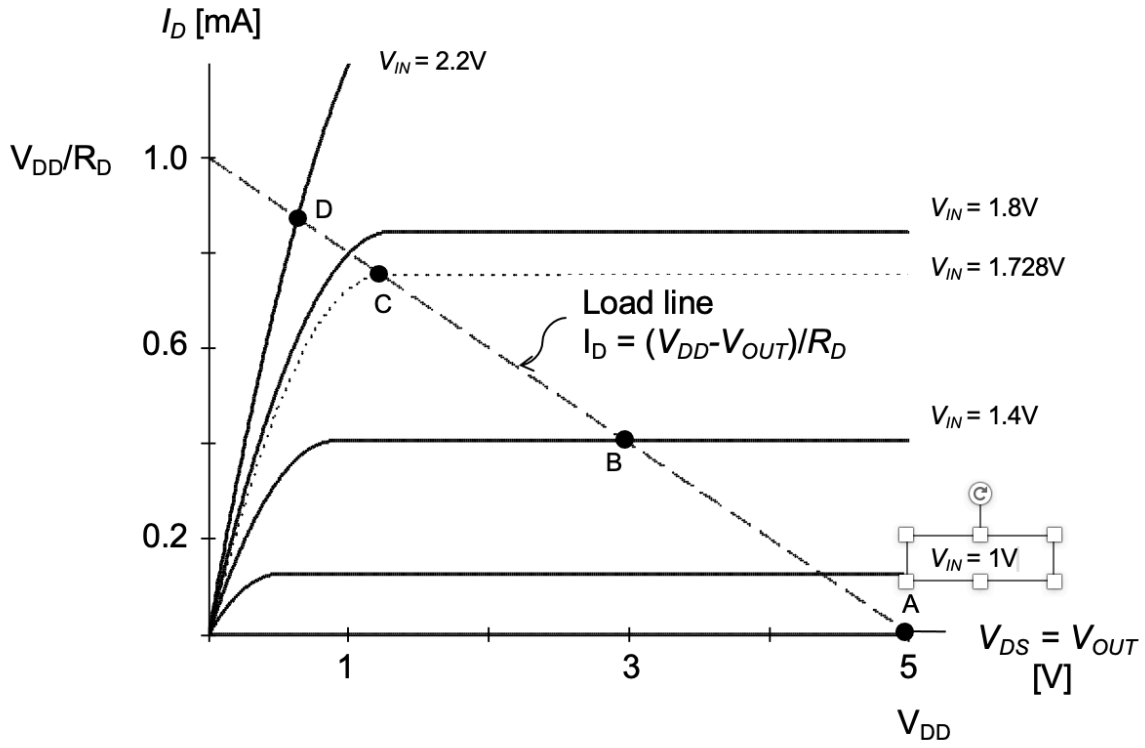


Figure 2.8.: Load line plot for the CS amplifier in Figure 2.7(a). Parameters: $V_{DD} = 5$ V, $R_D = 5$ kΩ, and $W/L = 20$.

The points A, B, C, and D marked in Figure 2.8 correspond to the points shown with the same annotation in Figure 2.7(b). Since the transistor drain characteristics are overlaid in Figure 2.8, it is easy to identify the operating regions that correspond to each point. For example, we can

immediately see that point B lies in saturation, since the intersect occurs in a region of constant drain current.

Example 2-3: Output Voltage Calculations for a Common-Source Voltage Amplifier

Construct a load line plot to verify the solution of Example 2-2(b) using $V_{DD} = 5\ V$, $R_D = 10\ \text{k}\Omega$, and W/L =10. Use $V_{IN} = V_{GS} = 1\ V$, $1.5\ V$, and $2\ V$ for the drain characteristic plot.

SOLUTION

The solution is shown in Figure 2.9. The load line is most easily drawn by connecting the points (0, $V_{DD}/R_D = 0.5\ mA$) and ($V_{DD} = 5\ V$, 0). The drain characteristics are drawn for the three given V_{GS} using the expressions of Table 2.1, by sweeping $V_{DS} = V_{OUT}$ from 0 V to 5 V. The intersect of the load line with the drain characteristic for $V_{IN} = 1.5\ V$ confirms the result already obtained in Example 2-2(b).

2.2.3. Biasing

After deriving the voltage transfer characteristic of our amplifier, we are now in a position to evaluate this circuit from an application standpoint. As we have discussed in Chapter 1, a common objective for a voltage amplifier is to create large output voltage excursions from small changes in the applied input voltage. With this objective in mind, it becomes clear that only a limited range of the transfer characteristic in Figure 2.7(b) is useful for amplification. For example, a change in the input voltage applied around point D in Figure 2.7(b) yields almost no change in the output voltage. In order to amplify small changes in V_{IN} into large changes in V_{OUT}, the transistor should be operated in the saturation region, i.e., in the vicinity of point B. The general concept of operating a circuit and its constituent transistor(s) around a useful operating point is called **biasing**.

Biasing generally necessitates the introduction of auxiliary voltages and/or currents that bring the circuit into the desired state. For the circuit considered in this section, proper biasing can be achieved by decomposing the input voltage into a constant component, and a component that represents the incremental voltage change to be amplified; this is illustrated in Figure 2.10(a). The incremental voltage component v_{in} could represent, for instance, the signal generated by a microphone or a similar transducer. The voltage V_{IN} is a constant voltage that defines the point on the overall transfer characteristic around which the incremental v_{in} is applied. We call V_{IN} the **input bias voltage** of the circuit.

Per IEEE convention, the total quantity in such a decomposition is denoted using a lowercase symbol and uppercase subscript, i.e., $v_{IN} = V_{IN} + v_{in}$ in our example. Similarly, the drain current is decomposed as $i_d = I_D + i_d$, where I_D is the current at the operating point, and id captures the current deviations due to the applied signal.

Figure 2.10(b) elucidates this setup further using the circuit's transfer characteristic. With $v_{in} = 0$, the output is equal to V_{OUT}, which is called the bias point or operating point of the output node. The bias point is sometimes also called the quiescent point (Q), since the corresponding voltage level corresponds to that of a "quiet" input. Note that V_{OUT} can be calculated by evaluating Equation 2.13, as done previously.

With some nonzero v_{in} applied, the output will now see an excursion away from the bias point. For example, applying a positive v_{in} will result in a negative incremental change v_{out} at the output.

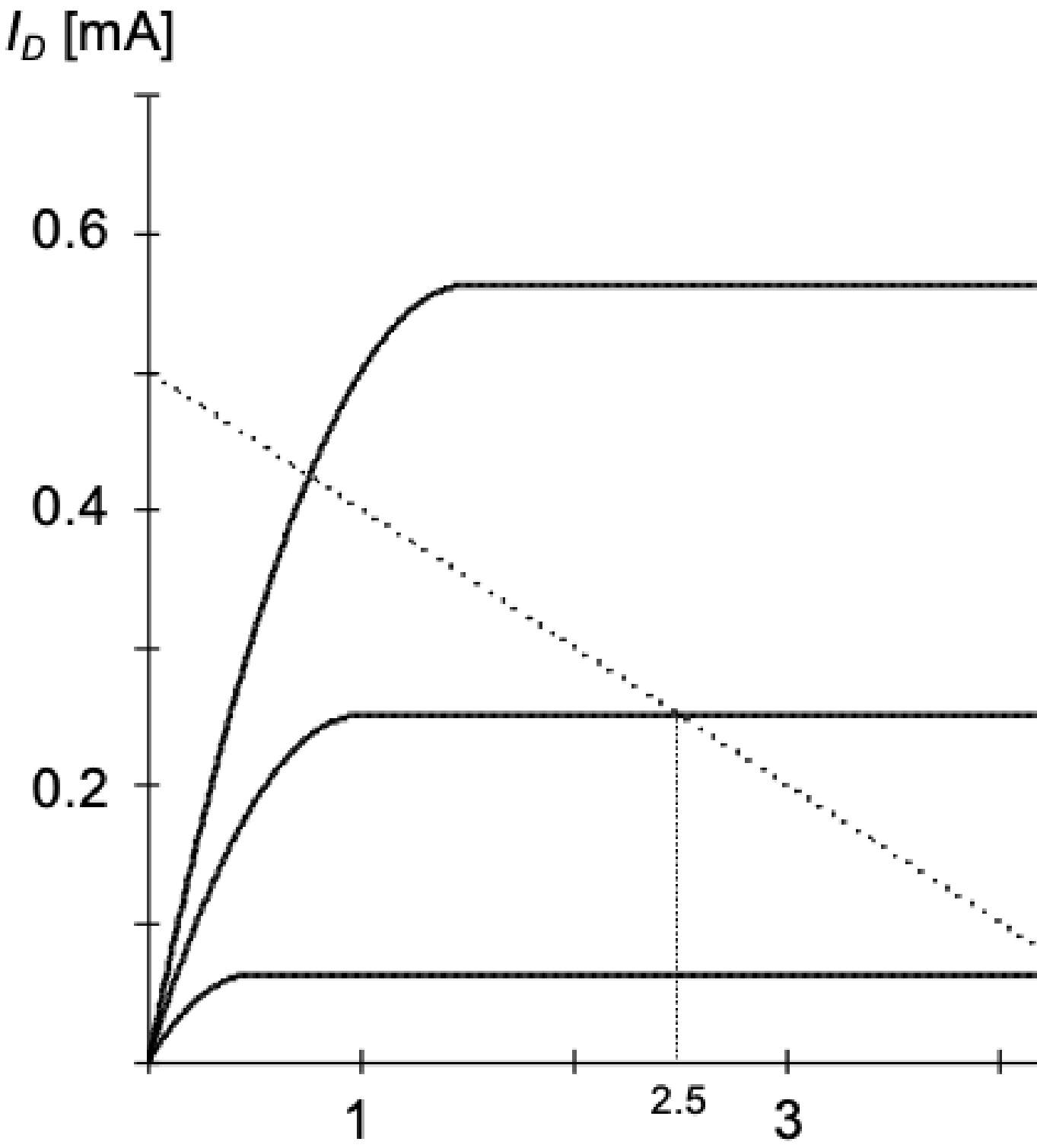


Figure 2.9.

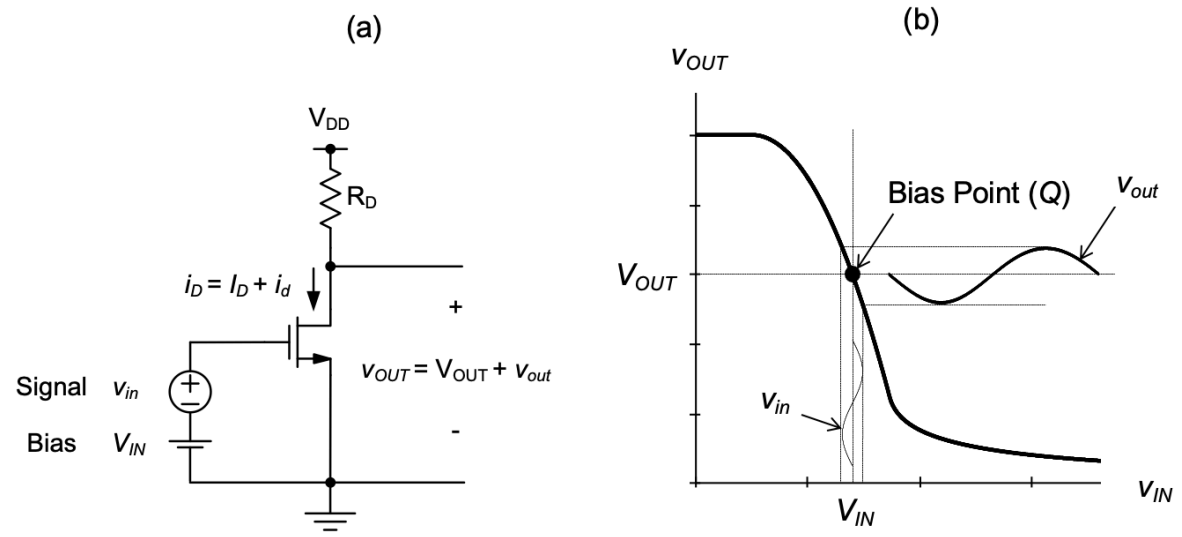


Figure 2.10.: (a) CS amplifier with input bias.(b) Transfer characteristic showing the bias point.

How can we compute v_{out} for a given v_{in}? Since Equation 2.13 must hold for the total quantities $v_{IN} = V_{IN} + v_{in}$ and $v_{OUT} = V_{OUT} + v_{out}$ we can write

$$V_{OUT} = V_{DD} - R_D \cdot \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(v_{IN} - V_{Tn})^2 \tag{2.17}$$

or

$$V_{OUT} + v_{out} = V_{DD} - R_D \cdot \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{IN} + v_{in} - V_{Tn})^2 \tag{2.18}$$

In order to simplify this expression, and since we are only interested in the change of v_{out} as a function of v_{in}, it is useful to eliminate the constant term from this expression, given by

$$V_{OUT} = V_{DD} - R_D \cdot \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{IN} - V_{Tn})^2 \tag{2.19}$$

After subtracting Equation 2.19 from Equation 2.18 and rearranging the terms, we obtain

$$v_{out} = -R_D\mu_n C_{ox}\frac{W}{L}(V_{IN} - V_{Tn}) \cdot v_{in}[1 + \frac{v_{in}}{2(V_{IN} - V_{TN})}] \tag{2.20}$$

Using the drain current expression of Equation 2.7, and by defining

$$V_{OV} = (V_{GS} - V_{Tn})|_Q = V_{IN} - V_{Tn} \tag{2.21}$$

the result can be further simplified and rewritten as

$$v_{out} = -\frac{2I_D}{V_{OV}} \cdot R_D \cdot v_{in}(1 + \frac{v_{in}}{2V_{OV}}) \tag{2.22}$$

where I_D is the transistor's drain current at the bias point, and V_{OV} is introduced as a symbol for the quiescent point gate overdrive voltage.

From the end result in Equation 2.22, we see that v_{out} is a nonlinear function of v_{in}. This is not surprising, since we are employing a transistor that exhibits a nonlinear I-V characteristic. While this derivation was relatively simple, the analysis of nonlinear circuits in general tends to be complex. Picture a circuit that contains several transistors, as for instance a cascade connection of several stages of the amplifier circuit considered here. Even with only a few nonlinear elements, most cases involving practical circuits with just moderate complexity tends to yield unwieldy expressions. A widely used solution to this problem is to approximate the circuit behavior using a linear model around its operating point, which we will discuss next.

2.2.4. The Small-Signal Approximation

Equation 2.22 is written in a format that suggests an opportunity for simplification. Provided that $v_{in} \ll v_{OV}$, the bracketed term is close to unity and we can write

$$v_{out} = -\frac{2I_D}{V_{OV}} \cdot R_D \cdot v_{in} = A_v \cdot v_{in} \tag{2.23}$$

where A_v is a constant voltage gain term that relates the incremental input and output voltages.

Interestingly, the term Av can also be found using basic calculus. Assuming that the incremental voltages represent infinitesimally small deviations in the total signal, we can rewrite Equation 2.23 as

$$dv_{OUT} = A_v \cdot dv_{IN} \tag{2.24}$$

and therefore

$$A_v = \left.\frac{dv_{OUT}}{dv_{IN}}\right|_Q = \left.\frac{dv_{OUT}}{dv_{IN}}\right|_{v_{IN}=v_{IN}} \tag{2.25}$$

where the derivative is evaluated at the circuit's operating point Q that is fully defined by choice of the input bias voltage V_{IN}. By applying Equation 2.25 to Equation 2.17, we find

$$\begin{aligned} A_v &= \frac{d}{dv_{IN}}\left[V_{DD} - R_D - \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(v_{IN} - V_{Tn})^2\right]\Bigg|_{v_{IN}=v_{IN}} \\ &= -R_D\mu_n C_{ox}\frac{W}{L}(v_{IN} - V_{Tn})\Bigg|_{v_{IN}=v_{IN}} \\ &= -R_D\mu_n C_{ox}\frac{W}{L}(V_{IN} - V_{Tn}) = -R_D\mu_n C_{ox}\frac{W}{L}V_{OV} \end{aligned} \tag{2.26}$$

Finally, using Equation 2.7, we find

$$A_v = -\frac{2I_D}{V_{OV}} \cdot R_D \tag{2.27}$$

which is the same result obtained previously. The voltage gain A_v can be interpreted graphically as shown in Figure 2.11. From Equation 2.25 and basic calculus we know that A_v is the slope of the tangent to the transfer characteristic at the point (V_{IN}, V_{OUT}), which is the operating point of the circuit.

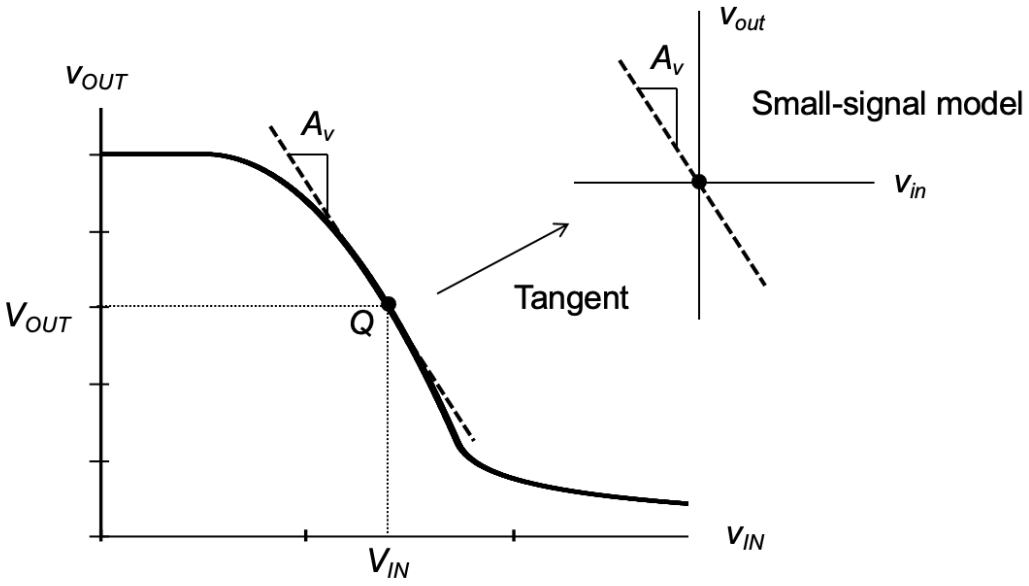


Figure 2.11.: Concept of small-signal voltage gain. In the small signal model, the input and output voltages are linearly related through A_v, the slope of the tangent at the operating point of the large-signal transfer characteristic.

In analog circuit nomenclature, A_v is called the **small-signal voltage gain** of the circuit; this emphasizes that this quantity is only suitable for calculations with "small" signals such that nonlinear effects are negligible. In the particular circuit considered here, "small" means $v_in \ll V_{OV}$, as seen from our analysis. The general concept of approximating circuit behavior by assuming small-signal excursions around an operating point is called **small-signal approximation**. In order to clearly distinguish a circuit transfer characteristic obtained through such an approximation from one that incorporates the nonlinear transistor behavior (e.g., Equation 2.17), the term **large-signal transfer characteristic** is typically used for the latter.

As we shall see in the remainder of this module, working with small-signal approximations greatly simplifies analog circuit analysis and design. The price paid for the approximation, however, is that the resulting equations by themselves cannot be used to reason about the circuit's behavior for large signals, as for instance signals where v_{in} is comparable to, or even greater than, V_{OV}. As illustrated in Figure 2.11, the small-signal approximation essentially creates a new coordinate system that linearly relates the input and output voltages. In this model, the output voltage follows the input

linearly, no matter how large the applied voltage is. In reality, considering the circuit's large signal transfer characteristic, signal clipping and strong waveform distortion can occur for large excursions and poorly chosen bias points. Examples of such cases are illustrated in Figure 2.12.

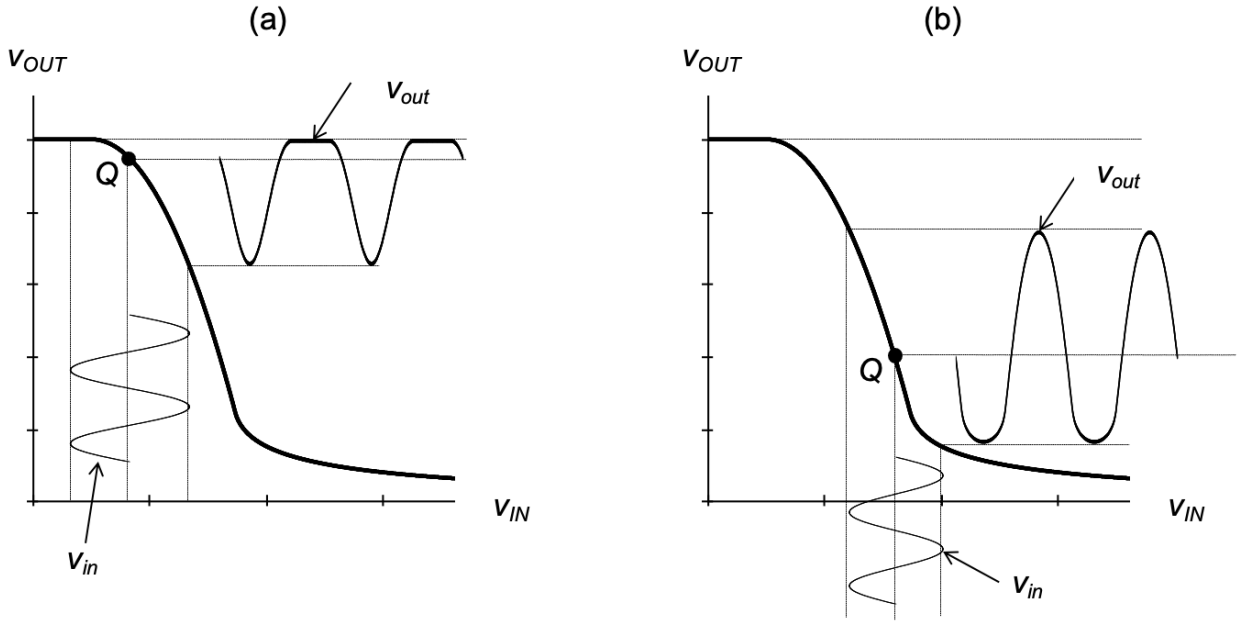


Figure 2.12.: Examples of signal clipping and distortion. (a) Output waveform is clipped due to supply voltage limit. (b) Output waveform drives the MOSFET into the triode region.

Example 2-4 : Signal clipping

Consider the circuit of Figure 2.10, using the parameters from Example 2-2(b): $V_{DD} = 5\ V$, $R_D = 10\ k\Omega$, $W/L = 10$, and V_{IN} is adjusted to 1.5 V, so that $V_{OUT} = 2.5\ V$ at the circuit's operating point. Calculate the most negative excursion that the incremental input voltage v_{in} can assume before the output is "clipped" to V_{DD} (as in Figure 2.12(a)).

SOILUTION

The circuit's output voltage is given by

$$v_{OUT} = V_{DD} - R_D \cdot \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{IN} + v_{in} - V_{Tn})^2$$

Clipping v_{OUT} to the supply voltage implies $v_{OUT} = V_{DD}$. This requires

$$0 = V_{IN} + v_{in} - V_{Tn}$$

$$v_{in} = -(V_{IN} - V_{Tn}) = -(V_{OV})$$

$$v_{in} = -(1.5V - 0.5V) = -1V$$

In words, applying a negative signal (v_{in}) at the input of magnitude larger than 1 V will cause the output to reach the supply voltage. Making v_i more negative will create a "plateau" in the output waveform as shown in Figure 2.12(a).

In a majority of analog circuits, it is sufficient to use the large-signal characteristic for bias-point and signal-range calculations. For all other purposes, as for instance voltage gain calculations, it is usually appropriate and justifiable to work with small-signal approximations. Without this clever split in the analysis, most analog circuits of only moderate complexity would not be amenable to hand analysis, simply because the nonlinear nature of the transistors would create prohibitively complex systems of non- linear equations.

Circuits that are designed to amplify small signals from a transducer are classical examples where the small-signal approximation works. Consider, for instance, the above-discussed amplifier circuit fed with an input signal from a radio antenna, which is often on the order of several hundred microvolts. As long as V_{OV} is chosen larger than several hundred millivolts, the small signal approximation will hold with reasonable accuracy. Other examples (not covered in this module) include amplifiers that rely on electronic feedback, which tends to minimize the signal excursions around a circuit's bias point (see Reference 2).

As a final remark, it should be noted that even if the input to a circuit is "small," the output will always show at least some amount of nonlinear distortion. In our basic amplifier, this distortion is caused by the bracketed term in Equation 2.22. In cases where even weak distortion is an issue, the designer often employs computer simulation tools to study the relevant behavior. From a design perspective, deviations from linearity can be minimized if needed. For the discussed common-source amplifier, this is seen from Equation 2.22: decreasing the ratio v_{in} / V_{OV}, either by reducing v_{in} or by increasing V_{OV} will result in improved linearity.

The exact analysis ofThe exact analysis of nonlinear distortion is beyond the scope of this module, and is typically treated only in advanced integrated circuit texts, as for instance Reference 3. We will focus here primarily on studying the relevant behavior of analog circuits using a linear small-signal abstraction, aided by basic bias-point and signal-range calculations.

2.2.5. Transconductance

The method of differentiating a circuit's large-signal transfer characteristic to obtain a small-signal approximation was straightforward for the simple one-transistor circuit discussed so far. Unfortunately, for a larger circuit it is usually much more difficult and often tedious to derive a complete transfer characteristic in the form of Equation 2.17.

A clever workaround that is predominantly used in analog circuit analysis is based on linearizing the circuit element-by-element around the operating point. This method is applied in three steps: (1) Compute all node voltages and branch currents at the operating point using the devices' large-signal model. (2) Substitute linear models for all nonlinear components and compute their parameters using the operating point information. (3) Compute the desired transfer function using the linear model obtained in step 2.

The biggest advantage of this method, called **small-signal analysis**, is that it avoids computing the large-signal transfer characteristic of the circuit, and instead defers the transfer function analysis until all elements have been approximated by linear models. The linearized models of nonlinear elements, such as MOSFETs, are typically called **small-signal models**.

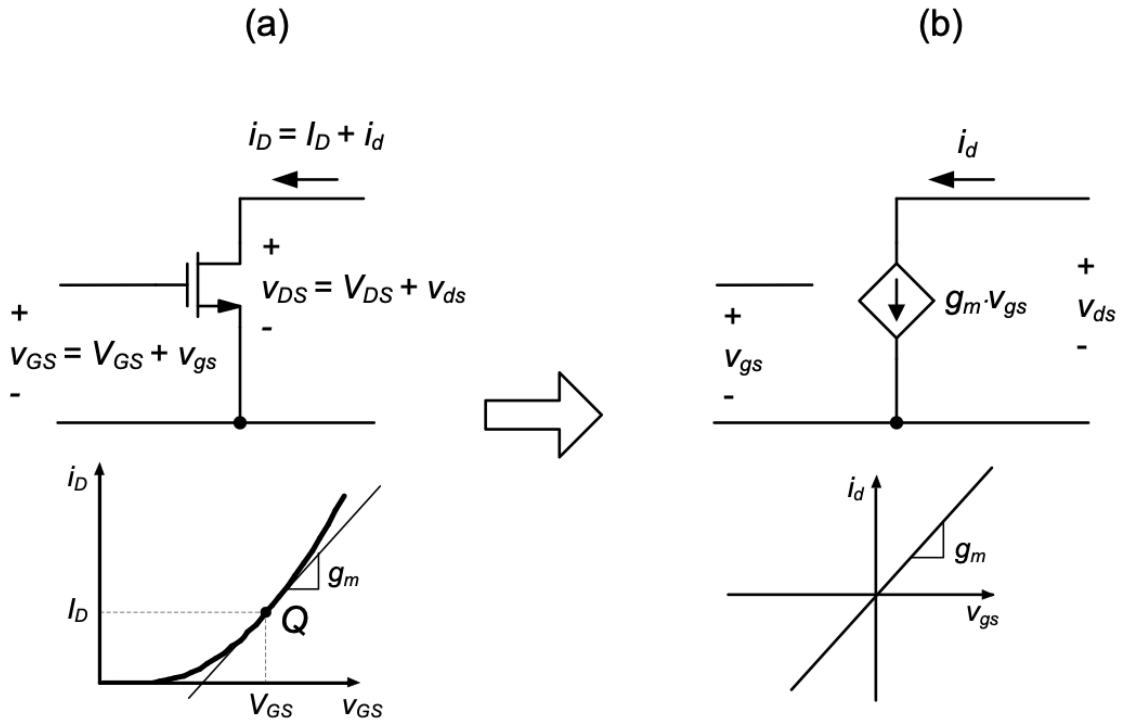


Figure 2.13.: (a) Large-signal transfer characteristic of an n-channel MOSFET in the saturation region. (b) Small-signal transconductance (g_m) at the operating point.

We will now illustrate the small-signal analysis approach by applying it to the basic common-source amplifier example covered so far in this chapter. Consider first the MOSFET device in Figure 2.7(a). In general, once the operating point of the transistor is known, the small-signal model is obtained by differentiating the large signal I-V relationships at this point. This is further illustrated in Figure 2.13, assuming that the MOSFET is biased in the saturation region. The proportionality factor that links the incremental drain current (i_d) and the gate-source voltage (v_{gs}) is given by the slope of the tangent to the large signal transfer characteristic at the bias point. This quantity is called **transconductance**, or g_m. Mathematically, we can write

$$g_m = \frac{i_d}{v_{gs}} = \left.\frac{di_D}{dv_{GS}}\right|_{v_{GS}=V_{GS}} \tag{2.28}$$

In order to find the transconductance for the saturation region of the device, we evaluate Equation 2.28 using Equation 2.7. This yields

$$g_m = \mu_n C_{ox}\frac{W}{L}(V_{GS} - V_{Tn}) = \mu_n C_{ox}\frac{W}{L}V_{OV} \tag{2.29}$$

Alternative forms of this expression are obtained by eliminating V_{OV} or $\mu_n C_{ox}\ W/L$ using Equation 2.7, which gives

$$g_m = \sqrt{2\mu_n C_{ox}\frac{W}{L}I_D} \tag{2.30}$$

or

$$g_m = \frac{2I_D}{V_{OV}} \tag{2.31}$$

All of the above equations can be used to calculate g_m; the choice of which equation is used depends on the given parameters. The physical unit for transconductance is $A/V = \Omega^{-1}$, or **Siemens** (S).

Once the transconductance is determined, we can insert the model of Figure 2.13(b) into the original circuit (Figure 2.7(a)) for further analysis. The resulting small-signal circuit equivalent is shown in Figure 2.14. No modeling modification is needed for the resistor R_D, as it is already assumed in Figure 2.7(a) that it follows a linear I/V law ($V = I \cdot R$). However, since the supply voltage is constant in the large-signal model, it must be replaced with 0 V or ground (GND) in the small-signal model. This is because the differentiation of a constant quantity yields zero.

Using the model of Figure 2.14, we now apply Kirchhoff's laws at the output and find

$$v_{out} = -g_m \cdot v_{in} \cdot R_D \tag{2.32}$$

Finally, by substituting Equation 2.31 we obtain

$$v_{out} = -\frac{2I_D}{V_{OV}} \cdot R_D \cdot v_{in} = A_v \cdot v_{in} \tag{2.33}$$

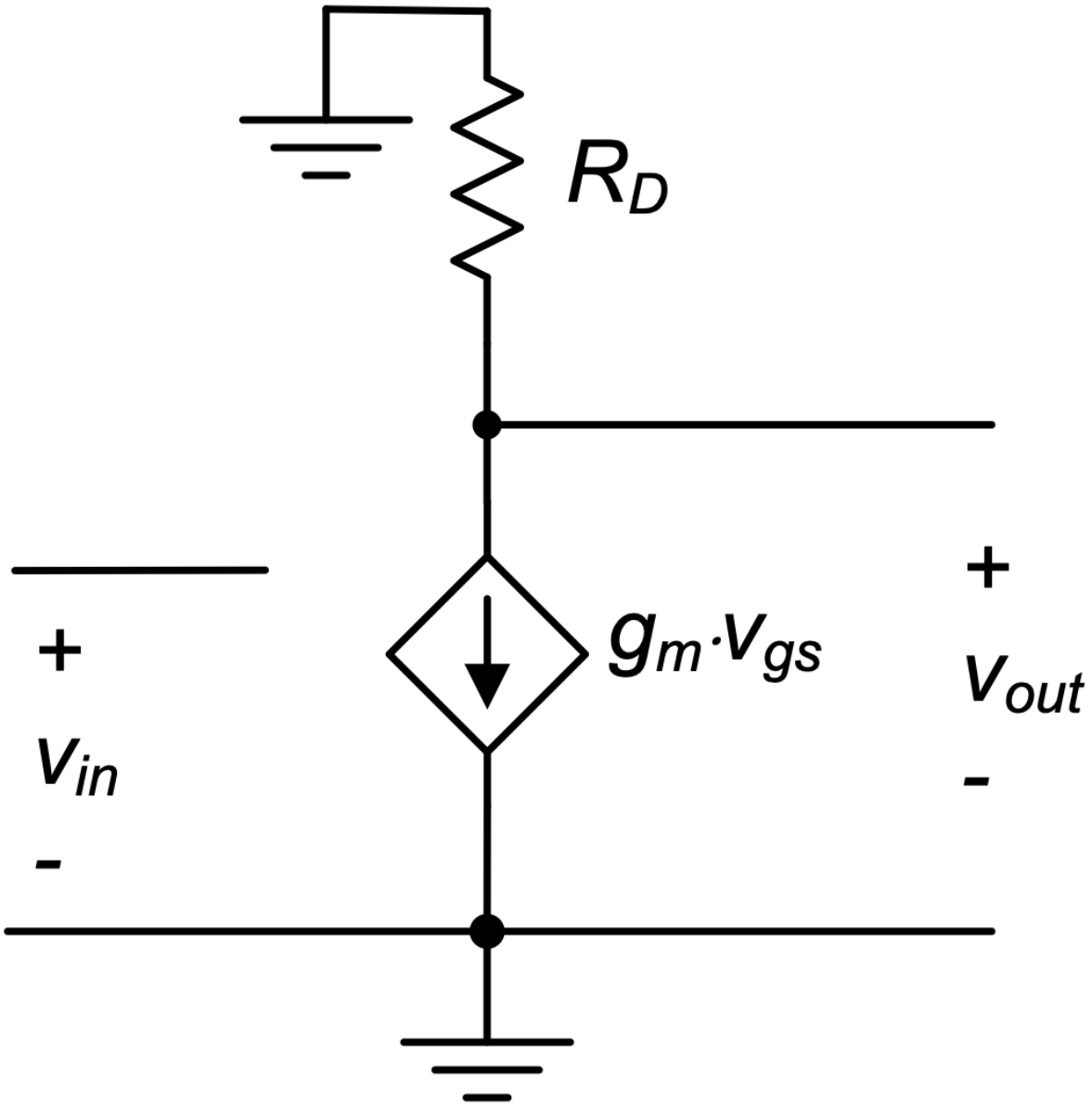


Figure 2.14.: Small-signal model of the circuit in Figure 2.7(a).

As expected, this result is equivalent to what was obtained by applying the small-signal approximation to Equation 2.22, and also by differentiating Equation 2.17 at the operating point. However, as indicated previously, the big advantage of working with a small-signal model for individual transistors is that this simplifies the analysis of larger circuits.

2.2.6. P-Channel Common-Source Voltage Amplifier

As shown in Figure 2.15(a), we can also build a CS amplifier using a p-channel device. Similar to the n-channel case, we can derive a large-signal transfer characteristic using Equation 2.11 and by applying KVL at the output node. The resulting plot is shown in Figure 2.15(b). Compared to the n-channel case (Figure 2.7(b)), one can show that the characteristic is flipped sideways (since $V_{SG} = V_{DD}$ - V_{IN}) and upside down (because I_D is negative for a p-channel transistor). Therefore, for small V_{IN} near 0 V, the device operates in the triode region and the output voltage is close to V_{DD}. For $V_{IN} = V_{DD}$, the device is off and the output is at 0 V, since $V_{SG} = 0$ and no current flows in the device.

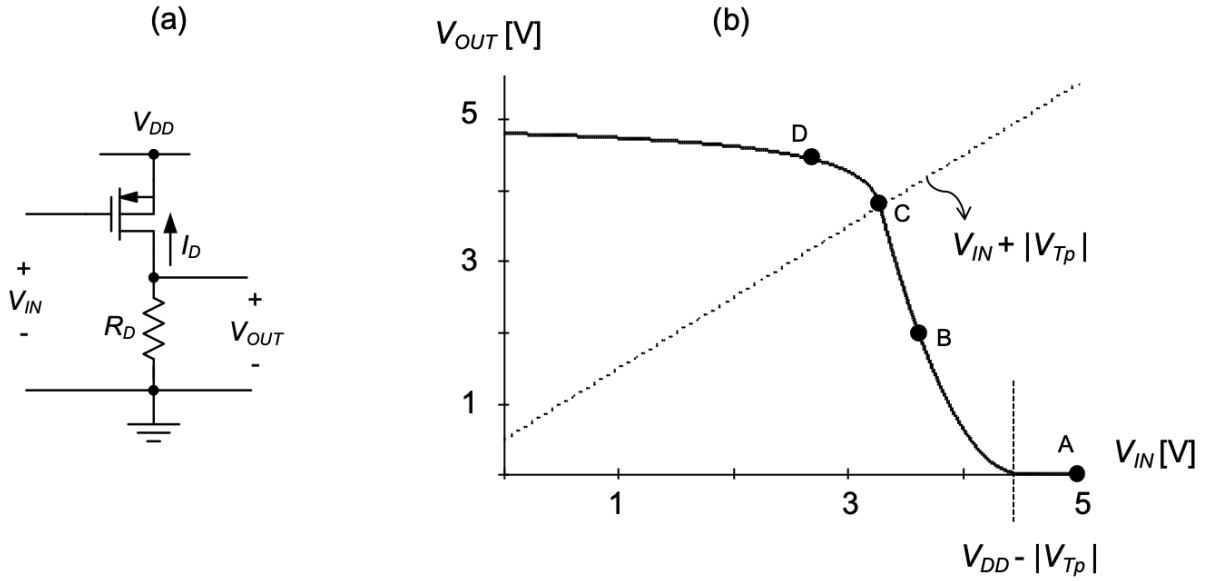


Figure 2.15.: (a) P-channel common-source amplifier schematic. (b) Voltage transfer characteristic for $V_{DD} = 5$ V, $R_D = 5$ $k\Omega$, and $W/L = 40$.

In terms of its small-signal model, it follows that the p-channel CS amplifier is identical to the n-channel version. This can be seen intuitively by comparing Figure 2.7(b) and Figure 2.15(b): in the region around point B, both transfer characteristics exhibit a negative slope and therefore will behave alike for small perturbations. We will show this formally in the following discussion.

We begin by inserting a small-signal model for the transistor as shown in Figure 2.16(a). Based on the positive variable convention of Equation 2.11, we define the transconductance for the p-channel transistor as

$$g_m = \frac{d}{dv_{SG}}((-i_D))\Big|_Q$$

$$= \mu_p C_{ox} \frac{W}{L}(V_{SG} + V_{Tp}) \tag{2.34}$$

$$= \frac{2(-I_D)}{V_{SG} + V_{Tp}} = \frac{2(-I_D)}{V_{OV}}$$

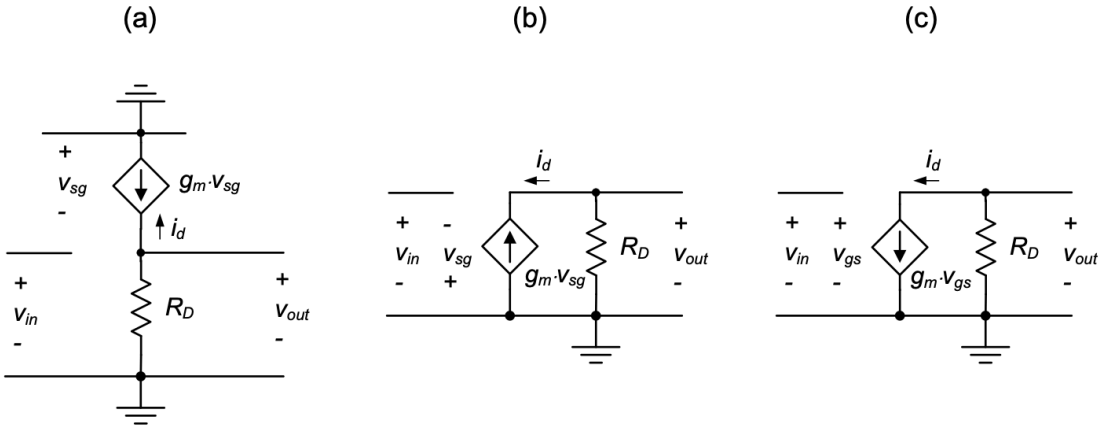


Figure 2.16.: Small-signal model of the p-channel common-source amplifier: (a) using the p-channel model directly, (b) a flipped version of (a), and (c) with sign changes applied to show equivalence with the n-channel model.

Note that through this expression, we have defined V_{OV} for the p-channel case as $V_{SG} + V_{Tp}$, which is a positive quantity for a p-channel device in the "ON" state.

Although we could solve for the small-signal voltage gain directly with the circuit shown in Figure 2.16(a) it is easier to flip the transistor 180° (Figure 2-15(b)) so that the circuit appears similar to the n-channel version. Since v_{sg}=–v_{gs} we can change signs at the input and the dependent current source and find that the p-channel common-source amplifier small-signal model is identical to the n-channel version as shown in Figure 2.16(c).

This result applies more generally to all the transistor configurations and model extensions that we will study in this module. Once the operating point parameters of a p-channel device have been determined (e.g., a calculation of g_m using Equation 2.34), it is perfectly valid to replace it with an n-channel equivalent. This is a very powerful and convenient result, since it allows us to focus on n-channel only configurations in small-signal analyses, without having to worry about the specific sign conventions of p-channels.

2.2.7. Modeling Bounds for the Gate Overdrive Voltage

According to Equation 2.31, $g_m/I_D = 2/V_{OV}$ tends to infinity as the gate overdrive voltage V_{OV} approaches zero. This implies that for a transistor that is "barely on," we can extract very large

transconductance values for only small bias currents. Unfortunately, this behavior is incorrect, and stems from limitations of the device model discussed in Section 2-1. As V_{OV} approaches zero, a more complex analysis is needed to predict the drain current and its derivative with respect to gate voltage see Reference 4.

Figure 2.17 plots the g_m/I_D characteristic of a MOSFET, and the expected behavior based on Equation 2.31. As we can see, for $V_{OV} < 150mV$, a large discrepancy exists between a physical device and the prediction based on the simple square-law model used in this treatment. In order to avoid unrealistic design outcomes due to this modeling limitation, we define a bound for the minimum allowed gate overdrive voltage for all circuits covered in this module

$$V_{OV} \geq V_{OVmin} = 150mV \tag{2.35}$$

Designing with a smaller V_{OV} would require a more elaborate model for hand calculations, which is beyond the scope of this module. The interested reader is referred to advanced material on this topic, available for example in References 4 and 5.

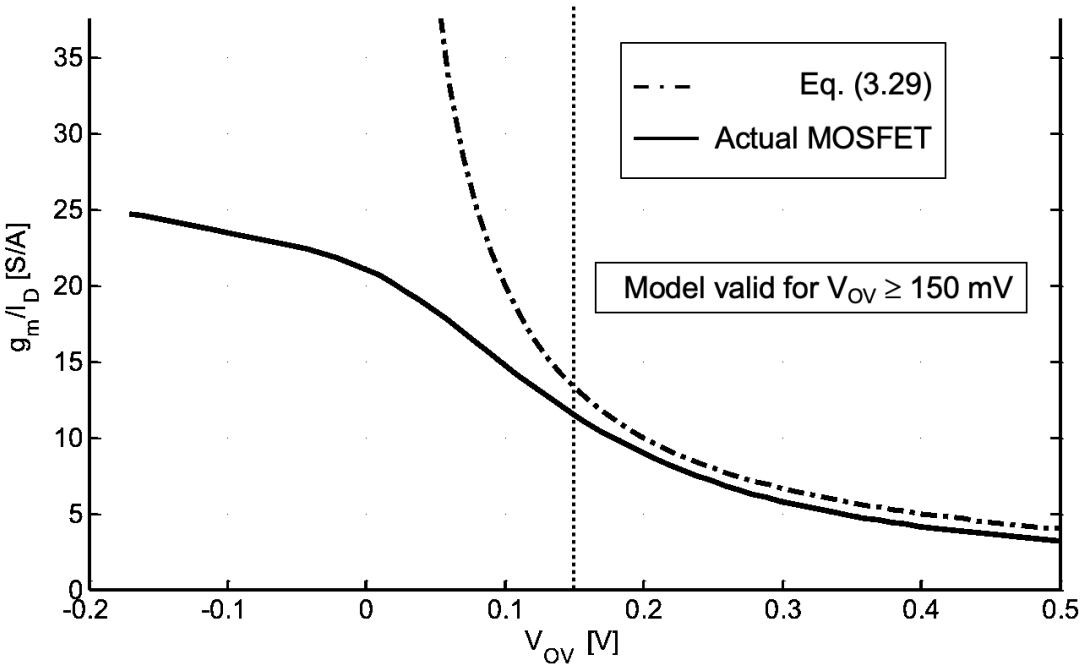


Figure 2.17.: Transconductance to current ratio predicted by Equation 2.31 and the behavior of an actual MOSFET.

2.2.8. Voltage Gain and Drain Biasing Considerations

In the basic common-source amplifier discussed so far, the drain resistor R_D serves a dual purpose: (1) it translates the device's incremental drain current (i_d) into a voltage (V_{out}), and (2) it supplies the quiescent point drain current (I_D) for the MOSFET. As we shall show next, this creates an

undesired link between the bias point constraints of the circuit and the achievable small-signal voltage gain of the amplifier. To see this, we rewrite Equation 2.36 as shown below

$$A_v = -\frac{2I_D}{V_{OV}} \cdot R_D = -2\frac{V_R}{V_{OV}} \tag{2.36}$$

where $V_R = V_{DD}–V_{OUT}$ is the voltage drop across RD at the operating point. This result leads to several interesting conclusions. First, note that the voltage gain of the amplifier is fully determined once V_{OV} and voltage drop across R_D are known. For example, if the circuit is biased such that $V_{OV} = 0.2\ V$ and $V_R = 2\ V$, we have $A_v = –20$; regardless of the particular W, L or $\mu_n\ C_{ox}$ of the employed MOSFET. Second, since the possible values for V_{OV} are lower-bounded (Equation 2.35) and V_R is upper-bounded (finite V_{DD}), there exists a maximum possible A_v that can be obtained

$$|A_{vmax}| = 2\frac{V_{Rmax}}{V_{OVmin}} = 2\frac{V_{DD} - V_{OVmin}}{V_{OVmin}} \cong 2\frac{V_{DD}}{V_{OVmin}} \tag{2.37}$$

In this result, it was assumed the transistor is biased at the edge of the triode region, a somewhat impractical, but appropriate limit case to consider. Evaluating the above expression for $V_{DD} = 5$ V and $V_{OVmin} = 150\ mV$ yields $|A_{vmax}|$ 67. Can we overcome this limit and change our amplifier such that it can achieve voltage gains beyond this value?

In order to investigate this, consider the load line illustrations shown in Figure 2.18. As explained in Section 2-2-2, the load line for our circuit is defined by the points (0, V_{DD}/R_D) and (V_{DD}, 0). From the location of these points, we see that the x-axis intercept of the load line is fixed, while the y-intercept moves lower with larger values of R_D. This reduces the slope of the load line, resulting in a larger small-signal voltage gain of the circuit. Furthermore, note that for a fixed quiescent point drain current I_D, larger R_D shifts the output bias point V_{OUT} to smaller values, i.e., closer to the edge of the MOSFET's triode region. This observation captures the result of Equation 2.37 in a graphical way: we cannot increase the small-signal voltage gain beyond a certain limit due the link between V_{OUT} and the chosen R_D.

A more ideal situation is depicted in Figure 2.19. If we could somehow create a load line that "rotates" about the desired operating point (as a function of R_D), the voltage gain could be set independently of V_{OUT}. A modified drain network that lets us achieve this behavior is shown in Figure 2.20(a). In this circuit, R_D is now connected to a voltage V_B (instead of the supply voltage V_{DD}) and an ideal current source I_B is used to provide a fixed current. In a realistic implementation circuit, I_B can be built, for example, using a p-channel MOSFET that operates in saturation. For the time being, we will neglect such implementation details, and postpone the discussion of current sources to Chapter 5.

With this new configuration, the relationship between i_D and v_{OUT} becomes

$$i_D = I_B = \frac{V_B - v_{OUT}}{R_D} \tag{2.38}$$

or

$$v_{OUT} = V_B + (I_B - i_D)R_D \tag{2.39}$$

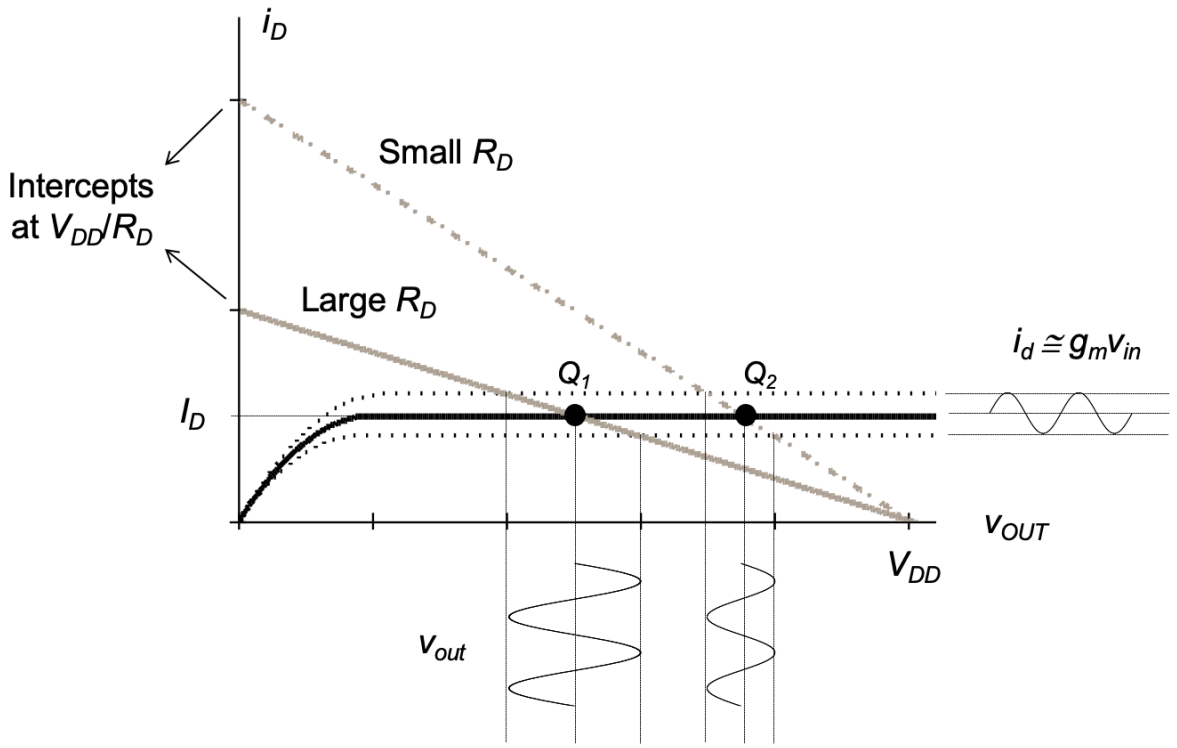


Figure 2.18.: Amplifier load lines for large and small values of R_D.

As we can see from these equations (also graphically shown in Figure 2.20(b)), at the point v_{OUT} = V_B, we have $i_D = I_B$, regardless of the value of R_D\$. Therefore, utilizing this point as the operating point of our amplifier precisely achieves the goal we have in mind. In particular, we wish to set $I_B = I_D$, the desired quiescent point drain current of the MOSFET, and $V_B = V_{OUT}$, the desired output operating point. With this choice, the role of the current source is to provide the MOSFET's bias current, while the resistor RD is responsible only for converting the incremental drain current into an incremental output voltage; no DC bias current flows in this element.

Since the voltage source V_B and the current source I_B aid in maintaining the circuit's bias point, we generally classify these elements as biasing sources. However, it is important to distinguish their function from the input bias voltage V_{IN}. V_{IN} directly sets the quiescent point gate-source voltage of the transistor and therefore fully defines the operating point on the MOSFET's I-V characteristic and the corresponding drain current. In the above-described scenario, I_B is adjusted to supply this same drain current, but does not define it. We therefore categorize I_B an **auxiliary bias** current that helps sustain, but does not set the quiescent point of the transistor. Since V_B defines the quiescent point output voltage of the circuit, we refer to this element as the **output bias voltage.**

Lastly, it is important to note that for the modified circuit in Figure 2.20(a), the previously derived small-signal model shown in Figure 2.14 still applies. This can be understood by differentiating Equation 2.39 at the operating point to find the small-signal equivalent of the network placed at the drain of the amplifier

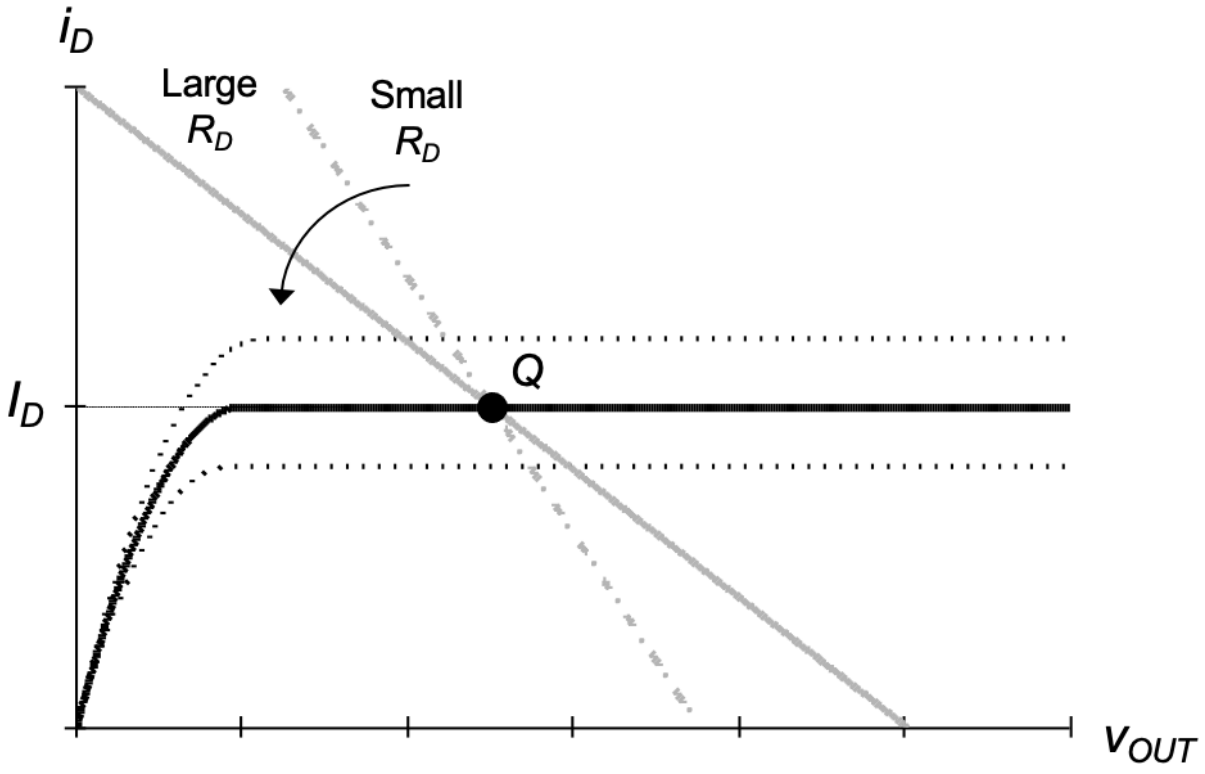


Figure 2.19.: Desired load line behavior.

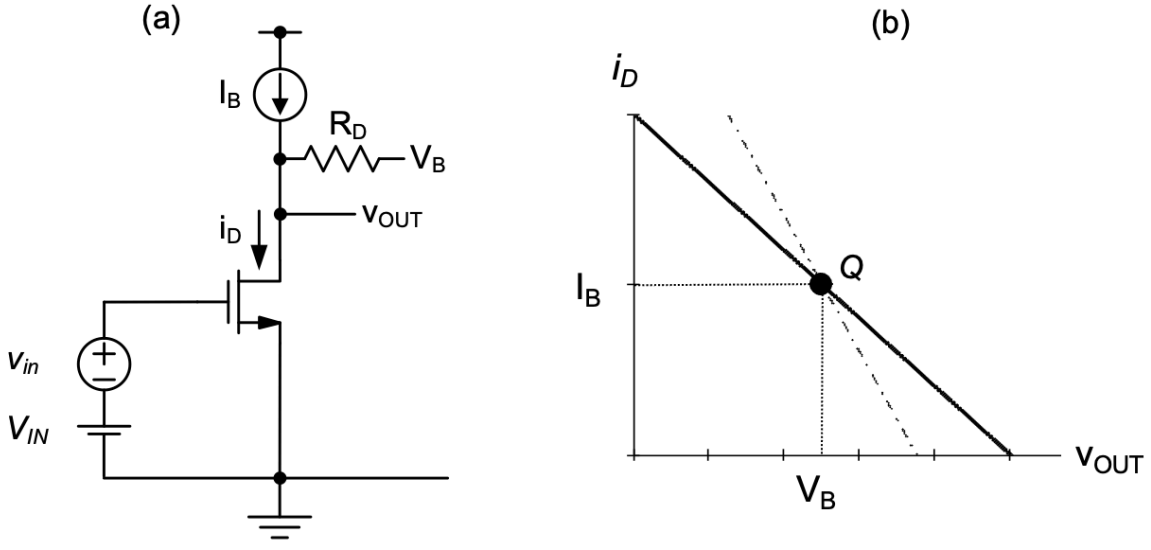


Figure 2.20.: (a) CS amplifier with modified drain biasing scheme. (b) Load line characteristic.

$$\left.\frac{dv_{out}}{di_D}\right|_Q = \frac{d}{di_D}(V_B + (I_B - i_D)R_D) = -R_D \tag{2.40}$$

As this result indicates, and as we have seen previously, any constant sources, such as V_B, I_B, etc., drop out of the small signal model, which captures only components that affect the incremental changes in currents and voltages around the circuit's operating point.

For the circuit in Figure 2.20(a), one might now be tempted to think that we can obtain an arbitrarily large voltage gain, as long as R_D is made very large. Unfortunately this is not the case for several reasons, the first of which stems from physical effects that we have not yet included in the MOSFET model. This aspect is further discussed in Section 2-3. In addition, there are practical limitations to the attainable voltage gain, discussed next.

2.2.9. Sensitivity of the Bias Point to Component Mismatch*

Consider a CS amplifier biased at the gate as done previously with a bias voltage V_{IN}, directly setting up the quiescent point drain current I_D (depending on W/L and other relevant device parameters). Our goal in using the drain bias network of Figure 2.20(a) is then to set $I_B = I_D$ and chose VB such the output is biased at a reasonable, desired V_{OUT}. Unfortunately, in practice, we can never achieve $I_B = I_D$ exactly; there will always be a non- zero $\Delta I = I_B - I_D$. This case is shown in Figure 2.21. As illustrated, the finite ΔI leads to a shift ΔV away from the desired output operating point V_B. This shift is proportional to R_D, since the current difference ΔI flows into R_D, creating the undesired ΔV.

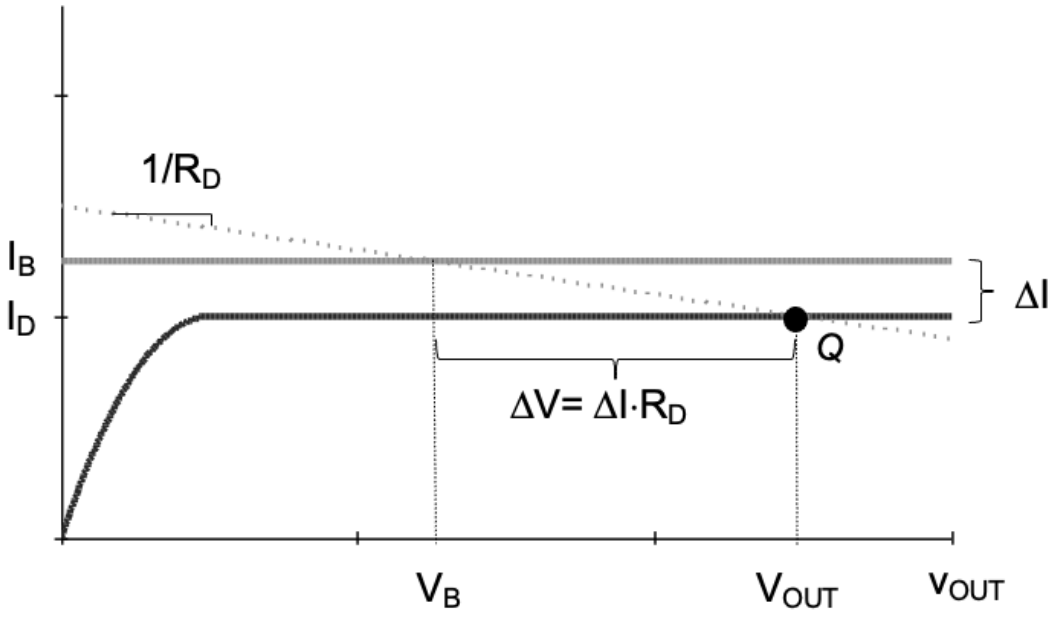


Figure 2.21.: Bias point shift due to mismatch in I_B and I_D.

Example 2-5: Output Bias Voltage Shift due to Current Mismatch

Consider the CS amplifier of Figure 2.20(a), biased at the gate such that V_{OV} = 500 mV. Assume I_D = 200 μA, and that R_D is chosen such that the amplifier achieves A_v = –400. Considering Figure 2.21, how much mismatch between I_B and I_D (in) can be tolerated such that V_{OUT} deviates from the intended bias point (V_B) by no more than $\Delta V = 500\ mV$? Repeat this calculation for A_v = –40 and –4.

SOLUTION

We begin by computing the transconductance of the MOSFET

$$g_m = \frac{2I_D}{V_{OV}} = \frac{2 \cdot 200\mu_A}{0.5V} = 800S$$

In order to achieve A_v = –400, we require R_D = 400/800 μS = 500$k\Omega$. Therefore,

$$\frac{\Delta I}{I_D} = \frac{\Delta V}{I_D R_D} < \frac{500mV}{200\mu A \cdot 500k\Omega} = 0.5\%$$

For A_v = –40 and A_v = –4, the result modifies to 5%, and 50%, respectively.

As expected, this result confirms that for larger $|A_v|$, the auxiliary bias current I_B must match the MOSFETs drain current more accurately. How precisely can we match these two currents? Unfortunately, answering this question in detail is beyond the scope of this module, and is the subject of advanced research papers such as Reference 6. Nonetheless, it can be said in general that matching currents, voltages, or any other electrical quantities in today's integrated circuits to better than 1% requires special care and understanding. In some cases, even 10% matching can be hard to guarantee. With this guideline in mind, it becomes clear that the circuit of Figure 2.20(a) may become impractical if we aim for too much voltage gain.

The general issue of properly dealing with variability in integrated circuit components is a complex topic that is still being actively researched. At the introductory level of this module, the main point that the reader should retain is that any circuit whose bias point relies on precisely matched components or high absolute accuracy in any electrical parameter may not be robust in the presence of component variability. In general, experienced circuit designers avoid situations that resemble the "balancing of a marble on the tip of a cone," i.e., circuits that will be overly sensitive to variations in component parameters. We will take up this point once more when discussing practical biasing circuits in Chapter 5.

As a final note concerning the issue of variability, it is worth mentioning that electronic feedback can help alleviate problems as the one analyzed in Example 2-5. Picture for example adding an auxiliary circuit to Figure 2.20(a) that somehow measures VOUT, and adjusts V_{IN} (and therefore I_D) until the desired output operating voltage is set. In such schemes, relatively large variations in I_B can be absorbed. Feedback circuits are not covered in this module, but are the subject of advanced texts such as Reference 2.

2.3. Channel Length Modulation

The MOSFET model used so far assumes that the drain current in the saturation region is independent of the drain-source voltage. This behavior corresponds to that of an ideal current source, which is generally non-physical. A more realistic output characteristic observed in real MOSFETs is shown in Figure 2.22. For a physical MOSFET, I_D tends to increase with V_{DS}; an effect that can be explained (to first-order) as a voltage dependent modulation of the channel length (see Section 2-3-1).

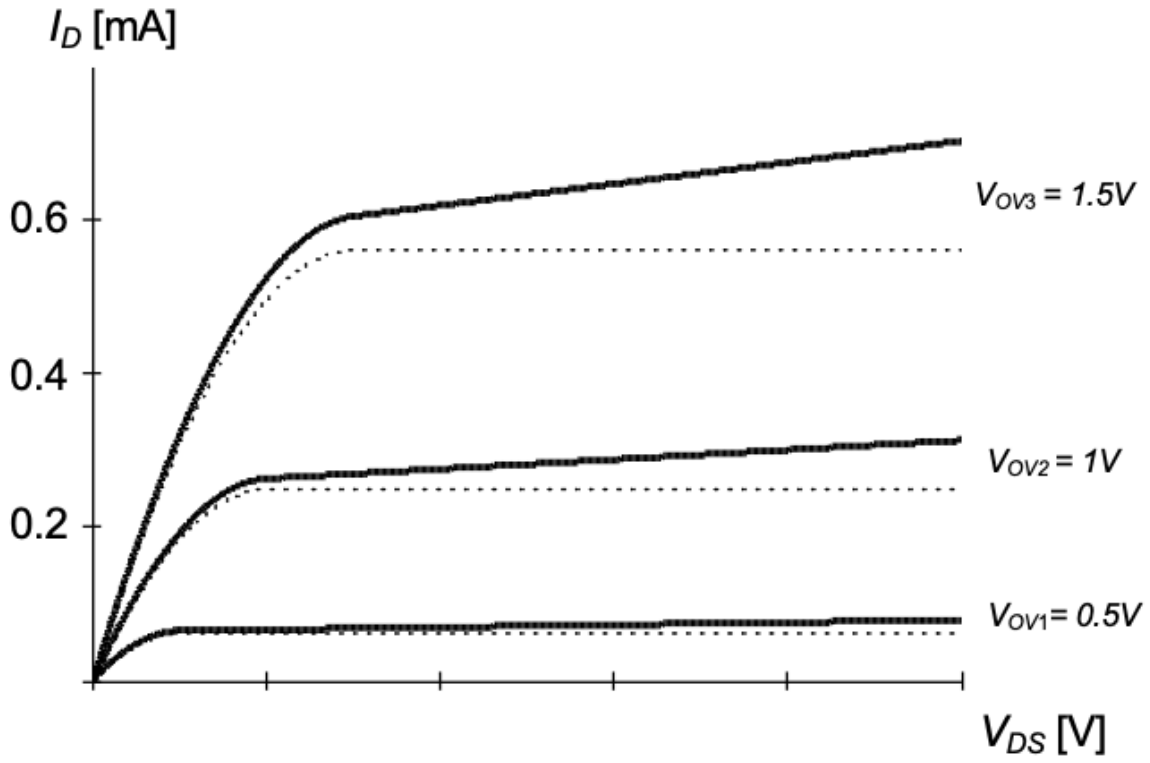


Figure 2.22.: Realistic n-channel I-V plots, incorporating the dependence of saturation drain current on V_{DS}. (solid lines), along with the first-order model assumed previously (dotted lines). Parameters: W = 20 μm, L = 2 μm.

When inserted into the circuit of Figure 2.20(a), the dependence of drain current on v_{DS} (v_{OUT}), will have an impact on the overall transfer characteristic of the amplifier, which we will thus consider in this section. For simplicity, let us first investigate the case of R_D –> ∞, i.e., no explicit drain resistance, and using only an ideal current source in the drain biasing network (see Figure 2.23(a)). In this case, the load line is horizontal at $I_B = I_D$, as shown in Figure 2.23(b). As before, the operating point Q is established at the point where the load line and the device's drain characteristic for the applied quiescent point input voltage (V_{IN}) meet.

Note that similar to the case considered in the previous section, the output operating point voltage in this circuit will shift by large amounts for relatively small changes in I_B (or MOSFET parameters). This issue must be addressed when this circuit is used in practice, for example by

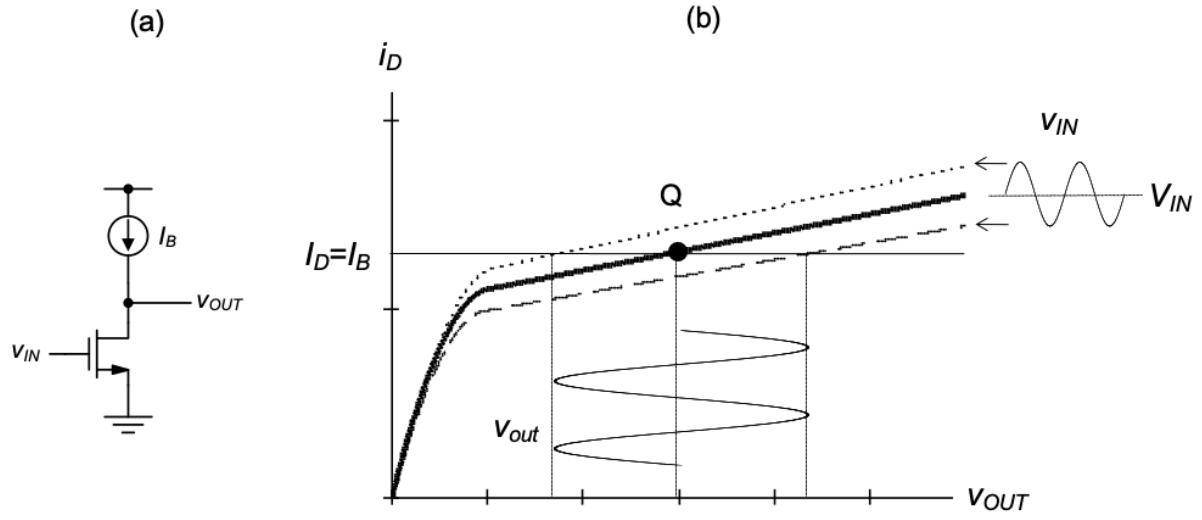


Figure 2.23.: (a) Intrinsic voltage gain stage. (b) Load line plot.

providing a feedback mechanism that adjusts I_B such that the desired V_{OUT} is maintained in the presence of component variations.

Assuming that a well-defined operating point has been established by some means, incremental changes in v_{IN} applied around the bias point Q will force the intersection of the horizontal load line with the MOSFET's drain characteristics to move sideways (since the drain current cannot change), creating a finite output voltage excursion vout. The magnitude of this voltage excursion, and thus the voltage gain of the circuit, depends on the slope of the MOSFET's drain characteristic in saturation. The voltage gain achieved in this configuration is commonly called the **intrinsic voltage gain** of the MOSFET, as it represents the voltage gain of the transistor by itself, without any added resistances in the drain bias network. By the same reasoning, the circuit of Figure 2.23(a) is often called the **intrinsic voltage gain stage**. The voltage gain and other parameters of more complex amplifier circuits are often directly related to the intrinsic voltage gain of their constituent transistors, giving this parameter a fundamental significance in circuit design.

2.3.1. The -Model

Unfortunately, the intrinsic voltage gain of a MOSFET cannot be predicted using the MOSFET model established so far. We will therefore extend the first-order MOSFET model to incorporate the dependence of the saturation current on the drain-source voltage. As a first step, we will describe the effect in terms of large-signal equations. Next, we will apply a small-signal approximation that makes it possible to capture the I_{Dsat}–V_{DS} dependence through a single resistor added to the MOSFET's small-signal model.

We begin by revisiting an approximation that was made in Section 2-1. In order to arrive at the constant drain current expression in saturation (Equation 2.7), it was assumed that ΔL, the distance from the pinch-off point to the drain, is negligible relative to the channel length L. In reality, this approximation is fine only as long we do not care about the $I_D - V_{DS}$ dependence seen

in Figure 2.22. Therefore, in order to get a quantitative handle on the MOSFET's intrinsic voltage gain, we must further investigate the impact of the physics at the drain side.

The simplest possible way to proceed is to factor ΔL into the existing derivation of Section 2-1. Instead of integrating Equation 2.5 over the length L, we use L–ΔL as the upper limit of the integral. Recall that L–ΔL is the actual location where the mobile charge vanishes, i.e., $Qn(y) = 0$. With this change we obtain the following expression for the saturation region.

$$I_D = \frac{1}{2}\mu_n C_{ox}\frac{W}{L-\Delta L}(V_{GS} - V_{Tn})^2 \tag{2.41}$$

Now, provided that ΔL is still small (but not negligible) relative to L, we can apply the following first-order approximation:

$$\frac{1}{L-\Delta L} \approx \frac{1}{L}(1 + \frac{\Delta L}{L}) \tag{2.42}$$

Next, note that ΔL must be a function of the drain voltage, since the depletion region widens with increasing reverse bias. This effect is commonly called **channel length modulation**. Unfortunately, an exact calculation of ΔL as a function of the terminal voltages involves solving the two-dimensional Poisson equation and leads to complex expressions. For simplicity, we assume that the fractional change in channel length is proportional to the drain voltage

$$\frac{\Delta L}{L} = \lambda_n V_{DS} \tag{2.43}$$

where λ_n is the channel length modulation parameter. Device measurements and simulations indicate that λ_n approximately varies with the inverse of the channel length. For the MOSFETs in this module, we will use

$$\lambda_n = \frac{0.1\mu m V^{(} - 1)}{L} \tag{2.44}$$

where L is in $\mu_n m$. Finally, we substitute Equation 2.44 and Equation 2.43 into Equation 2.41 and find a very useful approximation to the drain current in saturation, often called the λ-**model**:

$$I_D = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{GS} - V_{Tn})^2(1 + \lambda_n V_{DS}) \tag{2.45}$$

where V_{DS} $V_{DSsat} = V_{GS} - V_{Tn}$. This model has proven useful and sufficiently accurate for basic hand calculations, even though the physics related to the I_{DSsat}-V_{DS} dependence are in reality much more complex than discussed above. As long as λ_n is determined from measurements or accurate physical analysis, the model properly approximates a typical MOSFET's I-V characteristic to first-order. Higher-order models are typically not used in hand analysis, but find their use in advanced computer simulation models.

We now wish to incorporate the channel length modulation effect into the small-signal model of the MOSFET. An important new feature that must be considered in this task is that the drain current of Equation 2.45 now depends on two voltages, namely V_{DS} and V_{GS}. A common and appropriate

way of handling this situation for small-signal modeling is to approximate the incremental drain current around the operating point as the total differential (as frequently used in error analysis) due to both variables, i.e.

$$i_d = \left.\frac{\partial i_D}{\partial V_{GS}}\right|_Q \cdot v_{gs} + \left.\frac{\partial i_D}{\partial V_{GS}}\right|_Q \cdot v_{ds} \tag{2.46}$$

The above expression essentially treats v_{DS} as a constant when evaluating the derivative of i_D with respect to v_{GS}. Similarly, v_{GS} is assumed constant in the differentiation with respect to v_{DS}. This use of partial differentiation is justified and reasonably accurate as long as at least one of the following two conditions is met:

1. The excursion in the variable that is treated as a constant can be approximated as infinitesimally small and therefore negligible.
2. The excursion in the variable that is approximated as a constant is considerable, but nonetheless does not affect the derivative with respect to the second variable.

In the context of a common-source amplifier, for instance, the first condition applies to v_{GS}. Just as in the derivation of the simple small-signal model without V_{DS} dependence, we can argue that changes in v_{GS} are suitably modeled as "small" (relative to V_{OV}). The same condition cannot be applied to v_{DS} in general. Often times the output voltage, and therefore v_{DS}, see large excursions in amplifier circuits. In order for Equation 2.46 to be reasonably accurate, we must require the second condition, i.e., the derivative of i_D **w.r.t.** v_{GS} must not strongly depend on drain-source voltage. By inspection of Equation 2.45, we see that this condition is met as long as λ_n is small, which is typically the case for MOSFETs intended for use in amplifier stages.

To continue with our analysis, we rewrite Equation 2.46 as

$$i_d = g_m v_{gs} + g_o v_{ds} \tag{2.47}$$

where

$$g_m = \left.\frac{\partial i_D}{\partial V_{GS}}\right|_Q = \mu_n C_{ox}\frac{W}{L}(V_{GS} - V_{Tn})(1 + \lambda_n V_{DS}) = \frac{2I_D}{V_{GS} - V_{Tn}} = \frac{2I_D}{V_{OV}} \tag{2.48}$$

and

$$g_o = \left.\frac{\partial i_D}{\partial V_{GS}}\right|_Q = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{GS} - V_{Tn})^2\lambda_n = \frac{2I_D}{V_{GS} - V_{Tn}} = \frac{2I_D}{V_{OV}} \tag{2.49}$$

where the approximate end result assumes $\lambda_n V_{DS} \ll 1$, a condition that is often satisfied for long channels and moderate V_{DS}. For instance, assuming L = 2 μm and $V_{DS} = 2\ V$ gives $\lambda_n V_{DS} = 0.1$ which is much less than one.

In the above expressions, g_m is the transconductance of the MOSFET (as defined previously) and go is called the output conductance. The inverse of go is called the output resistance, $r_o = g_o^{-1}$. Graphically,the output conductance corresponds to the slope of the transistor's drain characteristic at the operating point, which is the derivative of i_D with respect to v_{DS}, while keeping the gate-source voltage constant (see Figure 2.24(a)). In the small-signal model of the MOSFET, the output conductance can be included as shown in Figure 2.24(b). This representation follows directly

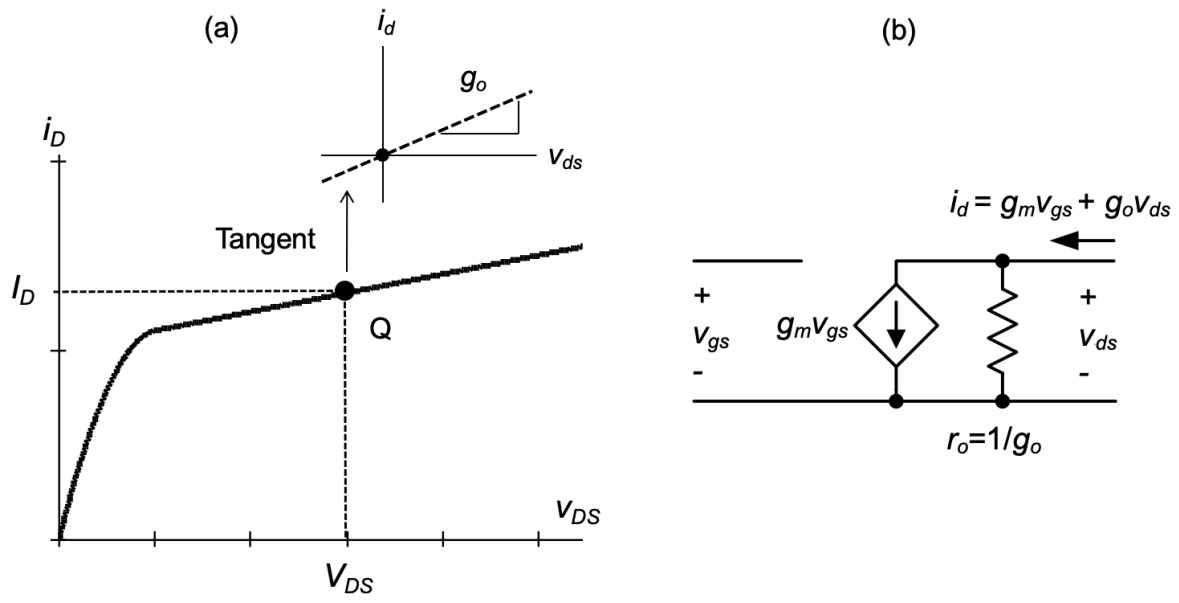


Figure 2.24.: (a) Graphical interpretation of g_o. (b) MOSFET small-signal model for the saturation region with output conductance (g_o).

from Equation 2.47, which represents Kirchhoff's Current Law equation for the drain node of the transistor.

Just as with the simple g_m-only small-signal model of Figure 2.13, the main idea for the usage of the extended model with go is to use the small-signal equivalent circuit of Figure 2.24(b) in a larger circuit. We will illustrate this using two examples of interest: the intrinsic voltage gain stage of Figure 2.23(a) and the common-source amplifier of Figure 2.20(a).

2.3.2. Common-Source Voltage Amplifier Analysis Using the λ-Model

For the intrinsic voltage gain stage, the small-signal model of Figure 2.24(b) corresponds directly to the small-signal model for the entire circuit with $v_{in} = v_{gs}$ and $v_{out} = v_{ds}$. Therefore, the voltage gain of the intrinsic gain stage is given by

$$A_v = -\frac{g_m}{g_o} = -g_m r_o \cong -\frac{2I_D}{V_{OV}} \cdot \frac{1}{\lambda_n I_D} = -\frac{2}{\lambda_n V_{OV}} \tag{2.50}$$

From this expression, we can see that the voltage gain can be increased by increasing L, which will decrease λ. Alternatively, the voltage gain can be increased by reducing V_{OV}, which corresponds to reducing the drain current I_D for a fixed aspect ration W/L. In this context, note that for $V_{OV} \to 0$, Equation 2.50 predicts infinite voltage gain. This non-physical outcome stems from the same issue already discussed in Section 2-2-7: for \$V_{OV} $\to$ 0, g_m approaches infinity for a fixed current in our simplistic square-law I-V model. As argued before, the usable range for V_{OV} must therefore be lower-bounded as specified in Equation 2.35. Assuming $V_{OV} = V_{OVmin} = 150\ mV$

and L = 1 μm, the intrinsic voltage gain of a MOSFET described by the parameters used in this module is approximately $2/(0.1 \cdot 0.15) = 133$.

Let us now consider the common-source amplifier of Figure 2.20(a). Including the finite output conductance from the λ-model, the small-signal model is modified as shown in Figure 2.25 and the voltage gain expression becomes

$$A_v = -g_m \left(\frac{1}{r_o} + \frac{1}{R_D} \right)^{-1} = -g_m R_{out} \tag{2.51}$$

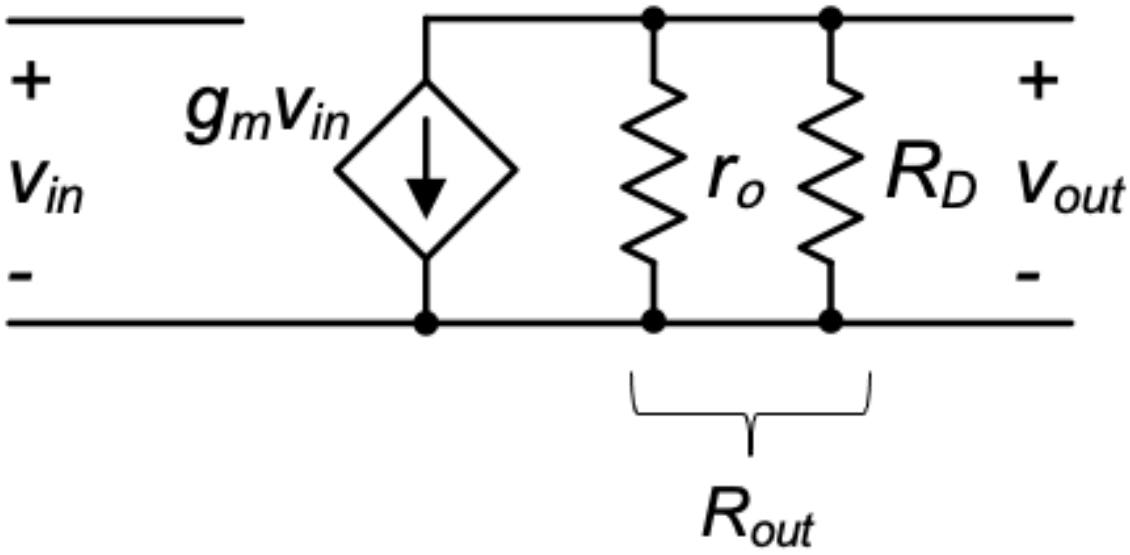


Figure 2.25.: Small-signal model for the circuit of Figure 2.20(a), with finite output conductance.

For $R_D \to \infty$, the small signal voltage gain approaches the intrinsic voltage gain as given by Equation 2.50. For $R_D \ll r_o$, we can approximate $A_v \cong -g_m R_D$. More generally, without even knowing the exact values of r_o and R_D, we can argue that as long as the desired gain is much less (in magnitude) than the intrinsic voltage gain, r_o can be neglected in the voltage gain calculation. To see this, we can rewrite Equation 2.51 as

$$\frac{1}{|A_v|} = \frac{1}{g_m r_o} + \frac{1}{g_m R_D}$$

$$g_m R_D = \frac{|A_v|}{1 - \frac{|A_v|}{g_m r_o}} \cong |A_v| \quad \text{for} \quad |A_v| \ll g_m r_o \tag{2.52}$$

Example 2-6: Analysis of a CS Amplifier Using the λ -Model

Consider the CS voltage amplifier of Figure 2.20(a) with $W = 80\ \mu m$, $L = 2\ \mu m$ and $R_D = 50\ k\Omega$. The gate is biased such that $V_{OV} = 500mV$ and V_B is set to 2 V, which is also the desired output operating point V_{OUT}. Compute the required bias current I_B and the small-signal voltage gain of the circuit using the λ-model. Repeat the small-signal voltage gain calculation for $R_D = 5\ k\Omega$.

SOLUTION

The bias current is found using

$$I_B = I_D = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{GS} - V_{Tn})^2(1 + \lambda_n V_{DS})$$

$$\frac{1}{2}\cdot 50\frac{\mu A}{V^2}\cdot\frac{80}{2}(0.5V)^2(1 + 0.05\cdot 2)$$

$$= 275\mu A$$

The corresponding transconductance and output resistance at the operating point are

$$g_m = \frac{2I_D}{V_{OV}} = \frac{2\cdot 275\mu A}{0.5V} = 1.1mS$$

$$r_o = \frac{1 + \lambda_n V_{DS}}{\lambda_n I_D} = \frac{1 + 0.05\cdot 2}{0.5V^{-1}\cdot 275\mu A} = 80k\Omega$$

According to Equation 2.51, the small signal voltage gain for $R_D = 50\ k\Omega$ is given by

$$A_v = -1.1mS(\frac{1}{80k\Omega} + \frac{1}{50k\Omega})^{-1} = -33.9$$

for $R_D = 5\ k\Omega$ we find

$$A_v = -1.1mS(\frac{1}{80k\Omega} + \frac{1}{5k\Omega})^{-1} = -5.18$$

Since the voltage gain for $R_D = 5\ k\Omega$ is much less than the intrinsic voltage gain of the transistor ($g_m r_o = 88$), it is appropriate to neglect r_o in this calculation. We can simply compute

$$A_v = -1.1mS\cdot 5k\Omega = -5.5$$

This result differs only by about 5.9% from the accurate calculation.

Another opportunity for useful engineering approximations in the application of the λ-model lies in the operating point calculation. We will illustrate this point through the example below.

Example 2-7: Approximate Operating Point Calculations

Recalculate I_D, g_m and r_o by approximating $\lambda V_{DS} \cong 0$ in the large-signal bias point calculations. Also recalculate A_v for $R_D = 50\ k\Omega$ using this approximation. Compare the results to the values obtained in Example 2-6.

SOLUTION

For $\lambda\ V_{DS}$ 0, the bias current is estimated as

$$I_B = I_D = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{GS} - V_{Tn})^2$$

$$= \frac{1}{2}\cdot 50\frac{\mu A}{V^2}\cdot\frac{80}{2}(0.5V)^2$$

$$= 250\mu A$$

The corresponding transconductance and output resistance at the operating point are

$$g_m = \frac{2I_D}{V_{OV}} = \frac{2\cdot 250\mu A}{0.5V} = 1mS$$

$$r_o = \frac{1}{\lambda_n I_D} = \frac{1}{0.5V^{-1}\cdot 250\mu A} = 80k\Omega$$

and the voltage gain becomes

$$A_v = -1mS(\frac{1}{80k\Omega} + \frac{1}{50k\Omega})^{-1} = -30.8$$

Relative to the accurate calculation from Example 2-6 (A_v = –33.9), this result is in error by only about 9.1%.

There are several reasons why it is commonly acceptable to neglect the λV_{DS} term in bias point hand calculations. First, without this approximation, the calculations can become cumbersome and lead to transcendental equations that are tedious and undesirable to solve in light of only a moderate percent-improvement in the obtained accuracy. If a more accurate result is desired, it can often be obtained more easily from computer simulations, which often follow a hand calculation in practice anyway. Last, one can argue that any circuit in which the operating point parameters strongly depend on λ may be impractical in the first place. The accuracy of the λ-model as far as absolute I-V values are concerned can only be approximate due to its empirical nature. For high accuracy analysis, much more complex models (such as the one described in Reference 7) must be used, carefully calibrated with physical measurements and subsequently evaluated in computer simulations. For the purpose of developing an introductory feel for circuits, however, the λ-model is still the most appropriate, mainly due its simplicity.

Table 2-3 summarizes the technology parameters introduced in this chapter.

Table 2.3.: Standard technology parameters for the λ-model.

Parameter	**n-channel MOSFET**	**p-channel MOSFET**
Threshold voltage	V_{Tn} = 0.5V	V_{Tp} = -0.5V
Transconductance parameter	$\mu_n C_{ox} = 50\mu A/V^2$	$\mu_p C_{ox} = 25\mu A/V^2$
Chanel length modulation parameter	$\lambda_n = 0.1V^{-1}/L$ (L in μ m)	$\lambda_p = 0.1V^{-1}/L$ (L in μ m)

2.4. Two-Port Model for the Common-Source Voltage Amplifier

The circuit shown in Figure 2.25 corresponds to a transconductance amplifier model (with voltage output) since we modeled the MOSFET as a voltage controlled current source, which is in line with its physical behavior in the saturation region. Alternatively, and since the circuit is meant to function as a voltage amplifier, we can equivalently model it using a native voltage amplifier two-port as shown in Figure 2.26. The reader can prove that for this model the open-circuit voltage gain is given by $A_v = -g_m R_{out}$, where R_{out} corresponds to the parallel connection of R_D and r_o. The native voltage amplifier model is sometimes preferred since it more directly expresses the intended function we assumed in this chapter. However, as we shall see in Chapter 3, the voltage amplifier two-port is no longer a convenient representation for the CS voltage amplifier when device capacitances are included.

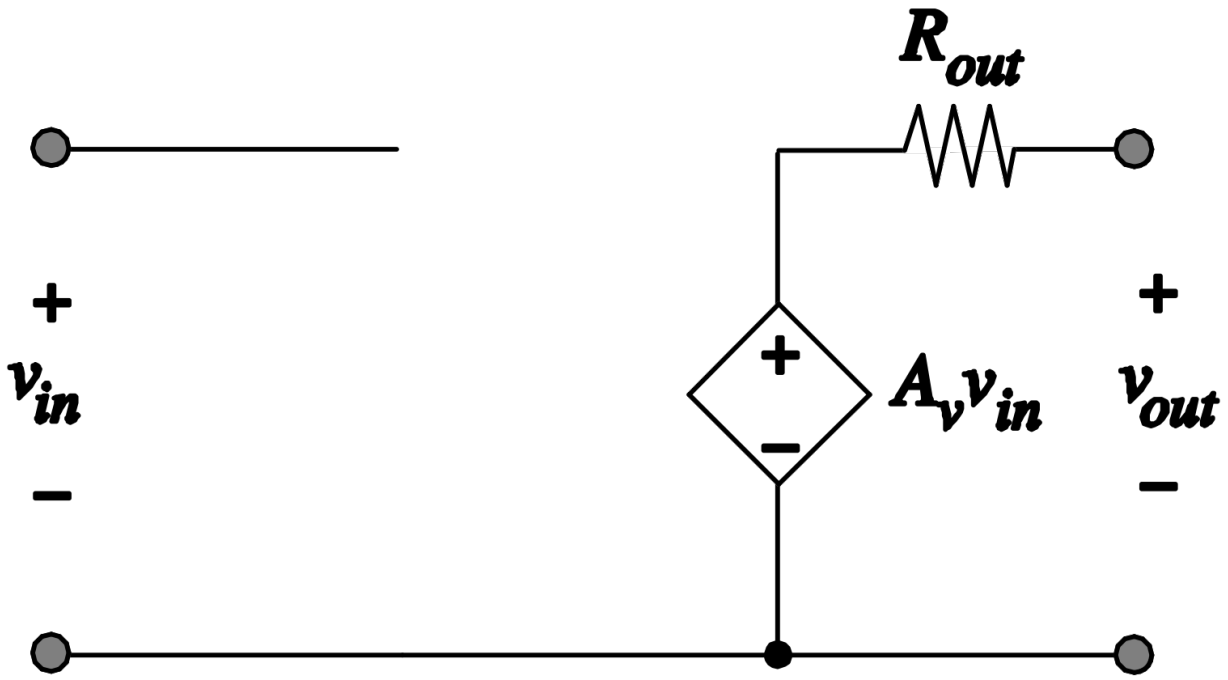


Figure 2.26.: Equivalent voltage amplifier-based model for the common-source amplifier circuit Figure 2.25.

Figure 2.27 shows the small-signal commonsource voltage amplifier model together with an input transducer and load resistance. Since the input resistance of the CS amplifier is infinite, no resistive division takes place at the input port and $v_{in} = v_s$. Thus, the overall voltage gain is computed as

$$A_v' = \frac{v_{out}}{v_s} = A_v \cdot \left(\frac{R_L}{R_L + R_{out}} \right) \tag{2.53}$$

The same expression can be found from the circuit of Figure 2.25 with R_L included across the output port. The reader is invited to prove this.

As a final note, we should emphasize that adding a load resistance may have implications on the bias point of the circuit. For example, if the load resistor carries a DC current, this current will

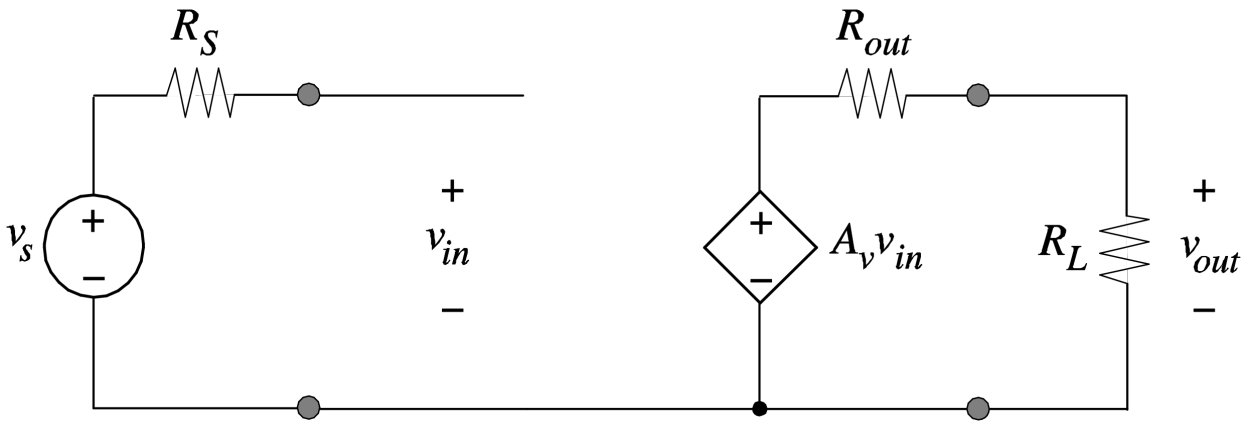


Figure 2.27.: Common-source amplifier model with transducer and load resistance.

affect the quiescent point output voltage of the amplifier. In this case, the load resistance must also be connected to the circuit in the operating point analysis where the small-signal parameters such as r_o are computed. Problem 2-17 looks at an example.

2.5. Summary

In this chapter we reviewed the basic I-V characteristics of a MOSFET and employed this device to construct examples of common-source voltage amplifiers. In deriving a model for the MOSFET, we concentrated on first-order effects that define the transistor's operation. The nonlinear nature of even the simplest device model dictates the use of small-signal approximations to enable analyses with manageable complexity. The presented methodology begins by finding the operating point of the transistor. Next, the small-signal equivalent is used to construct a linear small-signal model for further analysis.

We studied the basic common-source voltage amplifier with drain resistance and found that the voltage gain in this circuit is directly proportional to the voltage drop across the resistor, which imposes practical limits on the achievable voltage gain. A modified circuit based on an auxiliary bias current source was then analyzed as an alternative and used as a motivation to incorporate the effect of channel length modulation using the λ-model.

The most important concepts that you should have mastered are:

- Determining MOSFET drain currents in all regions of operation; determining the regions of operation based on the transistor's terminal voltages.
- Constructing transfer characteristics and load line plots for common-source stages for various drain bias configurations.
- Calculating the operating point of a CS stage and the MOSFET's small-signal parameters g_m and r_o.
- Drawing the small-signal model of a CS stage and calculating its voltage gain.

2.6. References

1. R. F. Pierret, *Semiconductor Device Fundamentals*, Prentice Hall, 1995.
2. P. R. Gray, P. J. Hurst, S. H. Lewis, and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*, 5th Edition, Wiley, 2008.
3. D. O. Pederson and K. Mayaram, *Analog Integrated Circuits for Communication: Principles, Simulation and Design*, 2nd Edition, Springer,
4.
5. Y. Tsividis, *Operation and Modeling of the MOS Transistor*, 2nd Edition, McGraw-Hill, New York, 1999.
6. P. G. A. Jespers, *The g_m/I_D Methodology, a Sizing Tool for Low-Voltage Analog CMOS Circuits*, Springer, 2010.
7. M. J. M. Pelgrom, et al., *"Matching properties of MOS transistors,"* IEEE J. of Solid-State Circuits, pp. 1433-1439, May 1989.
8. G. Gildenblat, et al., "PSP: An Advanced Surface-Potential-Based MOSFET Model for Circuit Simulation," *IEEE Trans. Electron Devices,* pp. 1979-1993, Sept. 2006.

2.7. Problems

Unless otherwise stated, use the standard model parameters specified in Table 2.3 for the problems given below. Consider only first-order MOSFET behavior and include channel length modulation (as well as any other second-order effects) only where explicitly stated.

P2.1 An n-channel transistor biased in saturation with $W/L = 10$ carries a drain current of 200 μA when $V_{GS} = 1.5\ V$ is applied. With $V_{GS} = 1\ V$, the current drops to 50 μA. Determine V_{Tn} and μC_{ox} of this transistor.

P2.2 Show that two MOS transistors in series with channel lengths L_1 and L_2 and identical channel widths can be modeled as one equivalent MOS transistor with length $L_1 + L_2$ (see Figure 2.28). Assume that M_1 and M_2 have identical parameters except for their channel lengths. Hint: there are (at least) two ways to solve this problem. One is through extensive algebra; the other is though physical insight and arguments based on the MOSFET cross-section.

P2.3 Derive an analytical expression for the input voltage that corresponds to the transition point (point C) between the saturation and triode regions in Figure 2.7(b).

P2.4 In Example 2-4, we calculated the most negative input excursion that the circuit in Figure 2.10 can handle before the output is clipped to the supply voltage. In this problem, calculate the most positive input excursion that can be applied before the MOSFET enters the triode region. Assume the same parameters as in Example 2-4: $V_{DD} = 5\ V$, $R_D = 10\ k\Omega$, W/L = 10, and V_{IN} is adjusted to 1.5 V, so that $V_{OUT} = 2.5\ V$ at the circuit's operating point.

P2.5 For the circuit shown in Figure 2.29, sketch v_{OUT} as a function of v_{IN}. Assume that v_{IN} varies from 0 to 5 V. There is no need to carry out any detailed calculations; simply draw a qualitative

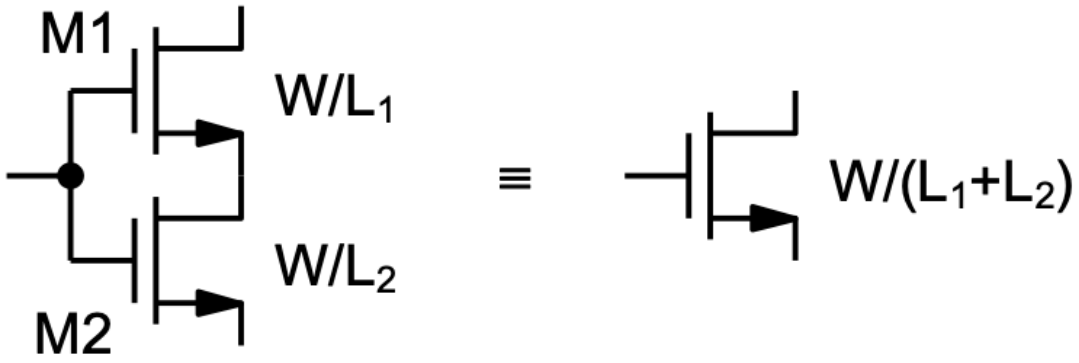


Figure 2.28.

graph and mark pertinent asymptotes and breakpoints (such as changes in the MOSFETs region of operation).

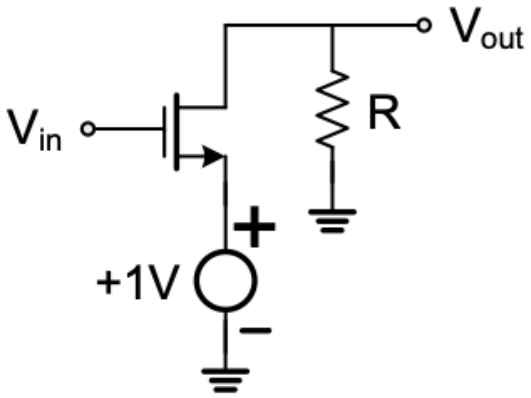


Figure 2.29.

P2.6 Derive the small-signal voltage gain given by Equation 2.27 through direct differentiation of Equation 2.17, i.e., apply 2.25.

P2.7 A field effect transistor built using a new (fictitious) material behaves "almost" exactly like a conventional MOSFET in silicon technology. The large signal I-V characteristic (in the saturation region) is given by

$$I_D = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{GS} - V_{Tn})^{2.5}$$

Assuming that $\mu C_{ox} W/L = 100\ \mu A/V^{2.5}$ and $I_D = 1\ \mu$A,

compute the transconductance of the device.

P2.8 The ratio of the small-signal drain current excursion and quiescent point drain current (i_d/I_D) is sometimes called the drain modulation index. Show that for a MOSFET, i_d/I_D is twice as large as the relative excursion in the gate-source voltage, v_{gs}/V_{OV}.

P2.9 Consider the CS amplifier shown Figure 2.10(a). Calculate the small-signal voltage gain assuming $V_{IN} = 1.5\ V$, W/L = 20, $R_D = 5\ k\Omega$, and $V_{DD} = 5\ V$. Be sure to check the device's region of operation.

P2.10 Consider the p-channel CS amplifier shown in Figure 2.15(a). Assuming $W/L = 20$, $R_D = 5\ k\Omega$, and $V_{DD} = 5\ V$, calculate the required quiescent point input voltage so that $V_{OUT} = 2.5\ V$. What is the small-signal gain of the circuit?

P2.11 Repeat Problem 2.9 with $R_D = 10\ k\Omega$. With this value, the MOSFET operates in the triode region. Compute the small-signal voltage gain by writing the large-signal relationship between the input and output for the triode region and subsequent differentiation at the operating point.

P2.12 Consider the cascade connection of two CS amplifiers as shown in Figure 2.30.

(a) Draw the small-signal equivalent model for this circuit.

(b) Calculate the small-signal voltage gains v_{out1}/v_{in} and v_{out2}/v_{in}. Assume that $V_{IN} = V_{DD}/2$ and that the device sizes are chosen such that the bias points of nodes v_{out1} and v_{out2} are also exactly at $V_{DD}/2$.

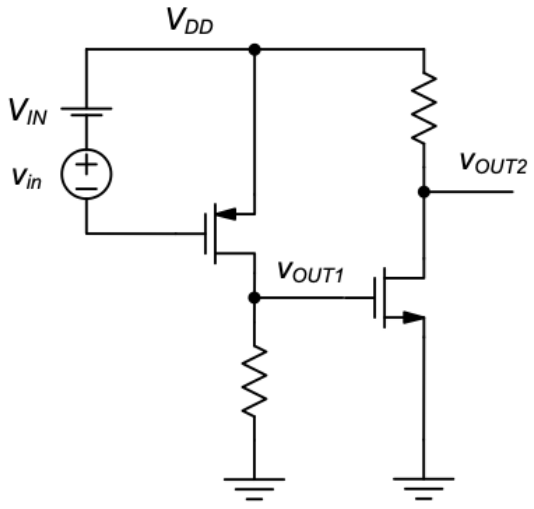


Figure 2.30.

P2.13 Repeat Problem 2.9 and include the effect of channel length modulation in the small-signal voltage gain calculation. Neglect channel length modulation in the bias point calculation. Quantify the percent difference in the calculated small-signal voltage gain compared to Problem 2.9. Assume that the channel length of the MOSFET is 2 μm.

P2.14 For the p-channel common-source amplifier shown in Figure 2.15(a)

(a) Given $W/L = 12\ \mu m/2\ \mu m$ and $R_D = 10\ k\Omega$, calculate V_{IN} such that V_{OUT} is 2.5 V. V_{DD} = 5V. Neglect channel length modulation in this calculation.

(b) What is the small-signal voltage gain, $A_v = v_{out}/v_{in}$? Include the effect of channel-length modulation in your calculation.

(c) To increase the voltage gain, you increase R_D to 100 $k\Omega$. Calculate the new small-signal voltage gain, A_v. You must re-bias the circuit so that $V_{OUT} = 2.5\ V$.

(d) We could also try to increase the voltage gain of the initial circuit by increasing W/L rather than R_D. Calculate the new A_v if W/L= 120 μm /2 μm and $R_D = 10\ k\Omega$. Be sure to re-bias the circuit so that $V_{OUT} = 2.5\ V$.

P2.15 Calculate the small-signal gain of the intrinsic gain stage shown in Figure 2.23(a), assuming W= 10 μm and using the parameters given below. In each case, explicitly calculate the gate overdrive voltage (V_{OV}), the transconductance (g_m) and the output resistance (r_o). Assume that the circuit is biased such that the MOSFET operates in the saturation region. Neglect V_{DS} dependence in the calculation of r_o .

(a) $I_D = 100\ \mu$A, $L = 2\ \mu$m

(b) $I_D = 50\ \mu$A, $L = 2\ \mu$m

(c) $I_D = 100\ \mu$A, $L = 4\ \mu$m

P2.16 Consider the CS amplifier of Figure 2.20(a) with the following parameters: $V_B = 2.5\ V$, $I_B = 500\,\mu$A, $W = 20\,\mu$m, $L = 1\,\mu$m, and $R_D = 5\ k\Omega$.

(a) Calculate the exact value of V_{IN} required such that $I_D = I_B$, and $V_{OUT} = V_B$. Include the effect of channel length modulation in your calculation.

(b) Using the value of V_{IN} found in part (a), re-compute V_{OUT} for the two cases when λ changes by +50% and -50%, respectively. Such discrepancies may be due to variations in the semiconductor process or simply due to uncertainty in the simplistic λ-model.

P2.17 Consider the CS amplifier shown in Figure 2.31 with the following parameters: $V_{DD} = 5\ V$, $V_{IN} = 1.8\ V$, W = 15 μm, L = 1 μm, $R_D = 5\ k\Omega$, and $R_L = 10\ k\Omega$.

(a) Calculate the output quiescent point voltage and drain current of the circuit, taking the connected load resistance R_L into account. Ignore channel-length modulation in this calculation.

(b) Calculate g_m and r_o using the parameters from part (a). Take V_{DS} dependence into account.

(c) Draw a two-port voltage amplifier model for the circuit and calculate the open-circuit voltage gain (A_v) and overall voltage gain (A_v') with the load resistor connected.

(d) Repeat parts (a) and (b) without considering the connected RL and recompute the two-port parameters of part (c). Summarize the observed differences.

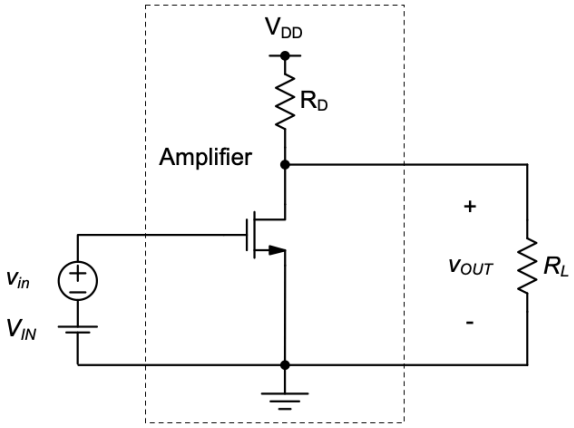


Figure 2.31.

3. Frequency Response of the Common-Source Voltage Amplifier

In the previous chapter, we have analyzed the common-source stage in terms of its static voltage transfer characteristic and did not consider any dynamic effects in the relationship between the circuit's input and output. The obtained results are therefore applicable only in the limit of slowly varying signals, and further analysis is needed to predict limits in the circuit's operating speed.

In most electronic circuits, the speed of operation is fundamentally limited by the presence of undesired capacitive elements. Therefore, for the purpose of including dynamic effects in the common-source voltage amplifier, we will expand the MOSFET model with its capacitive elements. In the spirit of the just-in-time modeling approach followed in this module, we first consider primary effects related to **intrinsic capacitance**, i.e., capacitance that is unavoidable and required for the operation of a MOSFET. We then refine our analysis to include **extrinsic capacitances**. Extrinsic capacitances are not required for the operation of a MOSFET, but nonetheless exist due to limitations or properties of a certain device structure or manufacturing process.

The analysis and inclusion of device capacitance will follow the small-signal modeling approach used in Chapter 2. That is, even though most MOSFET device capacitances are inherently non-linear, we will approximate them using linear elements at the MOSFET's operating point. At the various stages of the model development, we consider the dynamics of the amplifier for small-signal, sinusoidal inputs in the steady-state. Specifically, we evaluate the phase and magnitude of the amplifier's output signal to quantify its behavior as a function of frequency.

Even though the small-signal abstraction greatly simplifies the analysis of circuit dynamics, we will find that further simplifications and tools are needed to reason quickly and intuitively about the limiting effects. Therefore, this chapter includes a treatment of the **dominant pole approximation**, the **Miller theorem**, the **Miller approximation**, and the **open-circuit time constant** (OCT) analysis. These techniques are broadly applicable and useful for the analysis of a wide range of circuits, going far beyond the motivational common-source stage example treated in this chapter.

Chapter Objectives

- Review the basic concepts of frequency domain analysis.
- Extend the small-signal MOSFET model with intrinsic and extrinsic device capacitances.
- Derive the sinusoidal steady-state frequency response of the common-source stage at various levels of capacitance modeling and circuit abstraction.
- Review and develop tools and approximation methods that help simplify the frequency response analysis of a circuit: dominant pole approximation, Miller theorem, Miller approximation, and open-circuit time constant analysis.

3.1. Review of Frequency Domain Analysis

In this section, we will review important pre-requisite material using the RC circuit shown in Figure 3.1 as a driving example. Our objective is to gain insight into the circuit's behavior when a sinusoidal signal of a given frequency is applied at its input. Since the circuit consists of linear elements, it follows that the output can only contain a sinusoid at the same frequency that is applied. Therefore, all we need to determine is the amplitude and phase of the output sinusoid. Note that even though we restrict ourselves to sine waves, the analysis results are generally useful since arbitrary periodic signals can be constructed from a sum of sinusoids

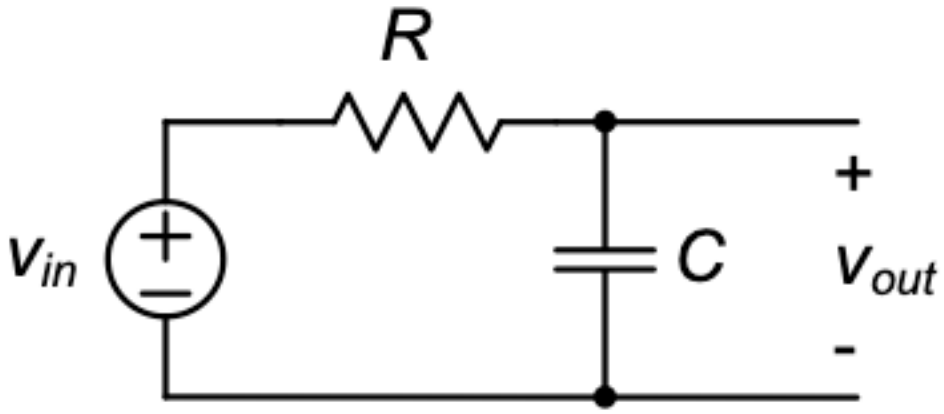


Figure 3.1.: RC circuit example.

From first principles, we could approach this problem by applying KCL and KVL, noting that the current flowing through a capacitor is given by $C \cdot dv/dt$. The result of this analysis is a linear differential equation that links $v_{in}(t)$ and $v_{out}(t)$. This equation can be solved for a sinusoidal input, yielding in general two components that make up the output. The first is called the transient part; it decays to zero for $t \to \infty$. The second is called the **steady-state** component, and it persists for all t. This latter component is what we are interested in.

A convenient shortcut to obtain the steady-state response is to work with **Laplace transform** models for each circuit element and to determine the transfer functions in the s-domain. Once an s-transfer function is created, the circuit's steady-state response to a sinusoidal input is found by letting $s = j\omega$ and by computing the phase and the magnitude of the output as a function of frequency (ω). The resulting characteristic is called the **frequency response** of the circuit and is usually plotted in the format of a **Bode plot**. The involved variables that capture how the magnitude and phase vary with frequency are called **phase vectors** or **phasors**. In this module,

the notation for phasors uses an uppercase variable name and lowercase subscripts such as V_{in} and I_{out}. We will now illustrate the flow of such an analysis using the RC circuit example

Example 3-1: Frequency Response of an RC Circuit

Find the magnitude and phase of the transfer function V_{out} /V_{in} in for the RC circuit in Figure 3.1.

SOLUTION

We begin by noting that in the s-domain, the reactance of a capacitor is given by $1/sC$. By applying the voltage divider rule, we can therefore write a transfer function that links v_{out} and v_{in} in the s-domain as follows

$$\frac{v_{out}}{v_{in}} = \frac{\frac{1}{sC}}{\frac{1}{sC} + R} = \frac{1}{1 + sRC}$$

In order to evaluate this transfer function for steady-state sinusoids, we let $s = j\omega$ and obtain

$$\frac{V_{out}}{V_{in}} = \left.\frac{v_{out}}{v_{in}}\right|_{s=j\omega} = \frac{1}{1 + sRC}$$

Following the rules for determining the magnitude and phase of a complex number, we obtain

$$\left|\frac{V_{out}}{V_{in}}\right| = \sqrt{\frac{1}{1 + (\omega RC)^2}}$$

and

$$\angle\frac{V_{out}}{V_{in}} = \tan^{-1}(-\omega RC)$$

From this result, we see that for $\omega RC \gg 1$ the sinusoid is attenuated and shifted by –90°, i.e.

$$\left|\frac{V_{out}}{V_{in}}\right| = \frac{1}{\omega RC} \qquad \angle\frac{V_{out}}{V_{in}} \simeq -90^\circ$$

For $\omega RC \ll 1$, the sinusoid is passed unattenuated and with no phase shift, i.e.,

$$\left|\frac{V_{out}}{V_{in}}\right| \simeq 1 \qquad \angle\frac{V_{out}}{V_{in}} \simeq 0^\circ$$

This result makes intuitive sense, since the capacitor carries a larger current for high frequencies, increasingly "shorting" the output port and attenuating the signal. At high frequencies, the phase approaches –90° due to the signal differentiation that takes place in the capacitor. Its current is given by $C \cdot dv/dt$, and differentiation of a sine wave yields a cosine wave that is –90° shifted in phase.

3.1.1. Bode Plots

In order to gain further insight from the magnitude and phase of a circuit, it is customary to plot the response in the form of a Bode plot, which shows the log of the magnitude versus the log of the frequency, and the phase angle versus the log of the frequency. In this representation, the magnitude and phase can be inspected over many orders of magnitude in frequency.

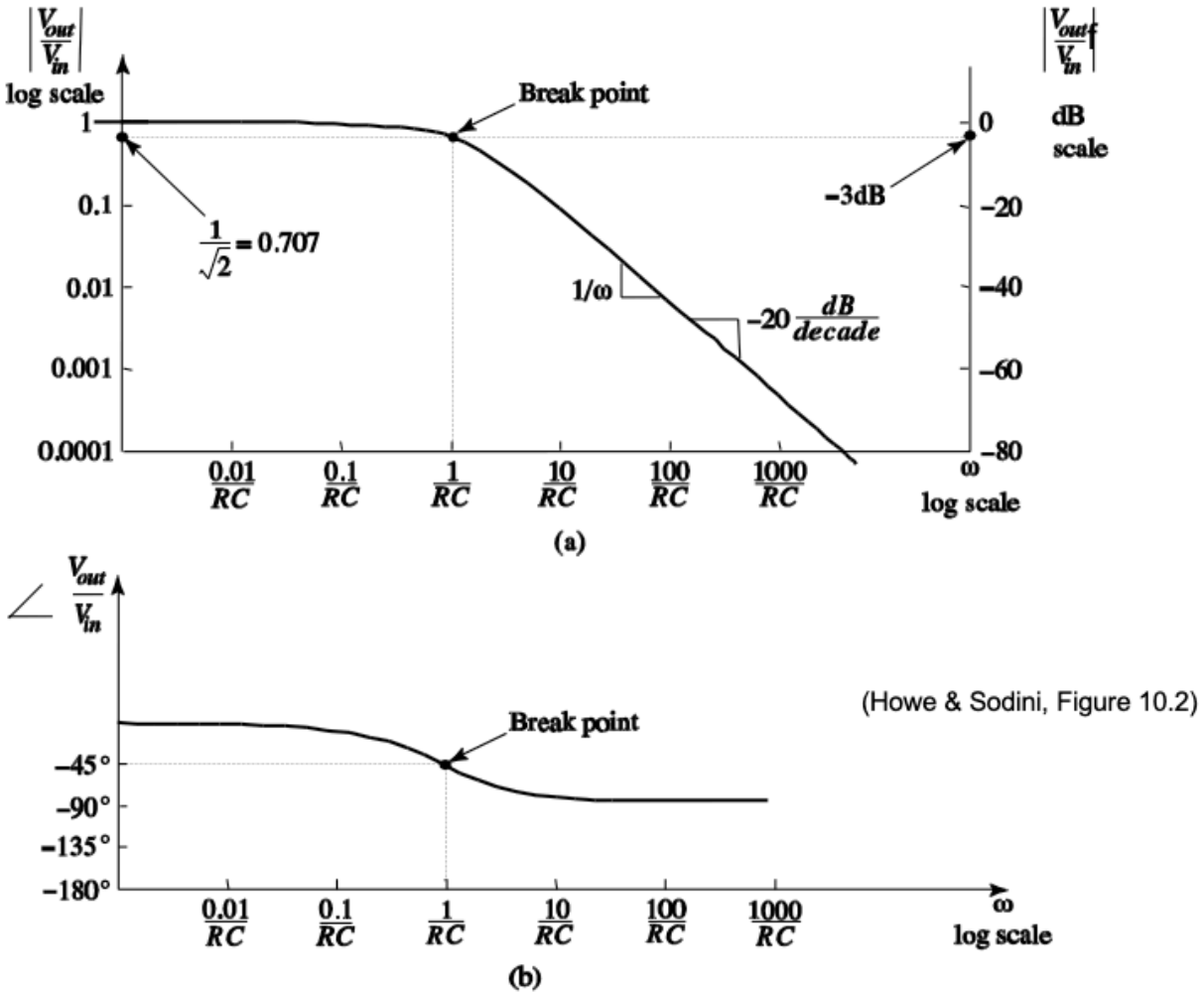


Figure 3.2.: Bode plot for the RC circuit example of Figure 3.1. (a) Log magnitude vs. log frequency. (b) Phase vs. log frequency.

A Bode plot for the circuit of Figure 3.1 is shown in Figure 3.2. A few interesting features can be identified from this plot as follows. First, recall from the analysis of the circuit that for very high frequencies, where $\omega \gg 1/RC$, the magnitude of the transfer function becomes inversely proportional to frequency. This is seen in the high-frequency region of the plot where the magnitude decreases by a factor of 10 for every factor of 10 increase in ω. Second, an interesting point in the Bode plot is where $\omega = 1/RC$, also called the **breakpoint frequency**. At the breakpoint, the magnitude is given by

$$\left|\frac{V_{out}}{V_{in}}\right| = \left|\frac{1}{1+j\omega RC}\right| = \left|\frac{1}{1+j}\right| = \frac{1}{\sqrt{2}} = 0.707 \tag{3.1}$$

and the phase is

$$\angle\frac{V_{out}}{V_{in}} = \tan^{-1}(-\omega RC) = \tan^{-1}(-1) = -45^\circ \tag{3.2}$$

It is customary to express the logarithmic magnitude scale on a Bode plot with a dimensionless unit called a **decibel** (dB). The magnitude of the ratio of voltages in units of dB is:

Ratio of voltages in decibels: $20\ \log\left|\frac{V_{out}}{V_{in}}\right|$ dB

Therefore, in terms of decibels (indicated on the right-hand y-axis in Figure 3.2) the magnitude falls at –20 dB/decade at high frequencies. Expressed in decibels, the magnitude of the voltage at the breakpoint frequency $\omega = 1/RC$ is $20\log(1/\sqrt{2}) \simeq -3$ dB.

The bandwidth of a circuit is a measure for the frequency range across which it exhibits only a small amount of attenuation. For a low-pass circuit (such as the RC circuit under investigation), the bandwidth is defined as the frequency for which the magnitude has dropped by a factor of $1/\sqrt{2}$ relative to its value at $\omega = 0$ (DC gain). Since $1/\sqrt{2}$ corresponds to -3 decibels, we refer to this quantity as the 3-dB bandwidth, or symbolically

$$\omega_{3dB} = \frac{1}{RC} \tag{3.3}$$

As an additional example, we will now look at the frequency response of the RC circuit with an additional resistor added in series with the capacitor C, as shown in Figure 3.3.

Example 3-2: RC Circuit with Additional Resistor

Find the magnitude and phase of the voltage transfer function for the circuit in Figure 3.3 and draw the corresponding Bode plot.

SOLUTION

By applying the voltage divider rule, we find

$$\frac{v_{out}}{v_{in}} = \frac{\frac{1}{sC} + R}{\frac{1}{sC} + R + R} = \frac{1+sRC}{1+2sRC}$$

Next, we let $s = j\omega$ and obtain

$$\frac{V_{out}}{V_{in}} = \left.\frac{v_{out}}{v_{in}}\right|_{s=j\omega} = \frac{1+j\omega RC}{1+2j\omega RC}$$

and finally

$$\left|\frac{V_{out}}{V_{in}}\right| = \sqrt{\frac{1+(\omega RC)^2}{1+2(\omega RC)^2}}$$

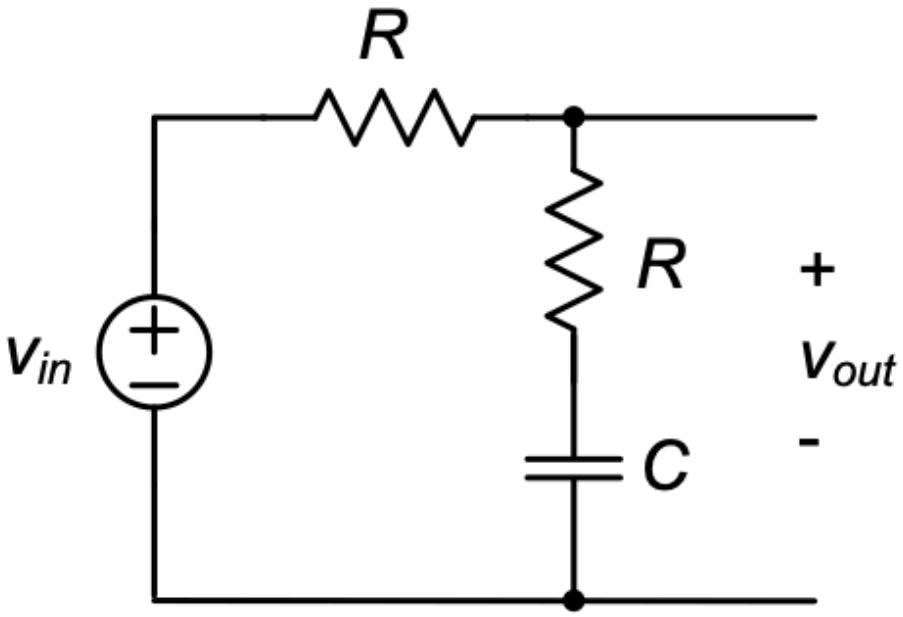


Figure 3.3.: RC circuit with series resistor.

and

$$\angle\frac{V_{out}}{V_{in}} = \tan^{-1}(-\omega RC) + tan^{-1}(-2\omega RC)$$

The Bode plot for these expressions is found in Figure 3.4. As we can see, the plot is similar to the previous example in terms of the low-frequency behavior and first breakpoint. There is, however, a second breakpoint beyond which the magnitude approaches a constant value of –6 dB (= 0.5), and the phase begins to return back to 0°. This behavior is intuitively understood by inspection of the circuit. At high frequencies, the capacitor becomes a short, essentially leaving a resistive voltage divider. Since the resistors are of equal value, the voltage attenuation approaches 0.5 at high frequencies. Similarly, the phase returns to 0° because the resistive division at high frequencies has no impact on the signal's phase.

3.1.2. Poles and Zeros

In linear system theory, **poles** and **zeros** are the s-values for which the value of the s-domain transfer function becomes infinity or zero, respectively. Since the behavior of a linear system is fully determined by the location of its poles and zeros, it is desirable to factor the transfer function in the following general format:

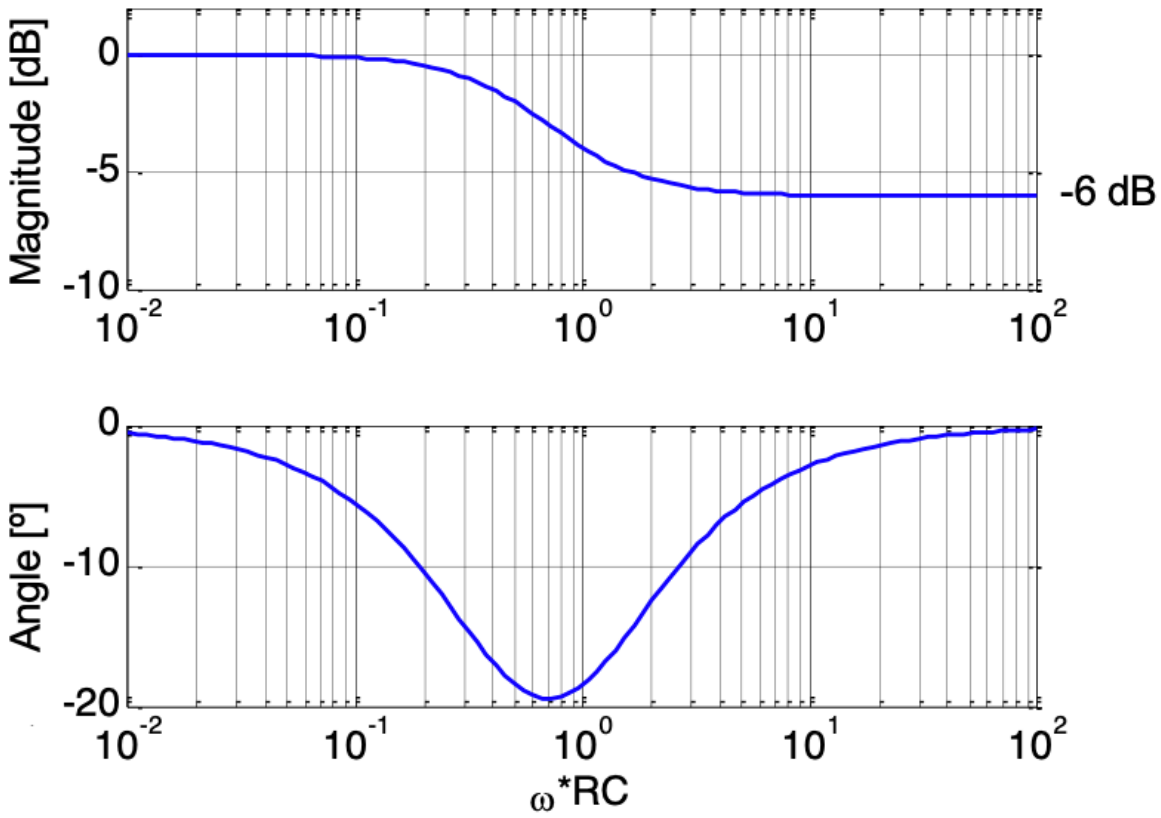

Magnitude [dB]
0
-5
-10
-6 dB
10^-2
10^-1
10^0
10^1
10^2
Angle [°]
0
-10
-20
10^-2
10^-1
10^0
10^1
10^2
ω*RC

Figure 3.4.

$$H(s) = K\frac{\left(1-\frac{s}{z_1}\right)\left(1-\frac{s}{z_2}\right)...\left(1-\frac{s}{z_m}\right)}{\left(1-\frac{s}{p_1}\right)\left(1-\frac{s}{p_2}\right)...\left(1-\frac{s}{p_n}\right)} \tag{3.4}$$

where K is a constant DC gain term, $p_1, p_2, ...p_n$ are the poles and $z_1, z_2, ..., z_m$ are the zeros. For example, the s-domain transfer function of Example 3-2 is given by

$$H(s) = \frac{v_{out}}{v_{in}} = \frac{\left(1-\frac{s}{z_1}\right)}{\left(1-\frac{s}{p_1}\right)} \tag{3.5}$$

where

$$z_1 = -\frac{1}{RC} \quad and \quad p_1 = -\frac{1}{2RC} \tag{3.6}$$

The reason why p_1 and z_1 are called poles and zeros can be understood from the plot in Figure 3.5, which evaluates Equation 3.5 using the complex argument $s = \sigma + j\omega$. At $s = p_1$, the magnitude of $H(s)$ becomes infinite, resembling the pole of a tent holding up the 2-dimensional sheet in this representation. Likewise, at $s = z_1$, the magnitude of $H(s)$ becomes zero; this could be viewed as pegs pinning down the tent at this particular location.

Since the steady-state magnitude response of the circuit is obtained by letting $s = j\omega$, it simply corresponds to the bold line marked at the front edge of the plot. In other words, evaluating $H(s)$ for the magnitude response corresponds to "walking" on the sheet of Figure 3.5 along the ω axis.

As we can see from Equation 3.6, the poles and zeros of the example considered here are (negative) real numbers. For arbitrary ratios of polynomials in s, the poles and zeros as expressed in Equation 3.4 can be complex numbers. For all circuits considered in this module, however, the poles and zeros will be real. Furthermore, all poles will be negative, as required for a stable system. The zeros encountered in this module can be either positive or negative as in 3.6. A negative zero is called a **left half plane (LHP) zero**, since it lies on the left side of the s-plane. A positive zero is called a **right half plane (RHP) zero**, since it lies on the right hand side of the s-plane.

When all the poles and zeros of a system are real, it is possible to create a set of rules that allow the construction of a bode plot by inspection. These rules are summarized in the next section.

3.1.3. Bode Plots of Arbitrary System Functions with Real Poles and Zeros

For the case of real negative poles and zeros, and letting $s = j\omega$, Equation 3.4 becomes

$$H(j\omega) = K\frac{\left(1+j\frac{\omega}{\omega_{z1}}\right)\left(1+j\frac{\omega}{\omega_{z2}}\right)...\left(1+j\frac{\omega}{\omega_{zm}}\right)}{\left(1+j\frac{\omega}{\omega_{p1}}\right)\left(1+j\frac{\omega}{\omega_{p2}}\right)...\left(1+j\frac{\omega}{\omega_{pn}}\right)} \tag{3.7}$$

where $\omega_{p1}, \omega_{p2}, ...\omega_{pn}$ are the pole frequencies and $\omega_{z1}, \omega_{z2}, ..., \omega_{zm}$ are the zero frequencies. For instance, in Example 3-2, we have

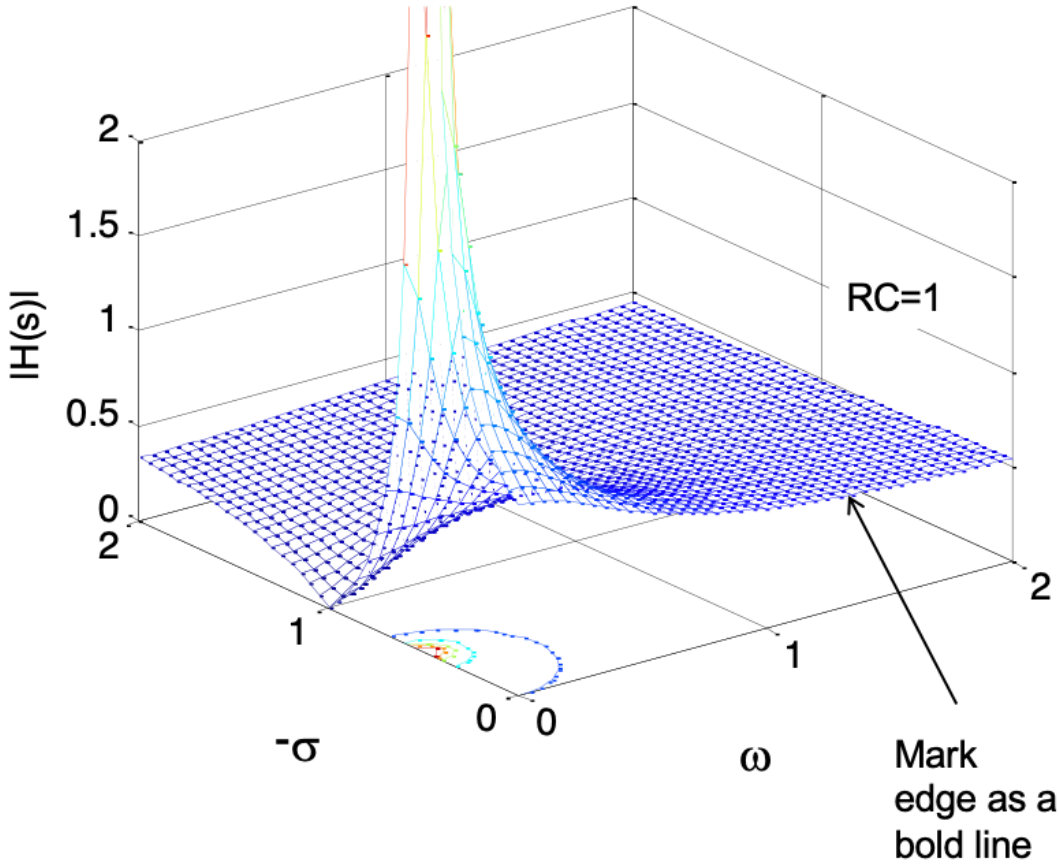


Figure 3.5.: 3-D plot of the magnitude of Equation 3.5, evaluated for $s = \sigma + j\omega$.

$$H(j\omega) = K\frac{\left(1+j\frac{\omega}{\omega_{z1}}\right)}{\left(1+j\frac{\omega}{\omega_{p1}}\right)} \tag{3.8}$$

where

$$\omega_{z1} = \frac{1}{RC} \quad and \quad \omega_{p1} = \frac{1}{2RC} \tag{3.9}$$

To determine the Bode plot from Equation 3.7, we must assess the effect of each binomial term on the magnitude and phase of the system function. If the frequency is such that $\omega \ll \omega_{zi}$ or ω_{pi}, then the respective binomial term will have little effect on the magnitude and phase of the system function, as it will simply multiply it by unity. On the other hand, if the frequency is such that $\omega \ll \omega_{zi}$ or ω_{pi}, the system function, magnitude, and phase will be altered. To see this, we evaluate the magnitude and phase of a general binomial term for a left half plane pole or zero and $\omega \gg \omega_i$

$$\left|1+j\frac{\omega}{\omega_i}\right| = \sqrt{1+\left(\frac{\omega}{\omega_i}\right)^2} \simeq \frac{\omega}{\omega_i} \tag{3.10}$$

$$\angle\left(1+j\frac{\omega}{\omega_i}\right) = tan^{-1}\left(\frac{\omega}{\omega_i}\right) \simeq 90^\circ \tag{3.11}$$

Therefore, if the binomial term is in the numerator of the generalized system function (corresponding to a LHP zero), the magnitude will be multiplied by ω/ω_i, and a phase angle of 90° will be added to the total phase. If the binomial term is located in the denominator (LHP pole), the magnitude will be multiplied by $1/(\omega/\omega_i)$ and a phase angle of 90° will be subtracted from the total phase. For a RHP zero, it follows that the magnitude will be multiplied by ω/ω_i, and a phase angle of 90° will be subtracted from the total phase.

When $\omega = \omega_i$, the magnitude and phase are

$$\left|1+j\frac{\omega}{\omega_i}\right| = |1+j| = \sqrt{2} \tag{3.12}$$

$$\angle(1+j) = 45^\circ \tag{3.13}$$

Therefore, if these binomial terms for the breakpoints are located in the numerator, the magnitude of the system function in the numerator is multiplied by $\sqrt{2}$ and a phase of 45° is added to (for a LHP zero) or subtracted from (for a RHP zero) the overall phase. If it is located in the denominator, the magnitude is multiplied by $1/\sqrt{2}$ and a phase of 45° is subtracted from the overall phase of the system function.

Given these results, a Bode plot can be constructed by referring to the following step-by-step procedure.

- Identify all the breakpoint frequencies ω_{pi} and ω_{zi} and list them in increasing order. Apply the following rules, beginning with the lowest breakpoint frequency.

- If the corresponding binomial term appears in the numerator of the system function, the magnitude slope will be increased by 20 dB/decade, when the frequency is greater than the breakpoint frequency.
- If the corresponding binomial term appears in the denominator of the system function, the magnitude of the slope will be reduced by 20 dB/decade when the frequency is greater than the breakpoint frequency.
- To plot the phase, we know that the binomial term will contribute $+45^\circ$ for a LHP zero, and -45° for a RHP zero at $\omega = \omega_i$. If it is in the denominator, it will contribute -45°. We assume that the $\pm 90^\circ$ phase changes linearly over the interval $0.1\omega i < \omega < 10\omega_i$.

Example 3-3: Bode Plot Construction

Construct a Bode plot for a system with the following parameters: $K = 100$, $\omega_{p1} = 10\ rad/s$, $\omega_{p2} = 100\ krad/s$, left half plane zero: $\omega_{z1} = 1\ krad/s$, right half plane zero: $\omega_{z2} = 10\ Mrad/s$.

SOLUTION

First we note that the DC gain $K = 100 = 40$ dB. Next we recognize that ω_{p1} is the lowest frequency term, creating a change of slope in the magnitude plot toward –20 dB/decade. The phase is 0° at the lowest frequency plotted, -45° at ω_{p1} and has reached -90° at approximately 10 ω_{p1}. Applying the given rules in a similar fashion to the remaining poles and zeros yields the Bode plot shown in Figure 3.6.

3.2. Frequency Response of the Common-Source Voltage Amplifier — First-Pass Analysis

We now wish to apply the analysis tools reviewed in the previous section to get a handle on the frequency response of the common-source voltage amplifier discussed in Chapter 2. Since the exact frequency behavior of this circuit is quite complex when taking all aspects into account, we partition this discussion into two steps. This section presents the first analysis step and uses the simplest possible model extension for the MOSFET that can be used to take capacitive effects, and thus frequency dependence, into account.

In the context of MOSFET capacitance modeling, it is useful to distinguish between intrinsic and extrinsic capacitances. Here, the term extrinsic refers to capacitances that are not needed to operate a MOSFET, but rather exist due to limitations or properties of a certain device structure or manufacturing process. As we shall see in Section 3-3, stray capacitances between the gate and source/drain terminals are examples of extrinsic capacitors. Intrinsic capacitance is unavoidable and required to operate the device. The oxide capacitance of a MOSFET falls into this category: without a capacitance between the gate and channel, no mobile charges can be induced ($Q = CV$), and the MOSFET would not function. In this section, we will look at frequency dependence effects due to the intrinsic capacitance only, beginning with a derivation of a circuit model that can be used to model this capacitance in the frequency response calculations.

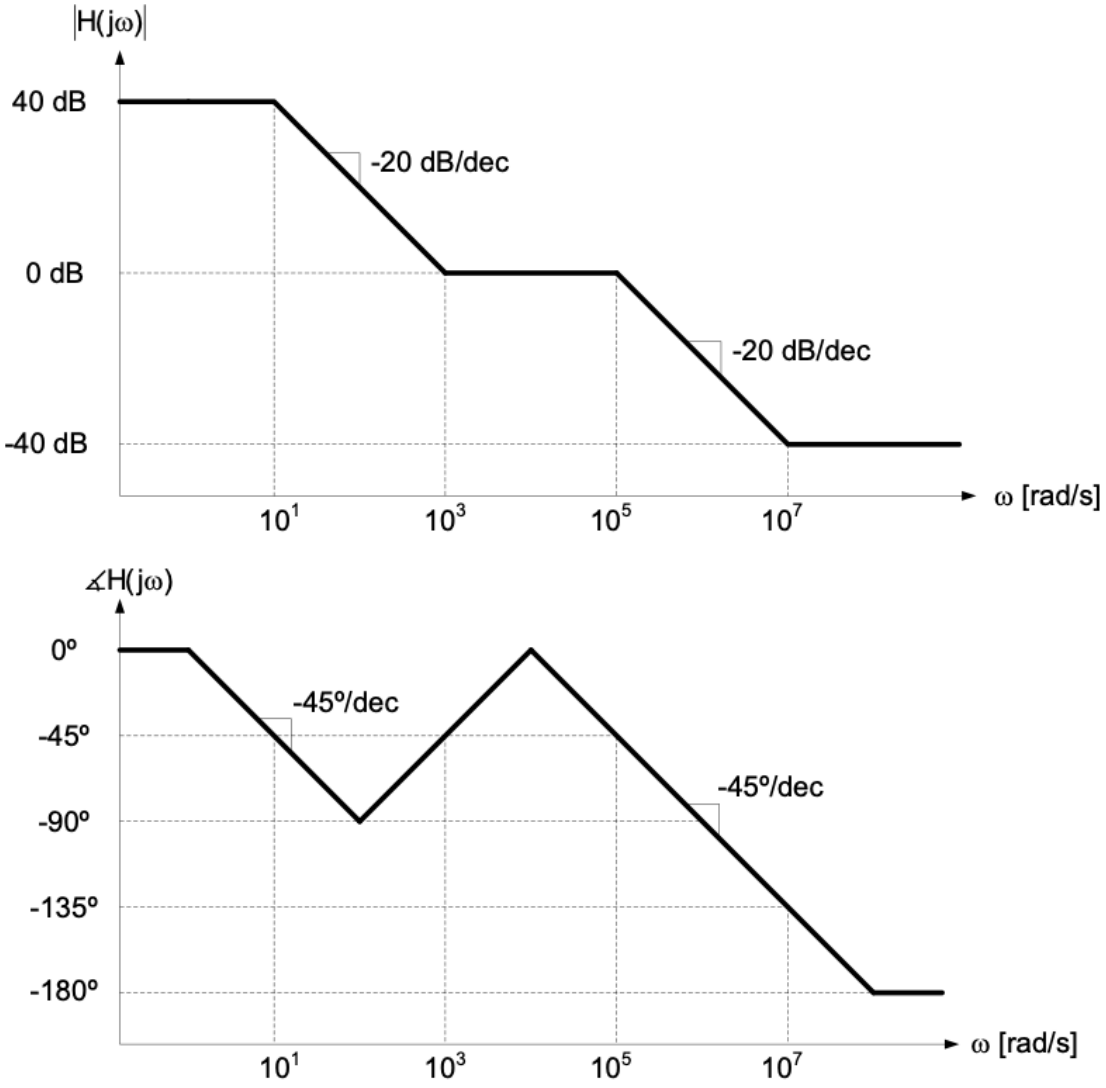

|H(jω)|
40 dB
0 dB
-40 dB
-20 dB/dec
-20 dB/dec
ω [rad/s]
10^1
10^3
10^5
10^7
∠H(jω)
0°
-45°
-90°
-135°
-180°
-45°/dec
-45°/dec
ω [rad/s]
10^1
10^3
10^5
10^7

Figure 3.6.

3.2.1. Modeling Intrinsic MOSFET Capacitance

Just as in the derivation of device transconductance and output conductance, the operating point must be considered when calculating small-signal capacitances. We begin by analyzing the intrinsic capacitance of a MOSFET in the triode region, with its cross-section shown in Figure 3.7(a). To first-order, the gate and the conductive channel can be viewed as a parallel plate capacitor, resulting in a gate-to-channel capacitance of

$$C_{gc} = WL\frac{\epsilon_{ox}}{t_{ox}} = WLC_{ox} \tag{3.14}$$

where WL is the capacitor plate area and C_{ox} is the oxide capacitance per unit area.

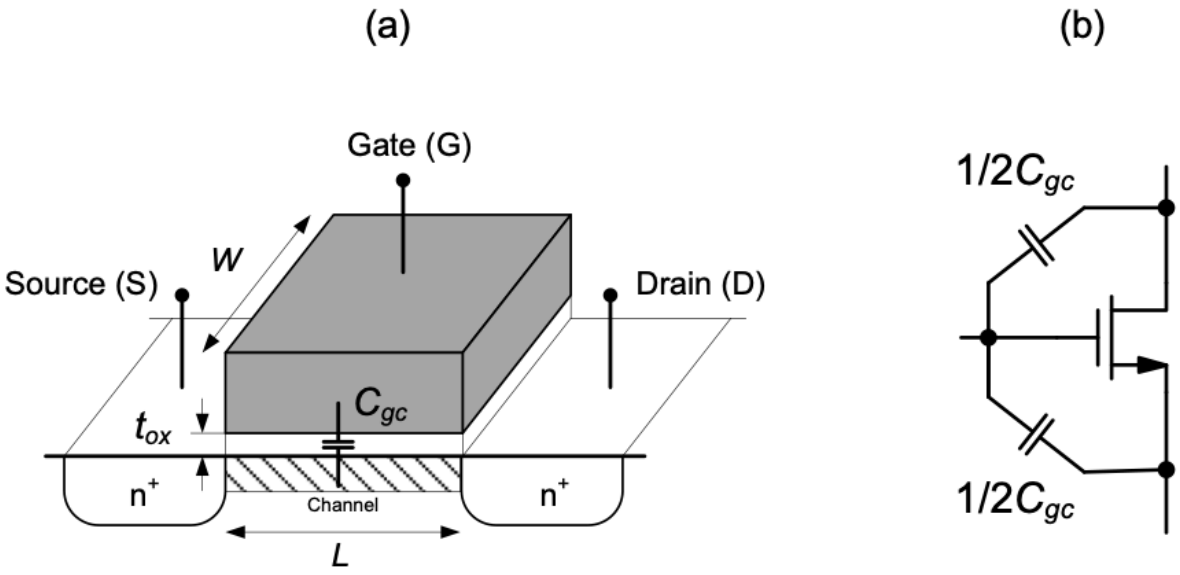


Figure 3.7.: (a) MOSFET cross-section showing the intrinsic capacitance between the gate and the channel (C_{gc}). (b) Capacitance model for the triode region.

If the source and drain were connected together, the small signal capacitance from the gate to source/drain would be equal to C_{gc} as given in Equation 3.14. How can we model the capacitance when source and drain are not connected, i.e., how is the capacitance distributed between the two terminals?

A common first-order approximation is to assign half of C_{gc} to the capacitance between the gate and the source and the remaining half between the gate and the drain. This is schematically illustrated in Figure 3.7(b). A qualitative argument that supports this approximation is that small changes in either the drain or source voltage must induce the same change in charge at the gate; therefore, the capacitance must be split equally.

A case that is more relevant to the analysis of a common-source stage is the behavior in the saturation region. For this case, we know that the conductive channel does not extend all the way from the source to the drain, but is pinched off at some coordinate L–ΔL. When the channel

is pinched-off, the drain potential (to first-order) no longer influences the charge under the gate. Therefore, the intrinsic capacitance from the gate to the drain is approximately zero in this region of operation.

In saturation, the channel charge is therefore controlled primarily by the potential between the gate and the source, and a significant capacitance is present between these two terminals. At first glance, one might expect that C_{gs} is equal to C_{gc}. However, this is not quite correct due to the pinch-off effect. Imagine applying a small voltage change to the source terminal. This will change the voltage across the oxide (and charge) near the source, but at the pinch-off point, the voltage across the oxide remains at V_{Tn}. This means that the capacitance in the saturation egion must be less than C_{gc}, because the charge does not see a uniform change as in the case of a simple parallel plate capacitor. Further analysis (see Reference 1) reveals that the capacitance between the gate and the source in the saturation region is given by

$$C_{gs} = \frac{2}{3}C_{gc} = \frac{2}{3}WLC_{ox} \tag{3.15}$$

The resulting small-signal MOSFET model that includes this capacitance is shown in Figure 3.8.

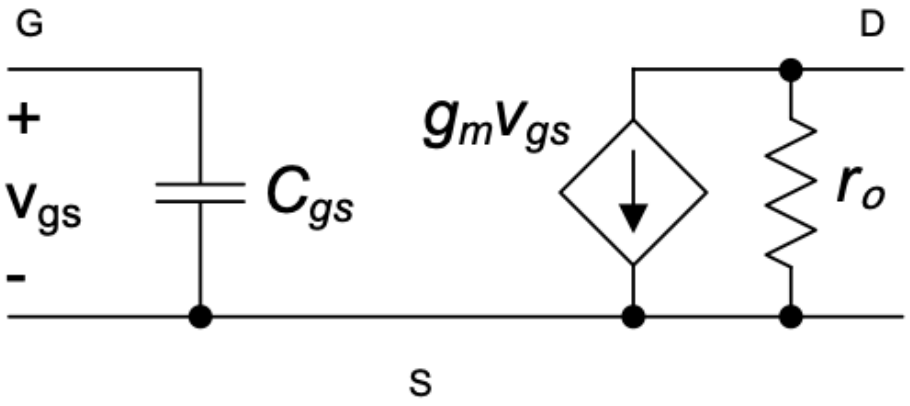


Figure 3.8.: MOSFET small-signal model for the saturation region, including the intrinsic gate capacitance.

3.2.2. Frequency Response with Intrinsic Gate Capacitance

To analyze the frequency response of the common-source amplifier with the intrinsic gate capacitance, we insert the model of Figure 3.8 into the small-signal circuit model of the amplifier, as shown in Figure 3.9.

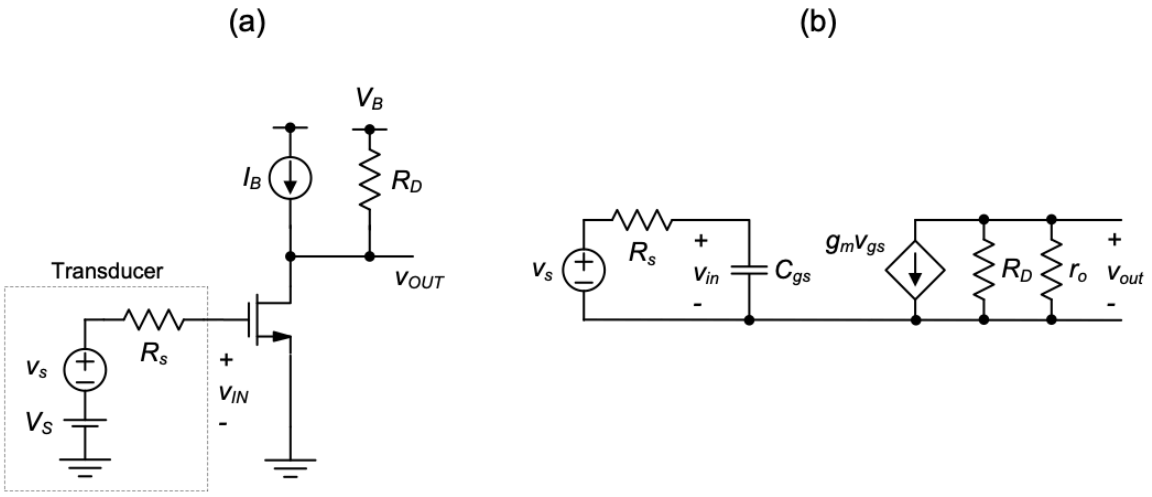


Figure 3.9.: (a) Common-source amplifier driven by a transducer with finite source resistance. (b) Small-signal model with the intrinsic gate capacitance.

Note that if the circuit were driven by an ideal voltage source at its input port (v_{in}), the added capacitance would have no effect on the circuit's operation. The ideal voltage source would provide any current that is needed to charge and discharge the gate capacitance without introducing any frequency dependence. The model in Figure 3.9 therefore considers a more realistic input source with finite resistance (R_s). At this point in the analysis, we purposely do not include any capacitive loading at the output of the amplifier, primarily to keep the first pass analysis simple and transparent.

In order to analyze the frequency response of the circuit in Figure 3.9, we first realize that the overall transfer function can be split into a product of two terms

$$\frac{v_{out}}{v_s} = \frac{v_{out}}{v_{in}} \cdot \frac{v_{in}}{v_s} \tag{3.16}$$

In this expression, the first term on the right-hand side corresponds to the DC voltage gain given in Equation 2.51, and is equal to $-g_m R_{out}$. The second term can be found by writing the voltage divider expression that relates node v_{in} to v_s

$$\frac{v_{in}}{v_s} = \frac{\frac{1}{sC_{gs}}}{\frac{1}{sC_{gs}} + R_s} = \frac{1}{1 + sR_sC_{gs}} \tag{3.17}$$

With this result, the complete s-domain transfer function from the input source to the output becomes

$$A_v(s) = \frac{v_{out}}{v_s} = \frac{-g_m R_{out}}{1 + sR_sC_{gs}} = \frac{A_{v0}}{1 + sR_sC_{gs}} \tag{3.18}$$

where $A_{v0} = A_v(0)$ is a generalized placeholder for the DC gain of the circuit. From this result, we see that the transfer function has a DC gain corresponding to the result of Chapter 2, and a single pole that is set by the source resistance and the intrinsic gate capacitance. As explained in Section 3-1, we can now evaluate this transfer function for steady-state sinusoids by letting $s = j\omega$. This will allow us to draw a Bode plot and compute the bandwidth of the circuit.

Example 3-4: Common-Source Amplifier Bandwidth Calculation

Calculate the 3-dB bandwidth for the amplifier in Figure 3.9, assuming $W = 20\mu m$, $L = 1\mu m$, $C_{ox} = 2.3\ fF/\mu m^2$ and $R_s = 50\ k\Omega$. Express the result in units of Hertz.

SOLUTION

For the given parameters, the gate-source capacitance is

$$C_{gs} = \frac{2}{3}WLC_{ox} = 30.67fF$$

The 3-dB bandwidth is

$$\omega_{3dB} = \frac{1}{R_s C_{gs}} = 652.3Mrad/s$$

and therefore

$$f_{3dB} = \frac{\omega_{3dB}}{2\pi} = 103.8MHz$$

An important question to ask at this point of the discussion is whether there is anything we can do to maximize the bandwidth of our amplifier. Assuming that we cannot change the source resistance R_s, the only remaining option is to minimize C_{gs}. This can be achieved by choosing a smaller transistor width or length [see Equation 3.15]. How will this affect the other performance metrics in the circuit? In the next subsection, we will show that there exists a direct tradeoff in the achievable bandwidth versus supply current for the circuit in consideration.

3.2.3. Tradeoff Between Bandwidth and Supply Current

Consider a design problem involving the circuit of Figure 3.9 and assume that the general objective is to maximize the circuit's 3-dB bandwidth while minimizing the transistor's drain current. For this analysis, we assume that R_s, R_{out} and A_{v0} are given through specifications, and that these parameters cannot be varied. This assumption is not atypical in practical circuit design. R_s might be fixed by the physical properties of the input transducer. R_{out} could be set by an output resistance requirement that allows the circuit to interface with subsequent circuit stages, while the DC gain A_{v0} could be determined by application requirements. Furthermore, for simplicity, we neglect channel-length-modulation in this analysis.

In order to study the tradeoff between bandwidth and current consumption, we will now write expressions for these quantities that rely on common parameters. For the 3-dB bandwidth, we begin by inserting Equation 3.15 into Equation 3.15 and obtain

$$\omega_{3dB} = \frac{3}{2} \cdot \frac{1}{R_s W L C_{ox}} \tag{3.19}$$

By using the following expression to eliminate C_{ox}:

$$g_m = \mu_n C_{ox} \frac{W}{L}(V_{GS} - V_{Tn}) = \mu_n C_{ox} \frac{W}{L} V_{OV} \tag{3.20}$$

and subsequently substituting $g_m = |A_{v0}|/R_{out}$, Equation 3.19 becomes

$$\omega_{3dB} = \frac{3}{2} \cdot \frac{\mu_n}{L^2} \cdot \frac{1}{|A_{v0}|} \cdot \frac{R_{out}}{R_s} \cdot V_{OV} \tag{3.21}$$

The above expression is now in a form that contains only technology parameters, design constraints (R_s, R_{out},and A_{v0}) and the gate overdrive voltage V_{OV} as a single design parameter. From this result, it is clear that in order to maximize bandwidth, we would like to use a technology that offers high mobility and short channels. The mobility is largely determined by material properties, while L is usually bounded by some $L = L_{min}$ that is specific to a certain process technology, for example, 1 μm for the transistors used in this module.

From Equation 3.21, we also see that we should maximize V_{OV}. However, a potential problem with this is due to the output signal range of the amplifier. As we know from Chapter 2, larger V_{OV} means that V_{DSsat} is also increased, and this means that the transistor enters the triode region at higher v_{OUT}. This could lead to clipping, as discussed previously.

An additional, and more fundamental issue relates to the current consumption of the circuit. To see this, we rewrite Equation 2.31 as

$$I_D = \frac{1}{2} \cdot g_m \cdot V_{OV} \tag{3.22}$$

and substitute $g_m = |A_v|/R_{out}$ to find

$$I_D = \frac{1}{2} \cdot \frac{|A_{v0}|}{R_{out}} \cdot V_{OV} \tag{3.23}$$

This result shows that a larger V_{OV} unfortunately requires a larger bias current for the transistor, and this is highly undesired in many applications, as for instance battery-powered devices.

While the above-observed tradeoff was discovered in the context of a particular circuit example, we will see throughout this module that the same tradeoff holds for all analog circuits. For a given technology and target specifications, current consumption directly scales with the circuit's 3-dB bandwidth requirements. An alternative, and more general way to capture the fundamental connection between supply current and bandwidth is to inspect the tradeoffs that pertain to the MOSFET in isolation of a specific circuit example, as discussed next.

To begin, note that the model in Figure 3.8 comes with "desired" and "undesired" elements and properties. The only aspect of the transistor that we value is its transconductance. The associated intrinsic capacitance and the supply current needed to create the transconductance are undesired.

Mathematically, we can identify the following figures of merit that capture the ratios between the desired and undesired quantities, in particular:

$$\frac{g_m}{I_D} = \frac{2}{V_{OV}} \propto \frac{1}{\sqrt{I}_D} \tag{3.24}$$

and

$$\frac{g_m}{C_{gs}} = \frac{3}{2} \cdot \frac{\mu_n}{L^2} \cdot V_{OV} \propto \sqrt{I}_D \tag{3.25}$$

The transconductance-to-current ratio, which is sometimes called the transconductance efficiency, deteriorates for larger V_{OV} (and larger I_D). On the other hand, the ratio of transconductance per intrinsic capacitance improves for larger V_{OV} (and larger I_D). This tradeoff is graphically illustrated in Figure 3.10. Note that as already pointed out in Section 2-2-7, the proportionality of g_m/I_D to $1/V_{OV}$ extends only down to a certain minimum gate overdrive, defined as V_{OVmin} in this module [see Equation 2.35].

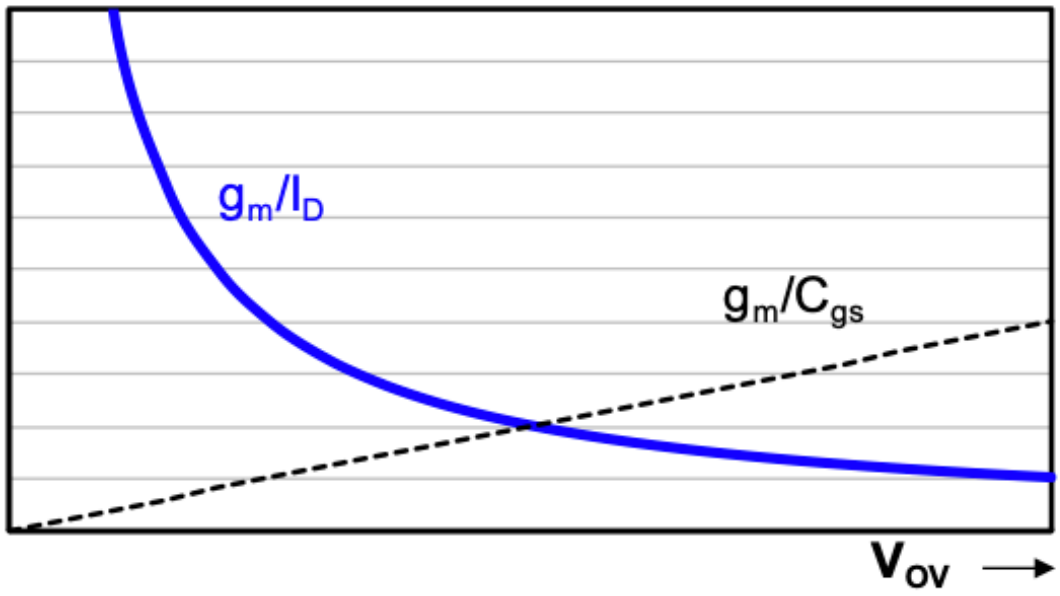


Figure 3.10.: Tradeoff between g_m/I_D and g_m/C_{gs}.

In essence, the gate overdrive voltage V_{OV} can be considered as a "knob" that lets us adjust the tradeoff between the two figures of merit. For a chosen V_{OV} and channel length, g_m/I_D and g_m/C_{gs} are fixed, and these parameters directly affect the speed and current consumption of the overall circuit. The gate overdrive V_{OV} has therefore been recognized by designers as an important parameter that affects most of the tradeoffs encountered in the optimization of a given circuit (see Reference 2). We will see examples of this throughout this module.

Interestingly, the product of the two figures of merit in Equation 3.24 and Equation 3.25 is given by

$$\frac{g_m}{C_{gs}} \cdot \frac{g_m}{I_D} = 3 \cdot \frac{\mu_n}{L^2} \tag{3.26}$$

From this result, it is clear that for high speed and low current consumption, the best we can hope for is a technology that provides high mobility and short channels. In this context, it is interesting to note that improvements in device engineering and manufacturing processes have provided tremendous improvements in manufacturable channel lengths. Since the 1970s, L_{min} has been improved from 10 μm to approximately 22 nm today; a $400x$ reduction!

3.2.4. Transit Frequency

The figure of merit given in Equation 3.25 is also known as the **transit frequency** of the transistor and coincidentally quantifies the frequency for which the magnitude of the transistor's current gain drops to unity. To determine the transit frequency, the transistor is operated in the common-source configuration and the input is driven by an ideal current source (see Figure 3.11). The output is short-circuited, and the current gain i_{out}/i_{in} is measured.

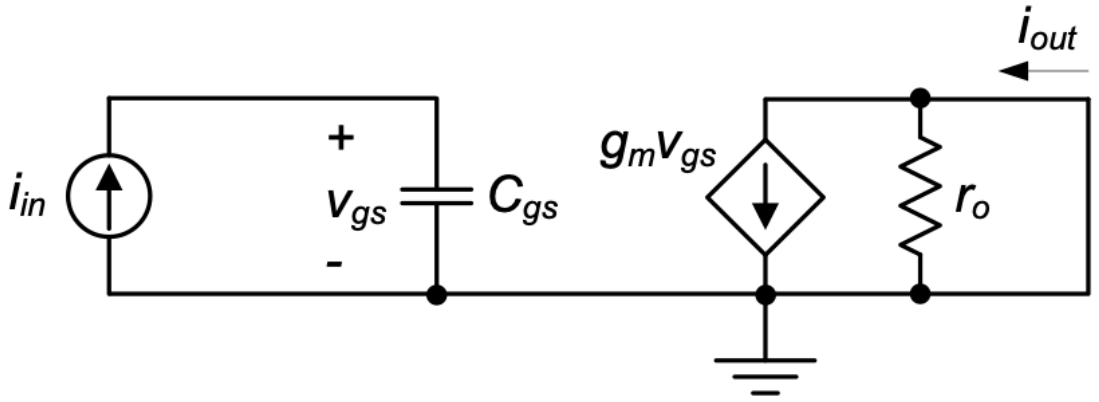


Figure 3.11.: Small-signal circuit model for finding the MOSFET's transit frequency.

From the circuit, it follows that

$$i_{out} = g_m \cdot v_{gs} = g_m \cdot \frac{i_{in}}{sC_{gs}} \tag{3.27}$$

Substituting $s = j\omega$ and rearranging yields

$$\frac{I_{out}}{I_{in}} = \frac{g_m}{j\omega C_{gs}} \tag{3.28}$$

The transit frequency then follows by setting

$$\left|\frac{I_{out}}{I_{in}}\right| = 1 = \frac{g_m}{\omega_T C_{gs}} \tag{3.29}$$

and therefore

$$\omega_T = \frac{g_m}{C_{gs}} \propto \sqrt{I_D} \tag{3.30}$$

The above quantity represents the transit frequency in rad/s. The symbol for the corresponding quantity in units of Hertz is $f_T = \omega_T/2\pi$.

The transit frequency gives the designer a feel for the maximum frequency at which a circuit can operate. The bandwidth of most practical circuit configurations is limited to a fraction of ω_T, often about one order of magnitude below.

3.3. Frequency Response of the Common-Source Voltage Amplifier—Second-Pass Analysis

We will now extend the results from the previous section to obtain a more accurate understanding of the frequency response of a realistic common-source amplifier. To begin, we will extend the MOSFET model to include extrinsic capacitances.

3.3.1. Modeling Extrinsic MOSFET Capacitance

Figure 3.12 shows the cross section of a MOSFET device for further study of its associated capacitive elements. The first component of extrinsic capacitance that we will consider is called **overlap capacitance**; it is due to overlap of the source and drain diffusions and the gate and the contribution of the fringe electric fields from the gate. The overlap capacitance C_{ov} is quantified as a linear capacitance proportional to the gate width, with units of $fF/\mu m$.

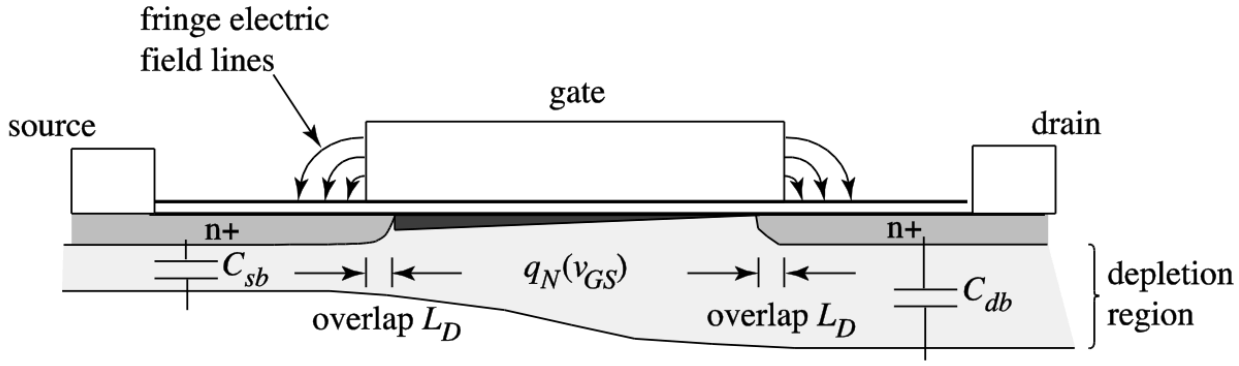
With overlap capacitance included, the total gate-source capacitance in saturation is the sum of Equation 3.15 and the overlap capacitance

$$C_{gs} = \frac{2}{3}WLC_{ox} + WC_{OV} \tag{3.31}$$

Since the drain has no influence on the channel charge, the only contribution to the gate-drain capacitance is C_{OV}

$$C_{gd} = WC_{OV} \tag{3.32}$$

In addition to the overlap capacitance, other extrinsic capacitance components are due to the reverse-biased junctions of the MOSFET. The drain-bulk and source-bulk capacitances C_{db} and C_{sb} indicated in Figure 3.12 originate from charge storage in the depletion regions between the drain



(Howe & Sodini, Figure 4.23)

Figure 3.12.: MOSFET cross section in saturation showing the overlap and fringe contributions to C_{ov}. The source-bulk and drain-bulk depletion capacitances are also shown qualitatively.

and source n+ regions and the p-type bulk. The following expressions can be used to estimate these capacitances (see Reference 1 for a derivation):

$$C_{db} = \frac{C_J \cdot AD}{(1 + V_{DB}/PB)^{MJ}} = \frac{C_{JSW} \cdot PD}{(1 + V_{DB}/PB)^{MJSW}} \tag{3.33}$$

$$C_{sb} = \frac{C_J \cdot AS}{(1 + V_{SB}/PB)^{MJ}} = \frac{C_{JSW} \cdot PS}{(1 + V_{SB}/PB)^{MJSW}} \tag{3.34}$$

In these expressions, V_{DB} and V_{SB} are the reverse bias voltages of the junctions at the operating point. Note that with increasing reverse bias, the values of the junction capacitances decreases. The geometry parameters used in the expressions are related to the layout of the transistor as shown in Figure 3.13.

- AD = Drain area
- AS = Source area
- PD = Perimeter of the drain diffusion (not including the edge under the gate)
- PS = Perimeter of the source diffusion (not including the edge under the gate)

All other parameters are defined in Table 3-1 along with the technology parameters introduced thus far.

The extrinsic capacitances discussed above are added to the MOSFET small signal model as shown in Figure 3.14. For completeness, this model contains an additional capacitance C_{gb} between gate and bulk. This capacitance is due to the overlap of the polysilicon gate onto the field oxide

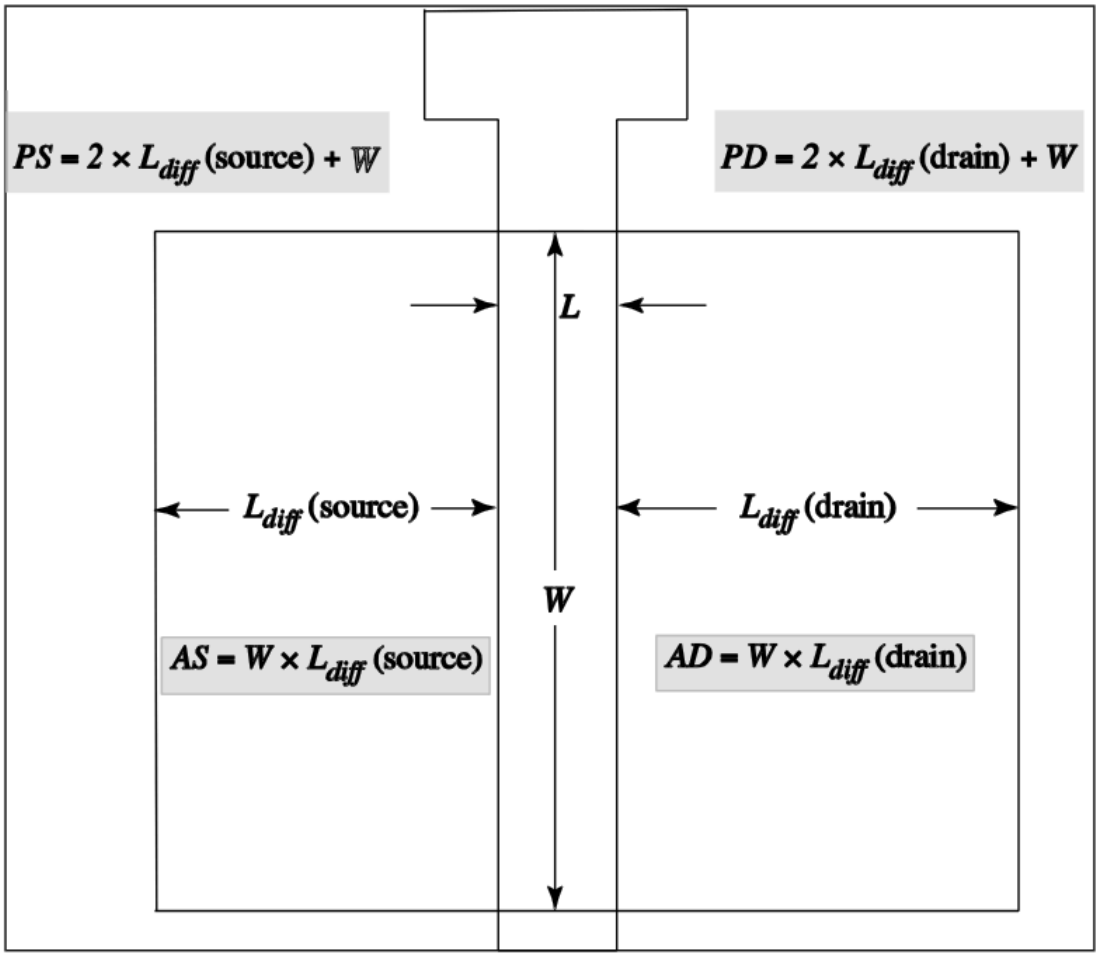


Figure 3.13.: Geometry parameters used for the calculation of junction capacitances.

region that isolates the MOSFET, as well as field lines from the gate terminating in the bulk of the transistor through the channel. This capacitance is usually small, and we will neglect it throughout this module.

Last, it is important to note that we have only modeled capacitances associated with the MOSFET, that is, the device without interconnections. The parasitic capacitances of the interconnections between MOSFETs can be a limiting factor and must be estimated from the layout and cross section for accurate analysis of a design. Off-chip wiring and package capacitances are also critical for evaluating the performance of any integrated circuit.

3.3.2. Transit Frequency with Extrinsic Capacitances

With extrinsic capacitances included in the model, the transit frequency expression of Equation 3.30 modifies to

$$\omega_T = \frac{g_m}{C_{gs} + C_{gd}} \tag{3.35}$$

This can be seen by inserting the model of Figure 3.14 into the test setup of Figure 3.14. C_{sb} and C_{db} are shorted to ground (assuming the bulk terminal is also grounded), while C_{gd} appears in parallel with Cgs.

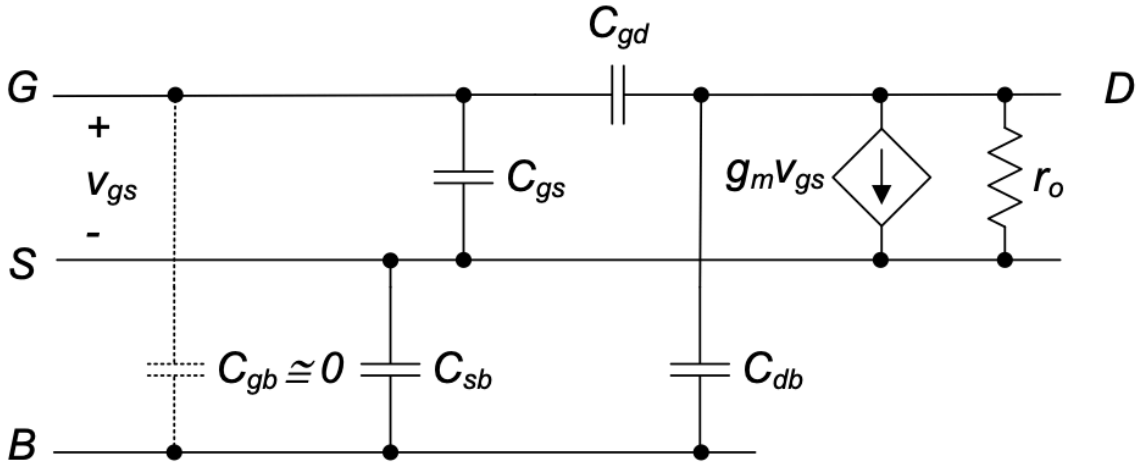


Figure 3.14.: Small-signal model for the n-channel MOSFET in saturation, including intrinsic and extrinsic capacitances.

Example 3-5: Mosfet Capacitance Calculation

Consider an n-channel MOSFET biased in saturation with $V_{DS} = 2.5\ V$, $I_D = 500\ \mu A$, $L = 1\ \mu m$, and $W = 20\ \mu m$. Determine all the capacitances in the small-signal model of Figure 3.14, except the gate-bulk capacitance C_{gb} that we consider negligible. Also calculate the transistor's transit frequency. Use the standard technology parameters defined in Table 3-1.

SOLUTION

Substituting C_{ox} and $C_{ov} = 0.5\ fF/\mu m$ into Equation 3.31, together with the MOSFET dimensions, we find

$$C_{gs} = \frac{2}{3}(20 \cdot 1\mu m^2)\left(2.3\frac{fF}{\mu m^2}\right) + 20\mu m\left(0.5\frac{fF}{\mu m}\right)$$

$$= 40.7fF$$

For the gate-drain capacitance, we obtain

$$C_{gd} = 20\mu m\left(0.5\frac{fF}{\mu m}\right) = 10fF$$

The remaining capacitances are the pn junction depletion capacitances C_{db} between the n+ drain and the substrate and C_{sb} between the n+ source and the substrate. Evaluating Equation 3.34, using the source junction bias voltage of $V_{SB} = 0\ V$ yields

$$C_{sb} = 19fF$$

The drain junction has a bias voltage of $V_{DB} = V_{OUT} = 2.5\ V$. Evaluating Equation 3.33 with this value and the given parameters gives

$$C_{db} = 11.6fF$$

Note that C_{db} is smaller than C_{sb} due to the larger reverse bias across the drain-bulk junction. To calculate the transit frequency, we first compute g_m using

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

$$= \sqrt{2 \cdot 50\frac{\mu A}{V^2} \cdot \frac{20}{1} \cdot 500\mu A} = 1mS$$

Therefore

$$f_T = \frac{1}{2\pi} \cdot \frac{g_m}{C_{gs} + C_{gd}} = \frac{1mS}{40.7fF + 10fF} = 3.14GHz$$

Table 3.1.: Standard technology parameters for the λ-model, with intrinsic and extrinsic capacitance parameters.

Parameter	**n-channel MOSFET**	**p-channel MOSFET**
Threshold voltage	$V_{Tn} = 0.5$V	V_{Tp} = -0.5V
Transconductance parameter	$\mu_n C_{ox} = 50\mu A/V^2$	$\mu_p C_{ox} = 25\mu A/V^2$
Chanel length modulation parameter	$\lambda_n = 0.1V^{-1}/L$ (L in μm)	$\lambda_p = 0.1V^{-1}/L$ (L in μm)
Gate oxide capacitance per unit area	$C_{ox} = 2.3fF/\mu m^2$	
Overlap Capacitance	$C_{ov} = 0.5fF/\mu m$	
Zero-bias planar bulk depletion capacitance	$C_{Jn} = 0.1fF/\mu m^2$	$C_{Jp} = 0.3fF/\mu m^2$
Zero-bias sidewall bulk depletion capacitance	$C_{JSWn} = 0.5fF/\mu m$	$C_{JSWp} = 0.35fF/\mu m$
Bulk junction potential	$PB = 0.95\ V$	
Planar bulk junction grading coefficient	$MJ = 0.5$	
Sidewall bulk junction grading coefficient	$MJSW = 0.33$	
Length of source and drain diffusions	$L_{diff} = 3\mu m$	

3.3.3. Frequency Response with Intrinsic and Extrinsic Gate Capacitances

To analyze the frequency response of the common-source amplifier with intrinsic and extrinsic capacitances, we insert the model of Figure 3.14 into the small-signal circuit model of the amplifier, as shown in Figure 3.15(a). Note that we have neglected C_{gb} and also discarded C_{sb}, since this capacitor has both terminals shorted to ground.

To simplify the full analysis of the amplifier, we redraw it as shown in Figure 3.15(b). We have taken the Norton equivalent at the input and combined the resistors at the input and output to reduce the number of terms carried in the algebra.

We begin the analysis by writing KCL at nodes 1 and 2

$$0 = -\frac{v_s}{R_s} + \frac{v_{gs}}{R_s} + v_{gs}sC_{gs} + (v_{gs} - v_{out})sC_{gd} \tag{3.36}$$

$$0 = g_m v_{gs} + sC_{gd}(v_{out} - v_{gs}) + \frac{v_{out}}{R_{out}} + sC_{db}v_{out} \tag{3.37}$$

Next, solving Equation 3.36 for v_{gs}, substituting into Equation 3.37, and rearranging yields

$$\frac{v_{out}}{v_s} = \frac{-g_m R_{out}\left(1 - s\frac{C_{gd}}{g_m}\right)}{1 + b_1 s + b_2 s^2} \tag{3.38}$$

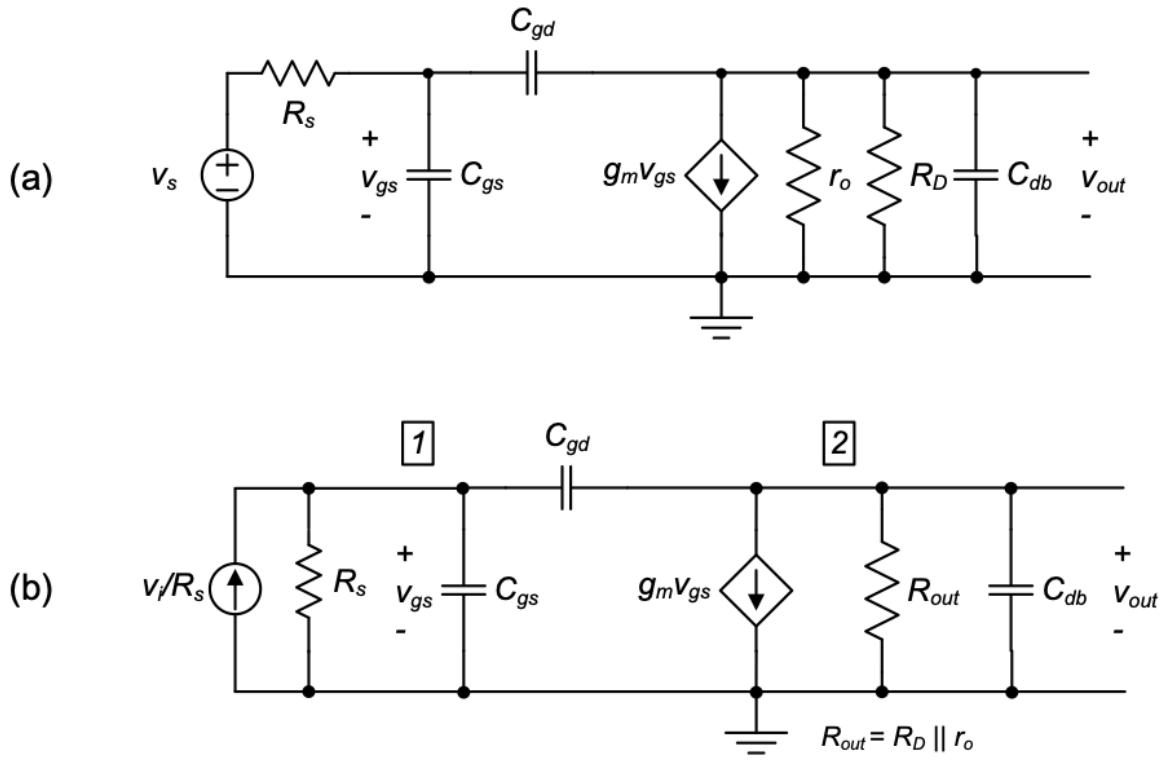


Figure 3.15.: Small-signal model of the common-source amplifier with both intrinsic and extrinsic capacitances included. With (a) Thevénin equivalent input source configuration, and (b) Norton equivalent input source configuration.

where

$$b_1 = R_s\left(C_{gs} + C_{gd}\right) + R_{out}\left(C_{db} + C_{gd}\right) + g_m R_{out} R_s C_{gd} \tag{3.39}$$

and

$$b_2 = R_s R_{out}\left(C_{gs}C_{gd} + C_{gs}C_{db} + C_{gd}C_{db}\right) \tag{3.40}$$

Although this result is algebraically complex, we can make a few preliminary observations about the terms in the numerator of Equation 3.38:

- At DC ($s = 0$, or all capacitors set to zero), the voltage gain of the circuit is $-g_m R_{out}$, as we already concluded from the low-frequency analysis in Chapter 2.
- The numerator contains a right half plane zero, $z_1 = g_m/C_{gd}$. Since obviously $C_{gd} < C_{gs}+C_{gd}$, we conclude [via comparison with Equation 3.35] that this zero occurs at frequencies beyond ω_T, and is therefore irrelevant in many practical scenarios.

The denominator of the transfer function is a second-order polynomial in s with complicated dependencies on all component values. All we can say at first glance from inspecting the denominator is that we expect to see two poles in the frequency response of this circuit, because it can (in principle) be factored into two binomial terms. Note that this factorization would yield an even more complicated expression.

The main issue with a result of this complexity is that it cannot be understood intuitively. Consequently, it is difficult to recognize the main parameters that are limiting the performance, which in turn prevents the designer from identifying ways to optimize the circuit. Even though Equation 3.38 is mathematically exact, we would rather like to work with an expression that sacrifices some accuracy and/or detail in return for transparency and focus on the main effects that limit the performance. In order to take steps in this direction, we begin by evaluating Equation 3.38 numerically, primarily to get a feel for the pole locations in a typical circuit.

Example 3-6: Magnitude Response of the Common-Source Amplifier

Evaluate and plot the steady-state magnitude response of Equation 3.38 numerically using the following transistor parameters: $g_m = 1\ mS$, $C_{gs} = 40.7\ fF$, $C_{gd} = 10\ fF$ and $C_{db} = 11.6\ fF$ (same as in Example 3-5). Assume $R_{out} = 5\ k\Omega$ and $R_s = 50\ k\Omega$. For comparison, also plot the magnitude response of Equation 3.18, i.e., considering only the intrinsic gate capacitance.

SOLUTION

The plots are generated by letting $s = j\omega$ in Equation 3.38 and Equation 3.18, and subsequently plotting the magnitude of the expression as a function of frequency. The result is show in Figure 3.16. From the plots, we conclude the following:

- In the response that uses intrinsic capacitance only, we see a pole at approximately 100 MHz; this number corresponds to the value obtained in Example 3-4.
- As expected, the response with both intrinsic and extrinsic capacitances exhibits two poles. More importantly, we see that one of the poles occurs at relatively low frequencies, while the other occurs at very high frequencies.

- The low-frequency pole of the case with extrinsic capacitance included lies significantly lower than 100 MHz. This tells us that extrinsic capacitance has a substantial impact on the bandwidth of this circuit.

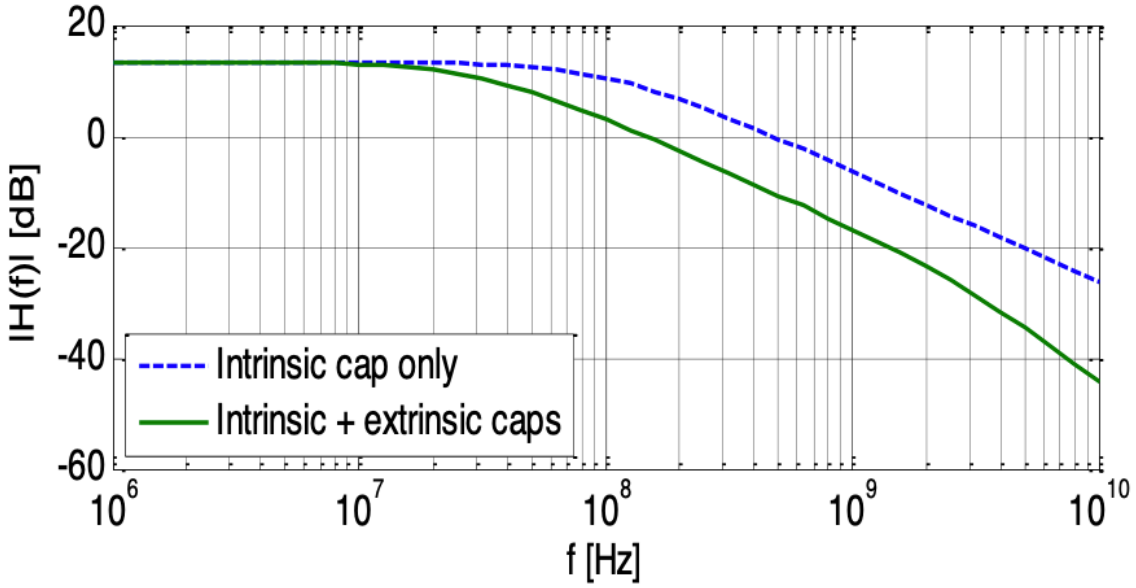


Figure 3.16.

From the result of this particular example, we see that the bandwidth of the common-source amplifier is primarily set by a single pole that lies far from any other breakpoint in the response. In this case, we call the bandwidth limiting pole of the circuit the dominant pole. When a dominant pole condition exists, we would like to work with an expression of the form

$$\frac{v_{out}}{v_s} = \frac{A_{v0}}{1 - \frac{s}{p_1}} \tag{3.41}$$

instead of evaluating Equation 3.38. In some sense, Equation 3.38 contains too much information about irrelevant features of the response that have no impact on the 3-dB bandwidth. A commonly used technique that allows us to simplify expressions of the form of Equation 3.38 is therefore discussed in the next sub-section.

3.3.4. The Dominant Pole Approximation

In general, the denominator of the transfer function given in Equation 3.38 can be factored into two binomial terms

$$\left(1 - \frac{s}{p_1}\right)\left(1 - \frac{s}{p_2}\right) = 1 - s\left(\frac{1}{p_1} + \frac{1}{p_2}\right) + \frac{s^2}{p_1 p_2} \tag{3.42}$$

Furthermore, we know from our numerical evaluation of the previous subsection that the magnitude of one of the poles is much larger than the other, i.e

$$|p_2| \gg |p_1| \tag{3.43}$$

and therefore

$$|p_2| \ll |p_1| \tag{3.44}$$

Consequently, we can eliminate the second term in s on the right hand side of Equation 3.42 and approximate

$$\left(1 - \frac{s}{p_1}\right)\left(1 - \frac{s}{p_2}\right) \simeq 1 - \left(\frac{s}{p_1}\right) + \frac{s^2}{p_1 p_2} \tag{3.45}$$

Now, comparing Equation 3.38 with Equation 3.45, we see that

$$-\frac{1}{p_1} = b_1 \tag{3.46}$$

and thus

$$p_1 = -\frac{1}{b_1}$$

$$= -\frac{1}{R_s[C_{gs} + C_{gd}] + R_{out}[C_{db} + C_{gd}] + g_m R_{out} R_s C_{gd}}$$

$$= -\frac{1}{R_s[C_{gs} + (1 + g_m R_{out})C_{gd}] + R_{out}[C_{db} + C_{gd}]} \tag{3.47}$$

This result gives us a relatively handy expression for the dominant pole in the common-source amplifier, and the bandwidth can be estimated using

$$\omega_{3dB} = \frac{1}{R_s[C_{gs} + (1 + g_m R_{out})C_{gd}] + R_{out}[C_{db} + C_{gd}]} \tag{3.48}$$

As opposed to Equation 3.38, Equation 3.48 is much more useful for evaluating which particular component of the circuit may limit the bandwidth. Specifically, the term $(1+g_m R_{out})C_{gd}$ looks like a potential problem. Whenever $g_m R_{out}$ is large (high gain), this term may dominate the denominator of Equation 3.48, and therefore limit the bandwidth. This is a very important conclusion, but unfortunately took us many lines of algebra (including the derivation of Equation 3.38, which was not shown in detail) to develop. A more desirable approach would hint with very little algebra that the aforementioned term may limit the bandwidth. Such an approach is possible via the application of the Miller theorem and the Miller approximation, discussed in the next subsection.

3.3.5. The Miller Theorem and the Miller Approximation

The **Miller theorem** is a general linear circuit theorem that can be used to replace an impedance connected between two circuit nodes by two impedances, connected from each terminal to ground. This is illustrated in Figure 3.17. The impedance Z in Figure 3.17(a) is replaced by the two impedances Z_1 and Z_2 in Figure 3.17(b). For the two circuits to be equivalent, it can be shown that

$$Z_1 = \frac{Z}{1 - A_{vM}} \quad and \quad Z_2 = \frac{A_{vM} Z}{A_{vM} - 1} \tag{3.49}$$

where $A_{vM} = V_2/V_1$ is the voltage gain across the impedance Z, also called the **Miller gain**.

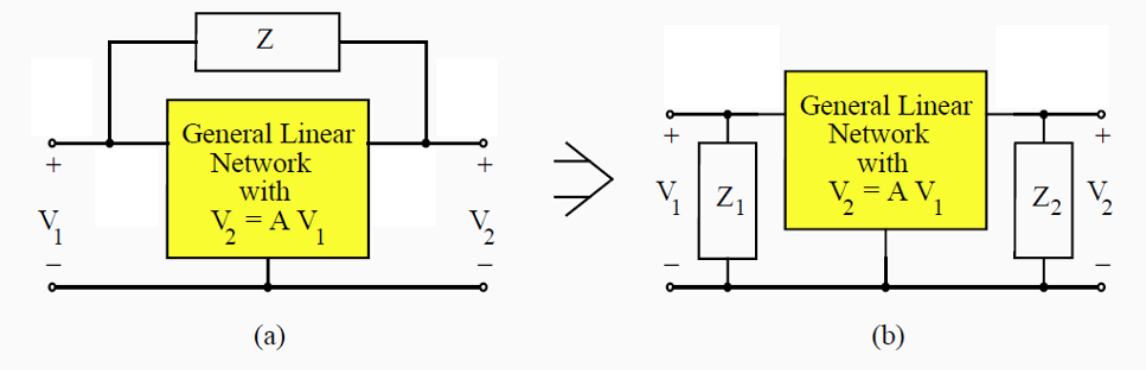


Figure 3.17.: Illustration of the Miller theorem.

The Miller theorem is useful for the simplification of a variety of circuits. In the context of the common-source amplifier analysis in this chapter, we will use the theorem to eliminate the coupling of the output and input through C_{gd}, and thereby arrive at a circuit that is easier to analyze and understand. Before applying the Miller Theorem to the full circuit model of Figure 3.15, we will first consider its application to an ideal voltage amplifier circuit with a coupling capacitance between the input and output, as drawn in Figure 3.18.

The goal of this example is to determine the effective shunt capacitance at the input port, when the signal is amplified by a gain of A_{vM} across the coupling capacitor C. Using $Z = 1/sC$, and $Z_{eff} = 1/sC_{eff}$ we can apply Equation 3.49 to find

$$\frac{1}{sC_{eff}} = \frac{\frac{1}{sC}}{1 - A_{vM}} \tag{3.50}$$

and therefore

$$C_{eff} = C(1 - A_{vM}) \tag{3.51}$$

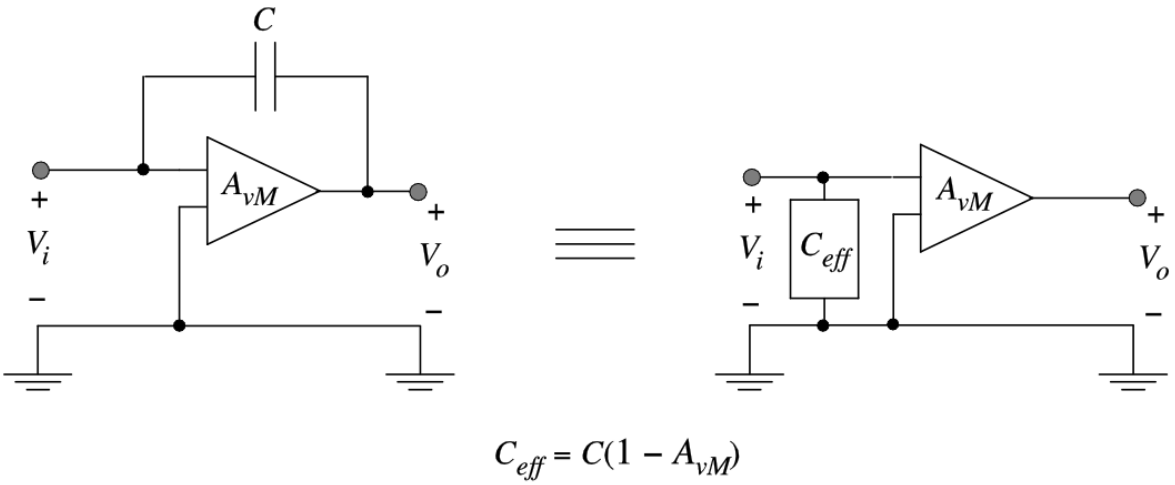


Figure 3.18.: Idealized voltage amplifier with coupling capacitance between its input and output.

If the voltage gain A_{vM} is a negative number (as in the case of a common-source amplifier), the capacitance C is "amplified" by the factor $(1 + |A_{vM}|)$. Intuitively, without relying on a complete proof of the Miller Theorem, this result can be understood by examining the voltages and currents of the capacitor C in Figure 3.18. The voltage across C is

$$v_C = v_{in} - v_{out} = v_i n(1 - A_{vM}) \tag{3.52}$$

and the current flowing into C from the input port is

$$i_{in} = i_C = sCv_c = sC(1 - A_{vM})v_i n$$

$$\frac{v_{in}}{v_{out}} = \frac{1}{sC(1 - A_{vM})} = \frac{1}{sC_{eff}} \tag{3.53}$$

In essence, the capacitance is multiplied due to the large swing at the amplifier output; this increases the voltage across the capacitor and therefore forces a correspondingly multiplied current into the input port.

This result applies qualitatively also to the common-source amplifier studied in this chapter—i.e., the negative gain of the amplifier causes an amplification of C_{gd}, which couples the input and output. However, a subtle difference is that the gain across the capacitor is not perfectly constant (as assumed above), but exhibits some frequency dependence.

To investigate, consider the circuit of Figure 3.19, which is the relevant section of the full common-source circuit needed to find the voltage gain across C_{gd}. Applying KCL at node 2 and solving for $A_{vM} = v_{out}/v_{gs}$ yields

$$\frac{v_{out}}{v_{gs}} = -g_m R_{out} \left(\frac{1 - s\frac{C_{gd}}{g_m}}{1 + sR_{out}(C_{db} + C_{gd})} \right) \tag{3.54}$$

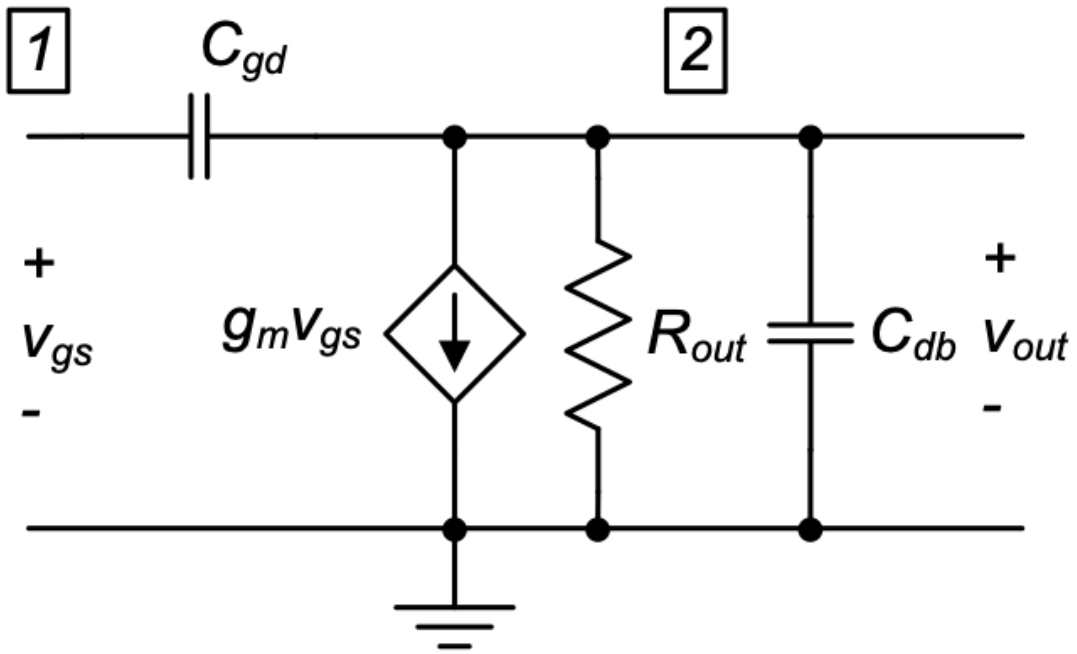


Figure 3.19.: Circuit to analyze the voltage gain across C_{gd}.

In this expression, the bracketed term contains a zero and a pole. The zero occurs beyond ω_T and can be safely discarded. The situation is somewhat different for the pole. If R_{out} is very large, or if an additional load capacitance is added to the circuit output (in parallel with C_{db}), the pole can occur at relatively low frequencies, making the gain across C_{gd} non-constant in the frequency range of interest. Provided that the pole in the bracketed term occurs outside the frequency band of interest, we can assume

$$\frac{v_{out}}{v_g s} \simeq -g_m R_{out} \tag{3.55}$$

This assumption is known as the Miller approximation, and it allows us to utilize the result from Equation 3.51, which assumed a constant gain across the capacitor in question.

To complete this discussion, we will now apply the Miller approximation to the model of the common-source amplifier in Figure 3.15. The result is shown in Figure 3.20. The capacitor C_{gd} is no longer connected between the input and output, but appears only across the input port, with its value multiplied by $(1 + g_m R_{out})$. From this model, the circuit bandwidth can be easily identified by inspection

$$\omega_{3dB} = \frac{1}{R_s[C_{gs} + (1 + g_m R_{out})C_{gd}]} \tag{3.56}$$

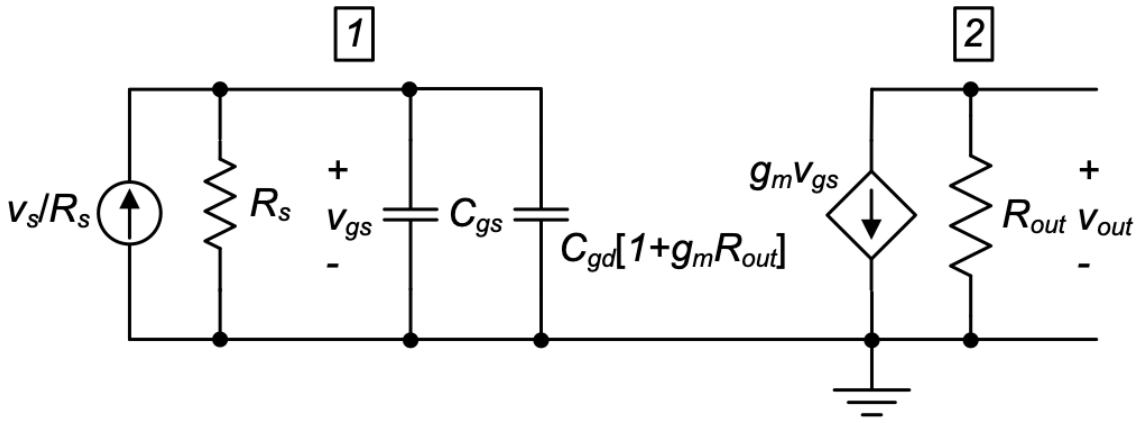


Figure 3.20.: Small-signal model of the common-source amplifier after applying the Miller approximation.

In comparison with Equation 3.48, this result is missing the term $R_{out}(C_{gd}+C_{db})$ in the denominator. This is not surprising and also inconsequential when the Miller Approximation is applied properly. As we pointed out above, the Miller approximation is justified only when this time constant is small in the first place, ensuring a constant Miller gain in the band where the dominant pole is expected to lie. Whenever the Miller approximation is applied, it must be verified that the neglected pole in the Miller gain occurs far beyond the frequency estimated by Equation 3.56. This leads to the following procedure for the proper application of the Miller approximation in common-source amplifiers:

1. Calculate the low-frequency gain across C_{gd} and draw the simplified circuit model (as in Figure 3.20) with the Miller-amplified shunt capacitance at the input.
2. Estimate the bandwidth of the circuit using Equation 3.56.
3. Calculate the frequency of the pole in Equation 3.54. If and only if this pole frequency is far beyond the frequency calculated in step 2, the Miller approximation result is valid.

In a typical common-source circuit without a large load capacitance as drawn in Figure 3.20, the Miller approximation typically holds. When a very large capacitor is connected to the output, the approximation becomes invalid and the dominant pole is set by the RC time constant formed at the output.

Example 3-7: Calculating the Common-Source Amplifier Bandwidth Using the Miller Approximation

Calculate the 3-dB bandwidth for the common-source voltage amplifier of Figure 3.15 using (a) the Miller approximation, and (b) the dominant pole approximation result of Equation 3.48. Parameters: $g_m = 1\ mS$, $C_{gs} = 40.7\ fF$, $C_{gd} = 10\ fF$, $C_{db} = 11.6\ fF$, $R_{out} = 5\ k\Omega$ and $R_s = 50\ k\Omega$ (same as in Example 3-6). Calculate the percent error in the result of part (a).

SOLUTION

a. Using the Miller approximation [i.e., Equation 3.56], we obtain

$$f_{3dB} = \frac{1}{2\pi} \cdot \frac{1}{R_s[C_{gs} + (1 + g_m R_{out})C_{gd}]}$$
$$= \frac{1}{2\pi} \cdot \frac{1}{50k\Omega[40.7fF + (1 + 1mS \cdot 5k\Omega)10fF]}$$
$$= 31.61MHz$$

b. Using Equation 3.48 we find

$$f_{3dB} = \frac{1}{2\pi} \cdot \frac{1}{R_s[C_{gs} + (1 + g_m R_{out})C_{gd}] + R_{out}[C_{db} + C_{gd}]}$$
$$= \frac{1}{2\pi} \cdot \frac{1}{50k\Omega[40.7fF + (1 + 1mS \cdot 5k\Omega)10fF] + 5k\Omega \cdot 21.6fF}$$
$$= 30.95MHz$$

The error in the result of part (a) is therefore

$$\frac{31.61 - 30.95}{30.95} = 2.1\%$$

The error of 2.1% seen in this example is acceptable and will in practice be overshadowed by uncertainty in the transistor model parameters.

3.3.6. Calculating the Non-Dominant Pole*

The reader may wonder how the non-dominant pole frequency can be calculated within the above-discussed framework. A common misconception is to assume that after applying the Miller approximation, the non-dominant pole can be simply found from the time constant in the output network, i.e., $R_{out}C_{db}$. This is incorrect, since the Miller approximation is not valid at the frequency where the non-dominant pole is located.

If a dominant pole condition exists, the proper way to estimate the non-dominant pole is by comparing the coefficients of Eqs. (3.45) and (3.38). Specifically, we utilize that

$$\frac{1}{p_1 p_2} = b_2 \tag{3.57}$$

and thus

$$p_2 = \frac{1}{p_1 b_2}$$
$$= -\frac{R_s[C_{gs} + C_{gd}] + R_{out}[C_{db} + C_{gd}] + g_m R_{out} R_s C_{gd}}{R_s R_{out}(C_{gs}C_{gd} + C_{gs}C_{db} + C_{gd}C_{db})} \tag{3.58}$$

To simplify, let us assume that $C_{gd} \ll C_{gs}$ and $C_{gd} \ll C_{db}$. Note that the latter assumption is not strictly true based on typical values for the technology assumed in this module (see Example 3-5). However, if a load capacitance is added to the circuit, the approximation is more easily justified, with Cdb replaced by $C_{db} + C_L$ (see Example 3-8), and we almost always have in practice $C_{gd} \ll C_{db} + C_L$. Thus, under the stated conditions, we can write

$$p_2 \simeq -\frac{R_s C_{gs} + R_{out} C_{db} + g_m R_{out} R_s C_{gd}}{R_s R_{out} C_{gs} C_{db}}$$

$$= -\left(\frac{1}{R_{out} C_{db}} + \frac{1}{R_s C_{gs}} + \frac{g_m}{C_{db}} \cdot \frac{C_{gd}}{C_{gs}} \right) \tag{3.59}$$

This approximate result indicates that the non-dominant pole lies at a frequency that is higher than $1/R_{out}C_{db}$, especially when g_m is large. Note that Equation 3.59 essentially represents a "parallel combination" of time constants (analogous to parallel connections of resistors)—that is, the smallest time constant in the expression sets the pole frequency.

3.4. Open-Circuit Time Constant Analysis

3.4.1. General Framework

In Section 3-3-4, we derived an approximate expression for the 3-dB bandwidth of a common-source voltage amplifier, assuming that a dominant pole condition exists. In this analysis, we found that the bandwidth is fully determined by the coefficient b_1 in the numerator of Equation 3.38.

The **open-circuit time constant (OCT) analysis** is a powerful and general technique that allows us to compute the term b_1 for arbitrary circuits, without the need to derive the full circuit transfer function with all high-order artifacts included. More importantly, it breaks the analysis into small and computationally manageable steps that provide insight about which circuit elements present the main bandwidth bottleneck. The step-by-step procedure for applying the OCT analysis method can be summarized as follows (see Reference 3 for a derivation)

1. Remove all but one capacitor in the circuit that is to be analyzed. Let us call this capacitor C_j.
2. Short all independent voltage sources and remove all independent current sources in the circuit.
3. Calculate the Thévenin resistance R_{Tj} seen by the capacitor C_j and compute the time constant $\tau_{j0} = R_{Tj}C_j$. Here, the subscript "o" is used to emphasize the open-circuit condition.
4. Repeat the above steps 1-3 for all remaining capacitors in the circuit.
5. The sum of all time constants is exactly equal to b_1. We can therefore estimate the circuit's 3-dB bandwidth using

$$\omega_{3dB} \simeq \frac{1}{b_1} \tag{3.60}$$

$$b_1 = \sum_{j=1}^{N} \tau_{jo} \tag{3.61}$$

where N is the total number of capacitors in the circuit. The τ_{jo} jo are called **open-circuit time** constants, because these were determined with all other capacitors open circuited.

Once a circuit is analyzed using the OCT method, we can see which of the individual open-circuit time constants is contributing most heavily to b_1. To increase the bandwidth, we can try to redesign the circuit by lowering the Thévenin resistance or the capacitor value of that time constant.

3.4.2. OCT Analysis of a Common-Source Stage

Consider the common-source amplifier shown in Figure 3.21(a) as an example to further understand the method of open-circuit time constants. We begin by considering C_{gs} and therefore remove all other capacitors and short the input source as shown in Figure 3.21(b). As evident from this circuit, the Thévenin resistance seen by capacitor C_gs is R_S and the individual time constant contribution from C_{gs} is

$$\tau_{gso} = R_S C_{gs} \tag{3.62}$$

Similarly, redrawing the circuit with only Cdb present will yield

$$\tau_{dbo} = R_{out} C_{db} \tag{3.63}$$

Next, we determine the individual time constant contribution from capacitor C_{gd}. To perform this calculation, we consider the circuit as redrawn in Figure 3.21(c). From this circuit, the Thévenin resistance seen across C_{gd} cannot be immediately determined by inspection. This is because of the g_m element, which couples the nodes to the left and right of the capacitance. We therefore resort to determining the Thévenin resistance from first principles, using a nodal analysis. As shown in Figure 3.22, we apply a test current source (i_t) and measure the resulting test voltage (v_t). Applying KVL and KCL, we find that

$$v_t = v_{gs} + R_{out}(g_m V_{gs} + i_t) \tag{3.64}$$

$$v_{gs} = i_t R_S \tag{3.65}$$

After substituting Equation 3.65 into Equation 3.64, we obtain

$$R_{Tgd} = \frac{v_t}{i_t} = R_S + R_{out}(g_m R_S + 1)$$

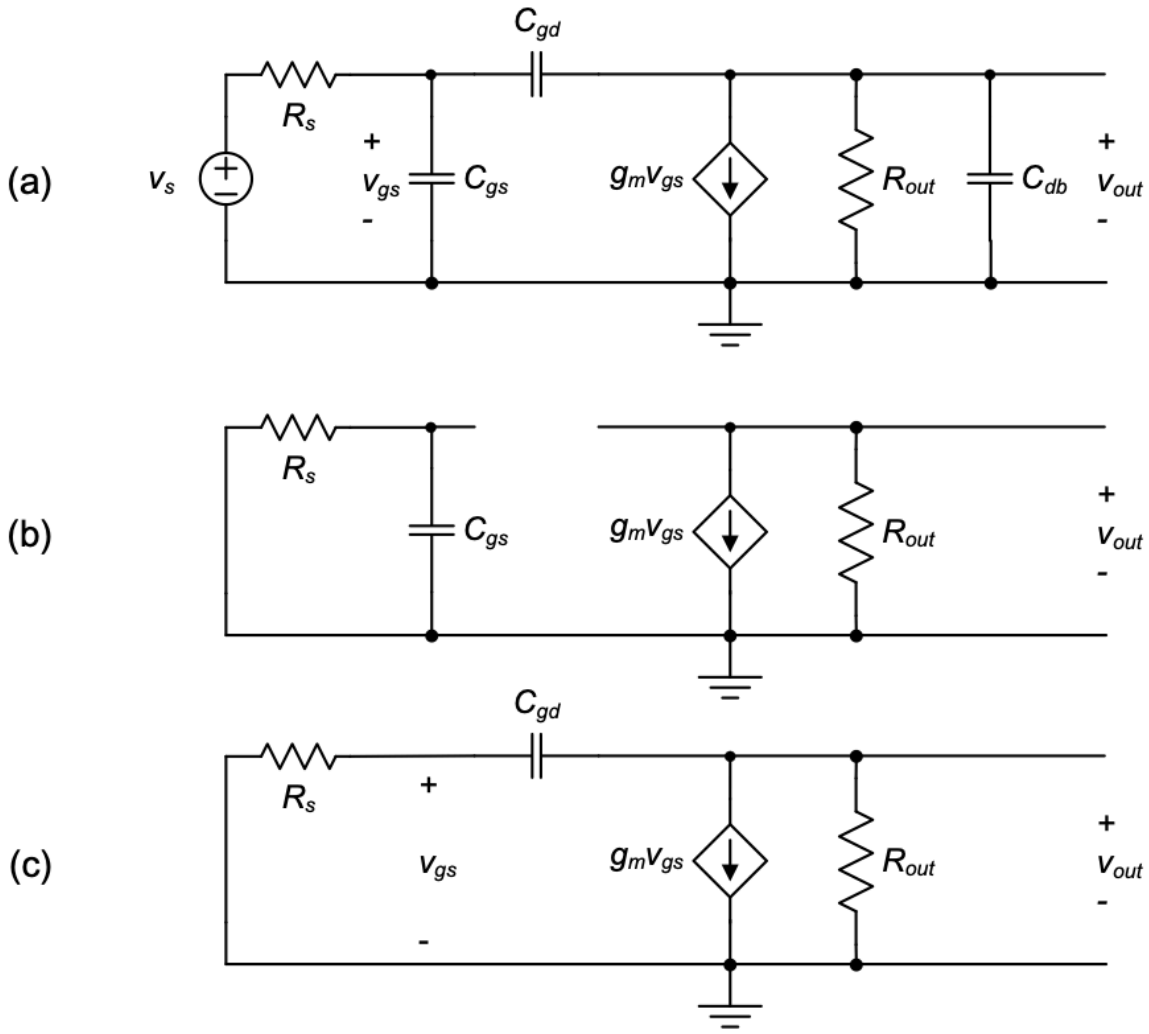


Figure 3.21.: OCT analysis for a common-source amplifier. (a) Complete circuit. (b) Circuit for finding τ_{gs}. (c) Circuit for finding τ_{gd}.

$$= R_S + R_{out} + g_m R_S R_{out} \tag{3.66}$$

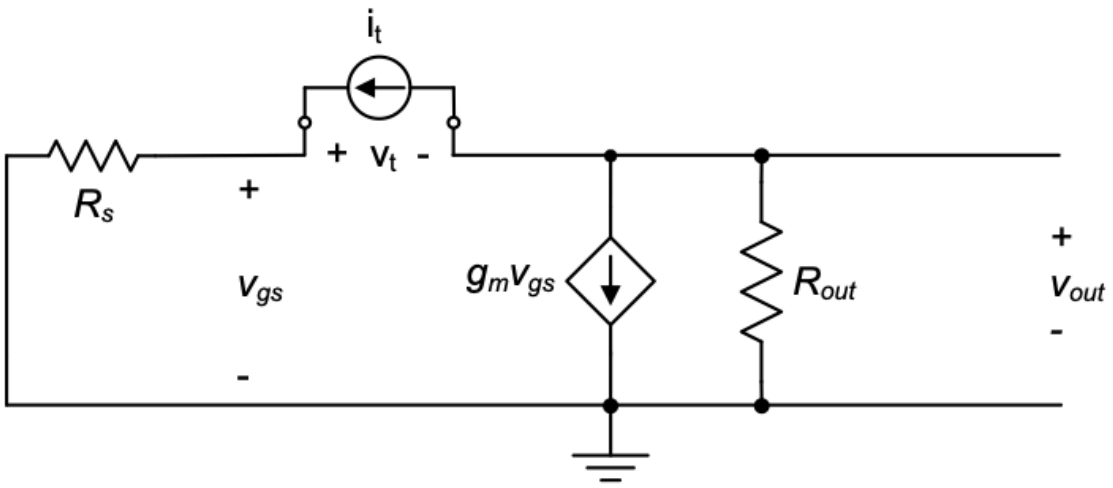


Figure 3.22.: Circuit to determine the Thévenin resistance seen across C_{gd}.

A common way to memorize this final result is "$R_{left} + R_{right} + g_m R_{left} R_{right}$," where R_{left} and R_{right} are the resistances seen to the left and right of the coupling capacitance C_{gd}, respectively. Using this result, the individual time constant resulting from C_{gd} is given by

$$\tau_{gdo} = R_{Tgd} C_{gd} = [R_S + R_{out} + g_m R_S R_{out}] C_{gd} \tag{3.67}$$

Next, we add the individual time constants from Equation 3.62, Equation 3.63, and Equation 3.67, which results in

$$\begin{aligned} b_1 &= R_S[C_{gs} + C_{gd}] + R_{out}[C_{db} + C_{gd}] + g_m R_{out} R_S C_g d \\ &= R_S[C_{gs} + (1 + g_m R_{out}) C_{gd}] + R_{out}[C_{db} + C_{gd}] \end{aligned} \tag{3.68}$$

Note that this result is identical to Equation 3.39, which was obtained from an exact nodal analysis of the complete circuit. This verifies that the method of open-circuit time constants is an exact analysis to determine the factor b_1, which multiplies the first-order term in s in the denominator of the generalized system function. As before, the resulting estimate of the 3-dB breakpoint frequency is therefore given by

$$\begin{aligned} \omega_{3dB} &= \frac{1}{b_1} \\ &= \frac{1}{R_S[C_{gs} + (1 + g_m R_{out}) C_{gd}] + R_{out}[C_{db} + C_{gd}]} \end{aligned} \tag{3.69}$$

It is important to remember that this result maintains good accuracy only if a dominant pole condition exists. As we showed in Section 3-3-4, this condition is required so that we can approximate $\omega_{3dB} \simeq 1/b_1$. Finally it is worth noting that Equation 3.69 shows that C_{gd} is effectively multiplied by the circuit's voltage gain; this corresponds to the Miller amplification effect discussed in the previous section.

A simple example where the dominant pole condition is not met is shown in Figure 3.23. The reader may prove that the exact transfer function of this circuit is

$$\frac{v_{out}}{v_s} = \frac{-g_m R}{(1+sRC)(1+sRC)} \tag{3.70}$$

and thus

$$|p_2| = |p_1| = \frac{1}{RC} \tag{3.71}$$

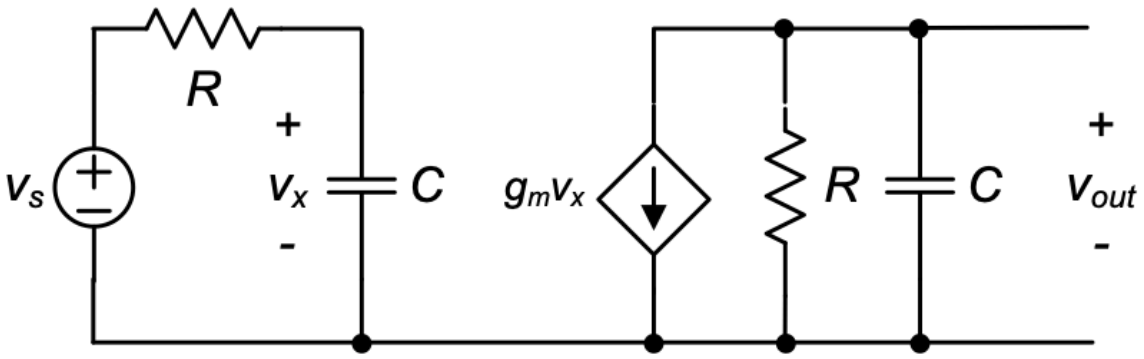


Figure 3.23.: Circuit example that violates the dominant pole assumption.

Therefore, we expect that the approximation of Equation 3.45 cannot be applied and $1/b_1$ will not be a good estimate for the circuit's bandwidth. It is now interesting to calculate the error that will result if the OCT method is nonetheless "blindly" applied.

In performing the OCT analysis, we see that the circuit in question has two open-circuit time constants equal to RC. The bandwidth estimate using OCT analysis is therefore

$$\omega_{3dB,OCT} \simeq \frac{1}{b_1} = \frac{1}{2RC} \tag{3.72}$$

On the other hand, we can find the exact 3-dB frequency of the circuit using

$$\frac{1}{\sqrt{2}} = \left| \frac{1}{(1+j\omega_{3dB}RC)(1+j\omega_{3dB}RC)} \right| \tag{3.73}$$

Solving for ω_{3dB} gives

$$\omega_{3dB} = \frac{1}{RC}\sqrt{\sqrt{2}-1} \simeq \frac{0.64}{RC} \tag{3.74}$$

The error in the OCT estimate is thus

$$\frac{0.5-0.64}{0.64} = -22\% \tag{3.75}$$

From this result, we can draw a few interesting conclusions. First, even though the dominant pole condition is grossly violated in the above example, the OCT analysis is not extremely far off from the exact result. Second, the OCT result is conservative in the sense that it tends to underestimate the circuit's bandwidth. This is desirable since the designer can rest assured that the bandwidth is at least as large as predicted by the OCT analysis. It can be shown that this latter property holds for arbitrary circuits whose poles lie on (or near) the real axis, and whose zeros occur beyond the estimated ω_{3dB}. This is the case for most circuits considered in this module. We will highlight exceptions where appropriate.

In summary, the reader should remember the following key points when applying the OCT analysis:

- In any circuit, the sum of the open-circuit time constants corresponds (exactly) to the term b_1, which multiplies the first-order term in the denominator of the circuit's s-domain transfer function.
- Under the following conditions, the bandwidth of the circuit can be approximated with good accuracy by $1/b_1$: (1) a dominant pole condition exists, (2) the transfer function contains only poles that lie on (or near) the real axis, and (3) the zeros in the transfer function occur beyond the bandwidth estimate in question.
- Even if no clear dominant pole condition exists, OCTs can be used to get a first-order feel for the bandwidth of a circuit. For instance, in a circuit with two identical real poles, the OCT bandwidth estimate is in error by –22%. As long as condition (2) above is met, the percent error will be negative and thus the estimated bandwidth is at least as large as the actual bandwidth (measured, e.g., using a circuit simulation).
- Open-circuit time constants, in general, do not necessarily correspond to the poles of a circuit. The OCT correspond to poles only in circuits that can be broken into decoupled RC sections, as is the case in the circuit of Figure 3.23.

Example 3-8 Common-Source Amplifier Bandwidth Estimate Using an OCT Analysis

Consider the circuit shown in Figure 3.24 and assume the following parameters: $W = 20\,\mu m$, $L = 1\,\mu m$, $I_B = 500\,\mu A$, $g_m = 1\,mS$, $C_{gs} = 40.7\,fF$, $C_{gd} = 10\,fF$, $C_{db} = 11.6\,fF$, $R_{out} = R_D \parallel r_o = 5\,k\Omega$ and $R_s = 50\,k\Omega$ (same as in Example 3-7). The value of the load capacitance is $C_L = 10\,pF$. Estimate the 3-dB bandwidth using an OCT analysis and propose a design modification that will increase the bandwidth by 20%. For this modification, you may not alter the circuit's DC gain, and R_S and C_L must be kept constant.

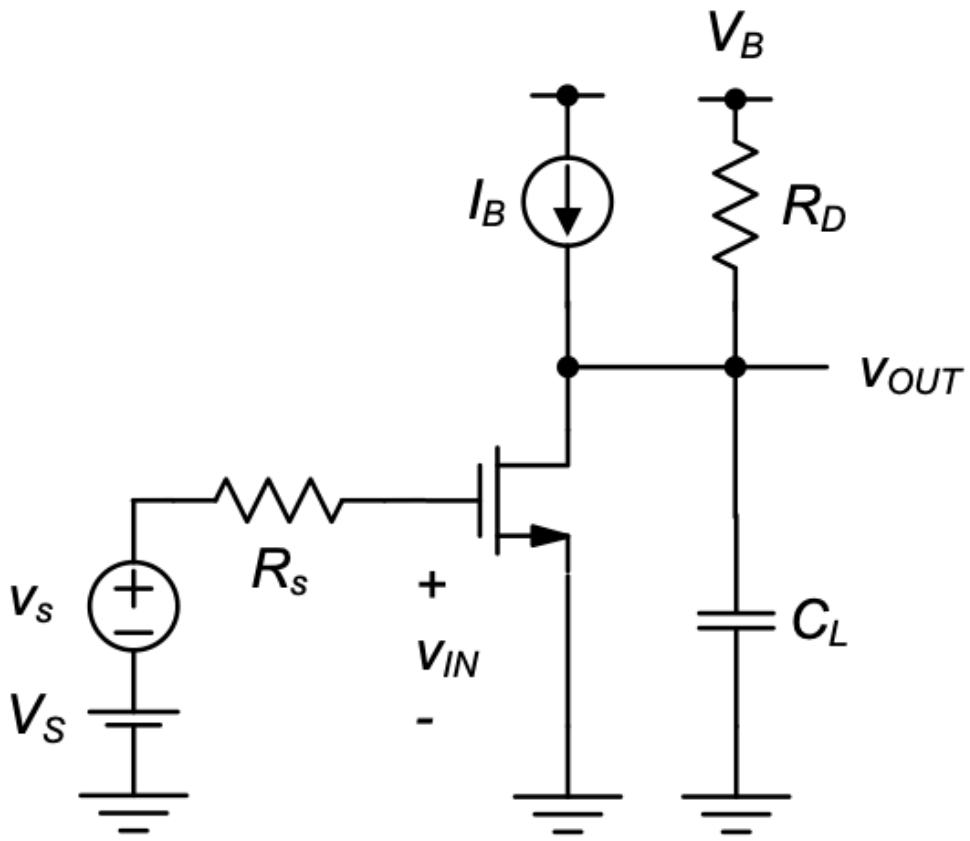


Figure 3.24.

SOLUTION

The circuit has three open-circuit time constants as expressed in Equation 3.62, Equation 3.63, and Equation 3.67, with the difference that C_L appears in parallel to C_{db}. The three OCT expressions are therefore

$$\tau_{gso} = R_S C_{gs}$$
$$\tau'_{dbo} = R_{out}(C_{db} + C_L)$$

$$\tau_{gdo} = R_{Tgd} C_{gd} = [R_S + R_{out} + g_m R_S R_{out}] C_{gd}$$

Evaluating these expression with the given numbers yields

$$\tau_{gso} = 50k\Omega \cdot 40.7fF = 2.035ns$$

$$\tau'_{dbo} = 5k\Omega(11.6pF + 10pF) = 50.01ns$$

$$\tau_{gdo} = [50k\Omega + 5k\Omega + 1mS \cdot 5k\Omega \cdot 50k\Omega]10fF$$

$$= 3.05ns$$

The bandwidth estimate is

$$f_{3dB} = \frac{1}{2\pi} \cdot \frac{1}{2.035ns + 50.01ns + 3.05ns} = 2.89MHz$$

In order to improve the bandwidth by 20%, it is clear that we must reduce the dominant open-circuit time constant τ'_{dbo}. Since C_L must remain unchanged, the only option is to reduce Rout. To first-order, reducing Rout to approximately 4 $k\Omega$ (a 20% reduction from the original value of 5 $k\Omega$) should get us close to the desired improvement. In order to keep the DC gain of the circuit constant, we now require a larger transconductance

$$g_m = \frac{A_{v0}}{R_{out}} = \frac{5}{4k\Omega} = 1.2mS$$

There are several ways to increase the transconductance of the MOSFET. (i) Keep the device width constant and increase the bias current I_D. An advantage of this option is that none of the device capacitances will change, thereby avoiding any counterproductive increase in the total time constant.(ii) Keep I_D constant and increase the device width W. This option has the advantage that the current consumption of the circuit will not increase. Finally, option (iii) is to increase both W and I_D by the same factor. This option has the advantage that the gate overdrive voltage V_{OV} remains unchanged, and hence the input bias voltage and output voltage swing are unaffected.

Since our primary focus in this example is to improve bandwidth, and current consumption and biasing considerations are secondary, we will apply option (i).

Using Equation 2.30, the new value of the required I_B is

$$I_B = I_D = \frac{g_m^2}{2\mu_n C_{ox}\frac{W}{L}} = \frac{(1.2mS)^2}{2 \cdot 50\frac{\mu A}{V^2} \cdot \frac{20}{1}} = 720\mu A$$

Note that this value is approximately 44% larger than the original bias current of 500 μA.

As a final verification step, we recompute the bandwidth estimate using the new value of Rout. The time constant τ_{gso} remains the same, while the change in τ_{gdo} is negligible. The dominant OCT modifies as follows

$$\tau'_{dbo} = 4k\Omega(11.6fF + 10pF) = 40.05ns$$

The modified bandwidth estimate is therefore

$$f_{3dB} = \frac{1}{2\pi} \cdot \frac{1}{2.035ns + 40.05ns + 3.05ns} = 3.53MHz$$

which is about 22% larger than the original bandwidth, satisfying our design intent.

3.4.3. OCT Extensions

The OCT analysis covered in this section is tailored toward finding the upper corner frequency in circuits that are limited by capacitive elements; this is the most common situation encountered in integrated circuit design. For completeness, it is worth mentioning that there exists a method of short-circuit time constants (see Reference 3), which aims at estimating the lower corner frequency of a circuit with a high-pass characteristic. This is useful for circuits that employ AC coupling of various forms.

In circuits that contain inductors, the additional time constants can be included by shorting all but one inductor at a time. The generalized framework that includes the consideration of both inductors and capacitors to estimate the upper corner frequency of a circuit is called **zero-value time constant analysis**. Finally, it is interesting to note that higher-order terms [such as b_2 in Equation 3.40] can be found using an OCT-like analysis. The interested reader is referred to Reference 4 for a comprehensive discussion of such methods.

3.4.4. Time Constants versus Poles

The distinction between open-circuit time constants and poles tends to be a source of confusion among circuit design students. We will therefore review the differences in this section using two examples.

Consider first the circuit of Figure 3.23. As we have shown above, this circuit has two open-circuit time constants, equal to RC. Also we found that this circuit has two poles, located at $-1/RC$. Thus, in this particular circuit, the poles coincide with the (reciprocals of the) time constants. The reason for this coincidence is that the two networks at the input and output are fully decoupled and represent simple first order RC sections. For such a topology, the circuit designer sometimes loosely speaks of a "pole at the input" and "pole at the output," which are directly set by the time constants of each network.

Consider now the circuit of Figure 3.25, which is the same as Figure 3.23, except that we have added an additional capacitor C between the input and output terminal. This circuit retains the two open-circuit time constants of the original circuit (equal to RC), but has an additional one due to the added capacitor, equal to $RC(2 + g_m R)$. On the other hand, the poles of this circuit can no longer be found by inspection. The transfer function has the form of Equation 3.38, with $b_1 = RC(4 + g_m R)$ and $b_2 = 3(RC)^2$. The two poles of the circuit are the roots of the denominator polynomial $1 + b_1 s + b_2 s^2 = 0$ and their value depends on the value of $g_m R$. Assuming $g_m R = 2$ as a numerical example, the roots, and therefore the poles become

$$p_{1,2} = -\frac{1}{RC}\left(1 \pm \sqrt{\frac{2}{3}}\right) \tag{3.76}$$

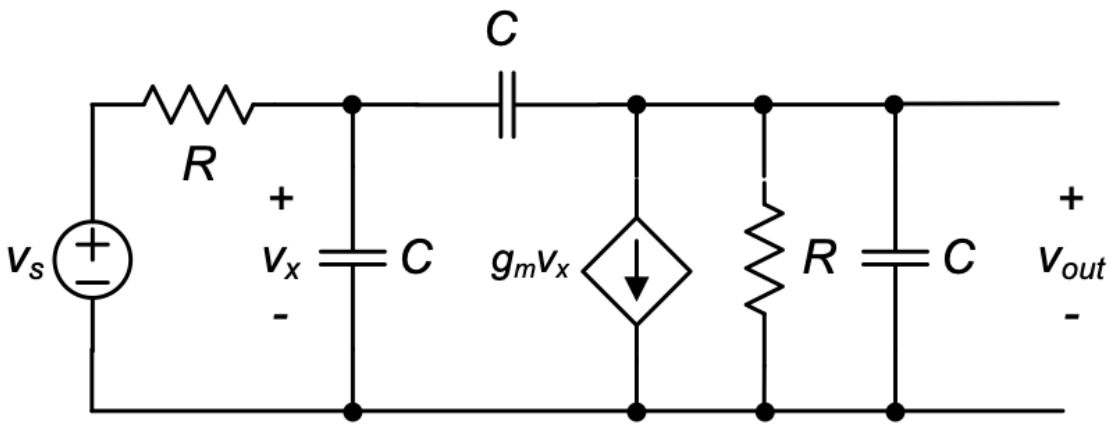


Figure 3.25.: Circuit example with three open-circuit time constants.

As we can see from this result, the poles do not coincide with any of the open-circuit time constants. More significantly, the number of poles (two) is not even equal to the number of open-circuit time constants (three).

As we can see from this result, the poles do not coincide with any of the open-circuit time constants. More significantly, the number of poles (two) is not even equal to the number of open-circuit time constants (three).

Finally, to fully close the loop between the two analysis techniques, note that the magnitude of the low-frequency pole in Equation 3.76 is approximately equal to $0.18/RC$. This is close to the 3-dB bandwidth predicted by the sum of the open-circuit time constants (for $g_m R = 2$): $1/(4RC + RC + RC) = 0.167/RC$, which, as expected, is slightly conservative.

3.5. High-Frequency Two-Port Model for the Common-Source Voltage Amplifier

To summarize, Figure 3.26 shows the most general two-port model for the common-source voltage amplifier [similar to Figure 3.15] with source and load networks included. The advantage of this model representation is that it is valid for arbitrary component values. The disadvantage is that analyzing a circuit based on this model leads to complex equations. Generally, one should use this model as the starting point for the analysis of more complex circuits that contain a CS amplifier (see Chapter 6). Then, whenever suitable, we can invoke simplifications such as the Miller approximation or open-circuit time constants to simplify the analysis.

Finally, note that the model of Figure 3.26 is not well suited for a translation into a native voltage amplifier model (using a voltage controlled voltage source) as done for the low-frequency circuit in Section 2-4. The capacitors connected to the output port would lead to a frequency-dependent open-circuit gain and output impedance (Z_{out} rather than R_{out}) that give a non-intuitive representation of the circuit. It is therefore preferred to describe this voltage amplifier using the transconductance model as shown.

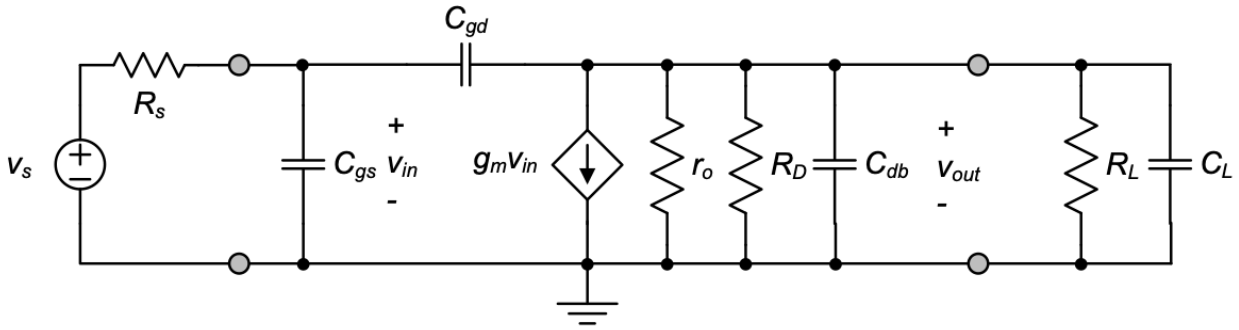


Figure 3.26.: General two-port model for the common-source voltage amplifier valid at high frequencies.

3.6. Summary

In this chapter we have reviewed the basic concepts of frequency domain analysis and introduced the intrinsic and extrinsic device capacitances of a MOSFET. Using the obtained small-signal model,

the frequency response of any circuit can be obtained from first principles using the following steps:

1. Derive the transfer function using a nodal analysis.
2. Let and solve for the magnitude of the $s = j\omega$ resulting expression.
3. Set the magnitude equal to $1/\sqrt{2}$ times the DC gain value, and solve for ω.

Unfortunately, this method is algebraically too complex for all but the most basic circuits. Consequently, we introduced several approximate methods and tools that are frequently used by analog circuit designers. These methods were developed using our driving example of a common-source voltage amplifier, but are widely used in other situations as well

- Provided that an exact (and potentially complicated) transfer function expression is available, the dominant pole approximation can be applied to arrive at a simplified bandwidth expression. In this approximation, it is assumed that a single pole dominates the response and sets the circuit's 3-dB bandwidth.
- The Miller approximation was used to obtain a quick estimate of the 3 -dB bandwidth specifically for the common-source voltage amplifier. Although it is not an exact calculation, it is very useful for determining an estimate of the bandwidth of the amplifier analytically. Furthermore, this analysis gave insight into the effect of "Miller-multiplication" of a capacitor that appears across a voltage gain path. This effect is found in a multitude of circuits, and understanding this mechanism is insightful for design.
- The method of open-circuit time constants is the most powerful and most broadly applicable technique discussed in this chapter. It provides an accurate answer for the circuit's bandwidth if a dominant pole condition exists. Even if the dominant pole condition is not strictly met, the method yields acceptable errors (on the conservative side) on the order of a few tens of percent, which is often acceptable in a first-order hand analysis. Finally, the method of open-circuit time constants is an excellent design tool since it assists in finding which capacitors and Thévenin resistances are dominating the dynamic performance.

3.7. References

1. R. F. Pierret, **Semiconductor Device Fundamentals**, Prentice Hall, 1995.
2. D. K. Shaeffer and T. H. Lee, "A 1.5 V, 1.5 GHz CMOS low noise amplifier," **IEEE J. Solid-State Circuits**, pp. 745–759, May 1997.
3. P. E. Gray and C. L. Searle, **Electronic Principles Physics, Models, and Circuits**, Wiley, 1969.
4. A. Hajimiri, "Generalized Time- and Transfer-Constant Circuit Analysis," **IEEE Trans. Circuits and Systems I**, Vol. 27, No. 6, pp. 1105–1121, June 2010.

3.8. Problems

Unless otherwise stated, use the standard model parameters specified in Table 3-1 for the problems given below. Consider only first-order MOSFET behavior and include channel-length modulation (as well as any other second-order effects) only where explicitly stated.

P3.1 Sketch the Bode plots (magnitude and phase) for the following transfer functions. Assume $R_iC_i \gg R_kC_k$ if $i > k$.

a. $[1/(1 + j\omega R_1C_1)][(1/(1 + j\omega R_2C_2))]$

b. $(j\omega R_3C_3)[(1 + j\omega R_4C_4)/(1 + j\omega R_5C_5)]$

c. $[(1 + j\omega R_6C_6)/(1 + j\omega R_8C_8)][(1 + j\omega R_7C_7)/(1 + j\omega R_9C_9)]$

P3.2 A system has a DC gain of 500, LHP zeros at 10 kHz and 1 MHz and LHP poles at 100 kHz, 10 MHz, and 100 MHz.

a. Write the s-domain transfer function that describes this system.

b. Draw a Bode plot for both the magnitude and phase of this system.

c. Switch the poles and zeros and repeat parts (a) and (b).

P3.3 Sketch the Bode plot for the magnitude, $|I_o/I_s|_{dB}$ and phase $\angle I_o/I_s$ of the circuit shown in Figure 3.27, given

a. $R_1 = 10\ k\Omega$, $R_2 = 100\ k\Omega$, $C = 1\ pF$

b. $R_1 = 0.1\ k\Omega$, $R_2 = 100\ k\Omega$, $C = 1\ pF$

c. $R_1 = 10\ k\Omega$, $R_2 = 100\ k\Omega$, $C = 10\ pF$

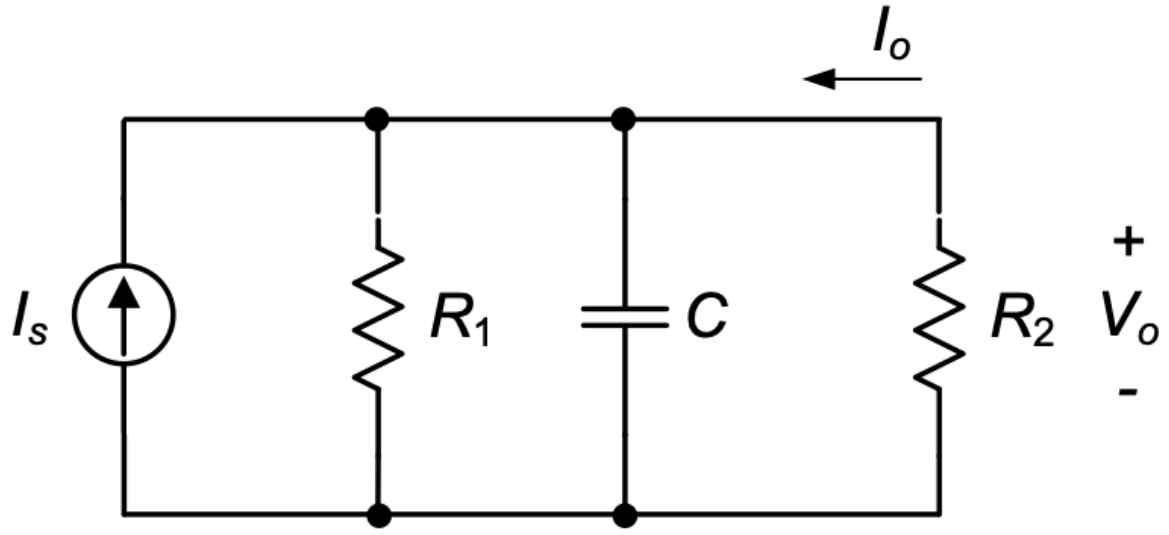


Figure 3.27.

P3.4 Repeat Example 3-5 for the following parameters (assuming $V_{DS} = 2.5\ V$). For each case, compute by which factor the transistor's transit frequency has changed relative to the value seen in Example 3-5.

a. $I_D = 500\ \mu A$, $L = 2\ \mu m$, and $W = 20\ \mu m$.
b. $I_D = 500\ \mu A$, $L = 1\ \mu m$, and $W = 40\ \mu m$.
c. $I_D = 1000\ \mu A$, $L = 1\ \mu m$, and $W = 40\ \mu m$.

P3.5 Repeat Example 3-5 for a p-channel MOSFET operating in saturation. Parameters: $V_{SD} = 2.5\ V$, $-I_D = 500\ \mu A$, $L = 1\ \mu m$, and $W = 20\ \mu m$. Compute the ratio of the transit frequency obtained in Example 3-5 and the value obtained for the p-channel device analyzed in this problem. What is the main parameter that is responsible for the lower f_T observed for the p-channel MOSFET?

P3.6 Calculate the drain-bulk capacitance of a 100 μm wide n-channel transistor for $V_{DB} = 2.5\ V$. Repeat the analysis for $V_{DB} = 1\ V$ and $V_{DB} = 4\ V$ and quantify by which factor the capacitance changes relative to the case of $V_{DB} = 2.5\ V$.

P3.7 Calculate the 3-dB bandwidth of the circuit shown in Figure 3.28. Note that both MOSFETs operate in the triode region. Parameters: $W_1 = 10\ \mu m$, $L_1 = 1\ \mu m$, $W_2 = 10\ \mu m$, $L_2 = 10\ \mu m$, $V_S = 2\ V$, $V_{DD} = 5\ V$. Consider only the intrinsic gate capacitance.

P3.8 Plot the magnitude of Equation 3.38 versus frequency using a software package such as $MATLAB^{®*}$ and find the exact value of the 3-dB frequency from the resulting graph. Parameters: $g_m = 1\ mS$, $C_{gs} = 40.7\ fF$, $C_{gd} = 10\ fF$, $C_{db} = 11.6\ fF$, $R_{out} = 5\ k\Omega$ and $R_s = 50\ k\Omega$ (same as in Examples 3-7. Compare the obtained number with the approximate results obtained in Examples 3-7.

> **Note**
>
> MATLAB is a registered trademark of The MathWorks, Inc.,
> 3 Apple Hill Road, Natick, MA.

P3.9 Calculate the frequency of the non-dominant pole of the circuit analyzed in Example 3-6.

P3.10 For $A_{vM} = 1$, Equation 3.51 predicts an effective input capacitance of $C_{eff} = 0$. Explain this result intuitively, in words. Hint: Consider the voltage waveforms at the input and output of Figure 3.18 for this particular case.

P3.11 In Examples 3-7, we saw that the bandwidth estimate obtained through the Miller Approximation was in close agreement with the result from the full analysis (incorporating a dominant pole approximation). In contrast, if we were to apply the Miller Approximation result of Equation 3.56 to Examples 3-8, we would find a large error in the resulting answer (convince yourself that this is true). Explain why it is not appropriate to use Equation 3.56 to estimate the bandwidth of the circuit in Examples 3-8.

P3.12 Consider the circuit of Figure 3.24 with the following parameters: $W = 100\ \mu m$, $L = 2\ \mu m$, $I_B = 3\ mA$, $V_B = 2.5\ V$, $C_L = 100\ fF$, $R_D = 1\ k\Omega$ and $R_s = 10\ k\Omega$.

(a) Estimate the required DC input bias V_S such that $I_D = I_B$ and $V_{OUT} = V_B$. Neglect channel-length modulation.

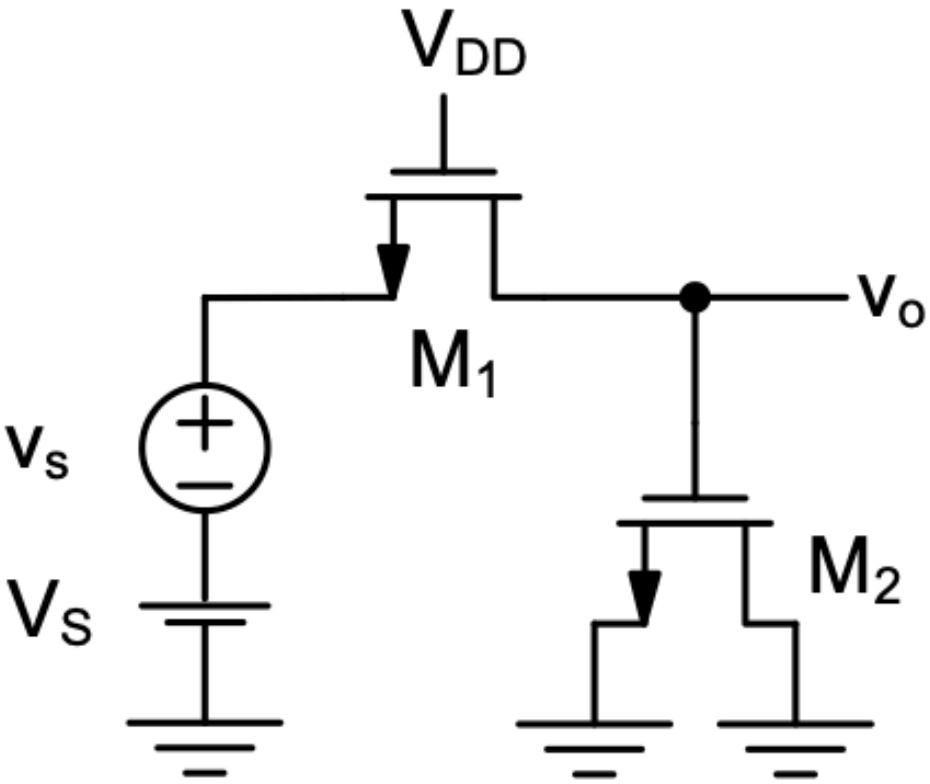


Figure 3.28.

(b) Calculate the MOSFET's transconductance and all device capacitances.

(c) Estimate the circuit's 3-dB bandwidth considering only the intrinsic gate capacitance.

(d) Estimate the circuit's 3-dB bandwidth using the Miller approximation.

(e) Estimate the circuit's 3-dB bandwidth using an OCT analysis.

P3.13 Consider the circuit shown in Figure 3.29.

(a) Write an analytical expression for the circuit's 3-dB bandwidth using an OCT analysis.

(b) Determine the exact analytical result for the circuit's 3-dB bandwidth.

(c) Compute the percent-error of the result in part (a), relative to the accurate result of part (b).

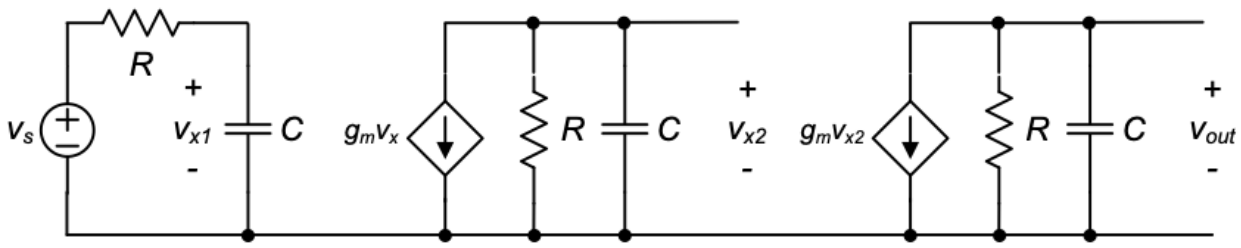


Figure 3.29.

P3.14 In this chapter, we saw that using the OCT method to estimate a circuits' bandwidth tends to be conservative. For example, in a circuit with two identical real poles (and no zeros), the bandwidth predicted using the OCT method is 22% lower than the actual bandwidth [see Equation 3.75]. Derive an analytical expression $f(n)$ that returns the percent error of the OCT analysis for a circuit with n identical real poles and no zeros. Note that $f(2) = -22\%$.

P3.15 Consider the common-source voltage amplifier of Figure 3.24. The goal of this design problem is to achieve a small-signal DC gain of -4 and a 3-dB bandwidth of 80 MHz. In addition, we wish to minimize the current consumption of the circuit. For simplicity in your calculations, neglect channel-length modulation and consider only the intrinsic gate capacitance. Assume the following parameters: $R_s = 2R_D$, $R_D = 5\ k\Omega$, $C_L = 1\ pF$.

a. Show that the required drain current I_D is related to the circuit's parameters and specifications as expressed below. In your analysis, approximate ω_{3dB} using an OCT analysis. Plot I_D as a function of V_{OV} for $L = 1\ \mu m$ and $L = 1.5\ \mu m$.

$$I_D = \frac{1}{2}\frac{C_L \cdot |A_{v0}| \cdot \omega_{3dB} \cdot V_{OV}}{1 - \frac{2}{3}\frac{L^2}{\mu_n V_{OV}} \cdot \frac{R_s}{R_D} \cdot |A_{v0}|\, \omega_{3dB}}$$

Note from this result that the choice of the gate overdrive voltage V_{OV} plays an important role in minimizing the required drain current.

b. From the expression and plots found in part (a), it is clear that the minimum channel length minimizes the current consumption of the amplifier. Explain in your words why this should be the case.

c. The drain current expression derived in (a) has a minimum for a certain value of V_{OV}. Calculate this value assuming $L = 1\ \mu m$ (minimum length). Also calculate R_D, the device width and drain current for the transistor at this optimum point.

d. Simulate the design using SPICE with the bias current and device geometries calculated in part (c). Measure the bandwidth of the circuit using an AC simulation. Since the SPICE transistor model contains extrinsic capacitances and finite output resistance, your circuit should fall short of the desired specs (despite the fact that we have used a conservative OCT estimate for ω_{3dB}). Calculate the percent discrepancies in the gain and bandwidth of the circuit.

e. Use a spreadsheet or math tool (Excel, $MATLAB^{®*}$, etc.) to setup the design equations for gain and bandwidth that include extrinsic capacitances and finite output conductance. With these additional modeling components added, it is difficult to derive a compact closed form solution as above. However, the setup in the spreadsheet will allow you to sweep the design parameters easily to find the new optimum that meets the gain and bandwidth specs. There are many different ways in which the spreadsheet can be structured. One is to use the width of the transistor as the main "knob" and calculate/iterate over all other parameters. The hand-calculated result can be used as an initial guess in this optimization. Use your spreadsheet to calculate the new bias current and device size that will meet the specs.

f. Simulate the refined design from (e) in SPICE and verify that you meet the desired specs. What is the obtained I_D, and how much larger is this value compared to the result from (c)?

> **Note**
>
> MATLAB is a registered trademark of The MathWorks, Inc.,
> 3 Apple Hill Road, Natick, MA.

P3.16 For the circuit shown in Figure 3.30, prove the following results, quantifying the Thévenin resistances seen between each pair of transistor terminals. Neglect the finite r_o of the MOSFET.

$$R_{gs} = \frac{R_G + R_S}{1 + g_m R_S}$$

$$R_{ds} = \frac{R_D + R_S}{1 + g_m R_S}$$

$$R_{gd} = R_G + R_D + G_m R_G R_D$$

where

$$G_m = \frac{g_m}{1 + g_m R_S}$$

P3.17 The circuit shown in Figure 3.31 is called a "source degenerated" common-source voltage amplifier. Analyze this circuit as indicated below.

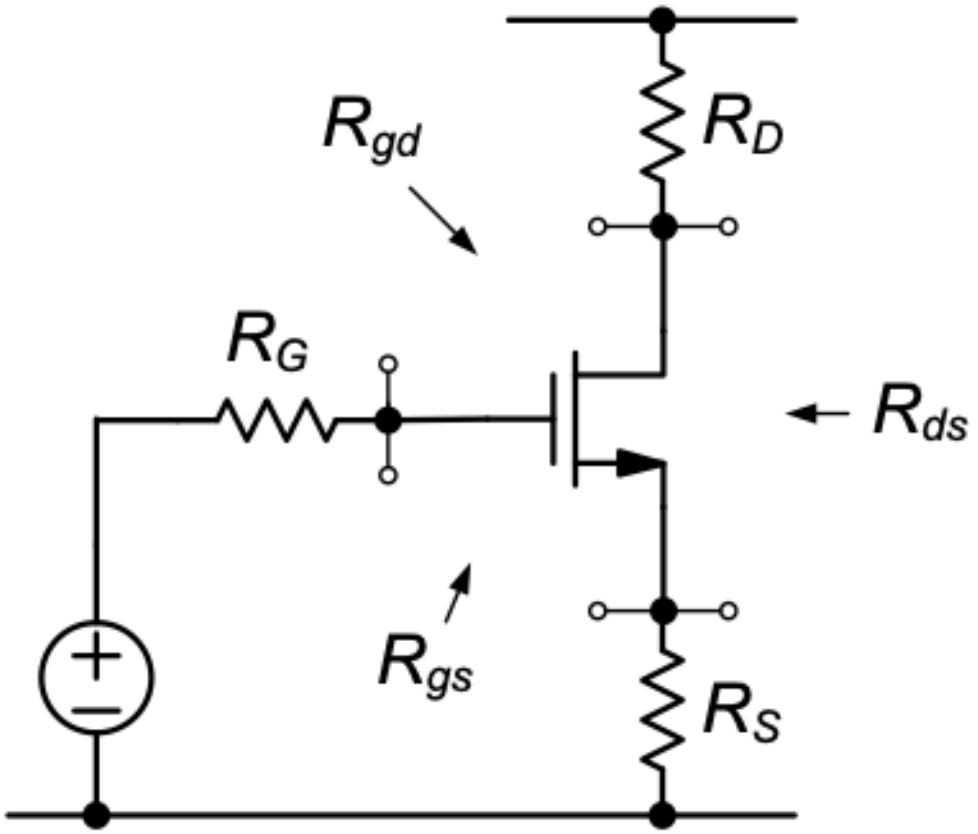


Figure 3.30.

a. Neglecting channel-length modulation and all capacitances in the circuit, show that the circuit's small-signal DC gain is given by

$$\frac{v_{out}}{v_{in}} \simeq -G_m R_D$$

where

$$G_m = \frac{g_m}{1 + g_m R_S}$$

is called the compound transconductance.

b. Using the results stated in Problem 3.16, estimate the DC gain and 3-dB bandwidth of the circuit assuming the following parameters: $R_G = 10\ k\Omega$, $R_S = 1\ k\Omega$, $R_D = 5\ k\Omega$, $I_D = 500\ \mu A$, $L = 1\ \mu m$, and $W = 20\ \mu m$. Consider only the two time constants contributed by C_{gs} and Cgd.

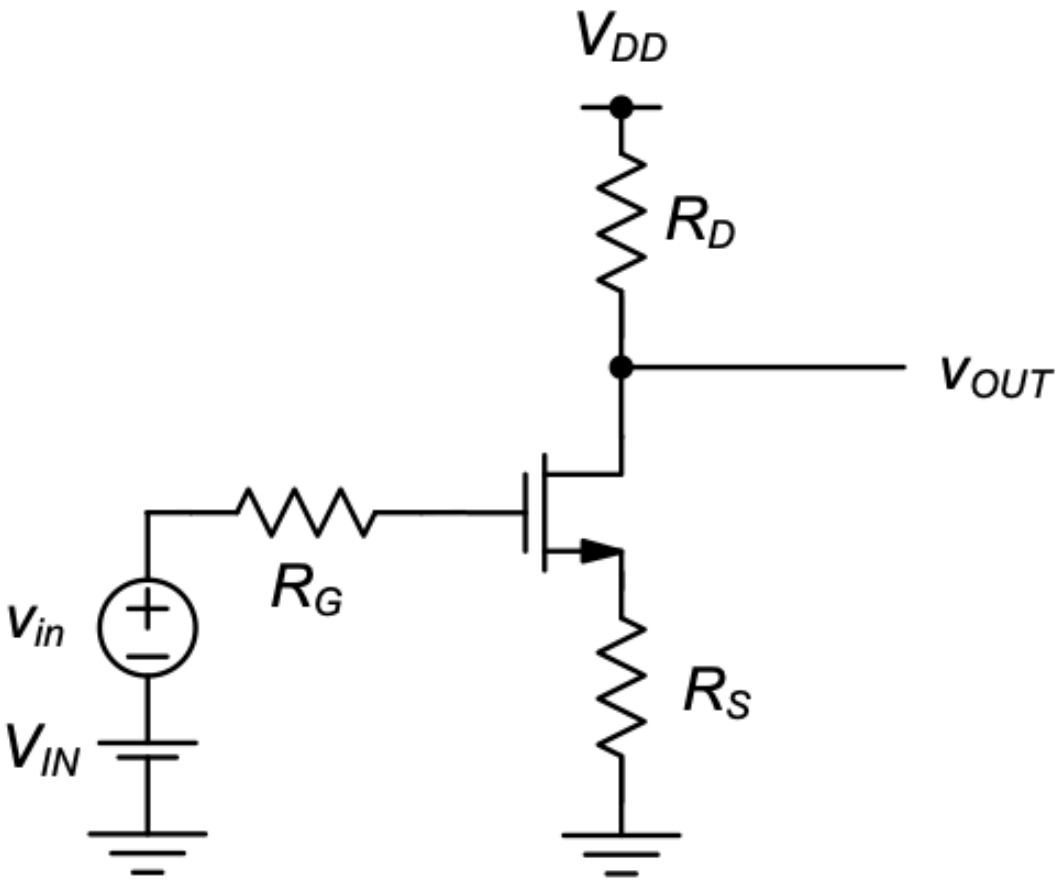


Figure 3.31.

4. The Common-Gate and Common-Drain Stages

TBD.

5. Biasing Circuits

The elementary transistor stages analyzed in the previous chapters rely on proper voltage and current biasing to function. So far, we have emulated these bias generators using ideal voltage and current sources. In this chapter, we will look at practical realizations of these elements using MOSFETs and passive components available within an integrated circuit.

While there exist numerous possibilities for setting up bias voltages and currents, we consider here a subset of options that have proven to be robust in practical circuits produced in volume. Many of the ideas and considerations that go into the design of bias circuitry are intimately related to the parameter variations seen in an integrated circuit process technology. For instance, threshold voltages cannot be accurately reproduced from fabrication run to fabrication run, and this mandates certain measures for desensitization to this parameter. In order to understand the rationale behind the proposed biasing circuits, we therefore include an overview of the basic variability issues that analog CMOS circuit designers must be aware of.

Chapter Objectives

- Review basic variability issues relevant for analog integrated circuits in CMOS technology.
- Discuss and analyze practical circuits that can establish the bias voltages and currents required to operate the elementary common-source, common-gate and common-drain stages.

5.1. Overview

Figure 5.1 provides and overview of the circuitry and topics that will be discussed in this chapter. The function of the shown circuits will be explained as we progress through this chapter. Following this introduction, we will investigate basic issues of process variation and device mismatch seen in a typical CMOS fabrication process. This review will help motivate some of the design choices made in later sections. Next, in Section 5.3, we investigate current mirror circuits, which are essential to distributing and generating bias currents in an integrated circuit.

The current that flows into a current mirror circuit is defined by an absolute current reference, for which there exist many different realizations. In Section 5.4, we will study one relevant example of a suitable circuit. In Section 5.5, we will then shift to the problem of bias voltage generation, as relevant for example in setting up the proper gate bias of a common-source or common-gate stage.

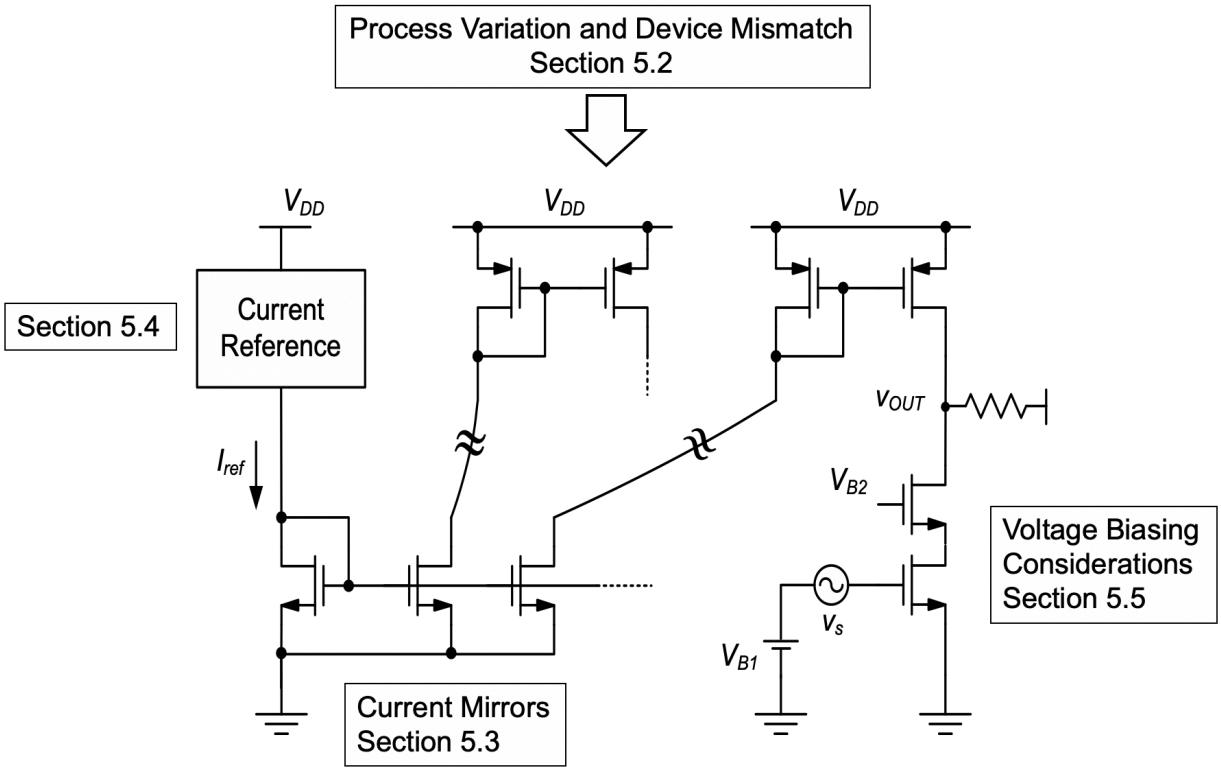


Figure 5.1.: Overview of biasing circuits and topics covered in this chapter.

5.2. Introduction to Process Variation and Device Mismatch

5.2.1. Process and Temperature Variations

In our analysis of elementary circuit configurations, we have so far implicitly assumed that the underlying component parameters (e.g., the threshold voltage of a MOSFET) are constant and accurately known. Unfortunately, this is not the case in reality. Especially in mass-produced integrated circuits, there are various forms of variability that result in parameter uncertainty due to imperfect fabrication, lifetime drift and influence of environmental conditions such as temperature and humidity. In a typical large semiconductor company, entire departments tend to focus on this issue, and there exists a wealth of related information that could easily fill multiple textbooks. As a result, the focus in this introductory module is to take a cursory look at only the basic issues, to the extent that this can help shape our thinking on how to arrive at practical and relatively insensitive circuit realizations.

The first issue that we will review in this section is related to variations arising from imperfect fabrication and temperature changes. In the context of fabrication imperfections, we will clearly distinguish between **global process variations** and **device mismatch**. The former term relates to variations that affect all devices on a chip uniformly, while the latter term refers to differences between nominally identical devices that are fabricated on the same chip (see Section 5.2.2).

Analog circuit designers often use the term **PVT variations** to refer to global variations in process, supply voltage, and temperature (see Table 5.1). The most basic way to capture global fabrication process variations is to define parameter sets that group the worst case outcomes as "slow," "nominal," and "fast" conditions. This nomenclature was adopted in the context of digital circuits (relating to the speed of a logic gate), but is also used among analog designers. The various parameter sets are often called process corners.

Table 5.1.: Examples of typical process, voltage and temperature (PVT) variations.

Process	The chip foundry defines three parameter sets for "slow," "nominal," and "fast" conditions.
Voltage	The chip's supply voltage is expected to vary by ±10%. For a nominal supply of 5 V, this means that all circuits must work for $V_{DD} = 4.5\ldots5.5$ V.
Temperature	Consumer products are typically expected to work in ambient temperatures ranging from 0…70° C. Circuits used in automotive applications must work reliably from –40…125° C.

Table 5.2 shows how some important integrated circuit parameters may vary across the three process parameter sets. Here, the nominal column contains the MOSFET parameters that we have assumed so far in this module (see Table 4-1). In the slow parameter set, the threshold voltage is increased and the transconductance parameters (C_{ox}) are reduced; this is the parameter combination that yields the slowest speed in a logic gate. The opposite is true for the fast parameter set.

Table 5.2.: Example of a slow, nominal and fast parameter set in a CMOS fabrication process. These parameters assume that the temperature is held constant at 25 ° C (room temperature).

Parameter	Slow	Nominal	Fast
V_{T0n} , $\lvert V_{T0p}\rvert$	0.7 V	0.5 V	0.3 V
$\mu_n C_{ox}$	40 $\mu A/V^2$	50 $\mu A/V^2$	60 $\mu A/V^2$
$\mu_p C_{ox}$	20 $\mu A/V^2$	25 $\mu A/V^2$	30 $\mu A/V^2$
R_{poly}	60 $\Omega/square$	50 $\Omega/square$	40 $\Omega/square$
R_{well}	1.4 $k\Omega/square$	1 $k\Omega/square$	0.6 $k\Omega/square$
C_{poly}	1.15 $fF/\mu m^2$	1 $fF/\mu m^2$	0.85 $fF/\mu m^2$

Table 5.2 also contains examples for parameter variations in passive IC components. R_{poly} and R_{well} are the sheet resistances of a resistor formed by a layer of polysilicon or n-well, respectively. C_{poly} is the capacitance parameter of a parallel plate capacitor formed by two layers of polysilicon. Advanced texts on integrated circuit design such as (Gray et al. 2009a) provide further information about the make-up of these and similar components.

The tabulated parameter variations do not take temperature variations into account; these must be added on top of the spread from fabrication. Table 5.3 lists a few typical temperature coefficients for each parameter. For example, if the operating temperature of a chip changes from 0°C to 70°C, the threshold voltage a MOSFET will shift by an additional –84 mV.

Table 5.3.: Typical temperature coefficients for integrated circuit device parameters.

Parameter	Temperature Coefficient
V_{T0n} , $\lvert V_{T0p}\rvert$	− 1.2 mV/°C
$\mu_n C_{ox}$	− 0.33 %/°C
$\mu_p C_{ox}$	− 0.33 %/°C
R_{poly}	+ 0.2 %/°C
R_{well}	+ 1 %/°C
C_{poly}	− 30 ppm/°C

The main take-home from the shown data is that in practice, the analog IC designer cannot view component parameters as constant numbers. His or her circuit must be immune to the level of variability described and function reliably across a large array of outcomes in process, voltage, and temperature. To show how significant these effects can be when neglected, the following example considers the impact of process variations on the bias point of a common-source amplifier.

Example 5-1: Impact of Process variations in a Common Source Amplifier

The circuit in Figure 5.2 was previously analyzed in Example 2-2(b) using nominal parameters (at room temperature). Given $V_{DD} = 5$ V, $R_D = 10$ kΩ, W/L = 10 and a desired output bias point of $V_{OUT} = 2.5$ V, we found that the input bias voltage should be set to $V_{IN} = 1.5$ V. Assuming $V_{IN} = 1.5$ V, recompute the circuit's operating assuming that the MOSFET parameters have shifted to the fast corner case given in Table 5.2. For simplicity, ignore variations in R_D and operating temperature.

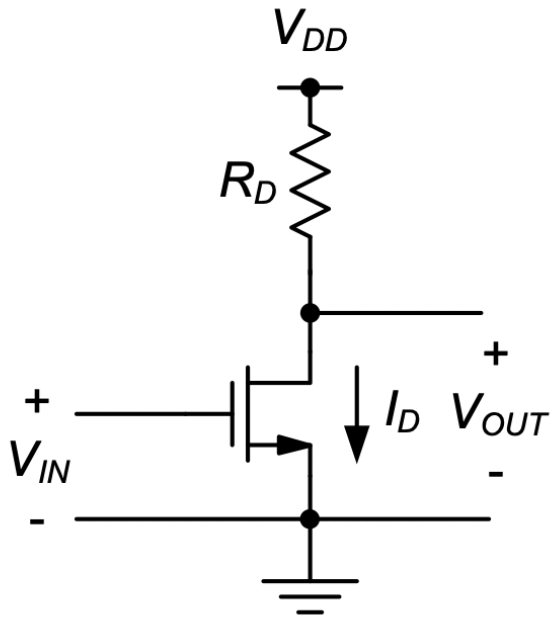


Figure 5.2.: Ex5-1

SOLUTION

Using the given parameters, we can directly compute:

$$\begin{aligned} V_{out} &= V_{DD} - \frac{1}{2}\mu_n C_{ox}\frac{W}{L}\Big(V_{IN} - V_{Tn}\Big)^2 R_D \\ &= 5V - \frac{1}{2}\cdot 60\frac{\mu A}{V^2}\cdot 10\cdot(1.5-0.3)^2\cdot 10k\Omega = 0.68V \end{aligned}$$

From this result, we see that the MOSFET no longer operates in saturation (since $V_{DS} = 0.68$ V $< V_{GS} - V_{Tn} = 1.2$ V). Using the MOSFET's equation for the triode region, we can compute $I_D =$ 408 A and $V_{OUT} = 918$ mV. This outcome differs substantially from the nominal operating point, and the circuit will essentially not function as intended for the fast corner conditions.

The main finding from the above example is that it will usually be impractical to bias the input of a common-source stage using a fixed bias voltage source. In practice, the integrated circuit designer generates bias voltages using circuits that will automatically adjust to corner-induced parameter spread and thereby make the circuit immune to process variations (see Section 5.5).

Generally speaking, a substantial amount of design time is usually spent on identifying biasing approaches that ensure a circuit's proper bias point across all possible operating conditions. In addition, once, the circuit is properly biased, the designer must verify that it maintains its key specifications across corners. A typical scenario is to guarantee a certain worst-case gain or bandwidth across all PVT scenarios.

As already mentioned, it is impossible to cover all aspects of robust design across PVT variations at the introductory level of this module. Nonetheless, having some of the basic knowledge established above will help us argue qualitatively about the practicality of the circuits discussed in this chapter, and ensure that they will at least have a chance to work in practice.

5.2.2. Mismatch

The process variations discussed in the previous section account for variability that affects all devices on a given chip equally. For example, all n-channel MOSFETs on a given chip may have slow parameters. Different from process variation, we use the term mismatch to capture variations between nominally identical devices, e.g., two MOSFETs of identical size on the same chip. Such variations are typically caused by line edge roughness, random doping fluctuations and similar effects.

Table 5.4.: Typical ranges of parameter mismatch for nominally identical, closely spaced components.

Parameter	Mismatch
V_{T0n} , $\lvert V_{T0p}\rvert$	5…30 mV
$\mu_n C_{ox}$, $\mu_p C_{ox}$	0.5…2%
R_{poly}	0.3…2%
C_{poly}	0.1…1%

Device mismatch typically follows Gaussian distributions and depends on device size and spacing (see Pelgrom, Duinmaijer, and Welbers 1989). For our purpose in this module, we will not expand upon the detailed theory behind this and instead consider only approximate numerical ranges that are typical for a technology as the one assumed in this module (see Table 5.4).

At first glance, we see from Table 5.4 that device mismatches are typically much smaller than global process variations.[1] For instance, the nominal threshold voltage for n-channel transistors can vary by ±200mV from fabrication run to fabrication run. However, within a specific fabrication outcome, the random threshold mismatch between two n-channels on the same chip is on the order of 10 mV.

This observation has a profound impact on the way integrated circuits are architected. That is, designers will usually try to exploit the fact that the components on the same chip show good matching. This contrasts with printed circuit board (PCB) design, where the designer often cannot rely on good matching between the available discrete components. Instead, PCB design can offer certain components with very high absolute accuracy across fabrication lots, such as 1%-precise resistors. As we know from Table 5.2, such levels of absolute accuracy are usually not available in an integrated circuit.

A classical example that exploits transistor matching is the so-called current mirror. This circuit is ubiquitous in integrated circuits, but infrequently used in PCB circuits. We will now analyze the **current mirror** as a first example of a biasing circuit that is insensitive to process variations

[1] This tends to hold true for technologies with feature sizes above 100 nm. For nano-scale devices, device mismatch can be comparable to process spread.

5.3. Current Mirrors

As we have seen in previous chapters, we would like to use current sources to setup the bias points for CS, CG, and CD stages. While we could in principle design individual, stand-alone current-source circuits each time we need a bias current, it is instead customary to work with only one (or a few) reference current generators on a given chip and "mirror" its current to the various locations where a bias current is needed. This is sketched out in Figure 5.1: a single reference current generator is used to feed a distribution network of current mirrors (to be discussed in this section), which then supplies bias currents to various circuit stages in a given chip or large sub-block.

In this section, we will discuss and analyze current mirror circuits at various levels of detail. We will begin by considering the most basic structure and perform a first-order analysis for this circuit. Next, we consider second-order error sources and look at an improved realization that invokes the cascode structure introduced in Section 4-5-1.

5.3.1. First-Pass Analysis of the Basic Current Mirror

Figure 5.3 shows the most basic realization of a current mirror using two identically sized n-channel MOSFETs. The circuit takes an input current I_{IN} and produces an output current I_{OUT}. Neglecting channel-length modulation for the time being, we can compute the gate-source voltage of M_1 using:

$$V_{GS1} = V_{IN} = V_{Tn} + \sqrt{\frac{2I_{IN}}{\mu_n C_{ox}\frac{W}{L}}} \tag{5.1}$$

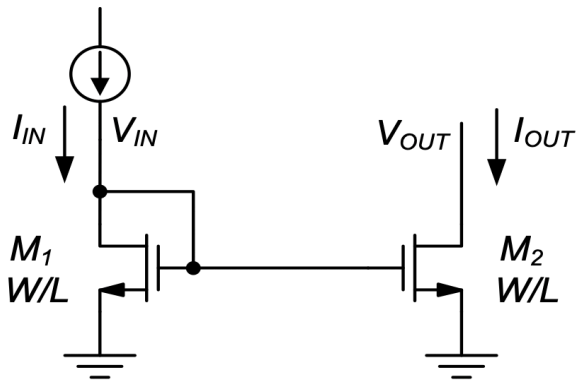


Figure 5.3.: Basic current mirror.

Since the gates and sources of the two MOSFETs are connected, we see that $V_{GS2} = V_{GS1}$, and therefore:

$$\frac{I_{OUT}}{I_{IN}} = \frac{I_{D2}}{I_{D1}} = \frac{\frac{1}{2}\mu_n C_{ox}\frac{W}{L}\left(V_{GS2} - V_{Tn}\right)^2}{\frac{1}{2}\mu_n C_{ox}\frac{W}{L}\left(V_{GS1} - V_{Tn}\right)^2} = 1 \tag{5.2}$$

Thus, the output current equals the input current (to first-order). In essence, the function of M_1 is to "compute" the gate-source voltage required for M_2 to supply the same current that is injected into M_1.

One important feature of this circuit is that it is immune to global process variations. From Equation 5.2, we see that absolute changes in V_{Tn} and $\mu_n C_{ox}$ that are common to M_1 and M_2 do not affect the current ratio. The circuit is affected only by mismatches in these parameters. However, as we have seen in Section 5.2.2, parameter mismatch tends to be small in integrated circuits.

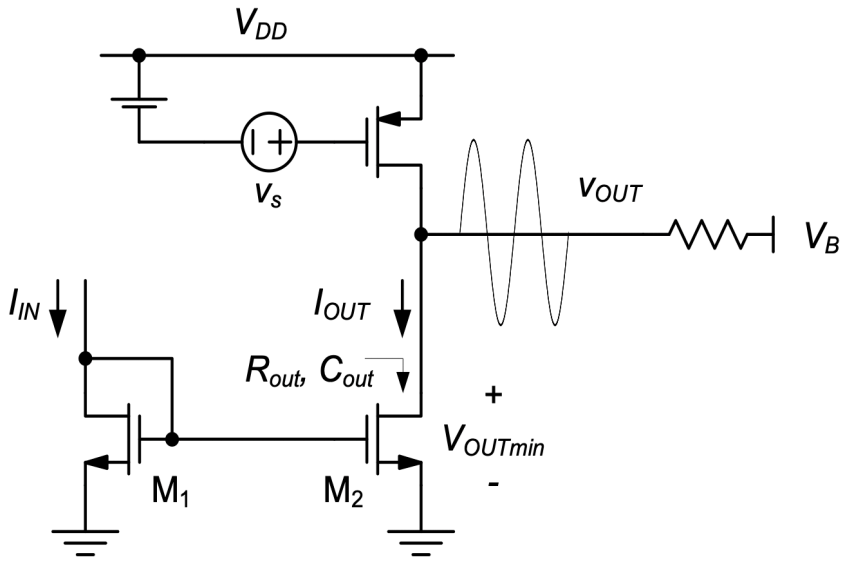


Figure 5.4.: Application of the basic current mirror in a common-source amplifier.

Figure 5.4 shows an application example of the basic current mirror in a p-channel common-source amplifier. This example is useful for identifying some general design objectives:

- We want to minimize the error in I_{OUT} so that the bias current of the common-source device is accurately set (see also Section 2-2-8).
- We want to minimize the voltage that is needed to keep M_2 in saturation (to allow for a large signal swing). We call this minimum voltage level the compliance voltage, V_{OUTmin}.
- We want to minimize the capacitance C_{out} that the current mirror contributes to the output node of the amplifier. This will help maximize the circuit's bandwidth.
- We want to maximize R_{out}, the resistance looking into M_2. A small R_{out} can substantially reduce the voltage gain of the circuit in some use cases.
- Lastly, it is desirable to scale the mirror's branch currents, that is, we want $I_{OUT} = K \cdot I_{IN}$, where typically $K > 1$. This helps reduce the overall current consumption of the circuit and provides flexibility in adjusting the current values within a larger distribution network.

These objectives tend to hold in general for all types of current mirror implementations and we will keep them in mind as we progress through the remaining subsections. For the time being, let us look into the scaling of branch currents. Essentially, we would like to accomplish

$$\frac{I_{OUT}}{I_{IN}} = \frac{I_{D2}}{I_{D1}} = \frac{\frac{1}{2}\mu_n C_{ox}\frac{W_2}{L_2}\Big(V_{GS2} - V_{Tn}\Big)^2}{\frac{1}{2}\mu_n C_{ox}\frac{W_1}{L_1}\Big(V_{GS1} - V_{Tn}\Big)^2} = K \tag{5.3}$$

where K is the current scaling factor. From the above expression, assuming that V_{Tn} and $\mu_n C_{ox}$ are exactly equal for both transistors, it follows that $K = (W_2/L_2)/(W_1/L_1)$. Thus, current scaling can be conveniently realized by scaling the MOSFETs' aspect ratios.

In practice, several guidelines exist on how exactly this scaling should be implemented. The first and most important guideline is that we should always maintain $L_1 = L_2$; the current scaling should be realized by scaling the widths rather than the lengths of the channels. This is preferable since the current in a modern MOSFET does not accurately scale with $1/L$. As already mentioned in Chapter 2, the $1/L$ proportionality in our equations is essentially due to the simplified physical model that we used in the derivation of the square-law expressions. For the 1- m technology assumed in this module, the deviation from the square law model is not as severe as for today's sub-100-nm transistors but still significant enough to avoid length scaling in current mirrors. Now, with $L_1 = L_2$, the current scaling factor is simply $K = W_2/W_1$, to first-order. In the next section, we will look at various second-order effects that cause K to deviate from the width ratio of the MOSFETs.

5.3.2. Second-Pass Analysis of the Basic Current Mirror

There exist several error sources in a current mirror that will affect its scaling factor. In general, we classify these error sources into two categories: systematic and random errors. Examples of systematic errors are

- Errors in transistor width ratios, for example due to mask misalignment or systematic etching imperfections.
- Differences in the drain-source voltages between M_1 and M_2, leading to current deviations caused by channel length modulation.
- Differences in the source potentials of M_1 and M_2 due to finite resistance in the interconnect (so-called "IR drop").

Examples of random errors are

- Random mismatches in device geometries, for example due to line edge roughness.
- Random mismatch in the transistors' threshold voltage or transconductance parameter.

In order to attain the best possible accuracy in a current mirror, the IC designer will typically try to minimize the impact of all of these errors. We will therefore analyze some of the most important effects and countermeasures in the following paragraphs. For simplicity, our analysis will consider each effect separately. Ultimately, however, the sum of all errors must be considered in practice.

To analyze the impact of systematic masking or etching errors, consider the specific example of a current mirror with a desired current ratio of two and a layout as shown in Figure 5.5. Here M_2 is drawn twice as wide as M_1 In an ideal situation, this would yield $K = 2$ based on the first order result of the previous subsection. In a typical IC process, however, masking or etching errors can lead to a systematic error in the width of a MOSFET, indicated as W in the shown layout. With this error, and neglecting any other imperfections for simplicity, we have

$$K = \frac{I_{OUT}}{I_{IN}} = \frac{2W_1 + \Delta W}{W_1 + \Delta W} = \frac{2 + \frac{\Delta W}{W_1}}{1 + \frac{\Delta W}{W_1}} \approx 2 - \frac{\Delta W}{W_1} \tag{5.4}$$

where the final approximation follows from a first-order Taylor expansion and holds for $\Delta W/W_1 << 1$.

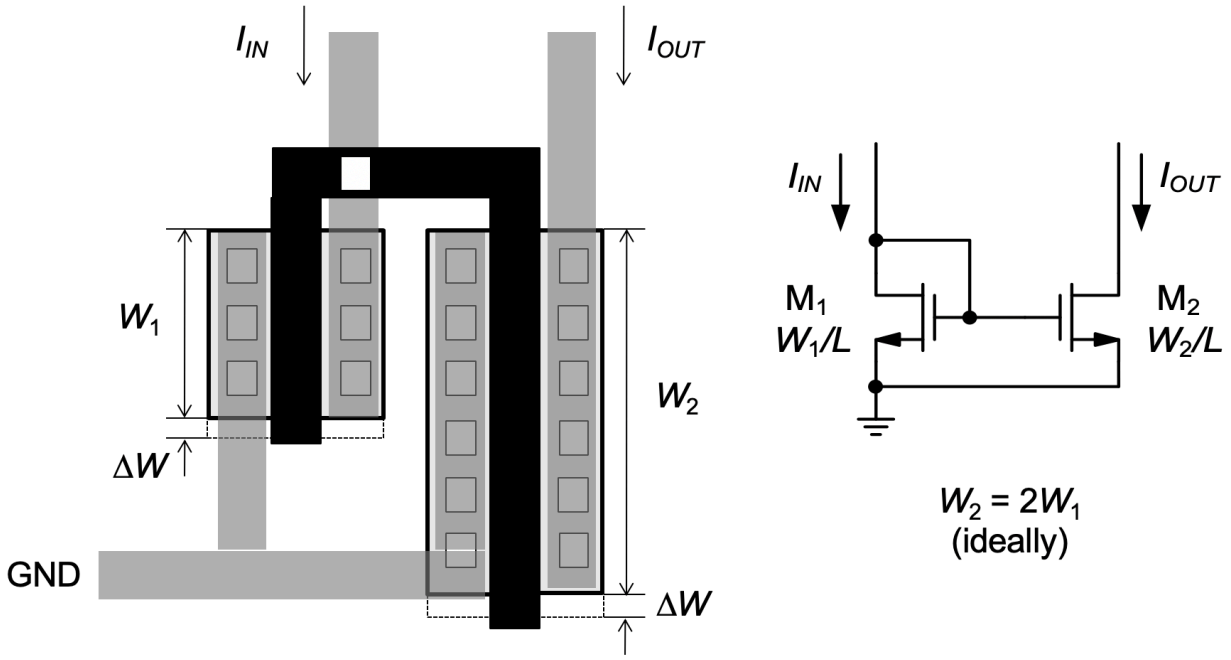


Figure 5.5.: Layout of a current mirror with a desired current ratio of two.

Especially for small transistors, the error term in Equation 5.4 can be significant. Therefore, it has become customary to adopt layout styles that eliminate issues due to W altogether. In the improved layout of Figure 5.6, M_2 is formed using two unit devices whose layout is identical to that of M_1. In this case,

$$K = \frac{I_{OUT}}{I_{IN}} = \frac{2W_1 + \Delta W}{W_1 + \Delta W} = 2 \tag{5.5}$$

and thus the circuit is insensitive to systematic width errors. Note that the idea of working with unit devices can be extended such that P unit devices are used for M_2 and Q unit devices are used to form M_1 This means that the mirror ratio $K = P/Q$ is restricted to rational numbers.

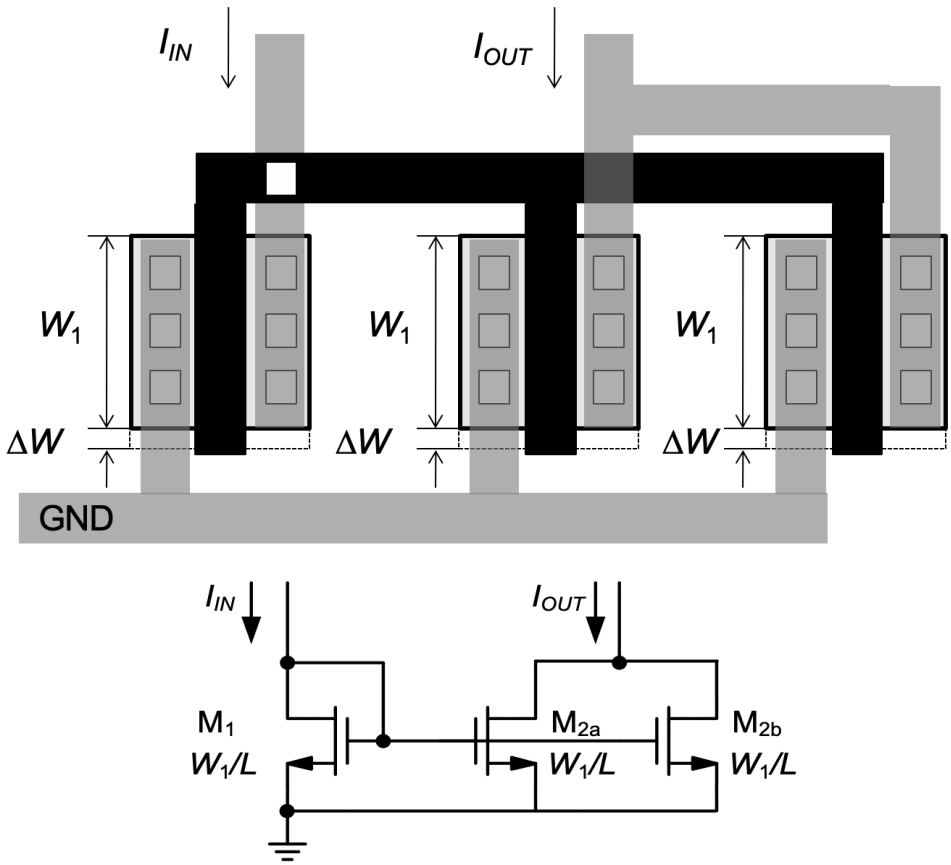


Figure 5.6.: Improved layout of a current mirror with a desired current ratio of two.

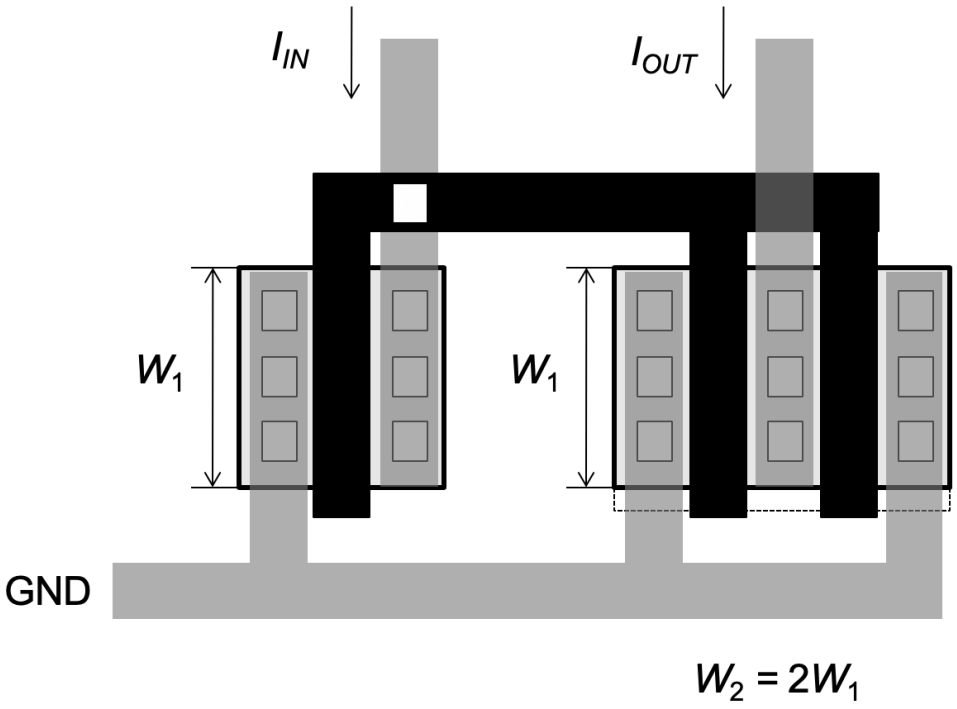


Figure 5.7.: Layout of a current mirror with shared drain regions.

A variant of the improved unit-device layout is shown in Figure 5.7. Here, the two unit transistors share a single drain region at the output node and therefore have a smaller output capacitance (C_{out} in Figure 5.4). This general idea is often applied when small C_{out} is desired in the particular use case of the current mirror. One disadvantage of the layout in Figure 5.7 is that the source/drain orientation of the rightmost channel are flipped. This can lead to residual systematic errors in process technologies that suffer from source/drain asymmetries. However, it can be shown that this error vanishes when an even number of unit devices are used for both M_1 and M_2 These and many other considerations are part of the knowledge base of experienced analog designers. The reader is referred to advanced literature on this topic for further information.

Another significant source of error in the current mirror ratio can result from differences in the voltages at the input and output nodes of the mirror. To see this, consider the current mirror example in Figure 5.8, which is assumed to have perfectly matched transistors of the same size. Even though the two transistors have identical output curves, their drain currents will differ whenever the input and output voltages do not match. Mathematically, we can analyze this effect by including channel-length modulation in the analysis. Specifically, since

$$I_{IN} = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}\Big(V_{GS1} - V_{Tn}\Big)^2(1 + \lambda_n V_{IN}) \tag{5.6}$$

and

$$I_{OUT} = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}\Big(V_{GS1} - V_{Tn}\Big)^2(1 + \lambda_n V_{OUT}) \tag{5.7}$$

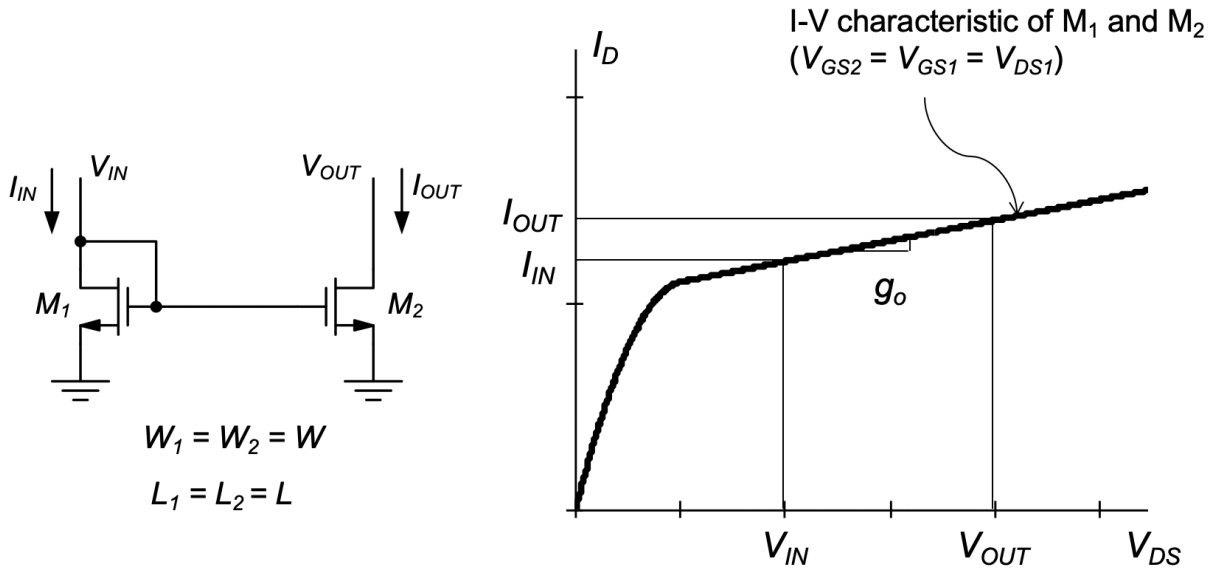


Figure 5.8.: Mirror error due to differences in drain-source voltage.

we have

$$\frac{I_{OUT}}{I_{IN}} = \frac{1 + \lambda_n V_{OUT}}{1 + \lambda_n V_{IN}} \tag{5.8}$$

From this result, we see that there are two ways to reduce errors in the current ratio. We can try to minimize the difference between V_{OUT} and V_{IN} as much as possible and/or reduce λ_n by using long-channel MOSFETs. Note also that reducing n is equivalent to reducing the small-signal output conductance g_o, which is simply the slope of the I-V curves in Figure 5.8 The smaller this slope, the smaller the difference between I_{OUT} and I_{IN}.

Example 5-2: Current Mirror Error Due to Drain-Source Voltage Difference

Consider the current mirror in Figure 5.8 . Assume $V_{OUT} = 2.5$ V and that the MOSFET width is chosen such that $V_{IN} = 1.5$ V. Calculate the percent error in the current ratio for L = 1 m and L = 3 m.

SOLUTION

For L = 1 m, we have \$ _n \$ = 0.1 V^{-1} [seeEq.(2.44)]. Using Equation 5.8, we find in this case

$$\frac{I_{OUT}}{I_{IN}} = \frac{1 + 0.1V^{-1} \cdot 2.5V}{1 + 0.1V^{-1} \cdot 1.5V} = 1.087 \tag{5.9}$$

The error in the current ration is 8.7%. Repeating the above calculation for $L = 3\mu m(\lambda_n = 0.033V^{-1})$, the error reduces to 3.2%.

Another example of a systematic error source that we will consider here is the voltage drop in the source connection of the mirror devices (see Figure 5.9). In the shown circuits, we assume for

simplicity that the two MOSFETs are identical and that $V_{OUT} = V_{IN}$, i.e., there is no error due to V_{DS} differences.

First consider the circuit of Figure 5.9(a), which takes the finite wiring resistance (R_{WIRE}) between the source terminals of M_1 and M_2 into account. The wire will carry some current, which is at the minimum equal to the drain current of M_1 flowing toward the ground node of the circuit. In a poorly constructed layout, the wire may also carry the current from another block (I_X) as shown. The total current in the wire is therefore $I_{WIRE} = I_{IN} + I_X$ and $V_{WIRE} = I_{WIRE} \cdot R_{WIRE}$. By applying KVL in Figure 5.9(a) we see that $V_{GS2} = V_{GS1} + V_{WIRE}$. Therefore, we can use the equivalent model of Figure 5.9(b) for further analysis.

Now, assuming that V_{WIRE} is relatively small, we can think about this voltage as a small-signal perturbation around the operating point of M_2 ($I_{D2} = I_{OUT} = I_{IN}$). Therefore, we can write

$$I_{OUT} = I_{IN} + g_m V_{WIRE} \tag{5.10}$$

and

$$\frac{I_{OUT}}{I_{IN}} = 1 + \frac{g_m}{I_{IN}} V_{WIRE} = 1 + 2\frac{V_{WIRE}}{V_{OV}} \tag{5.11}$$

where g_m and V_{OV} are the transconductance and quiescent point gate overdrive (V_{GS}–V_{Tn}) of the MOSFETs, respectively. To see that this error source can be quite significant, consider the case of $V_{OV} = 200mV$ and $V_{WIRE} = 10mV$. The resulting error in the mirror ratio is 10%. In practice, the designer will mitigate voltage drop issues by (1) minimizing the distance between M_1 and M_2 (to minimize R_{WIRE}), (2) avoid any excess current (I_X) in the source connection between M_1 and M_2 , and (3) work with reasonably large gate overdrive voltages (V_{OV}) in current mirrors.

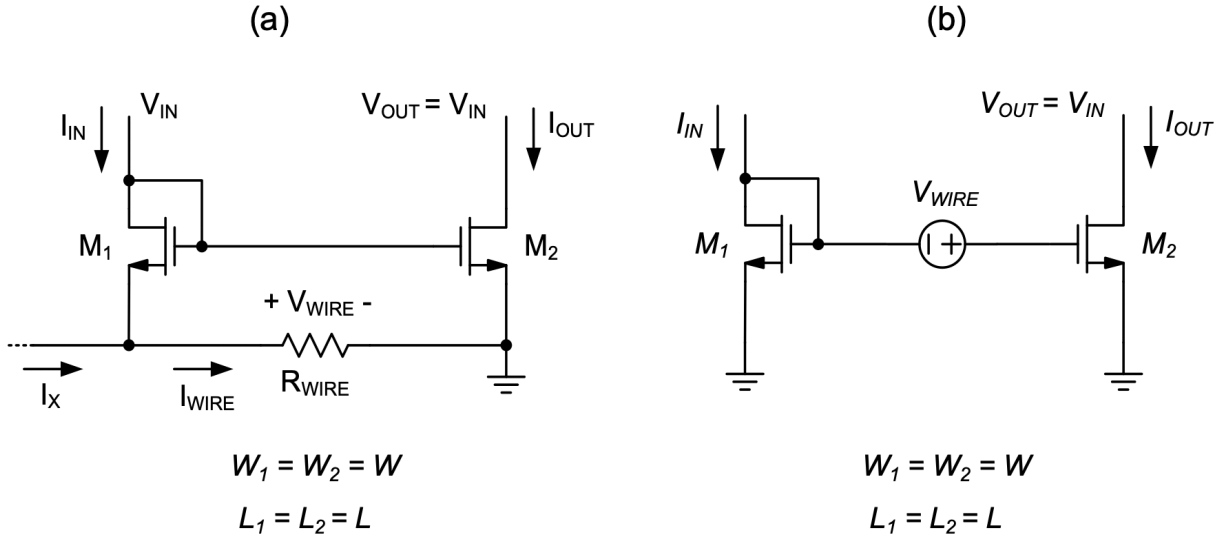


Figure 5.9.: Mirror error due to voltage drop in the source connection.

As a final step in this subsection, let us now consider a few examples of random mismatch effects in current mirrors, and specifically mismatch in the transistors' threshold voltages and transconductance parameters. The case of threshold voltage mismatch can be modeled exactly as shown in Figure 5.9(b), but V_{WIRE} is now replaced with ΔV_{Tn}, the threshold voltage mismatch between M_1 and M_2. Therefore, we can write in this case

$$\frac{I_{OUT}}{I_{IN}} = 1 + 2\frac{\Delta V_{Tn}}{V_{OV}} \tag{5.12}$$

The conclusion from this expression is similar to what we have already stated above. To minimize errors due to threshold voltage mismatch, the designer must work with reasonably large values of V_{OV}. Since $\Delta V_{Tn} \approx 10mV$ is not unusual in CMOS technology (see Section 5.2.2), it follows that it is rather difficult to guarantee highly accurate mirror ratios. Even if we make $V_{OV} = 1V$, the corresponding error is still 2%. For the case of transconductance parameter mismatch, we can write

$$\frac{I_{OUT}}{I_{IN}} = \frac{\frac{1}{2}(\mu_n C_{ox})_2 \frac{W}{L}\Big(V_{GS2} - V_{Tn}\Big)^2}{\frac{1}{2}(\mu_n C_{ox})_1 \frac{W}{L}\Big(V_{GS1} - V_{Tn}\Big)^2} = \frac{(\mu_n C_{ox})_2}{(\mu_n C_{ox})_1} \tag{5.13}$$

where $(\mu_n C_{ox})_{1,2}$ are the transconductance parameters of the two MOSFETs, and all other parameters are assumed to be equal. Thus, for typical mismatch in the transconductance parameter on the order of 1% (see Section 5.2.2), it is often the case that this particular error is overshadowed by mismatches in the MOSFETs' threshold voltages.

5.3.3. Multiple Current Sources and Sinks

The basic current mirror concept discussed so far can be utilized to provide multiple current outputs that either source a current from V_{DD} or sink a current into ground. A circuit that uses p-channel devices to create multiple currents sourced from V_{DD} is shown in Figure 5.10.

If we also require current sinks, the circuit approach shown in Figure 5.11 can be used. Here, the output current from device M_1 is used as a reference current for the n-channel mirror composed of M_1 and M_2. Note that a direct application for this circuit would be the CS-CD amplifier of Figure 4-28; it requires one current source from V_{DD} and one current sink.

Neglecting all error terms, and assuming equal channel lengths, the value of the DC current I_{OUT1} is equal to

$$I_{OUT1} = \frac{W_1}{W_R} I_{REF} \tag{5.14}$$

From this current we have derived a current source and current sink with devices M_2 and M_4. Ideally, these currents are

$$I_{OUT2} = \frac{W_2}{W_R} I_{REF} \tag{5.15}$$

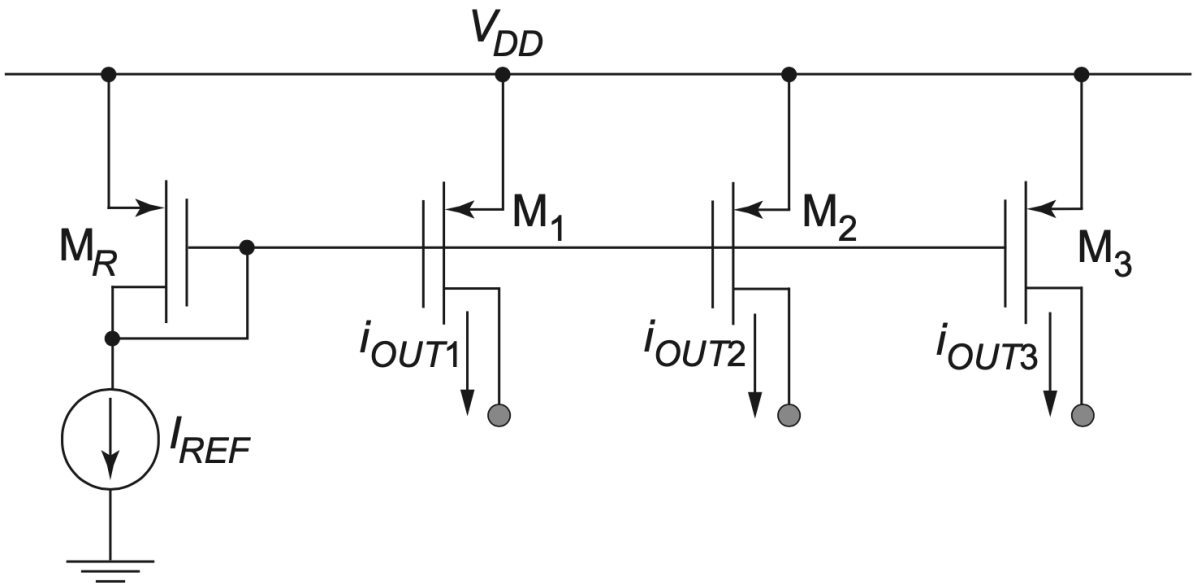


Figure 5.10.: PMOS current mirror with multiple outputs.

$$I_{OUT4} = \frac{W_4}{W_R} I_{OUT1} = \left(\frac{W_4}{W_3} \cdot \frac{W_1}{W_R}\right) I_{REF} \tag{5.16}$$

Example 5-3: Current Sources/Sinks

Design current sources with DC current values of 10 A and 20 A and current sinks with DC current values of 10 A and 40 A. The small-signal source resistance of all current sources and sinks should be at least 1 MΩ. The compliance voltage of both current sources and sinks must be less than 0.5 V. You are given one reference current source of 10 A with which you can derive the others.

SOLUTION

A suitable topology for this design is shown in Figure 5.12. We begin this design by realizing that in order to meet the compliance voltage requirement, we need $V_{GS} = V_{SG} \leq 1V$. This defines the value of $(W/L)_R$.

$$V_{GS} = V_{Tn} + \sqrt{\frac{I_{REF}}{\frac{1}{2}\mu_n C_{ox}\frac{W}{L}}} \Rightarrow \left(\frac{W}{L}\right)_R = 1.6$$

If we set $(W/L)_1 = (W/L)_2 = 1.6$, then $I_{D1} = I_{D2} = 10\mu A$. To make $I_{D3} = 40\mu A$, let $(W/L)_3 = 4(W/L)_2 = 6.4$. The p-channel devices are sized the same way.

$$V_{SG} = 1V = -V_{Tp} + \sqrt{\frac{I_{REF}}{\frac{1}{2}\mu_p C_{ox}\frac{W}{L}}} \Rightarrow \left(\frac{W}{L}\right)_4 = 3.2$$

To make $I_{D5} = 10\mu A$ and $I_{D6} = 20\mu A$, we use $(W/L)_5 = 3.2$ and $(W/L)_6 = 6.4$.

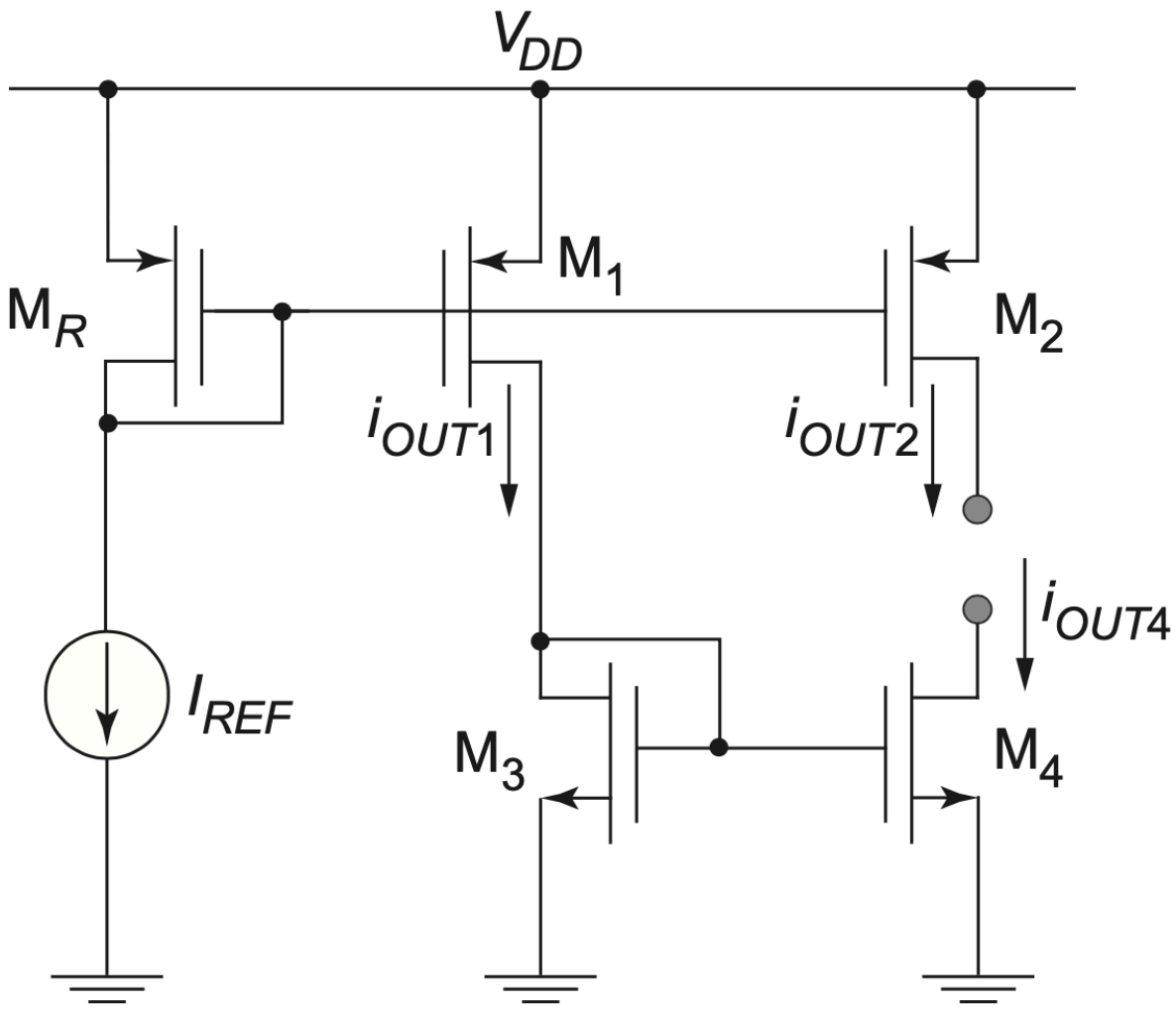


Figure 5.11.: Circuit to produce a current source M_2 and current sink M_4.

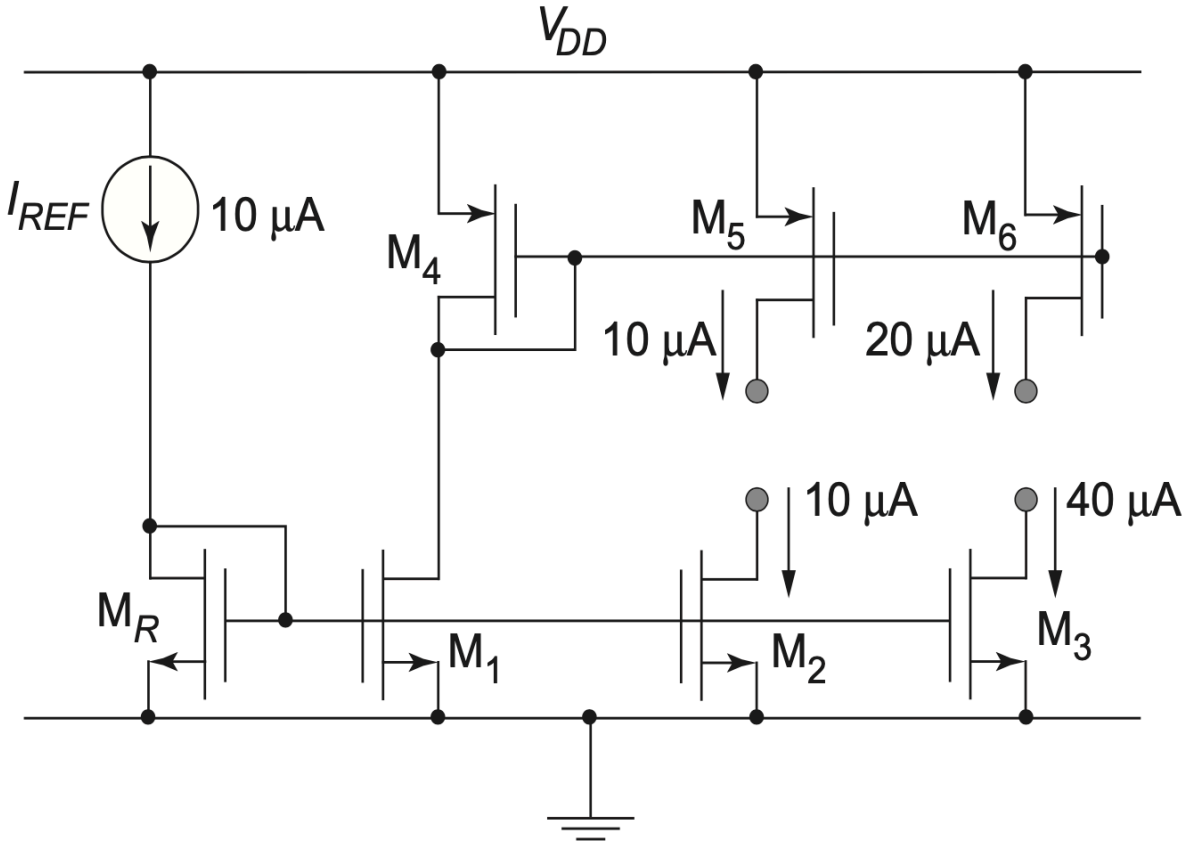


Figure 5.12.: Ex5-3

Now we can check the small-signal source resistances. For $I_D = 10\mu A$ and $\lambda_n = \lambda_p = 0.1\mu m V^{-1}/L$, minimum length $L = 1\mu m$ will already satisfy the requirement of $r_o = 1M\Omega$. For $I_D = 20\mu A$, we need $L = 2\mu m$ and for $I_D = 40\mu A$, we need $L = 4\mu m$. Since we want all of the n-channels to have the same length, and all of the p-channels to have the same length, we arrive at the following design choice (all values in m): $(W/L)_R = (W/L)_1 = (W/L)_2 = 6.4/4$ and $(W/L)_3 = (4 \times 6.4)/4$. For the p-channels: $(W/L)_4 = (W/L)_5 = 6.4/2$, and $(W/L)_6 = (2 \times 6.4)/2$. As indicated through the multipliers, the layout of M_3 and M_6 should consist of multiple unit devices.

5.3.4. Cascode Current Mirror

As we have seen in the previous subsection, the accuracy of the current ratio in the basic current mirror is affected by a number of undesired effects. The cascode current mirror discussed in this section improves on a subset of these issues. Specifically, as we shall see, it is less sensitive to differences between V_{IN} and V_{OUT} and correspondingly also provides a much larger output resistance (R_{out}). The most basic realization of a cascode current mirror is shown in Figure 5.13(a). The output branch of this circuit stacks two MOSFETs in a cascode configuration (see Section 4-5-1). To compute the output resistance of this circuit, we consider the small-signal model of the circuit in Figure 5.13(b).

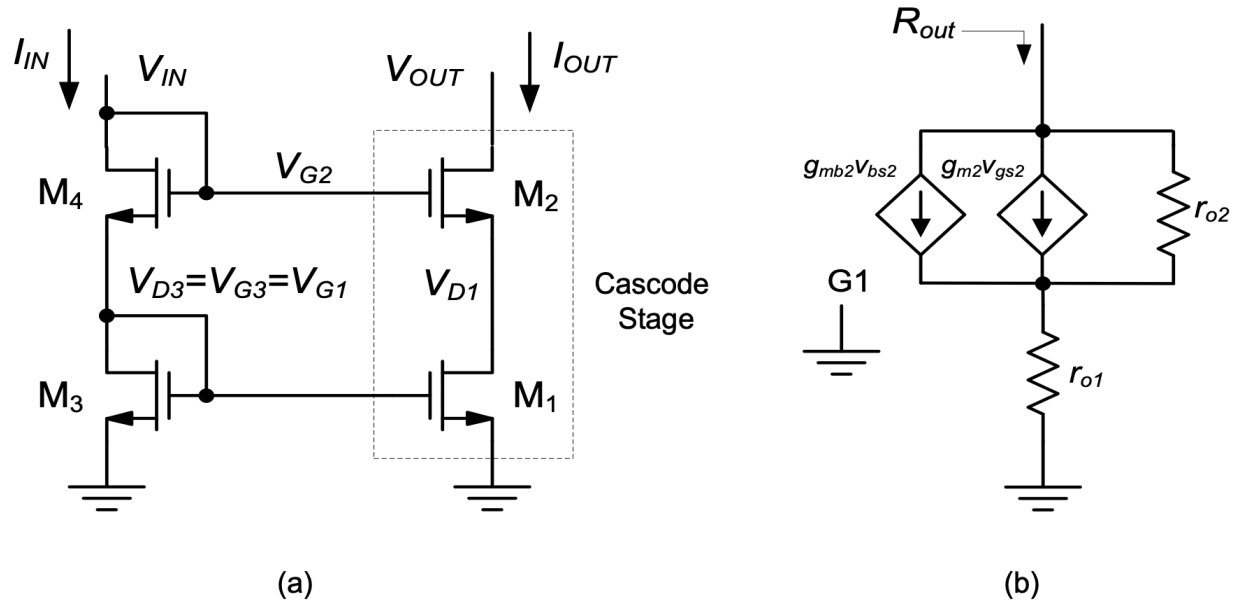


Figure 5.13.: Basic cascode current mirror. (a) Complete circuit. (b) Small-signal circuit model for the output branch.

Note that this circuit resembles the common-gate model of Figure 4-9, with r_s replaced by r_{o1}, which is the output resistance of the bottom transistor M_1. Consequently, R_{out} is given by Eq. (4.21), which is repeated here with the proper variable substitutions ($r_s \rightarrow r_{o1}, r_o \rightarrow r_{o2}$ and $g'_m \rightarrow g'_{m2}$, where $g'_{m2} = g_{m2} + g_{mb2}$)

$$R_{out} \approx r_{o2}[1 + g'_{m2} r_{o1}] \approx r_{o2} g'_{m2} r_{o1} \tag{5.17}$$

Thus, the output resistance of this structure is very large, which implies that any changes in the output voltage will not affect the output current significantly. Mathematically, we can view any disturbance in V_{OUT} as a small signal quantity, v_{out}. The resulting disturbance in the output current, i_{out} is simply v_{out}/R_{out}, which is small for large R_{out}.

Even though we know that R_{out} is large from the above quantitative result, it is useful to develop a qualitative feel for why this must be the case. To investigate, Figure 5.14 shows the output branch of the cascode current mirror for further inspection.

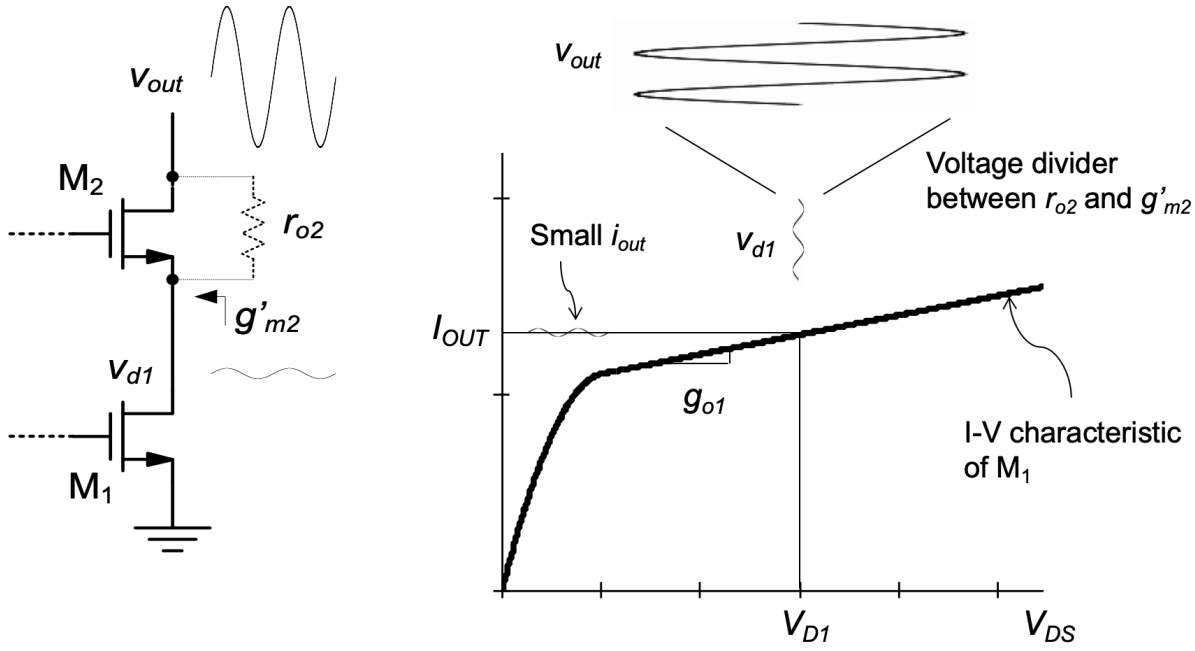


Figure 5.14.: Qualitative inspection of the output branch in a cascode current mirror.

In this drawing, we apply an output perturbation and consider the voltage swing at the drain of M_1. Since the resistance at the drain node of M_1 is low ($\approx 1/g'_{m2}$), the output voltage perturbation appears highly attenuated at this node (the attenuation is approximately given by the ratio of $1/g'_{m2}$ and r_{o2}). Consequently, the drain current of M_1, which is equal to the output current, sees only a very small voltage perturbation. In essence, M_2 shields the current mirror transistor M_1 from the output disturbance; the drain voltage of M_1 is "pinned" by the low-resistance node created by M_2.

Shielding effect of the cascode

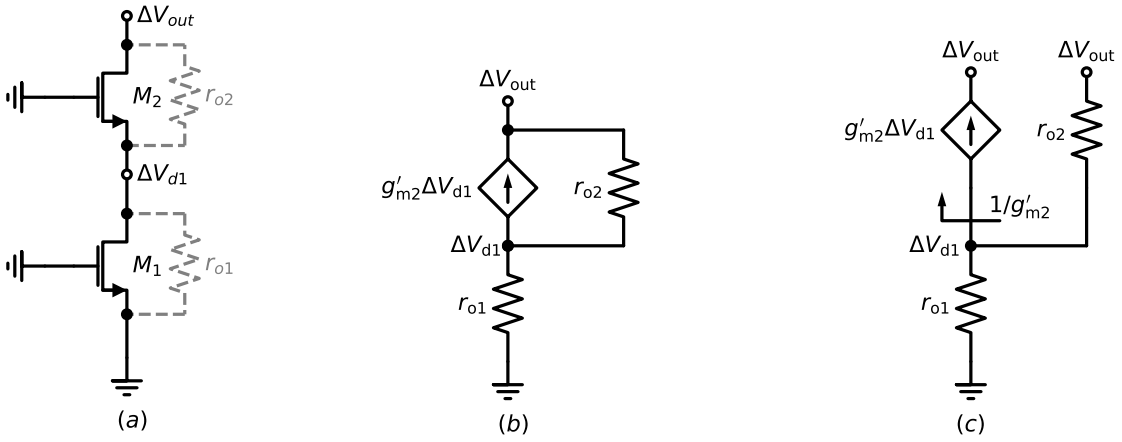


Figure 5.15.: (a) cascode with voltage perturbation applied at its output (b) & (c) equivalent small signal model

Applying KCL we can write

$$\frac{\Delta V_{out} - \Delta V_{d1}}{r_{o2}} - g'_{m2} \cdot \Delta V_{d1} = \frac{\Delta V_{d1}}{r_{o1}}$$

$$\frac{\Delta V_{out} - \Delta V_{d1}}{r_{o2}} = \Delta V_{d1} \cdot \left(g'_{m2} + \frac{1}{r_{o1}} \right)$$

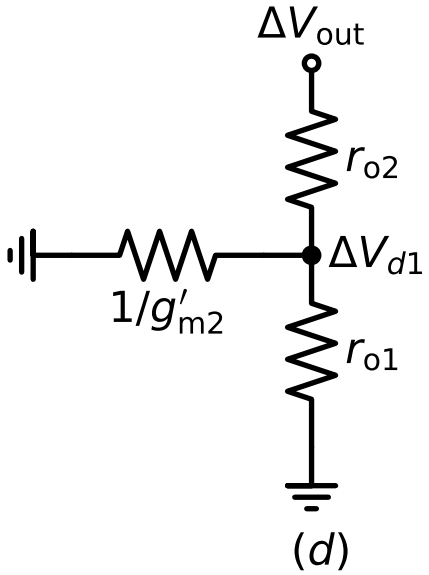


Figure 5.16.: (d) voltage divider between r_{o2} and $1/g'_{m2} \parallel r_{o1}$

For $g'_{m2}r_{o1} \gg 1$

$$\frac{\Delta V_{out} - \Delta V_{d1}}{r_{o2}} \approx \Delta V_{d1} \cdot g'_{m2}$$

The equation above reperesents a voltage divider between r_{o2} and $1/g'_{m2}$

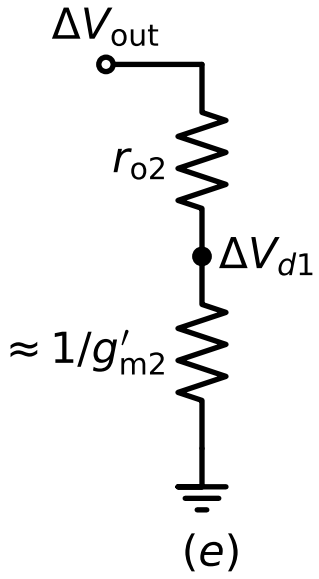


Figure 5.17.: (e) voltage divider between r_{o2} and $1/g'_{m2}$

While the circuit in Figure 5.13 is insensitive to changes in V_{OUT}, it is important to realize that any difference in the drain voltages of M_1 and M_3 will still lead to a (potentially large) systematic error. Similar to Equation 5.8, we can write

$$\frac{I_{OUT}}{I_{IN}} = \frac{I_{D1}}{I_{D3}} = \frac{1 + \lambda_n V_{DS1}}{1 + \lambda_n V_{DS3}} \tag{5.18}$$

For this reason, the circuit is purposely constructed such that nominally $V_{DS1} = V_{DS3}$. Assuming that $I_{OUT} = I_{IN}$ and that M_2 and M_4 are identical, applying KVL to the circuit of Figure 5.13 reveals

$$V_{D1} = V_{D3} + V_{GS4} - V_{GS2} = V_{D3} \tag{5.19}$$

Thus, the circuit of Figure 5.13 effectively eliminates this important shortcoming of the basic current mirror.

Unfortunately, the benefits of the cascode current mirror do not come for free. Specifically, notice that the circuit's output compliance voltage (V_{OUTmin} is significantly larger than that of a basic current mirror. In a basic current mirror (Figure 5.3), we have $V_{OUTmin} = V_{DSsat2} = V_{OV2}$, which is the gate overdrive voltage of the MOSFET in the output branch. For the cascode current mirror, we can investigate the situation by considering Figure 5.18, which graphically illustrates all voltage levels and voltage drops. Here, we assume for simplicity that all threshold (V_{Tn}) and gate overdrive voltages (V_{OV}) are identical. With this assumption, the voltage at the drain of M_1 is $V_{Tn} + V_{OV}$.

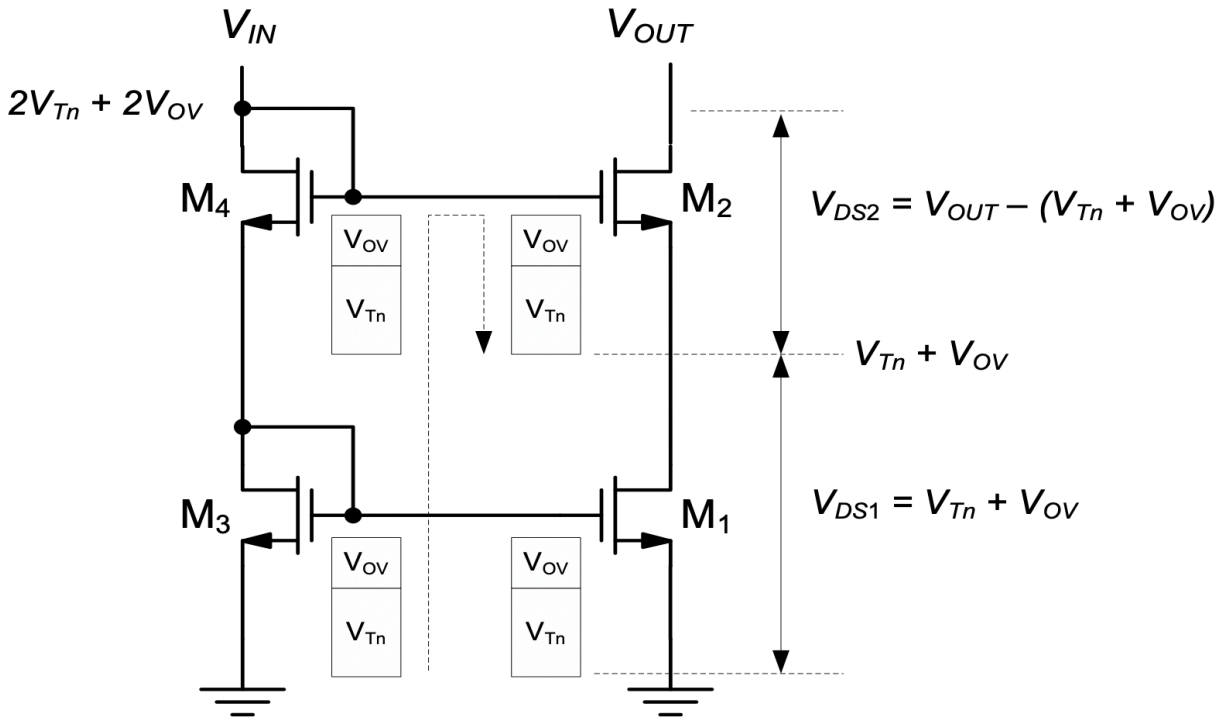


Figure 5.18.: Analysis of the compliance voltage in the basic cascode current mirror.

This implies that M_1 will always be in saturation, since the drain-source voltage exceeds V_{OV} by some margin (equal to V_{Tn}). In order for M_2 to operate in saturation, we require

$$V_{DS2} = V_{OUT} - (V_{Tn} + V_{OV}) > V_{DSsat2} = V_{OV} \tag{5.20}$$

and thus

$$V_{OUT} > V_{Tn} + 2V_{OV} \tag{5.21}$$

which means $V_{OUTmin} = V_{Tn} + 2V_{OV}$. Note that for typical values of V_{OV} and V_{Tn}, the compliance voltage of the cascode current mirror can become quite large, e.g., 0.5 V + 1 V = 1.5 V, taking away a significant amount of signal swing from the available voltage supply range (consider for example Figure 5.4).

5.3.5. The High-Swing Cascode Current Mirror

In applications where the large compliance voltage of the circuit in Figure 5.13 is problematic, an alternative scheme, called **high swing cascode current mirror** can be used. We will develop this circuit from the previous solution using a few intermediate steps.

First, consider the output branch of a cascode current mirror as shown in Figure 5.19. In the annotation of this circuit, it is assumed that V_{G2} is set up such that M_1 operates at the edge of saturation, i.e., $V_{DS} = V_{DSsat} = V_{OV}$. In this case, we require

$$V_{OUT} - V_{OV} > V_{DSsat2} = V_{OV} \tag{5.22}$$

and thus

$$V_{OUT} > 2V_{OV} \tag{5.23}$$

which means $V_{OUTmin} = 2V_{OV}$, corresponding to a substantial improvement over Equation 5.21.

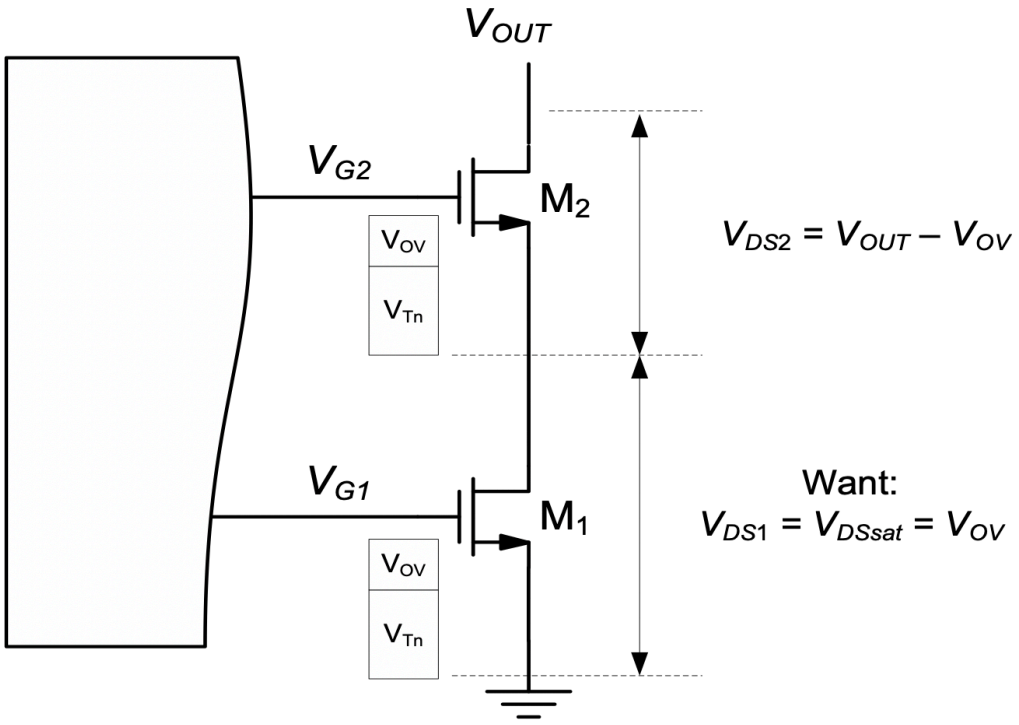


Figure 5.19.: Desired output branch biasing to achieve the minimum possible compliance voltage.

The question that remains is how exactly V_{G2} should be generated to achieve this improvement. To investigate, we first compute the required value of V_{G2} by applying KVL in Figure 5.19

$$V_{G2} - V_{DS1} + V_{GS2} = V_{Tn} + 2V_{OV} \tag{5.24}$$

It turns out that many options exist for setting V_{G2} to the above-calculated value. The most basic option is shown in Figure 5.20. Here, an extra current branch is introduced to bias the added transistor M_6. In a practical implementation, this current typically originates from an extra branch added to a PMOS current mirror in the overall biasing network. The key idea in this setup is that M_6 is sized to one-quarter the width used for M_2.

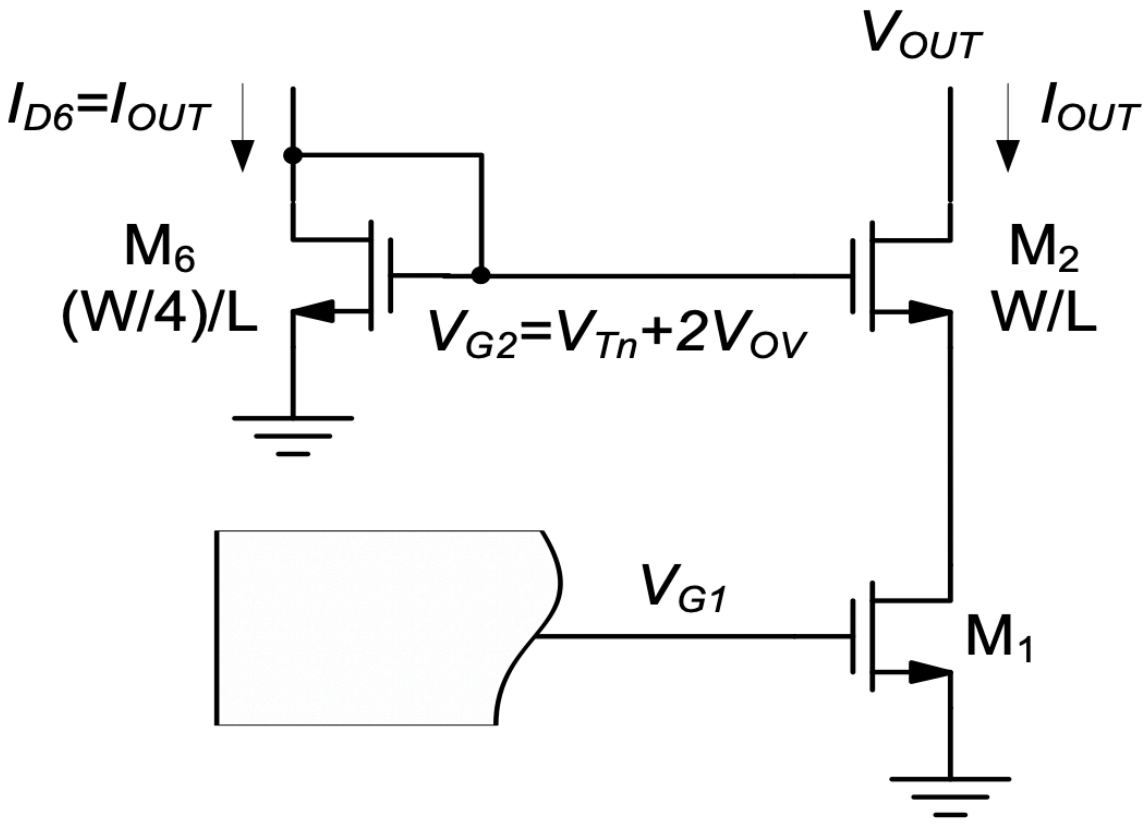


Figure 5.20.: Basic circuit for cascode voltage generation.

With this sizing, we have

$$\begin{aligned} V_{G2} = V_{GS6} &= V_{Tn} + \sqrt{\frac{2I_{D6}}{\mu_n C_{ox}\frac{W/4}{L}}} \\ &= V_{Tn} + 2\sqrt{\frac{2I_{OUT}}{\mu_n C_{ox}\frac{W}{L}}} = V_{Tn} + 2V_{OV} \end{aligned} \tag{5.25}$$

which achieves the desired objective.

In practice, the designer will usually not want to bias M_1 exactly at the edge of saturation, but rather leave some margin. This can be achieved by sizing the width ratio smaller than 1/4. Table 5.5 shows the resulting margins for a few integer ratios. Choosing a sizing ratio of 1/6 often yields a reasonable compromise between compliance voltage and circuit robustness in a practical circuit.

Table 5.5.: V_{DS1} as a function of the ratio $k = W_6/W_2$ in the circuit of Figure 5.21.

$\mathbf{k = W_6/W_2}$	$\mathbf{V_{DS1}}$	$\mathbf{V_{DS1} - V_{OV}}$ (Margin)
1/4	V_{OV}	0
1/5	1.24 V_{OV}	0.24 V_{OV}
1/6	1.45 V_{OV}	0.45 V_{OV}
1/7	1.64 V_{OV}	0.64 V_{OV}
1/8	1.83 V_{OV}	0.83 V_{OV}
1/9	2 V_{OV}	V_{OV}

In order to complete the high-swing cascode current mirror circuit, we still need to design the circuit's input branch. The most obvious (but non-preferred) solution for the input branch is shown in Figure 5.21(a).

This circuit suffers from the problem that $V_{DS1} \neq V_{DS3}$, and therefore a systematic error is introduced in the mirror ratio (see Equation 5.18). An elegant solution to this problem is shown in Figure 5.21(b), where M_4 has been added to replicate the gate-source voltage drop of M_2, such that $V_{DS1} = V_{DS3}$. Just as in the circuit of Figure 5.21(a), the gate voltage of M_3 self-adjusts to the point where M_3 carries the injected current (I_{IN}). M_4 merely acts as a current buffer, passing all of the input current to M_3.

The final circuit of Figure 5.21(b) has been widely used in practice and is insensitive to process variations, such as global shifts in threshold voltage, Nonetheless, there are two remaining issues with this circuit that are worth mentioning. First, it is sometimes inconvenient to provide the extra current source used to bias M_6. Problems P5.5 and P5.6 look into alternative solutions that do not require an extra input current source, but still achieve low compliance voltage in the output branch. The second issue stems from the backgate effect. In our analysis above, we have neglected the fact that the threshold voltage of M_2 will be larger than that of M_6. This is because the source of M_6 is connected to ground (and thus $V_{SB6} = 0$), while the source potential of M_2 is positive (and thus $V_{SB2} > 0$). As a result, assuming a sizing ratio of 1/4, V_{DS1} is more accurately given by

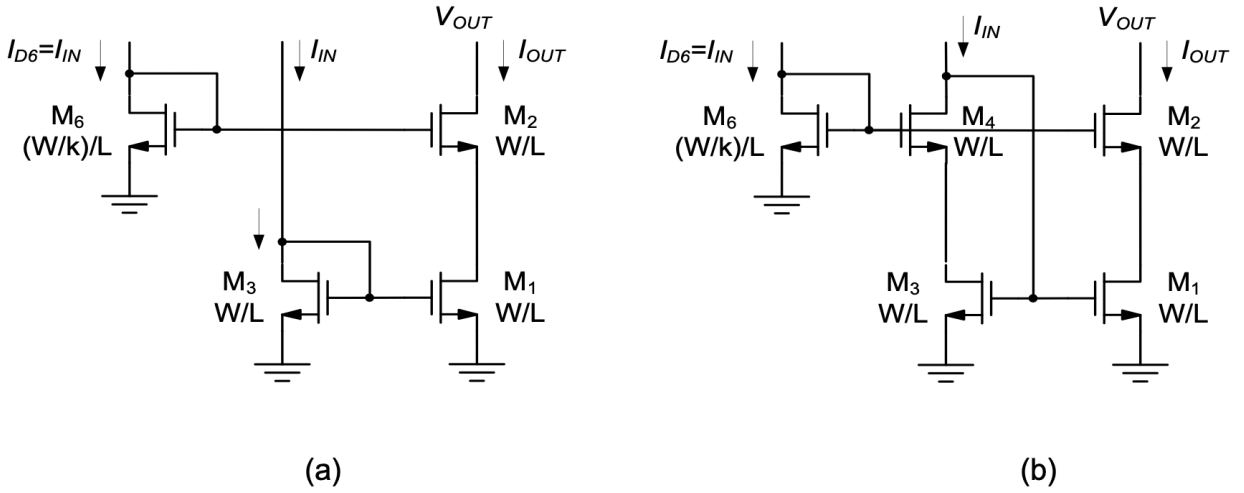


Figure 5.21.: High-swing cascode current mirror with input branch included. (a) Non-preferred solution that suffers from a systematic mirror ratio error. (b) Solution that avoids the systematic mirror ratio error.

$$\begin{aligned} V_{DS1} &= V_{GS6} - V_{GS2} \\ &= V_{Tn0} + 2V_{OV} - \Big[V_{Tn}(V_{SB2}) + V_{OV}\Big] \\ &= V_{OV} - \Big[V_{Tn}(V_{SB2}) - V_{OV}\Big] = V_{OV} - \Delta V_{Tn} \end{aligned} \tag{5.26}$$

where ΔV_{Tn} is a positive quantity that causes M_1 to enter the triode region, unless sufficient margin is provided. In practice, the designer can use computer simulations to ensure that sufficient saturation margin is guaranteed. Another option is to change the circuit to mitigate this problem at its root. The thought process that leads to the alternate solution is illustrated in Figure 5.22.

The biasing transistor M_6, as discussed previously, is redrawn in isolation in Figure 5.22(a). Figure 5.22(b) shows an equivalent circuit that breaks M_6 into four transistors, each with an aspect ratio of W/L. Assuming that the ideal square law model holds, the series connection of these transistors behaves like a MOSFET with aspect ratio of W/(4L), or (W/4)/L (see Problem P2.2).

Consequently, V_{G2} must be equal to $V_{Tn}+2V_{OV}$, as in the original circuit of Figure 5.22(a). Furthermore, notice that the transistor M_{6a} in Figure 5.22(b) must operate in the saturation region (since it is diode-connected). This means that this MOSFET's gate-source voltage is equal to $V_{Tn}+V_{OV}$, and the potential at its source node is equal to V_OV.

Next, in Figure 5.22(c), the three bottom transistors are lumped into a single device, again based on the argument that a device with an aspect ratio of W/(3L) can be replaced with one that has (W/3)/L. Note that the combined transistor (M_{6b} in Figure 5.22(c)) operates in the triode region, since its drain-source voltage (V_{OV}) is smaller than V_{GS6b}–$V_{Tn} = 2V_{OV}$. Of course, all of the above conceptual arguments can be validated quantitatively, by carrying out a first-principle analysis using MOSFET I-V equations.

The main advantage of the circuit in Figure 5.22(c) becomes apparent when it is inserted back into the cascode current mirror, as shown in fig-5.18. Since M_{6a} has the same W/L and carries

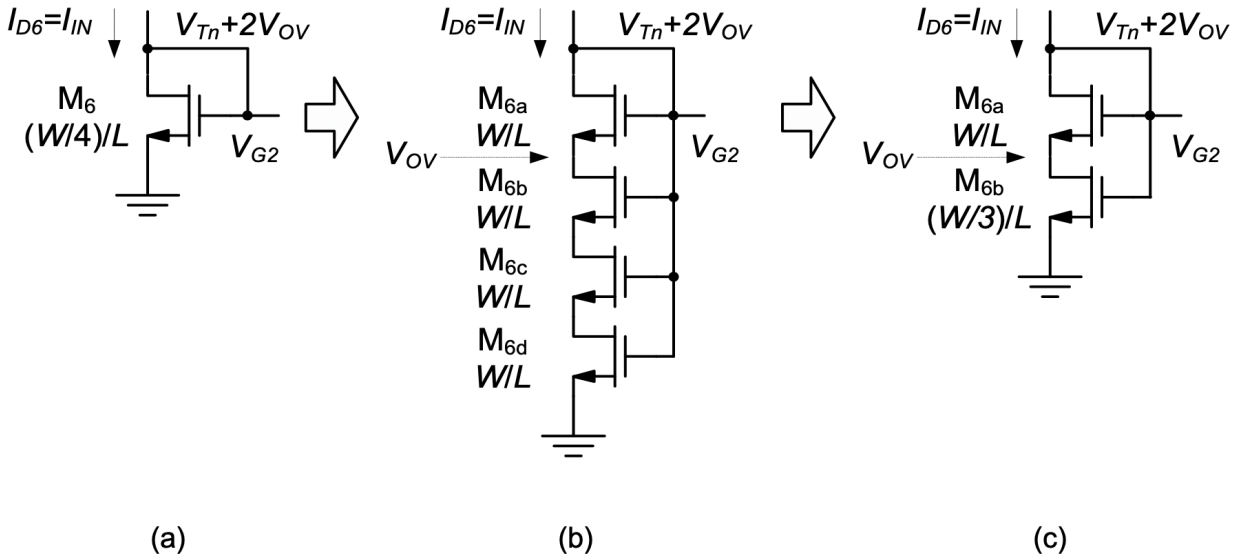


Figure 5.22.: Conceptual steps for replacing the W/4 cascode biasing device with a two-transistor compound circuit

the same current as M_2, the source potential of these transistors is identical (neglecting channel length modulation). Hence, the error term due to backgate effect that we saw in Equation 5.26 is suppressed, since $V_{SB2} = V_{SB6a}$.

$$\begin{aligned} V_{DS1} &= V_{DS6b} + V_{DS6a} - V_{GS2} \\ &= V_{OV} + V_{Tn}(V_{SB6a}) + V_{OV} - \Big[V_{Tn}(V_{SB2}) + V_{OV}\Big] \\ &= V_{OV} \end{aligned} \tag{5.27}$$

Finally, note that even though the circuit of Figure 5.23 provides a somewhat less error-prone setup for the generation of V_{G2}, the designer will still want to leave margin and back off from the ideal W/3 sizing for M_{6b}. Table 5.5 lists the margin for various integer choices larger than 3 (see also problem P5.4).

Example 5-4: Design of a Cascode Current Mirror

The cascode current mirror in Figure 5.23 is configured such that $I_{OUT} = I_{IN}$. In this example, we wish to design a similar current mirror that sets $I_{OUT} = 4I_{IN} = 400\mu A$. M_1 and M_2 are to be sized such that $V_{OV} = 200mV$ and using a channel length of 2 m. M_{6b} should be sized such that m = 5. Given these specifications, determine all transistor widths and node voltages. Also compute the circuit's output compliance voltage and output resistance (R_{out}). Neglect channel-length modulation in bias point calculations.

Table 5.6.: V_{DS1} as a function of width scaling factor m for W_{6b} in Figure 5.23.

m	$\mathbf{V_{DS1}}$	$\mathbf{V_{DS1} - V_{OV}}$ (Margin)
3	V_{OV}	0

m	$\mathbf{V_{DS1}}$	$\mathbf{V_{DS1} - V_{OV}}$ (Margin)
4	1.24 V_{OV}	0.24 V_{OV}
5	1.45 V_{OV}	0.45 V_{OV}
6	1.64 V_{OV}	0.64 V_{OV}
7	1.83 V_{OV}	0.83 V_{OV}
8	2 V_{OV}	V_{OV}

SOLUTION

The width of M_1 and M_2 is found by solving

$$I_{OUT} = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}V_{OV}^2$$

for W, and inserting the given numbers and technology parameters. This yields $W_1 = W_2 = 800\mu m$. To implement the current ratio $I_{OUT}/I_{IN} = 4$, we require $W_3 = W_4 = W_{6a} = W_1/4 = 200\mu m$ and $W_{6b} = W_{6a}/5 = 40\mu m$.

The voltage V_{G1} is simply $V_{Tn} + V_{OV} = 0.7V$. With m=5, we know from Table 5.6 that $V_{D1} = V_{D3} = V_{D6b} = 1.45V_{OV} = 290mV$. To compute V_{G2} accurately, we must first estimate the threshold voltage of M_{6a} using

$$V_{Tn}(V_{SB}) = V_{T0n} + \gamma_n \cdot \left(\sqrt{2\phi_f + V_{SB}} - \sqrt{2\phi_f}\right)$$

Evaluating the above equation with $V_{SB6a} = V_{D6b} = 290mV$ gives $V_{T6a} = 590mV$. Therefore, $V_{G2} = 1.45V_{OV} + V_{T6a} + V_{OV} = 1.08V$. The circuit's output compliance voltage is $V_{OUTmin} = V_{D1} + V_{OV} = 0.49V$. The circuit's output resistance is given by $R_{out} \cong ro1 \cdot g'_{m2}r_{o2}$ (see Section 4-5-1). Therefore, we compute

$$g_{m2} = \frac{2I_{D2}}{V_{OV}} = \frac{2 \cdot 400\mu A}{200mV} = 4mS$$

$$\begin{aligned} g'_{m2} &= g_{m2}\left(1 + \frac{\gamma_n}{2\sqrt{2\phi_f + V_{SB}}}\right) \\ &= 4mS\left(1 + \frac{0.6V^{-1}}{2\sqrt{0.8V + 0.29V}}\right) = 5.15mS \end{aligned}$$

and

$$r_{o1} \approx r_{o2} \approx \frac{1}{\lambda_n I_{D1}} = \frac{1}{0.05V^{-1} \cdot 400\mu A} = 500k\Omega$$

These numbers lead to $R_{out} \cong 12.88M\Omega$. The schematic in Figure 5.24 summarizes the results obtained in this example.

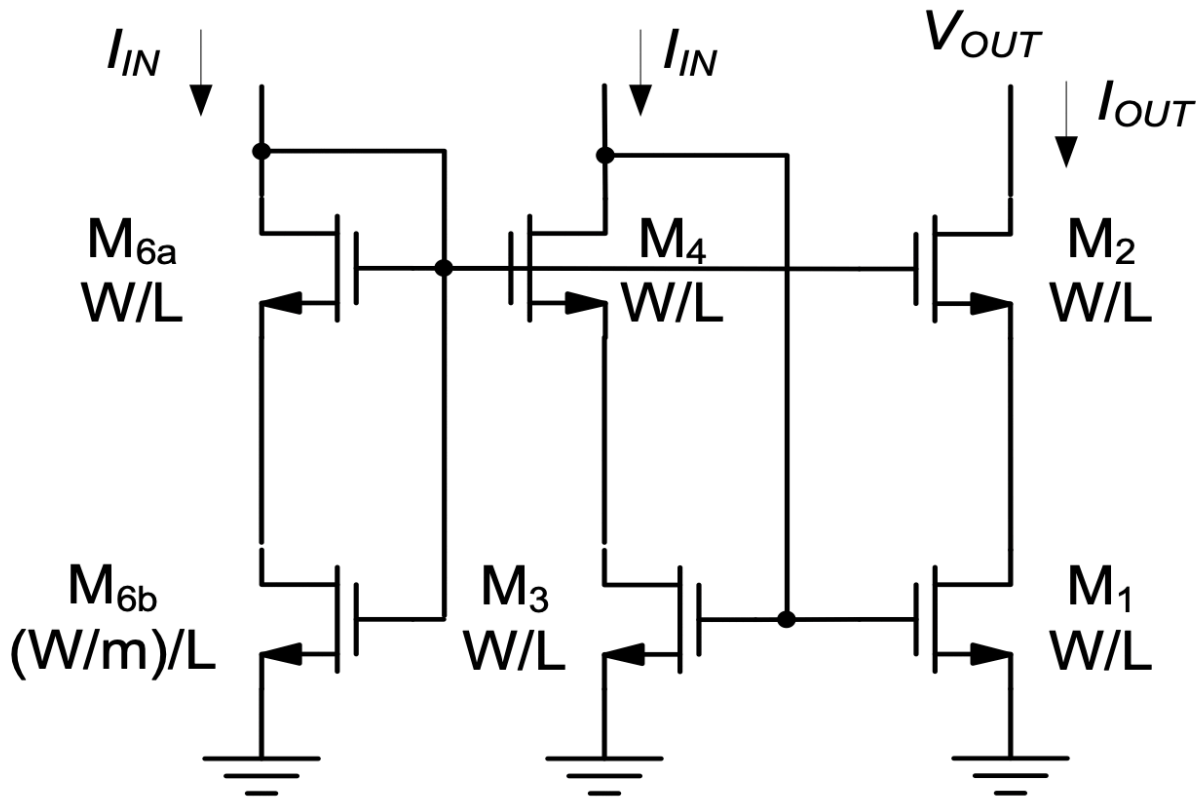


Figure 5.23.: Complete high-swing cascode current mirror using a triode device (M_{6b}) for cascode biasing.

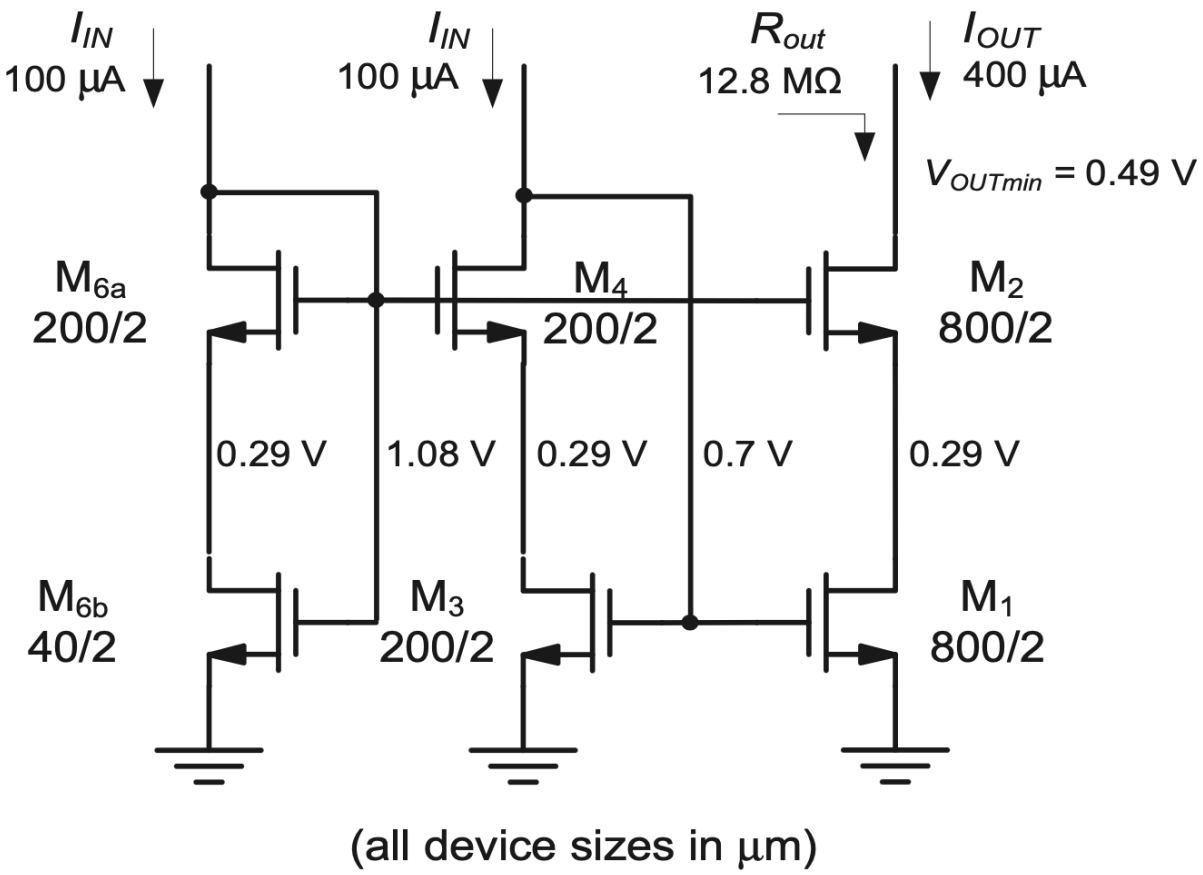


Figure 5.24.: Ex5-4

5.4. Current References

The current mirror circuits discussed in the previous section are useful for replicating and distributing bias currents within a sub-circuit or an entire chip. Ultimately, however, the currents that are being distributed must originate from some form of a reference current generator (see Figure 5.1). Over the years, a wide variety of current references have been developed, each having specific pros and cons for the intended application. Within the scope of this introductory module, we will consider only two examples, primarily as a starting point for further reading and to complete the picture on how a complete biasing network within a larger chip might be constructed. For a more comprehensive discussion, the reader is referred to advanced texts such as (Gray et al. 2009a).

We begin by considering the most basic of all possible reference generator circuits, shown in Figure 5.25. This circuit is essentially a current mirror, with its input branch tied to the supply via a resistor.

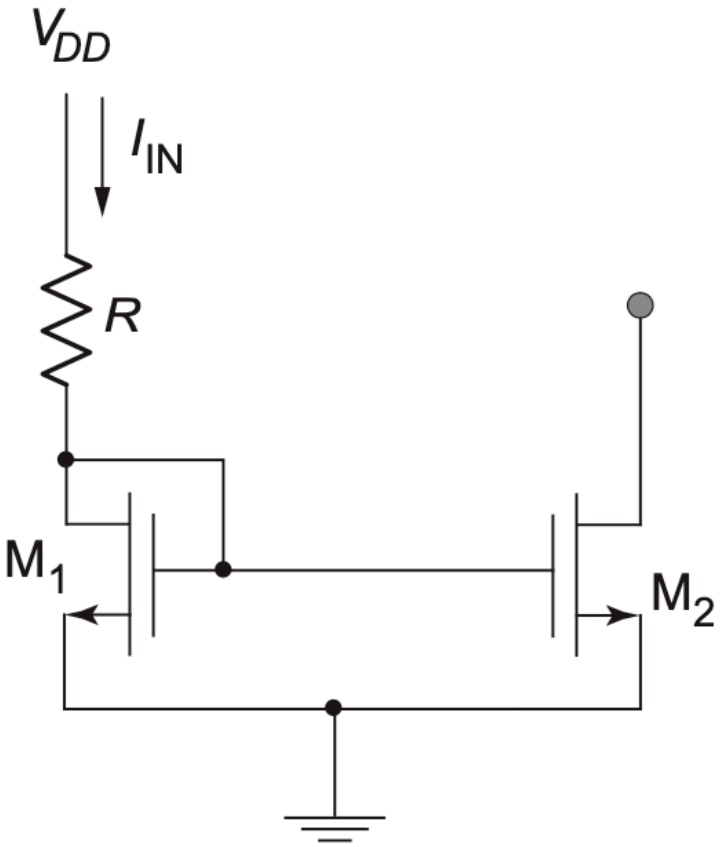


Figure 5.25.: Simple supply-referenced current generator.

In this circuit, we have

$$I_{OUT} \approx I_{IN} = \frac{V_{DD} - V_{Tn} - V_{OV}}{R} \tag{5.28}$$

From this result, noting that typically $V_{DD} \gg V_{Tn}$, and $V_{DD} \gg V_{OV}$, we see that the current is roughly proportional to the supply voltage. Given the variations in supply voltage that a robust

circuit must withstand (see Table 5.1), this solution is often not suitable for all but relatively primitive and low-performance circuits. What we desire is a current generator that is (to first-order) insensitive to supply variations. The so-called self-biased constant-gm current generator discussed next is an example of an improved circuit that is frequently used in practice. To understand the self-biased constant-gm current generator, consider first the circuit shown in Figure 5.26(a).

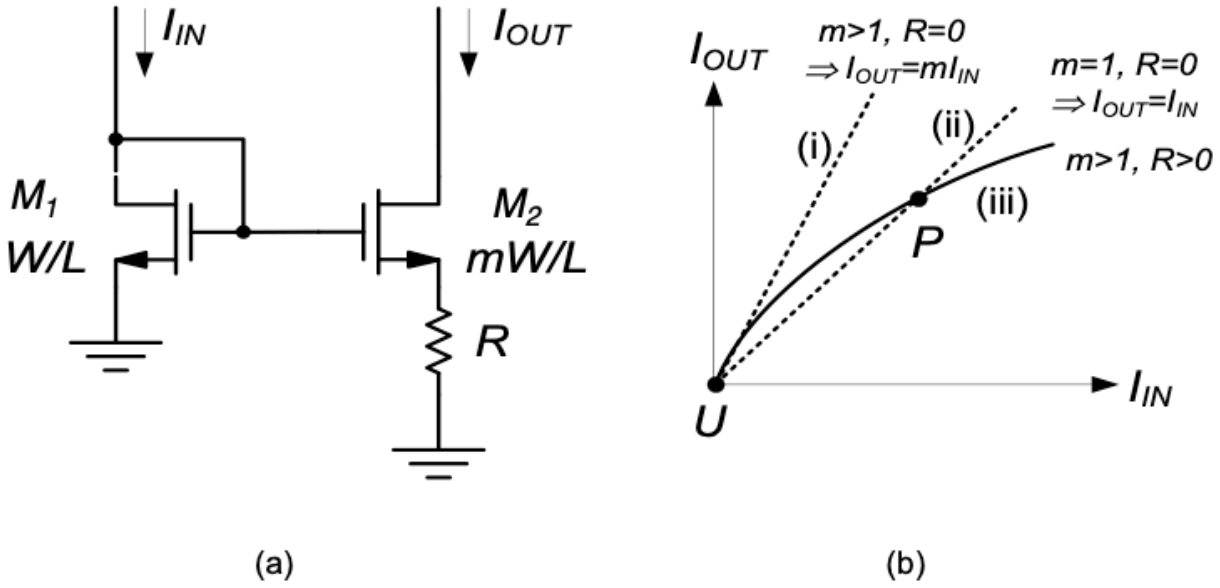


Figure 5.26.: (a) Core building block of a constant-gm current generator. (b) Current transfer characteristics for various scenarios.

This is a current mirror-like circuit with a resistor R added in the source of M_2. Assuming that M_2 is scaled m times wider than M_1, and letting $R = 0$ for the time being, we know that I_{OUT} is approximately equal to $m \cdot I_{IN}$. This is illustrated using the dashed line (i) in the graph of Figure 5.26(b). Line (ii) is included for reference, corresponding to $m = 1$, i.e., $I_{OUT} = I_{IN}$. Now, assuming $m > 1$ and $R > 0$, we know that I_{OUT} must be smaller compared to case (i) with $R = 0$. This is because the voltage drop across R reduces the gate-source voltage of M_2 and consequently results in smaller I_{OUT}. As I_{IN} increases, I_{OUT} [curve (iii)] bends away further and further from line (i) and ultimately intersects with line (ii). While it is possible to derive a closed-form equation of this curve (see problem P5.7), we focus our attention on point P. A particularly interesting property of point P is that it defines an absolute current level that (to first-order) depends only on the MOSFET sizes and R, i.e., it is independent of the supply voltage. In order to build a current reference that utilizes this point, a few extra transistors must be employed, as shown in Figure 5.27. First focus on M_3 and M_4. These transistors form a current mirror that forces $I_{OUT} = I_{IN}$, which is necessary for operation at point P (see Figure 5.26(b)).

Unfortunately, simply forcing $I_{OUT} = I_{IN}$ does not guarantee that the circuit operates at P. There exists another (undesired) point where $I_{OUT} = I_{IN} = 0$, labeled U in Figure 5.26(b). If only M_1–M_4 were present in this circuit, it would not be clear which operating point the circuit will chose when the supply voltage is turned on. The outcome may depend on second-order effects, such as parasitic capacitive coupling, and on how quickly the supply ramps up.

In order to guarantee that the circuit will eventually operate at point P, the designer will always

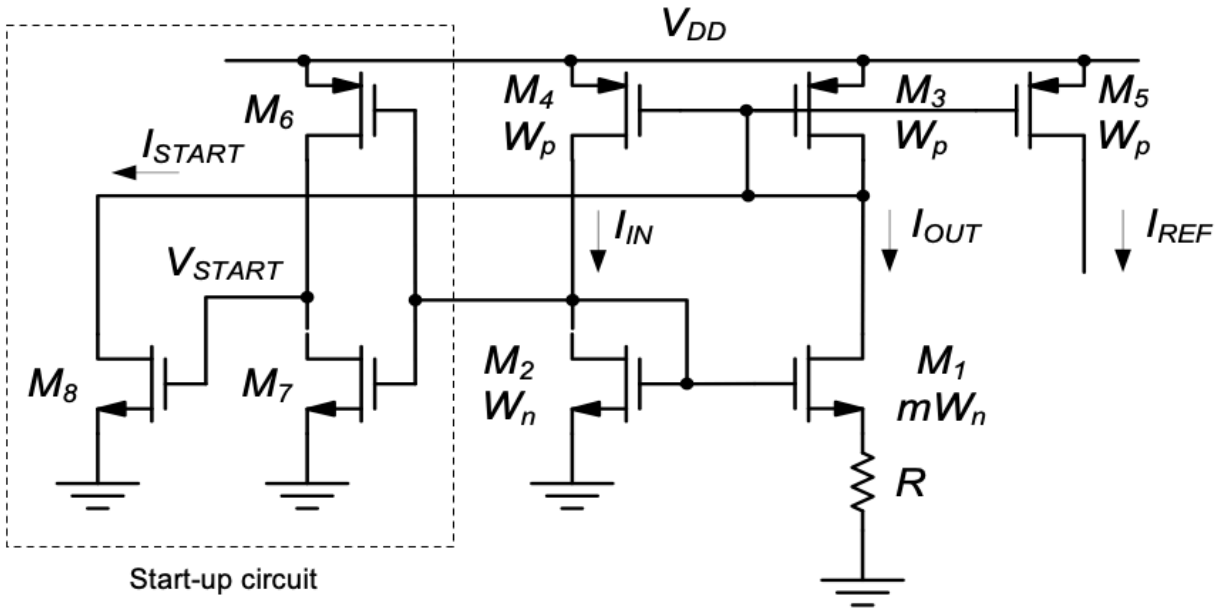


Figure 5.27.: Complete self-biased constant-gm current reference circuit.

include a so-called start-up circuit. This circuit is formed by M_6–M_8 in Figure 5.27.

To understand the operation of the start-up circuit, consider first the case where the circuit starts up in point U, i.e., $I_{OUT} = I_{IN} = 0$. This condition necessitates that $V_{GS2} < V_{Tn}$, since no drain current is flowing in M_2. With $V_{GS2} < V_{Tn}$, M_7 will be off and M_6 will be on, pulling the voltage at node V_{START} toward V_{DD}. Thus, M_8 will turn on and force a drain current into M_3, which will subsequently be mirrored into M_4 and M_2. Therefore, the circuit has no choice but to leave point U and ultimately arrive at P, which is the only other possible DC operating point.

Once point P is reached we have $V_{GS2} > V_{Tn}$ by some overdrive voltage, typically a few hundred millivolts. For the given V_{GS2} in this point, M_7 must be sized large enough so that V_{START} lies near ground, and no current flows in M_8 ($I_{START} = 0$). In a typical design M_7 is much larger than M_6, resulting in a so-called low-threshold inverter.

Our final task is to compute the current $I_{REF} = I_{IN} = I_{OUT}$ in Figure 5.27. We begin by applying KVL around the gate-source voltages of M_1 and M_2.

$$I_{REF}R = V_{GS2} - V_{GS1} \tag{5.29}$$

Neglecting backgate effect, i.e., assuming equal threshold voltages for M_1 and M_2 (for simplicity), Equation 5.29 becomes

$$I_{REF}R = V_{OV2} - V_{OV1} \tag{5.30}$$

Now, since for a MOSFET

$$V_{OV} = \sqrt{\frac{2I_D}{\mu_n C_{ox}\frac{W}{L}}} \tag{5.31}$$

and M_1 is m times wider than M_2, Equation 5.30 can be rewritten as

$$I_{REF} = \frac{V_{OV2}\Big(1 - \frac{1}{\sqrt{m}}\Big)}{R} \tag{5.32}$$

Finally, eliminating V_{OV2} using Equation 5.31 and solving for I_{REF} gives

$$I_{REF} = \frac{2(\sqrt{m}-1)^2}{m} \cdot \frac{1}{\mu_n C_{ox}\frac{W_n}{L_n}R^2} \tag{5.33}$$

This equation is primarily useful for setting the absolute current level in the circuit, and at first glance does not seem to have any special structure. A much more important result from the above analysis follows from considering the transconductance of M_2, given by

$$g_{m2} = \frac{2I_{D2}}{V_{OV2}} = \frac{2I_{REF}}{V_{OV2}} = \frac{2\Big(1 - \frac{1}{\sqrt{m}}\Big)}{R} \tag{5.34}$$

As we see from this result, g_{m2} depends only on the resistance R and the scaling factor m, i.e., the transconductance will not be affected by MOSFET process and temperature variations. In a way, the circuit "recomputes" I_{REF} such that the transconductance is held constant to the value given by Equation 5.34. This is the reason why this circuit is typically called a constant-gm reference generator, as mentioned earlier. It should be noted, of course, that not only the transconductance of M_2 is held constant when this circuit is used. Any other MOSFET that utilizes I_{REF} or a copy of this current will behave similarly.

In practice, the device type used to implement resistor R should be chosen with care. When implemented on-chip, the designer will often opt for a highly doped polysilicon resistor that has relatively small process variations and a small temperature coefficient (see Table 5.2 and Table 5.3). Alternatively, the resistance is sometimes placed off-chip, where it can be realized, for example, with a 1%-accurate and low temperature coefficient metal film resistor.[2]

As a final note, we should emphasize that the foregoing analysis neglected many second-order effects, such as channel-length modulation and back-gate effect. In practice, these effects can have some bearing on the circuit's accuracy and therefore leave room for improvements (such as including cascode transistors). The interested reader will find many articles on this topic in analog circuit literature and advanced texts, such as (Gray et al. 2009a).

[2]When the resistor is placed off-chip, the designer must take great care to avoid stability issues. Using feedback circuit analysis techniques, it can be shown that even relatively small amounts of parasitic capacitance at the source node of M_1 can make the circuit oscillate.

5.5. Voltage Biasing Considerations

In addition to bias currents, building a complete ana- log circuit will require the generation of various bias voltages needed to operate common-source, common-gate and common-drain stages (see for example V_{B1} and V_{B2} in Figure 5.1). In this section we will discuss an exemplary subset of solutions that have found their use in practice.

As we have already seen in Example 5-1, the com- mon-source stage is very sensitive to variations in its input bias voltage. As a result, a majority of practical CS circuits are embedded in feedback networks that regulate the input bias voltage to the proper value, thereby absorbing process variations and mismatch effects. Since feedback is beyond the scope of this module, the input biasing techniques suggested for the common-source stage are meant to be applied only to a subset of applications where the circuit's voltage gain is low (typically < 10), and the amplifier is utilized "open-loop," without a feedback network. This complication typically does not exist for CG and and CD stages, and the proposed circuits are therefore more or less universally applicable.

5.5.1. Voltage Biasing for a Common-Source Stage

Due to the voltage gain of a CS amplifier, its input bias voltage usually cannot be set to a fixed voltage without causing prohibitive sensitivities to component variations and mismatch. Thus, it is important to design the bias circuitry with variability in mind and construct solutions that can track or absorb any significant deviations from nominal parameter conditions. Especially for common-source stages, soltions applied in practice often involve the use of feedback or differential circuit topologies (see Gray et al. 2009a). Since these topics are beyond the scope of this module, we will concentrate here only on a few basic ideas that can be understood with the prerequisites established so far.

Specifically, we will focus in this subsection on a few possible solutions to the problem encountered in Example 5-1. The main problem in this example was that the input bias voltage was held constant, while the threshold voltage and other parameters in the circuit changed due to process variations. Ideally, we would like to "automatically compute" the input bias voltage of the stage such that it tracks the required value across process corners.

A first option that accomplishes this is shown in Figure 5.28. Here, M_1 is the MOSFET that implements the common-source amplifier and v_s and R_s model a transducer that generates the voltage we wish to amplify. The transistor M_2 is a **replica device** that computes the proper gate-source voltage required to carry the current I_B. Note that this overall arrangement resembles a current mirror, which we have already determined to be insensitive to process variations. If V_{Tn} or $\mu_n C_{ox}$ change, the gate-source voltage of M_1 (V_B) adjusts so that this transistor's drain current remains equal to I_B. This means (to first-order) that no current flows into the resistive divider formed by R_1 and R_2. These resistors can be sized to establish the desired output quiescent point and voltage gain. For example, for $R_1 = R_2$ and $V_{DD} = 5V$, we have $V_{OUT} = 2.5V$, approximately independent of process and temperature.

While the above-discussed circuit will work robustly, it has one big limitation in that both transducer terminals must be accessible and compatible with the bias voltage desired for M_1. One possibility for overcoming this constraint is to employ AC coupling (see Figure 5.29). AC coupling means that the transducer signal is coupled into the circuit via a capacitor. In the circuit of Figure 5.29

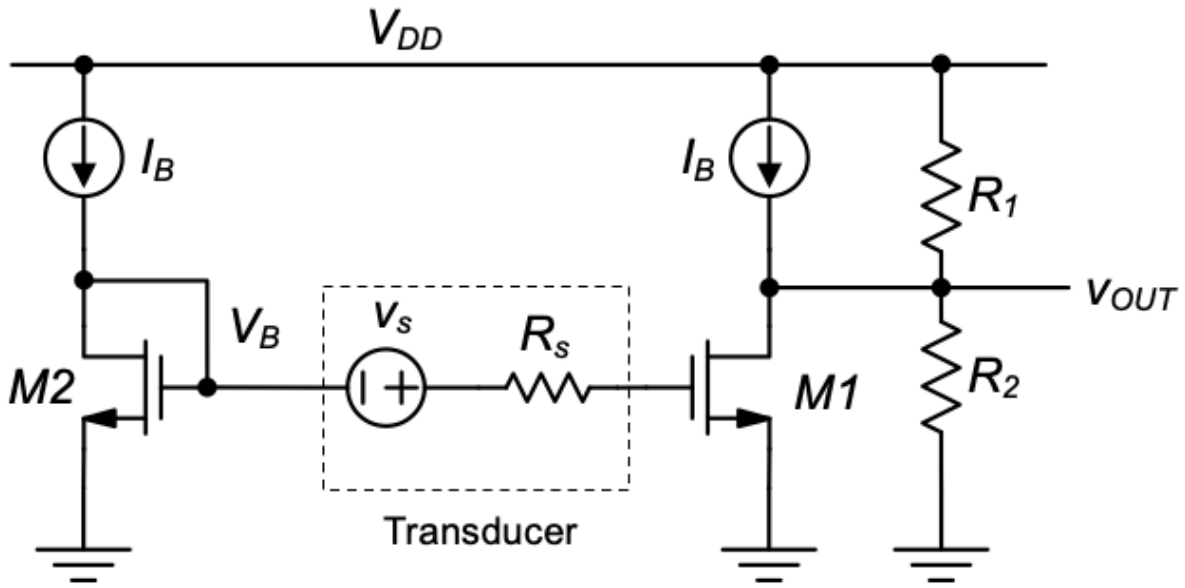


Figure 5.28.: Replica biasing approach for setting up the quiescent point gate-source voltage of a common-source stage.

R_{large} and C_{large} form a first-order high pass filter with corner frequency $\omega_c = 1/R_{large}C_{large}$ (neglecting the resistance $1/g_{m2}$, which is in series with R_{large}).

To avoid filtering the signal, ω_c must be chosen smaller than the smallest frequency of interest. For instance, if we are interested in amplifying a 20 Hz signal (the lower end of the audio frequency spectrum), we need $R_{large}C_{large} > 1/(2\pi \cdot 20Hz) \cong 4ms$. Assuming we can comfortably integrate resistances up to $100k\Omega$ on our chip, this means that $C_{large} > 4nF$. Such a large capacitance is typically impractical for integration on chip and would have to be realized as an external component.

A shortcoming of the circuit in Figure 5.29 is that the resistors R_{large} and R_s form a voltage divider, which can be detrimental when R_s is very large. Figure 5.30 shows an alternate approach in which the transducer can be directly connected to the MOSFET gate. In this circuit, the bias current I_B is injected into the drain of the common-source transistor (M_1) and extracted again using the current mirror formed by M_2 and M_3. The bias point voltage at node X is given by V_B–V_{GS1}, which places constraints on the minimum required value for V_B. Note however, that V_B does not have to be accurately set or track process variations; as long as M_1 and M_2 are in saturation, node X tracks (DC) changes in V_B and the circuit remains properly biased. As far as the signal is concerned, the capacitor C_{large} establishes an AC ground at the source of M_1 beyond the high-pass corner frequency of the circuit. Just as in the previous circuit, it can be shown that the AC coupling capacitor must take on large values to enable the passing of low frequencies through the circuit.

As we have seen from the previous examples, achieving proper biasing together with the processing of low-frequency signals in a basic common-source stage comes with some undesired constraints and restrictions. Many of these issues can be mitigated when the signal is present in the form of a current, originating for example from a common-gate stage that is driving the common-source amplifier. We will see an example of such a circuit in Chapter 6.

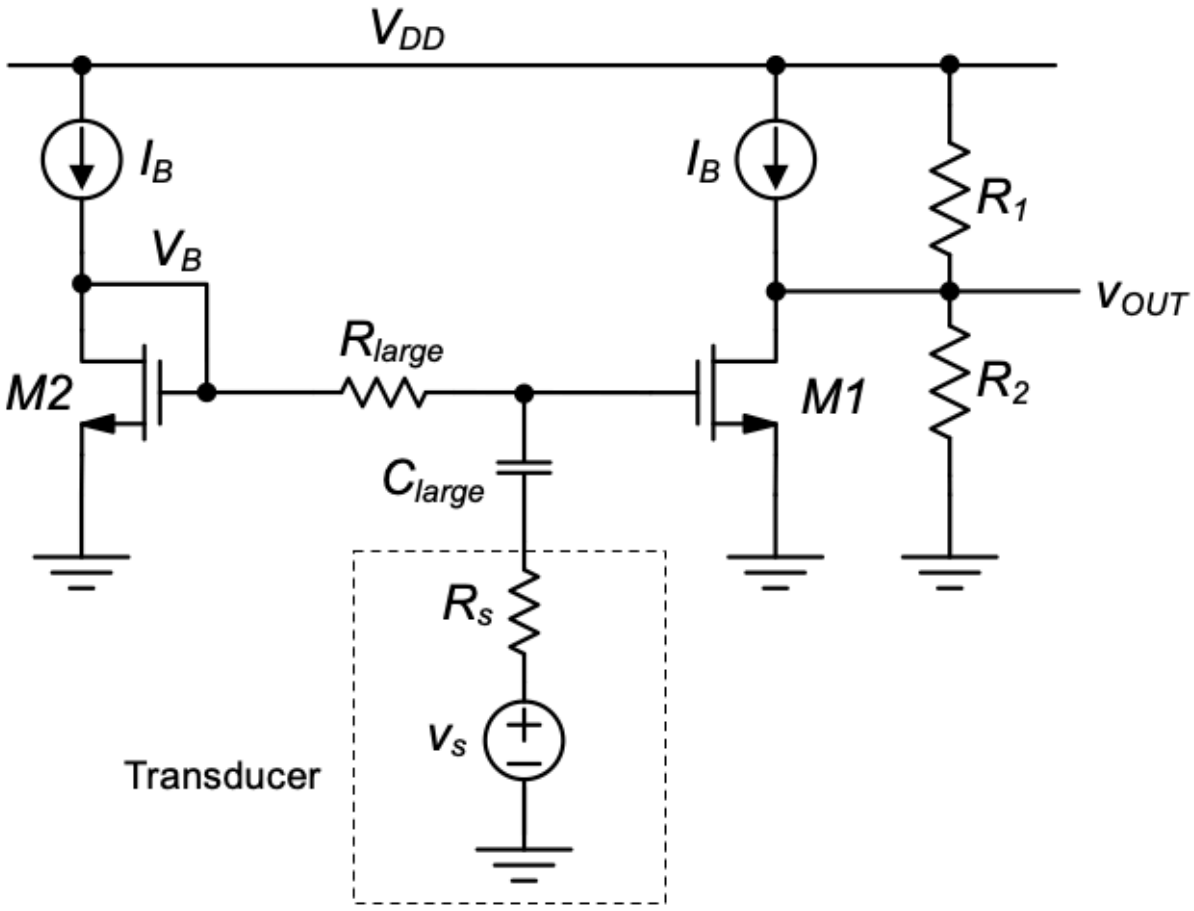


Figure 5.29.: Replica biasing approach using AC coupling.

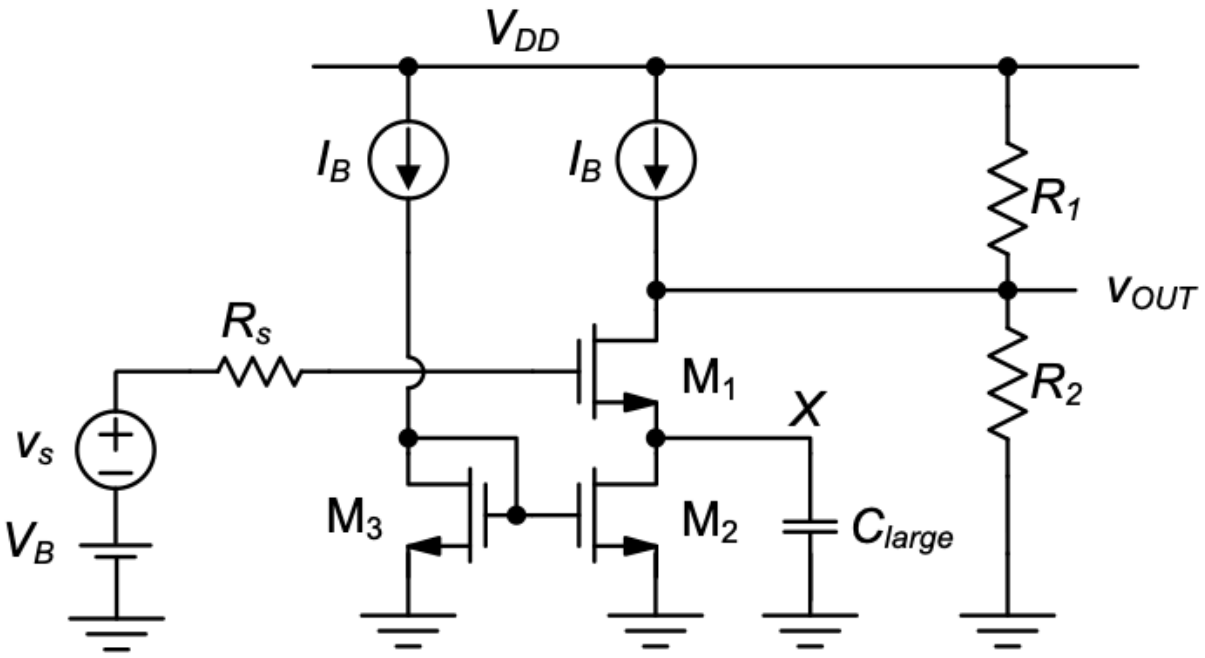


Figure 5.30.: Biasing-approach using AC coupling at the source of a common-source stage.

5.5.2. Voltage Biasing for a Common-Gate Stage

Compared to a common-source stage, setting up the bias voltage for the gate of a common-gate stage is usually less intricate. To see this, we consider two classical usage examples shown in Figure 5.31. In Figure 5.31(a) the common-gate device M_2 is utilized in a cascode stage. Since a cascode stage is often designed for large voltage gain, a typical objective is to maximize the available output voltage swing. Consequently, the gate bias voltage of M_2 is set up in the same way as in the high-swing cascode current mirror discussed in Section 5.3.4, which means the drain-source voltage of M_1 is set to V_{DSsat1} plus some margin for robustness and tolerance to mismatches. A reasonable margin is achieved by using $m = 5$ (see Table 5.6) in the sizing of M_{3b}.

Figure 5.31(b) shows an example where a common-gate stage is used to interface to a photo diode. The signal current generated in the photodiode passes through M_2 and causes a proportional voltage swing at the output. In this circuit, the output swing is usually not very large, and thus the gate bias voltage for M_2 is not tightly constrained by voltage swing requirements. Typically, the gate voltage is set such that the photo diode is biased at a suitable reverse bias. This is accomplished by sizing R_1 and R_2 appropriately.

In both of the circuits in Figure 5.31, variations in the transistor parameters (such as V_{Tn}) will cause the overall operating point of the circuits to shift. How- ever, unlike the common-source stage of Example 5-1, these circuits are not very sensitive to such shifts. For instance, if the threshold voltage of M_2 in Figure 5.31(b) changes by 100 mV, all this means is that the reverse bias voltage of the diode will change by approximately the same amount. If properly designed (with margins), this won't cause the circuit to fail or behave improperly. This strongly contrasts the situation with the circuit of Example 5-1, where such changes in the threshold voltage can have detrimental effects on the stage's operation.

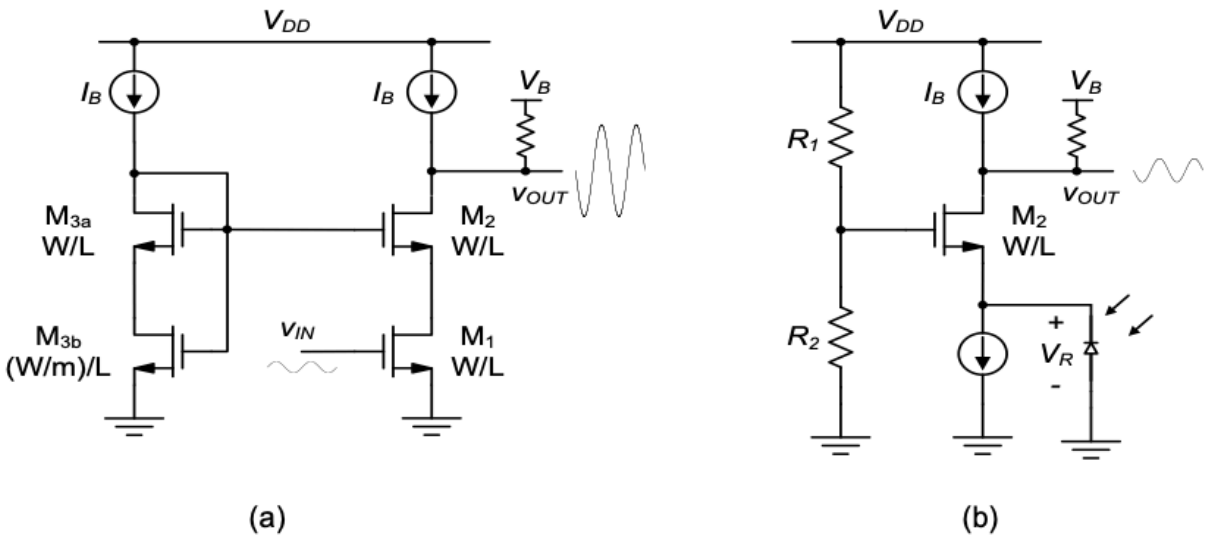


Figure 5.31.: Voltage biasing in two usage cases of a common-gate stage.

5.5.3. Voltage Biasing for a Common-Drain Stage

In a common-drain stage, the input and output voltages at the quiescent point are directly coupled. As shown in Figure 5.32, $V_{OUT} = V_{IN}–(V_{Tn} + V_{OV})$. Proper voltage biasing in a common-drain stage boils down to making sure that the input and output quiescent point voltages are compatible with the circuits that are connecting to the stage input and output. As in a common-gate stage, variability in transistor parameters often does not have detrimental effects as long as a proper margin is included in the design.

In some applications, the shift between the input and output quiescent point is undesired. In this case, a p-channel common-drain stage can be used to provide a shift in the opposite direction (see Figure 5.33). In this circuit, M_1 can be sized such that the quiescent points V_{IN} and V_{OUT} are approximately equal.

When a common-drain stage is employed primarily to shift quiescent points, the designer calls this circuit a level shifter. Level shifters are generally useful to interface two stages that are otherwise incompatible in terms of their ideal quiescent point out- put/input voltages.

5.6. Summary

In this chapter, we have surveyed general considerations and basic circuits related to the voltage and current biasing of elementary transistor stages. We have seen that the variability inherent to CMOS process technology influences the design and architecture of these support circuits and ultimately determines whether a certain biasing scheme can be deemed practical. We analyzed the basic current mirror and its cascoded variant with respect to their nonidealities and articulated some of the most important design guidelines. As an example of a reference current generator, we looked at the so-called constant-gm biasing circuit and analyzed its first-order behavior. Finally,

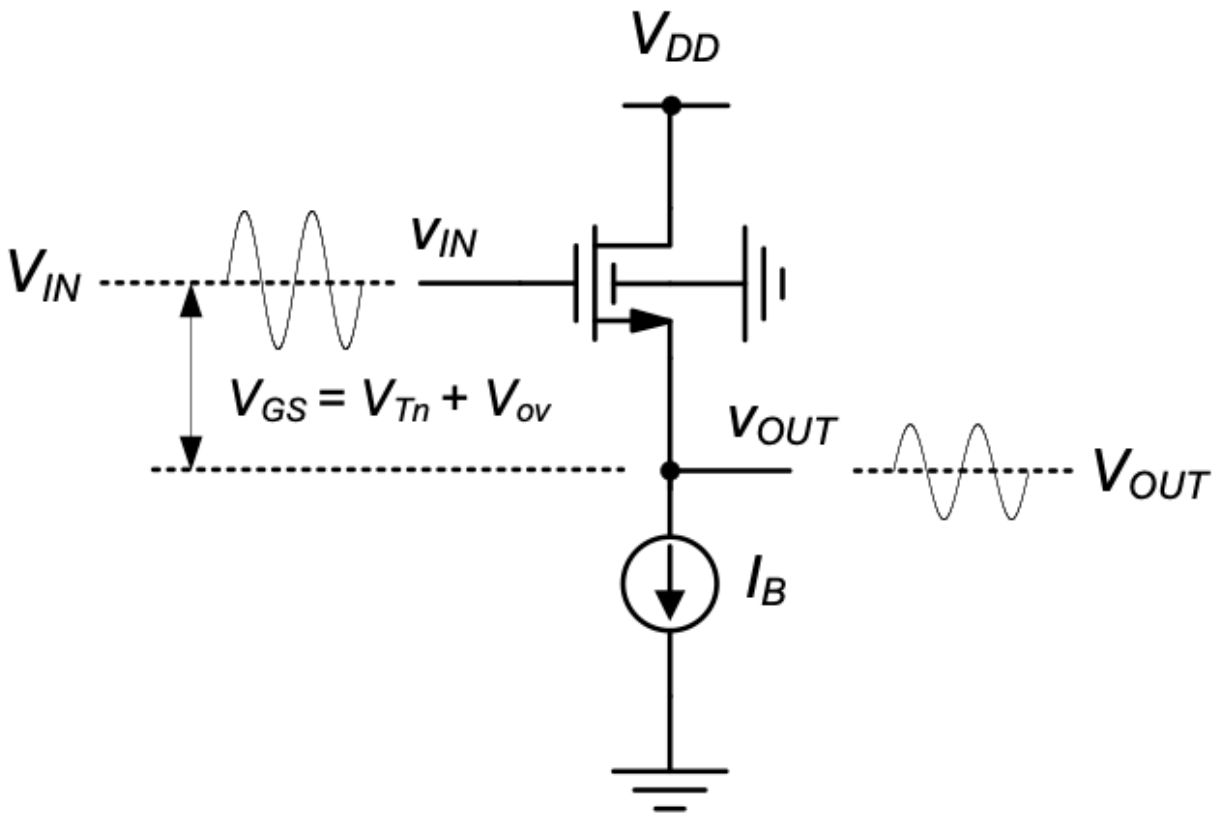


Figure 5.32.: Relationship between the input and output quiescent points in a common-drain stage.

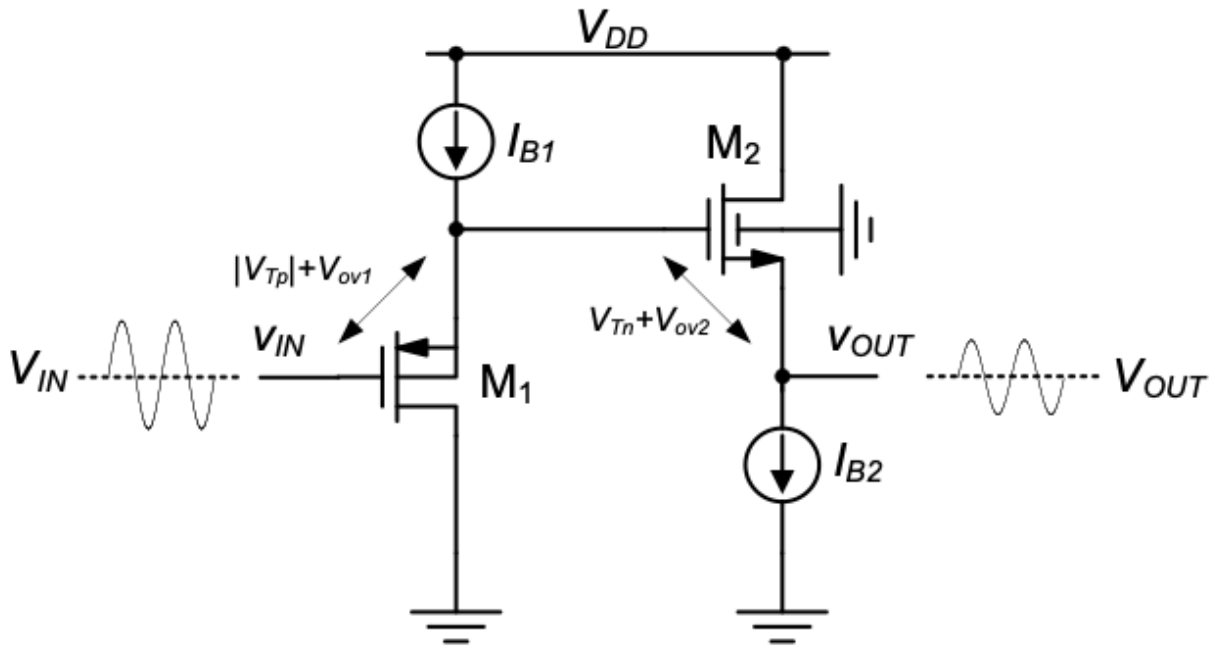


Figure 5.33.: Back-to-back common-drain stages to realize equal input and output quiescent point voltages.

this chapter looked into the problem of voltage biasing for the three elementary stage configurations. We determined that in lieu of feedback, biasing a common-source stage properly is most challenging and must be considered with care and knowledge of relevant process variation and mismatch effects. While most of the presented ideas and circuits were presented in the context of simple application examples, they generally also apply to more complex circuit designs studied in advanced literature.

5.7. Problems

Unless otherwise stated, use the standard model parameters specified in Table 4-1 for the problems given below. Consider only first-order MOSFET behavior and include channel-length modulation (as well as any other second-order effects) only where explicitly stated.

P5.1 Consider the bias current generator circuit of Figure 5.25. Parameters: $R = 4k\Omega, W/L = 20$

(a) Compute the current I_{IN} assuming nominal MOSFET parameters and supply voltage ($V_{DD} = 5V$).
(b) Recompute I_{IN} for slow MOSFET parameters (see Table 5.2) and $V_{DD} = 4.5V$. Repeat for fast parameters and $V_{DD} = 5.5V$.
(c) What are the percent errors of the currents found in part (b), relative to the nominal current computed in (a)?

P5.2 In Example 5-1, we showed analytically that changing the MOSFET parameters from nominal to fast pushes the transistor into the triode region. Construct a load line plot that shows this

graphically. That is, draw the output curves of the MOSFET (I_D versus V_{DS}) for the two corner cases and show how the intersect with the load line shifts when fast parameters are assumed. Be sure to neglect channel length modulation.

P5.3 Set up a suitable analysis that allows you derive the values given in Table 5.5. Set up an equation that computes V_{DS1} as a function of k and V_{OV}.

P5.4 Set up a suitable analysis that allows you derive the values given in Table 5.6. Set up an equation that computes V_{DS1} as a function of m and V_{OV}.

P5.5 The circuit in Figure 5.34 can be used to achieve high-swing cascode biasing without an extra input current branch. Given the annotated bias point voltages, what is the proper W/L ratio for M_4 that achieves the minimum output compliance voltage? Express the desired $(W/L)_4$ as a multiple k of W/L.

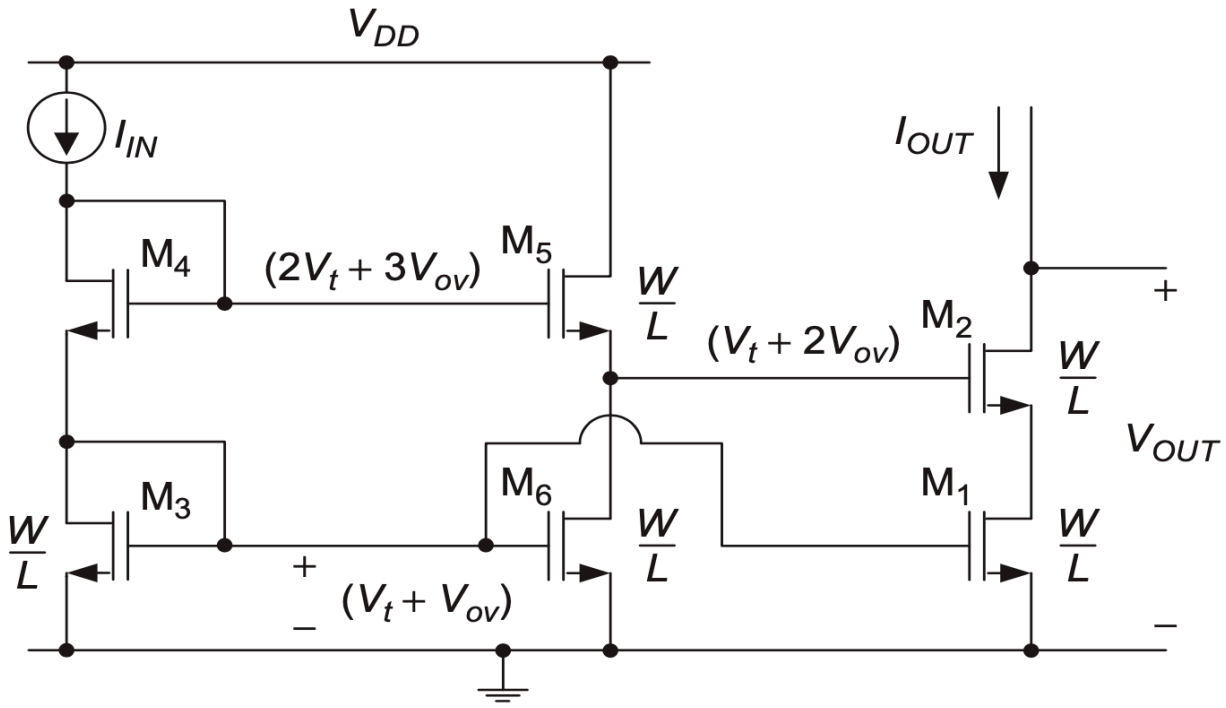


Figure 5.34.: P5-5

P5.6 The circuit in Figure 5.35 is called a "Sooch" cascode current mirror. It uses one single branch for setting up all bias voltages for a high-swing cascode current mirror. Given the annotated bias point voltages, what is the proper W/L ratio for M_5 that achieves the minimum output compliance voltage? What is the minimum required voltage across the input branch (V_{IN})?

P5.7 Derive a closed-form expression for curve (iii) in Figure 5.26(b) Verify graphically that the intersect with line (ii) corresponds to the current level given in Equation 5.33. Assume the following parameters: $m = 4$, $W/L = 25$, $R = 2k\Omega$. Be sure to neglect channel-length modulation and back-gate effect.

P5.8 For the circuit of Example 5-1, compute the proper V_{IN} that would need to be applied in the fast parameter case such that the output bias voltage remains the same as in the nominal case.

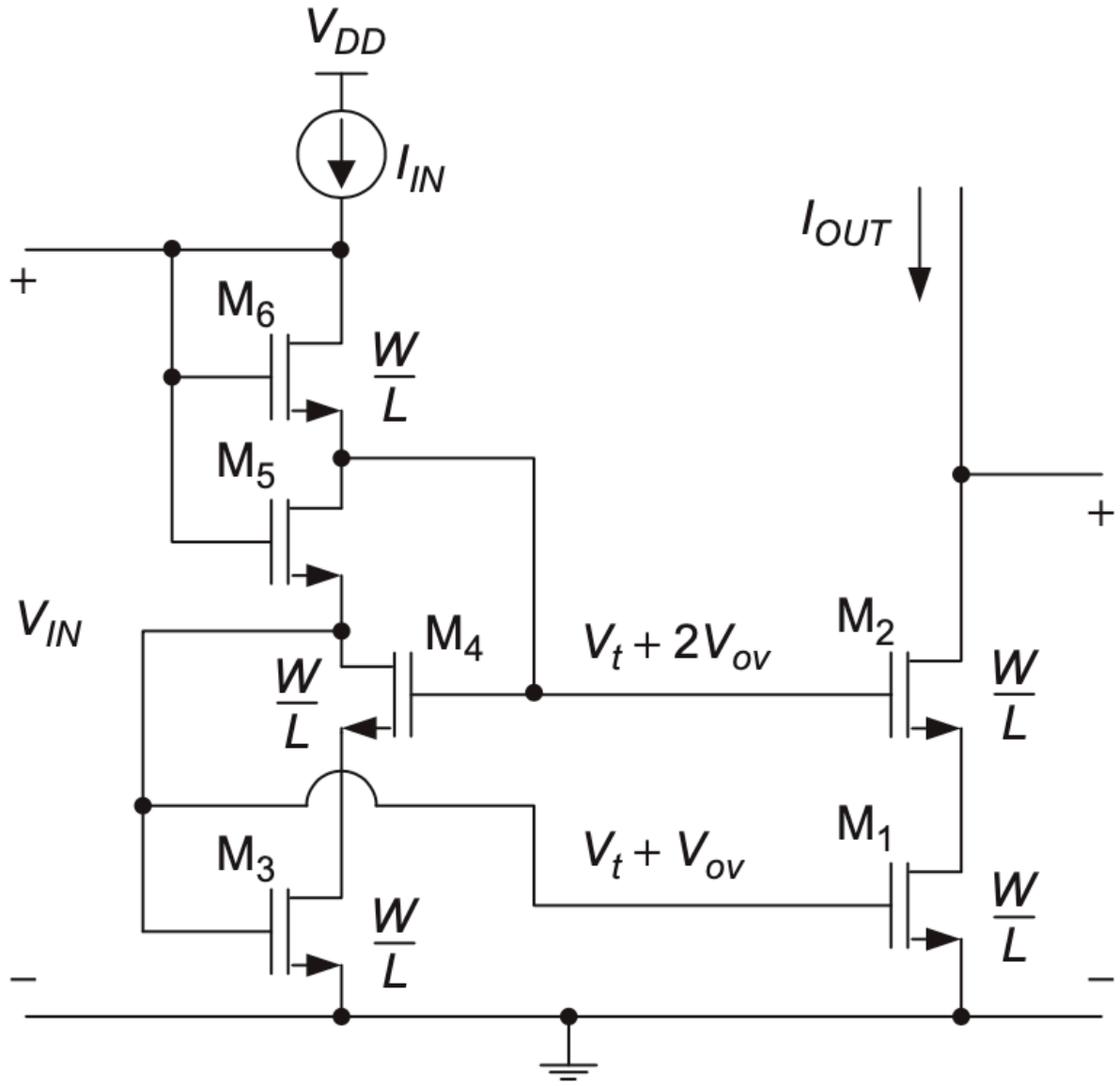


Figure 5.35.: P5-6

Assume nominal conditions for supply temperature and R_D. In this biasing condition, what is the voltage gain, and by which percentage has it changed relative to the nominal case?

P5.9 Figure 5.36 shows a cascode current source consisting of M_{1A} and M_{1B}, and a single transistor current source consisting of M_2. Assume that the cascode current source is optimally biased, i.e., V_{B1B} is chosen such that $V_{DS1a} = V_{DS1a,sat} = V_{OV1a}$. /backAssume also that both current sources supply the same current I_O. Neglect backgate effect.

(a) Find relationships between W_1, L_1 and W_2, L_2 such that both current sources have the same parasitic output capacitance, and the same output compliance voltage V_{Omin} that keeps all the devices saturated. For simplicity, assume $\lambda = 0$ in this part of the analysis. Note: The parasitic capacitance at the drain of M_2 is given by $C_{db} + C_{gd}$. Similarly, assume that the output capacitance of the cascode current source is approximately equal to $C_{db} + C_{gd}$ of M_{1b}. (In the cascode current source, the effect of other capacitances referred to the output node is negligible.)
(b) Using the result from part (a), show that the expression given below must hold. R_{O1} and R_{O2} are the output resistances of each current source, as indicated in Figure 5.36.

$$\frac{R_{O1}}{R_{O2}} = \frac{g_{m1} r_{o1}}{4}$$

(c) Calculate V_{Omin}, R_{O1} and R_{O2} for $I_O = 100\mu A$ and $(W/L)_1 = 10\mu m/2\mu m$. [Use the relationships between device sizes from part (a).]

5.10 The circuit shown in Figure 5.37 is a so-called self-biased, V_{Tn}-referenced current generator. Assuming $(W/L)_1 = 50$ and $(W/L)_3 = (W/L)_4 = (W/L)_5$, find the value for R so that $I_{OUT} = 100\mu A$. Assume $\lambda = 0$ and neglect backgate effect.

P5.11 For the circuit shown in Figure 5.38, ignore the backgate effect and finite output resistance unless otherwise stated. All devices have identical widths and lengths and operate in saturation (W/L = 50) and $I_{ref} = 200\mu A$.

(a) Calculate R such that the drain-source voltage of M_1 is 1.5 times its gate overdrive, i.e., $V_{DS1} = 1.5 V_{OV1}$.
(b) Suppose that due to random mismatch, the threshold voltage of M_1 is 10 mV larger than that of all the other transistors. What is the percent error in I_{OUT} caused by this mismatch? Use appropriate small-signal approximations in your calculation.
(c) Suppose now that the threshold voltage of M_2 is increased by 10 mV while all other thresholds are at their nominal value. What is the percent error in I_{OUT} caused by this mismatch? In this calculation, include the effect of the finite output resistance for M_1, assuming $g_m r_o = 50$.

P5.12 In the circuit of Figure 5.39, V_{G1} is adjusted such that $I_{D1} = 50\mu A$. The W/L ratio of M_1 and M_2 is equal to 4, while M_3 is sized such that $W/L = 1/2$, and $V_{DD} = 5V$.

(a) Ignoring backgate effect, compute the minimum and maximum values of V_{G2} for which all transistors remain in saturation.
(b) Repeat part (a) with backgate effect included.

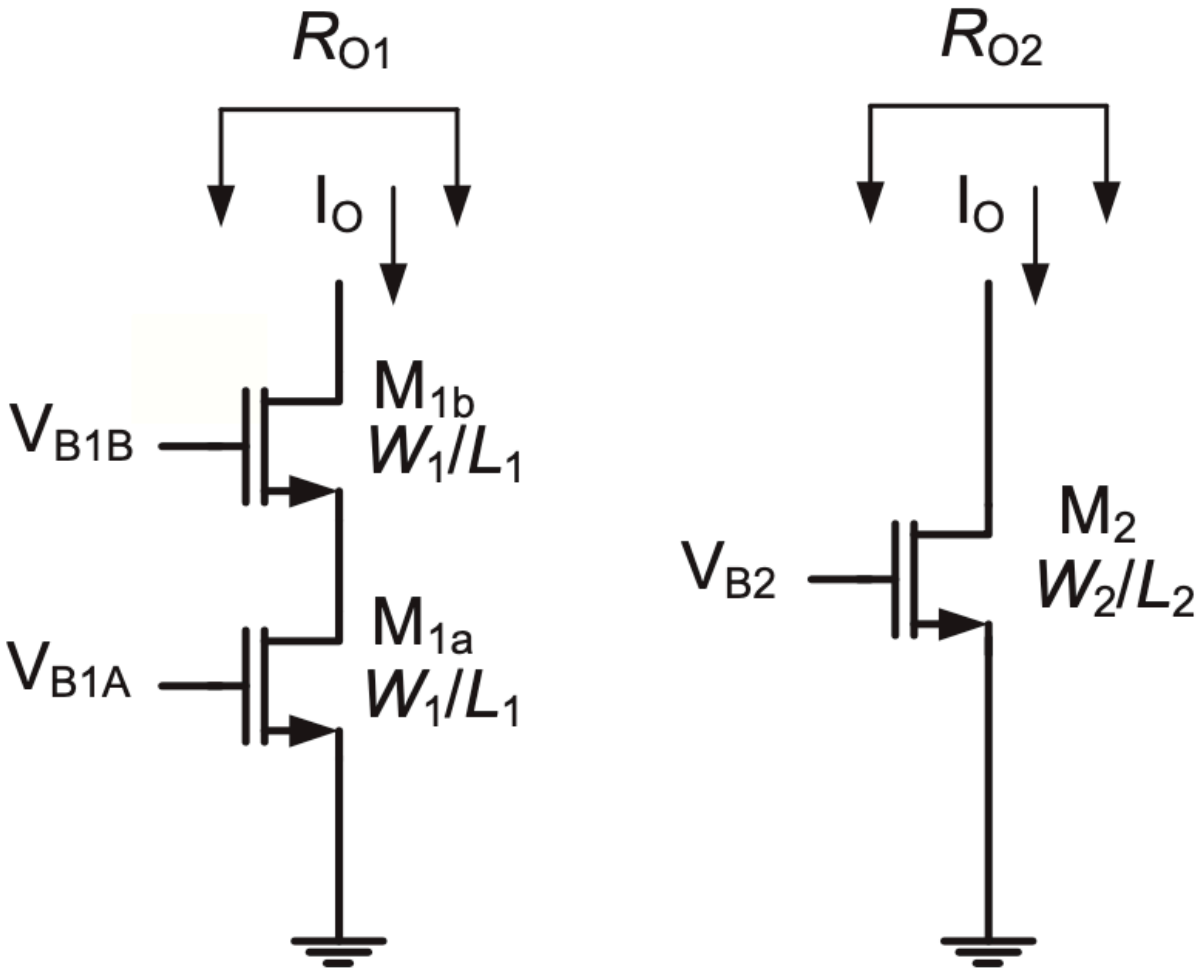


Figure 5.36.: P5-9

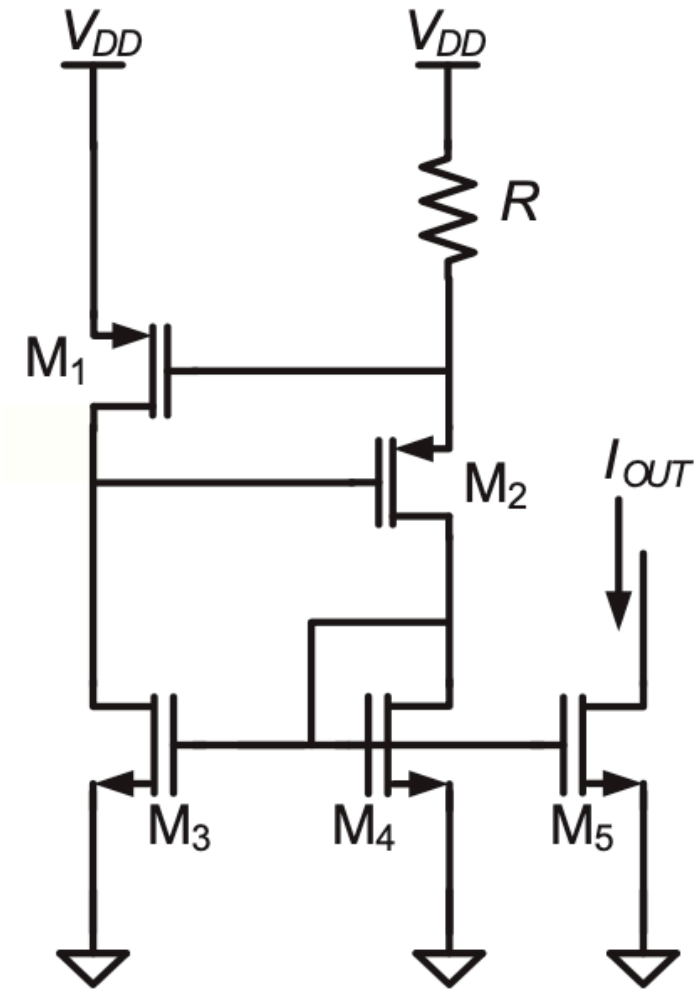


Figure 5.37.: P5-10

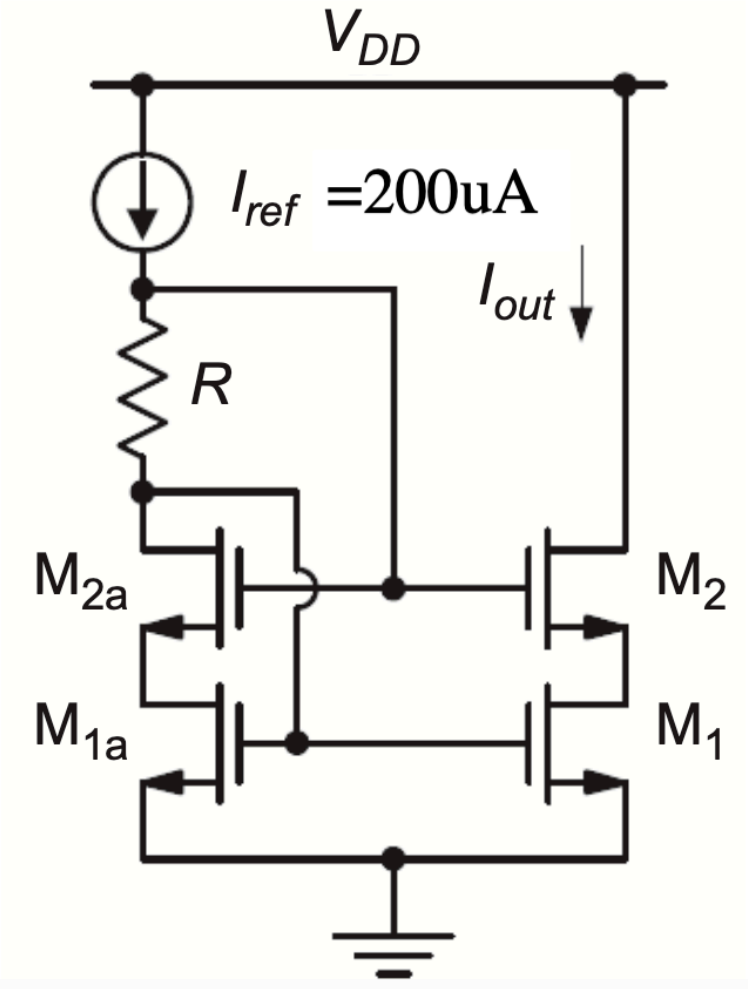


Figure 5.38.: P5-11

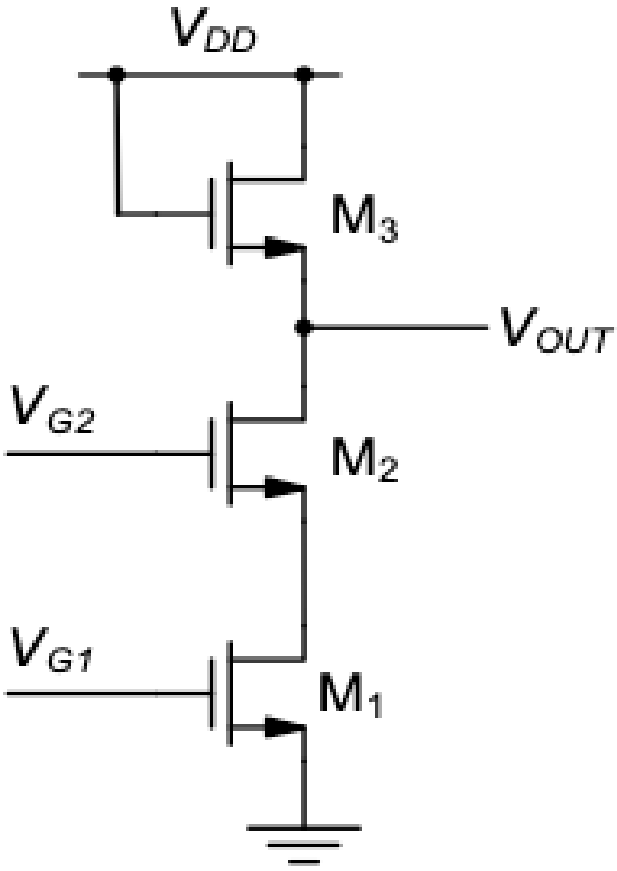


Figure 5.39.: P5-12

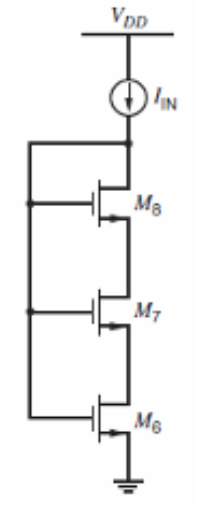


Figure 5.40.: P5-13

P5.13 For the circuit of Figure 5.40, compute the width ratios W_8/W_6 and W_7/W_6 such that $V_{DS8} = V_{DS7} = V_{OV6}$. Assume that all channel lengths are identical. Ignore the backgate effect and channel-length modulation.

P5.14 For the circuit of Figure 5.27, size the transistors $M_1 - M_4$ and R such that $I_{OUT} = 100\mu A$, $V_{OV} = 300mV$. Assume that all channel lengths are equal to 3 m. Neglect backgate effect and channel length modulation.

P5.15 Design a cascode current mirror circuit using n-channel devices assuming the following specifications. The circuit should take an input current of 10 A and generate three outputs at 20 A, 50 A, and 100 A, respectively. The output compliance voltage should be no larger than 800 mV and the gate overdrive of the transistors should be designed as large as possible (for immunity to mismatch), while maintaining a reasonable saturation margin. Draw the complete circuit diagram, including all device sizes. This problem does not have a unique solution.

P5.16 Draw a layout (using any tool you prefer) for the circuit designed in Example 5-3.

Part II.

Learn to Walk — Real Transistors, Noise, Mismatch, and Distortion

6. Introduction

TBD.

7. Introduction to Feedback

7.1. Preliminaries

In electronics the term **feedback** refers to the situation where a signal **sensed** (or derived or sampled or measured) from the output port of an amplifier is **returned** to the input port, where it is **combined** with the externally applied input signal to create a new signal to be processed by the amplifier itself.

- when the returned signal is added to the external signal, we have positive feedback (a.k.a. regenerative feddback)
- when the returned signal is subtracted we have negative feedback (a.k.a. degenerative feedback)

Negative feedback was conceived in 1928 by Harold Black.

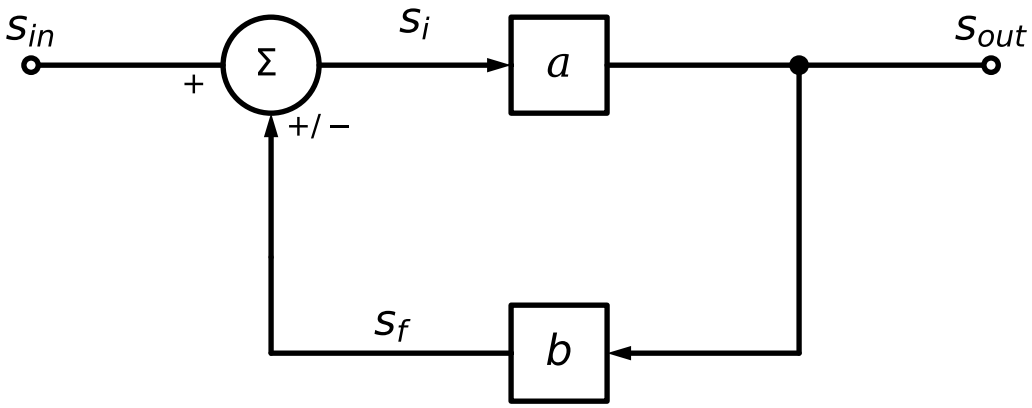


Figure 7.1.: Block diagram of a feedback system.

7.2. Systematic Feedback Analysis Frameworks

- Two-port Analysis
  - Proposed by Harold Black
  - Helpful for gaining basic intuition about feedback
  - Map the feedback network onto one of four topologies
    - voltage-voltage (serie-shunt), current-voltage (shunt-shunt), voltage-current (series-series), current-current (shunt-series), depending on the desired input/output quantities

  - In practice, the feedback network "b" (and possibly the forward amplifier "a") are not unilateral, so mapping arbitrary circuits into a generic "ab" block diagram that uses two-ports is not obvious
  - The feedback network causes loading at the input and output of the basic amplifier
    * model the feedback network as an ideal two-port and absorb the loading effects into the forward amplifier. This procedure is often quite tedious.

  - Some feedback circuits cannot be modeled using two-ports
    * e.g. bias circuits with feedback loops tend to have only one port

- Return Ratio Analysis
  - Proposed by Hendrik Bode
  - "Asymptotic" method. It does no attempt to break the circuit into pieces
  - The closed-loop properties of the feedback circuit are described in terms of the *return ratio* of a dependent source in an active device
  - The block diagram used to analyze the closed loop properties of the feedback circuit is extended to handle the feedforward through the feedback network
  - Often easier than two-port analysis

7.3. Negative Feedback

- Basic Idea
  - trade off gain for other desirable properties
- Desirable properties
  - desentisize gain
  - extend the bandwidth
  - control input and output impedance
  - reduce NL distorsion
- Drawbacks (undesired properties)
  - reduced gain
  - under certain circumstances the negative feedback in an amplifier can become positive and of such a magnitude as to cause oscillations (instability)

NOTE: it should not be implied that positive feedback always lead to instability

7.4. General negative feedback structure

Basic Terminology:

s_{in} = input signal applied to the feedback system

s_{out} = output signal from the feedback system

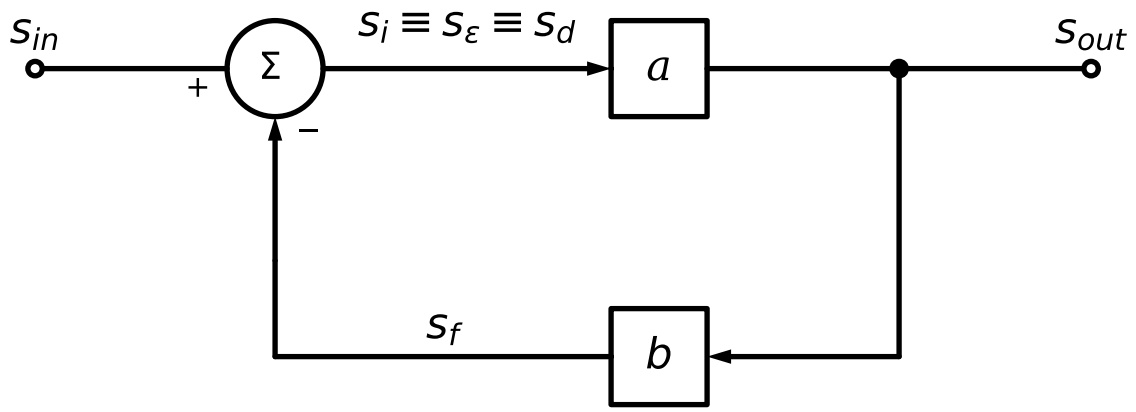


Figure 7.2.: Block diagram of a negative feedback system.

$a = \frac{s_{out}}{s_i}$ = open-loop gain = gain of "basic" amplifier = gain of feedforward amplifier

b = feedback factor

$a \cdot b = L$ = loop gain (it is always unit-less)

$1 + ab = 1 + L$ = amount of feedback

$s_i \equiv s_\epsilon \equiv s_d = s_{in} - s_f$ = input to "basic" amplifier = error = difference

$A = \frac{s_{out}}{s_{in}}$ = closed-loop gain = gain of feedback system

Ideal feedback assumptions

- the feedback network does not load the basic amplifier output
- the feedback network does not load the basic amplifier input
- the feedback network is unilateral
- the basic amplifier is unilateral

Closed-loop gain expression

The feedback network measures (or samples or senses) the output signal s_{out} and provides (returns) a feedback signal s_f (that is related to s_{out} by the **feedback factor** b)

$$\begin{aligned} s_{out} &= a \cdot s_i \\ s_f &= b \cdot s_{out} \\ s_\epsilon \equiv s_d &\equiv s_i = s_{in} - s_f \end{aligned}$$

$$A = \frac{s_{out}}{s_{in}} = \frac{s_{out}}{s_i + s_f} = \frac{1}{\frac{s_i}{s_{out}} + \frac{s_f}{s_{out}}} = \frac{1}{\frac{1}{a} + b} = \frac{a}{1 + ab} =$$
$$= \frac{1}{b} \cdot \frac{1}{1 + 1/(ab)} = \frac{1}{b} \cdot \frac{1}{1 + \frac{1}{L}} = \frac{1}{b} \cdot \frac{L}{1 + L} \approx \frac{1}{b}$$

The most important benefit coming from the use of negative feedback is under the condition $L \gg 1$:

- for $L \gg 1$ the actual gain $A(\approx 1/b)$ is virtually independent of the open-loop gain a. The open-loop gain a is typically an extremely inaccurate parameter, and varies significantly subject to drift with temperature, supply voltage, fabrication process, and DC biasing conditions.

Although physically unattainable the limit $L \to \infty$ represent the ideal condition:

$$A_{ideal} = \lim_{L \to \infty} A = \frac{1}{b}$$

Key result: when the loop gain L is large, the closed loop gain A approaches the ideal closed loop gain A_{ideal}, which is equal to $1/b$

To achieve gain we need $b \leq 1$.

$b \leq 1$ is easy to implement. We can for example use a wire ($b = 1$), or a resistive divider (ratiometric) or a capacitive divider (ratiometric).

Example 1 - voltage buffer

$v_f = v_{out} \Rightarrow b = v_f/v_{out} = 1$

$v_{out} = a \cdot (v_{in} - v_{out})$

$v_{out}(1 + a) = a \cdot v_{in}$

$\frac{v_{out}}{v_{in}} = \frac{a}{1+a} \approx 1$

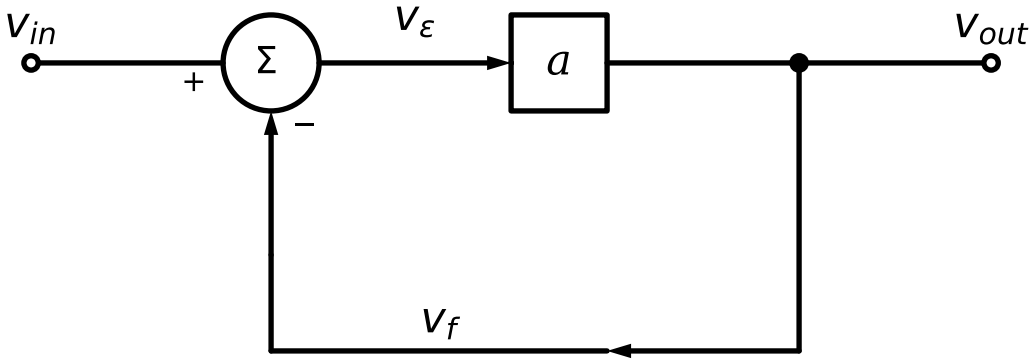


Figure 7.3.: Block diagram of a voltage buffer.

Example 2 - non-inverting voltage amplifier

$b = \frac{v_f}{v_{out}} = \frac{R_1}{R_1 + R_2}$

$\frac{v_{out}}{v_{in}} \approx \frac{1}{b} = 1 + \frac{R_2}{R_1}$

If we regard $1/L$ as an error term, then L gives a measure of how close the actual gain A is to the ideal gain $A_{ideal} = 1/b$.

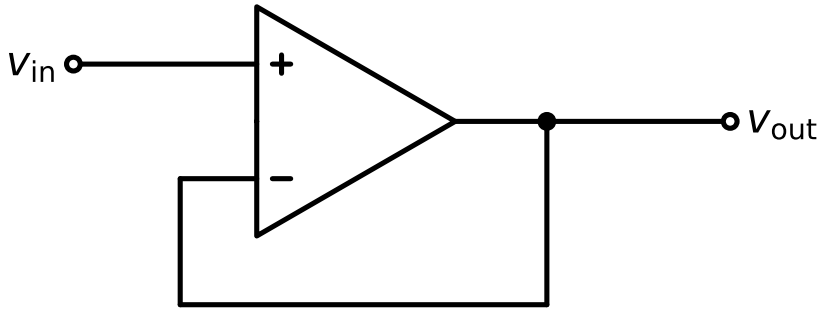


Figure 7.4.: op amp based voltage buffer

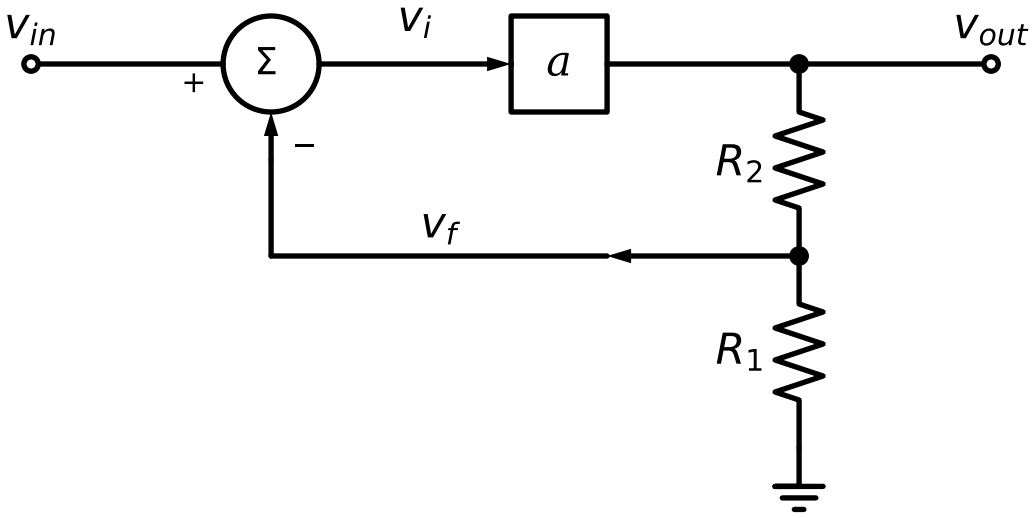


Figure 7.5.: Block diagram of a non inverting voltage amplifier (series-shunt feedback)

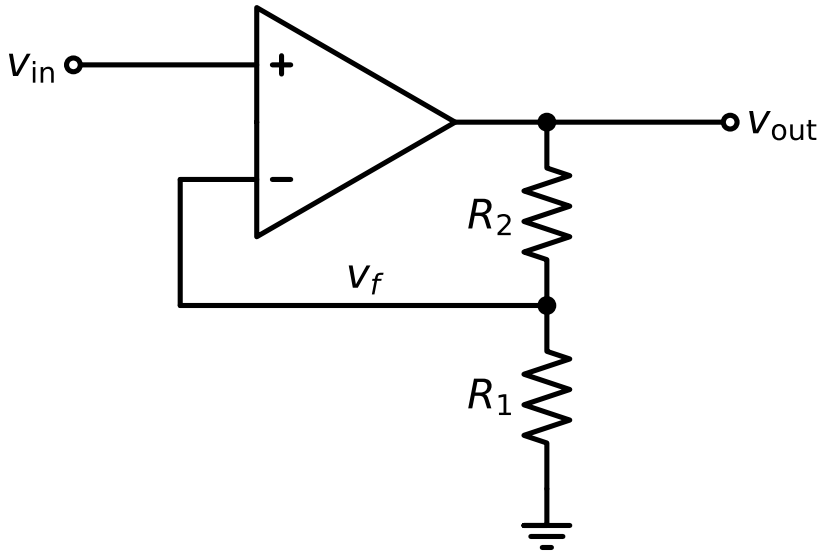


Figure 7.6.: op amp based non inverting voltage amplifier

$$A = \frac{1}{b} \cdot \frac{1}{1 + \frac{1}{L}} \approx \frac{1}{b} \cdot \left(1 - \frac{1}{L}\right)$$

The % gain error between actual gain A and ideal gain A_{ideal} is usually defined as follows:

$$gainerror\% = 100 \times \frac{A - A_{ideal}}{A_{ideal}} = 100 \times \frac{A_{ideal} \cdot \left(\frac{1}{1 + 1/L}\right) - A_{ideal}}{A_{ideal}} =$$
$$= 100 \times \frac{-1}{1 + L} \approx 100 \times \frac{-1}{L}$$

Example 3 - op amp based non inverting voltage amplifier

KLC at node v_{out}:

$$\frac{v_N - v_{out}}{R_2} = \frac{v_{out} - a_D \cdot (v_{in} - v_N)}{r_o}$$

KCL at node v_N:

$$\frac{v_N - v_{in}}{r_i} + \frac{v_N}{R_1} + \frac{v_N - v_{out}}{R_2} = 0$$

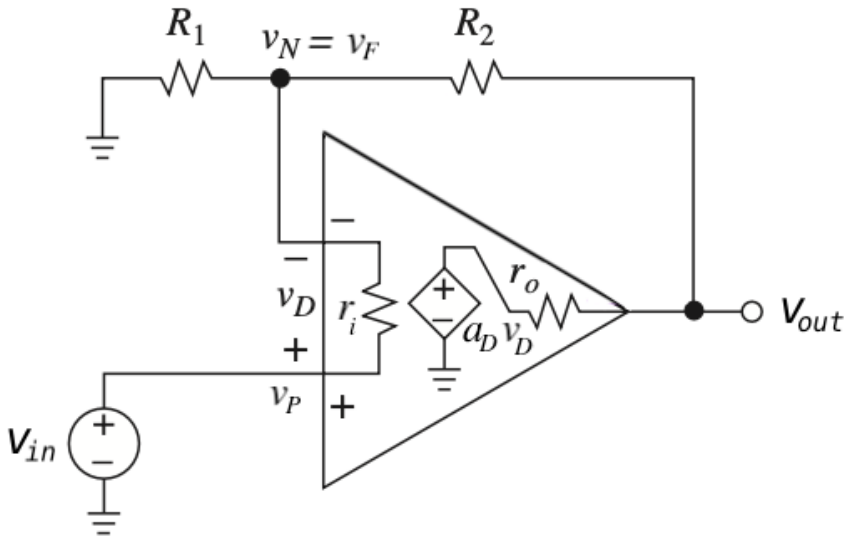


Figure 7.7.: Non inverting voltage amplifier with non ideal op amp (Franco 2015)

For $a_D = 10^6$, $r_i = 10M\Omega$, $r_o = 10\Omega$, $R_1 = 20k\Omega$ and $R_2 = 80k\Omega$:

$$A_{ideal} = \frac{R_1 + R_2}{R_1}$$

$$A = \frac{v_{out}}{v_{in}} = \frac{a_D \cdot r_i(R_1 + R_2) + R_1 r_o}{a_D \cdot R_1 r_i + r_o(R_1 + r_i) + R_1 R_2 + r_i(R_1 + R_2)} = 4.999975$$

$$gainerror = \frac{4.999975 - 5}{5} = -5 \cdot 10^{-6} = -5ppm$$

The exact analysis is very tedious. A better way to obtain the same insight is to use the simpler approach shown in example 4.

7.5. The error signal s_ϵ and the feedback signal s_f

From Figure 7.2

$$s_\epsilon = \frac{s_{out}}{a} = \frac{A \cdot s_{in}}{a} = \frac{\not{a}}{1 + L} \cdot \frac{s_{in}}{\not{a}}$$
$$= \frac{s_{in}}{1 + L}$$

also

$$s_f = b \cdot s_{out} = b \cdot A \cdot s_{in} = \not{b} \cdot \frac{1}{\not{b}} \frac{1}{1 + 1/L} \cdot s_{in}$$
$$= \frac{s_{in}}{1 + 1/L}$$

These results show that for a large loop gain (ideally for $L \to \infty$), the error signal becomes very small (ideally $s_\epsilon \to 0$), causing the feedback signal s_f to closely follow the input signal s_{in} ($s_f \to s_{in}$).

If the feedback signal closely tracks the input signal, then the output signal is also a close "replica" of the input signal, and therefore feedback is also effective in reducing distorsion due to small signals. Low frequency distorsion is caused by changes in the slope of the basic-amplifier transfer characteristic (i.e. changes in the small signal gain of the basic-amplifier).

7.6. Benefits of negative feedback

7.6.1. Gain Desensitivity

To investigate the effect that a variation on the open-loop gain a causes on the closed-loop gain A we differentiate A w.r.t. a

$$A = \frac{a}{1 + ab}$$

$$\frac{dA}{da} = \frac{1 + ab - ab}{(1 + ab)^2} = \frac{1}{(1 + ab)^2}$$

$$\Delta A = \frac{\Delta a}{(1 + ab)^2}$$

$$\frac{\Delta A}{A} = \frac{\Delta a}{A} \cdot \frac{1}{(1 + ab)^2} = \frac{\Delta a}{\frac{a}{1 + ab}} \cdot \frac{1}{(1 + ab)^2} = \frac{\Delta a / a}{(1 + ab)}$$

Thanks to feedback, a fractional change in the gain of the basic amplifier is reduced by a factor $1 + L = 1 + ab$ in the closed-loop. Conceptually, the loop gain L can be be made arbitrarily large. The loop gain is a key parameter and as we will see later it plays an important role also in bandwidth and impedance calculations.

7.6.2. Effect of feedback on non-linearity

Amplifiers are made up of transistors. Since, transistors are non-linear devices, the transfer charateristic of any practical amplifier is non-linear. Assume a basic amplifier that besides the desired linear relationship, also exhibits a quadratic and cubic relationship between its input and output.

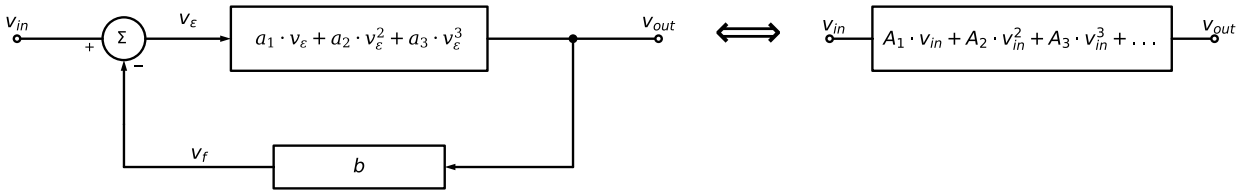


Figure 7.8.: Effect of negative feedback on non-linearity

$$v_{out} = a_1(v_{in} - bv_{out}) + a_2(v_{in} - bv_{out})^2 + a_3(v_{in} - bv_{out})^3 \tag{7.1}$$

$$v_{out} = A_1 v_{in} + A_2 v_{in}^2 + A_3 v_{in}^3 + ... \tag{7.2}$$

Substituting Equation 7.2 into Equation 7.1 we get:

$$\begin{aligned} v_{out} =& a_1(v_{in} - bA_1 v_{in} - bA_2 v_{in}^2 - bA_3 v_{in}^3 - ...) + \\ & a_2(v_{in} - bA_1 v_{in} - bA_2 v_{in}^2 - bA_3 v_{in}^3 - ...)^2 + \\ & a_3(v_{in} - bA_1 v_{in} - bA_2 v_{in}^2 - bA_3 v_{in}^3 - ...)^3 \end{aligned}$$

and equating back into Equation 7.2:

for the linear term

$$\begin{aligned} & A_1 v_{in} = a_1 v_{in} - a_1 b A_1 v_{in} \Leftrightarrow A_1(1 + a_1 b) = a_1 \\ & A_1 = \frac{a_1}{1 + a_1 b} \end{aligned} \tag{7.3}$$

or:

$$\begin{aligned} &A_1 = a_1 - a_1 b A_1 \Leftrightarrow A_1 = a_1(1 - bA_1) \Leftrightarrow \\ &1 - bA_1 = \frac{A_1}{a_1} = \frac{1}{1 + a_1 b} \end{aligned} \tag{7.4}$$

for the quadratic term

$$\begin{aligned} &A_2 v_{in}^2 = -a_1 b A_2 v_{in}^2 + a_2 \left(v_{in}^2 + b^2 A_1^2 v_{in}^2 - 2bA_1 v_{in}^2\right) \Leftrightarrow \\ &A_2(1 + a_1 b) = a_2(1 - bA_1)^2 \Leftrightarrow A_2 = a_2 \frac{(1 - bA_1)^2}{1 + a_1 b} \Leftrightarrow \\ &A_2 = \frac{a_2}{(1 + a_1 b)^3} \end{aligned} \tag{7.5}$$

for the cubic term

$$\begin{aligned} &A_3 \cdot v_{in}^3 = -a_1 b A_3 v_{in}^3 + a_3(1 - bA_1)^3 v_{in}^3 - 2a_2 b A_2(1 - bA_1) v_{in}^3 \Leftrightarrow \\ &A_3(1 + a_1 b) = a_3(1 - bA_1)^3 - 2a_2 b A_2(1 - bA_1) \Leftrightarrow \\ &A_3 = \frac{a_3(1 - bA_1)^3 - 2a_2 b A_2(1 - bA_1)}{(1 + a_1 b)} \Leftrightarrow \\ &A_3 = \frac{a_3}{(1 + a_1 b)^4} - \frac{2a_2^2 b}{(1 + a_1 b)^5} \end{aligned} \tag{7.6}$$

The linear term a_1 (as expected) is reduced by $1 + L$, where $L \equiv a_1 b$

The quadratic term a_2 is reduced by $(1 + L)^3$

The cubic term a_3 is reduced by $(1 + L)^4$, but ther is also an extra term due to the interaction with the second-order term a_2

If at different signal levels, the basic amplifier's small-signal gain varies, since feedback reduces the overall gain variations w.r.t. the gain variations of the basic amplifier, then feedback also reduces distorsion.

Note:

- feedback reduces distorsion without reducing the output voltage range. The overall gain is reduced, but additional gain can be provided with a preamplifier that operates with smaller signal swings, and therefore less distorsion.
- feedback does not help with saturation. In the case an amplifier's output shows hard saturation (i.e., the output become independent of the input), the incremental gain of the basic amplifier $\to 0$, so the amount of feedback also $\to 0$, and as a result negative feedback cannot improve the situation.

7.6.3. Effect of feedback on bandwidth

As an example, assume the transfer function of the basic amplifier is the following:

$$a(j\omega) = \frac{a_0}{1 + \frac{j\omega}{\omega_p}}$$

Then, the closed-loop transfer function is

$$A(j\omega) = \frac{a(j\omega)}{1 + a(j\omega) \cdot b} = \frac{a_0}{1 + a_0 b} \cdot \frac{1}{1 + \frac{j\omega}{\omega_p} \cdot \frac{1}{(1 + a_0 b)}}$$

The gain is reduced by $1 + L_0$, but the bandwidth is increased by $1 + L_0$, where $L_0 = a_0 b$.

The product of gain and bandwidth remains constant.

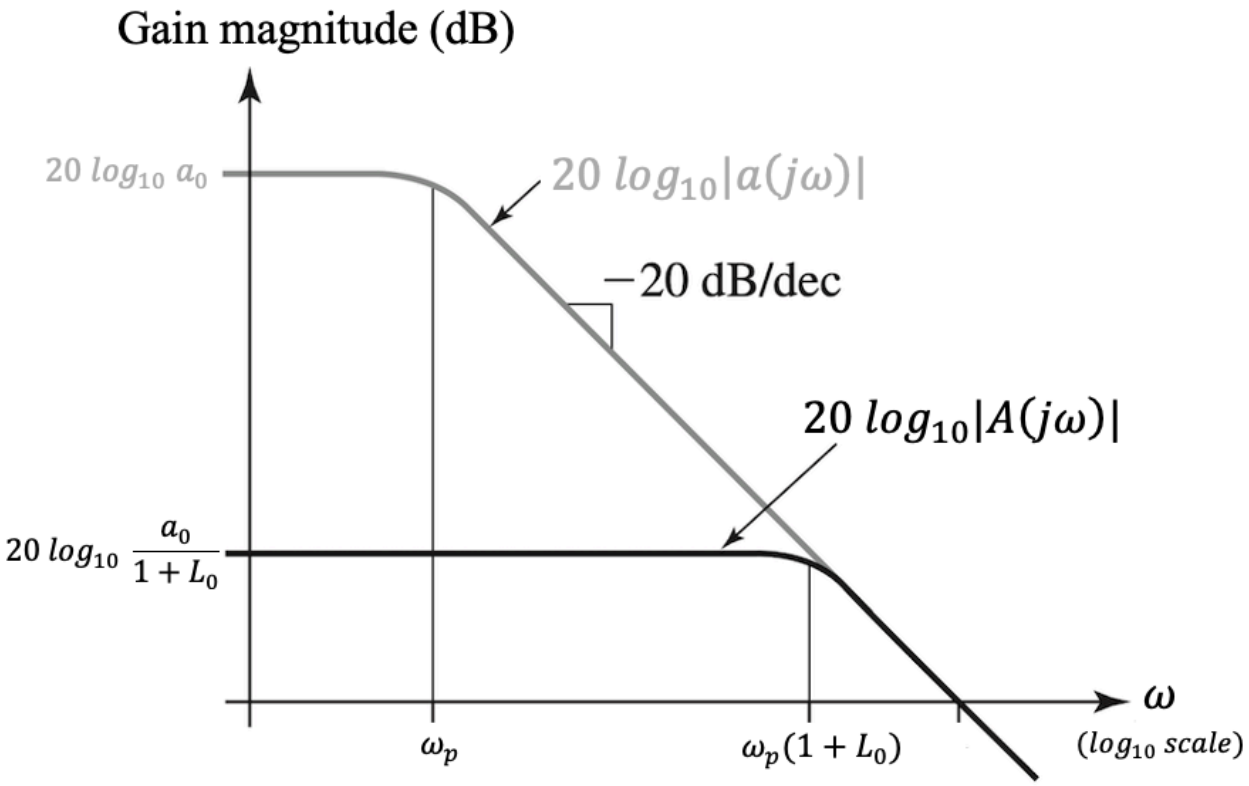


Figure 7.9.: Gain magnitude vs. frequency for the basic amplifier and the feedback amplifier

7.6.4. Effect of feedback on input/output resistances

In actual applications the input and output resistances (a.k.a. terminal resistances) of the amplifier play a key role. When an amplifier is driven by a non ideal source and drives an output load, the input resistance forms a divider with the source's resistance, and the output resistance forms a

divider with the load, therefore reducing the overall gain from source to load. This reduction is usually referred as loading effect. Negative feedback modifies the terminal resistances in ways that tend to reduce the impact of loading.

7.6.4.1. Series-shunt (voltage-voltage) feedback configuration - Voltage Amplifier

Let's consider the application of negative feedback around a voltage amplifier. In this case feedback is performed sampling the output voltage and then feeding back a voltage that is a scaled version of the output voltage "into" the external input voltage source driving the amplifier. Note that the operation of voltage sensing at the output is performed in parallel, or shunt (when we measure a voltage we place the voltmeter in parallel, never in series), while to combine the external input voltage source and the voltage returned by the feedback network we connect them in series (to connect two voltage sources we place them in series, never in parallel).

To focus on the effect of negative feedback on r_i and r_o of the basic amplifier, we assume there is no loading effect on both ports of the basic amplifier. To neglect loading at the input's of the basic amplifier, we assume that the sources v_{in} and bv_{out} have *zero* series resistances. To neglect loading at the output port of the basic amplifier, we assume the output port of the basic amplifier is left open and the input port of the feedback network has *infinite* resistance.

Note: for the series-shunt feeback configuration, b is unitless.

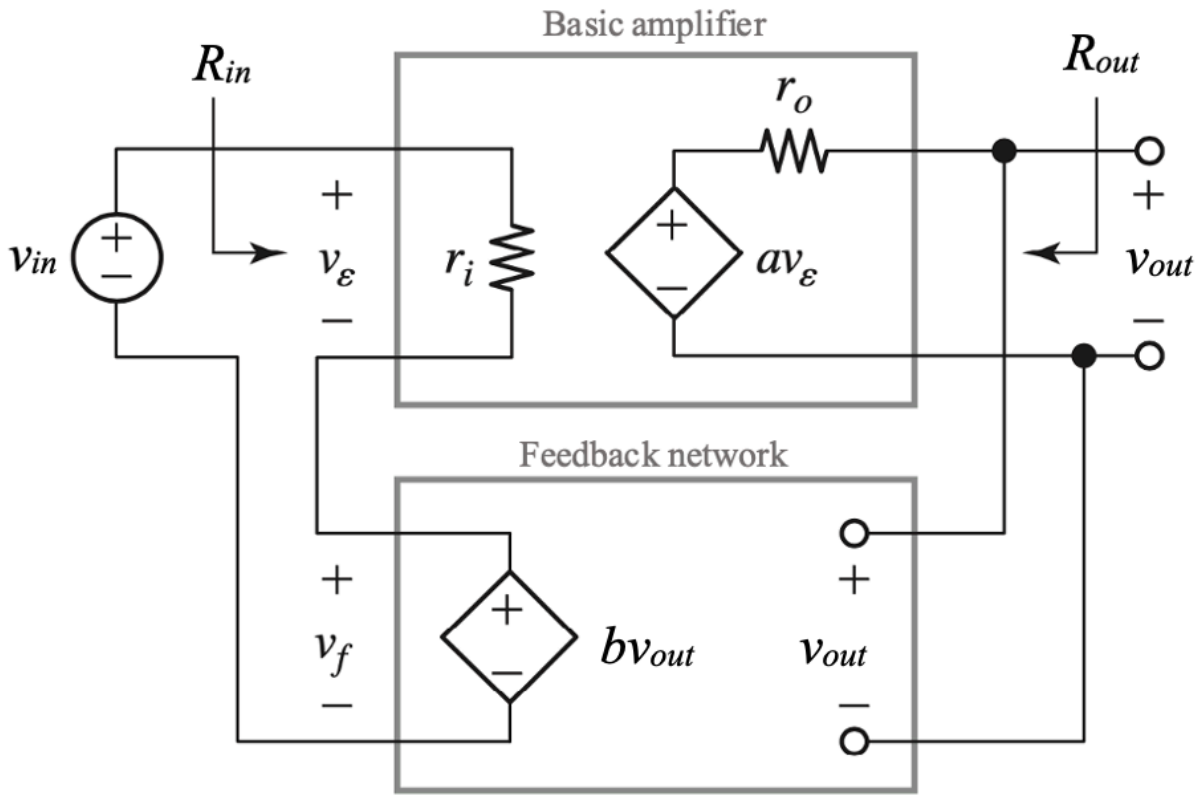


Figure 7.10.: The ideal series-shunt configuration or voltage amplifier (Franco 2015)

It can be shown:

$$R_{in} = \frac{v_x}{i_x} = r_i(1 + L)$$

$$R_{out} = \frac{v_t}{i_t} = \frac{r_o}{1 + L}$$

For $L \to \infty$ the series-shunt configuration gives $R_{in} \to \infty$ and $R_{out} \to 0$

7.6.4.2. Shunt-shunt (current-voltage) feedback configuration - Transimpedance Amplifier

Let's consider the application of negative feedback around a transimpedance amplifier. In this case feedback is performed sampling the output voltage and then feeding back a current that is a scaled version of the output voltage "into" the external input current source driving the amplifier. Note that to combine the external input current source and the current returned by the feedback network we connect them in parallel or shunt (to connect two current sources we place them in parallel, never in series).

To neglect loading at the input of the basic amplifier, we assume that the sources i_{in} and bv_{out} have *infinite* parallel resistances. To neglect loading at the output port of the basic amplifier, we assume the output port of the basic amplifier is left open and the input port of the feedback network has *infinite* resistance.

Note: for the shunt-shunt feeback configuration, b has dimensions of $1/\Omega$.

It can be shown:

$$R_{in} = \frac{r_i}{1 + L}$$

$$R_{out} = \frac{r_o}{1 + L}$$

For $L \to \infty$ the shunt-shunt configuration gives $R_{in} \to 0$ and $R_{out} \to 0$

7.6.4.3. Series-series (voltage-current) feedback configuration - Transconductance Amplifier

Let's consider the application of negative feedback around a transconductance amplifier. In this case feedback is performed sampling the output current and then feeding back a voltage that is a scaled version of the output current "into" the external input voltage source driving the amplifier. Note that the operation of current sensing at the output is performed in series, or shunt (when we measure a current we place the ampmeter in series, never in parallel).

To neglect loading at the input of the basic amplifier, we assume that the sources v_{in} and bi_{out} have *zero* series resistances. To neglect loading at the output port of the basic amplifier, we assume the output port of the basic amplifier is shorted and the input port of the feedback network has *zero* resistance.

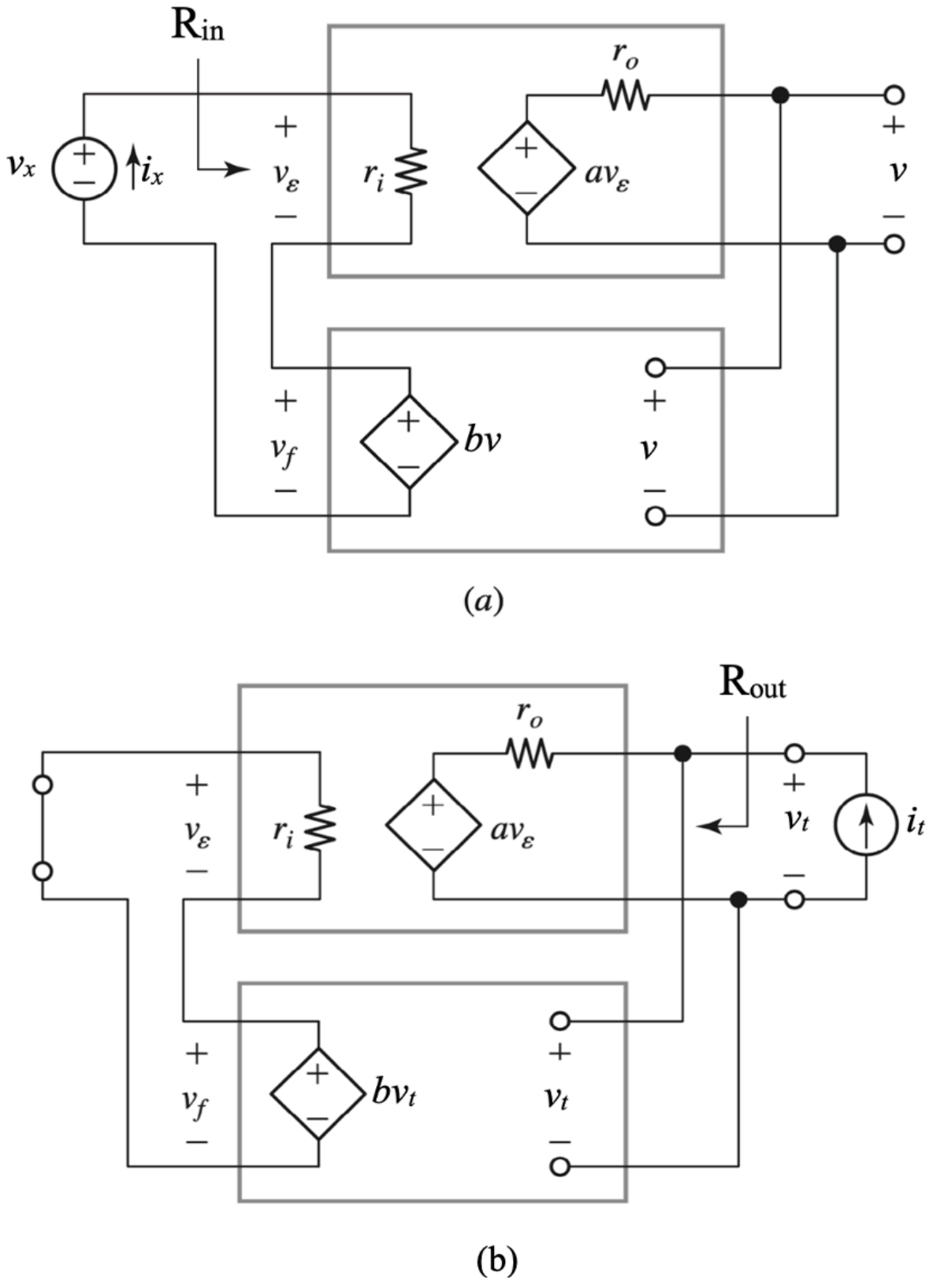


Figure 7.11.: (a) test circuit for the calculation of the input impedance and (b) test circuit for the calculation of the output impedance

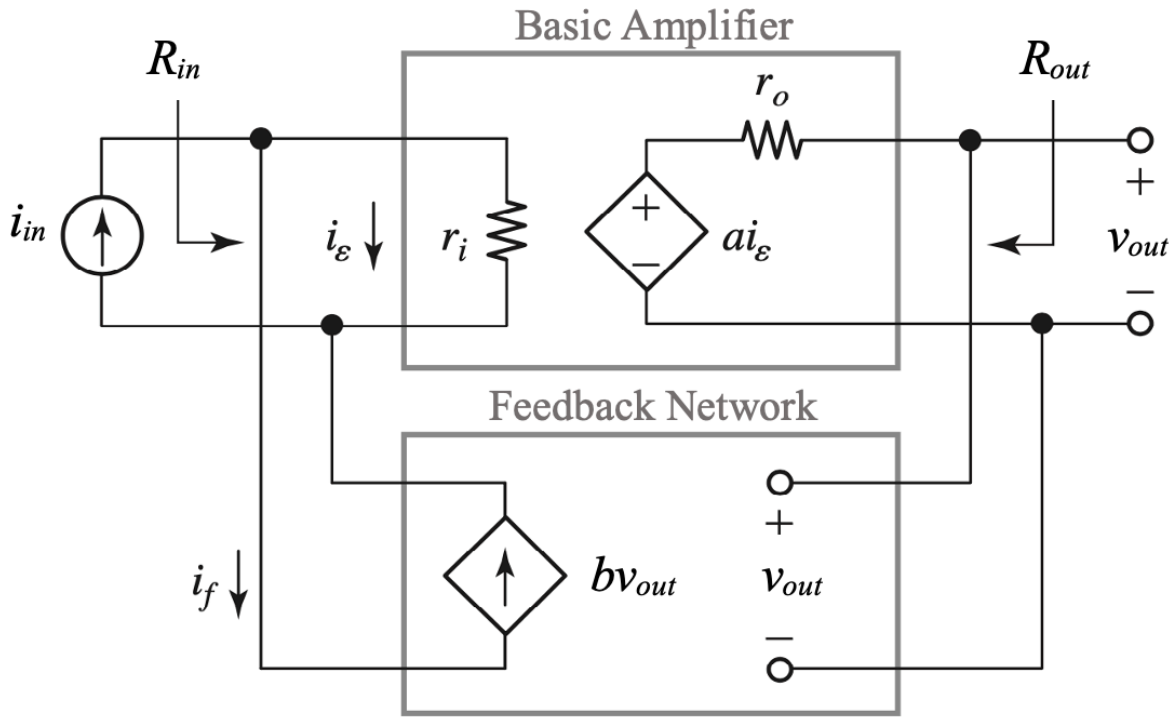


Figure 7.12.: The ideal shunt-shunt configuration or transimpedance amplifier (Franco 2015)

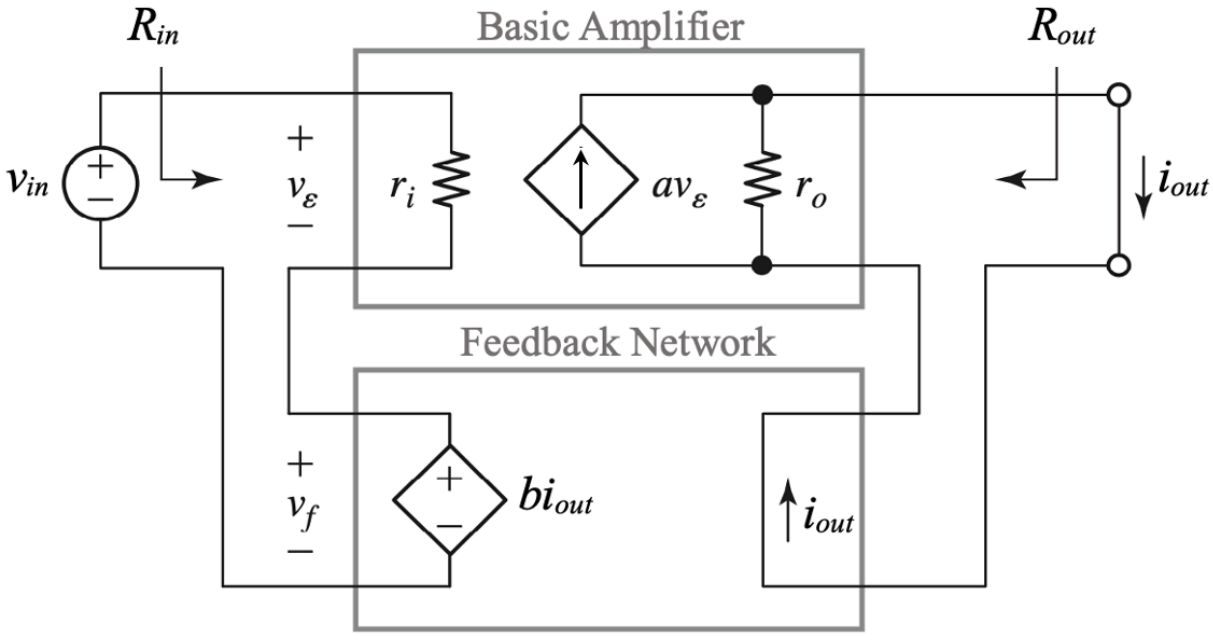


Figure 7.13.: The ideal series-series configuration or transconductance amplifier (Franco 2015)

Note: for the series-series feedback configuration, b has dimensions of Ω.

It can be shown:

$$R_{in} = r_i(1 + L)$$

$$R_{out} = r_o(1 + L)$$

For $L \to \infty$ the series-series configuration gives $R_{in} \to \infty$ and $R_{out} \to \infty$

7.6.4.4. Shunt-series (current-current) feedback configuration - Current Amplifier

Let's consider the application of negative feedback around a current amplifier. In this case feedback is performed sampling the output current and then feeding back a current that is a scaled version of the output current "into" the external input current source driving the amplifier.

To neglect loading at the input of the basic amplifier, we assume that the sources i_{in} and bi_{out} have *infinite* parallel resistances. To neglect loading at the output port of the basic amplifier, we assume the output port of the basic amplifier is shorted and the input port of the feedback network has *zero* resistance.

Note: for the shunt-series feedback configuration, b is unitless.

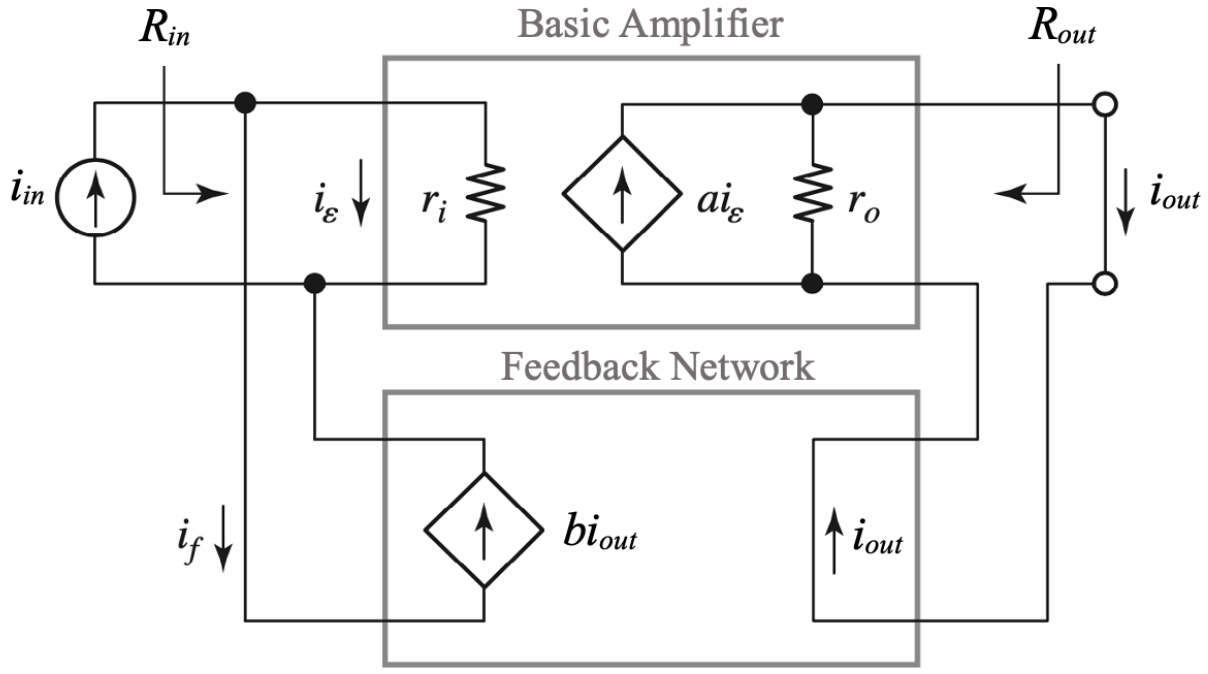


Figure 7.14.: The shunt-series configuration or current amplifier (Franco 2015)

It can be shown:

$$R_{in} = \frac{r_i}{1+L}$$

$$R_{out} = r_o(1+L)$$

For $L \to \infty$ the shunt-series configuration gives $R_{in} \to 0$ and $R_{out} \to \infty$

7.7. Feedback analysis of op amp circuits

In typical op amp circuits r_i and r_o are negligible compared to the resistances used in the feedback network.

7.7.1. Non Inverting configuration

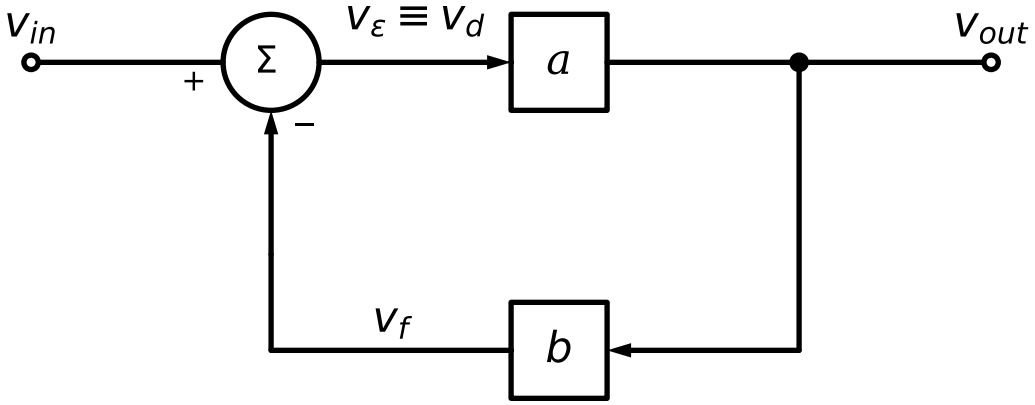


Figure 7.15.: Block diagram of negative feedback system.

For identifying a and b directly from the circuit, we can set $v_{in} = 0$ and note that with $v_{in} = 0$

$i_{out} = \dfrac{v_{out}}{R_2 + R_1||r_i} = \dfrac{a_D v_D}{r_o + R_2 + R_1||r_i}$

$v_f = -v_D = i_{out} \cdot (R_1||r_i)$

$v_{out} = i_{out} \cdot \left(R_2 + R_1||r_i\right)$

So,

$b = \dfrac{v_f}{v_{out}} = \dfrac{R_1||\not{r_i}}{R_1||\not{r_i} + R_2} \approx \dfrac{R_1}{R_1 + R_2}$

$a = \dfrac{v_{out}}{v_D} = a_D \cdot \dfrac{R_2 + R_1||\not{r_i}}{R_2 + R_1||\not{r_i} + \not{r_o}} \approx a_D$

In practice, for identifying a and b, it is better to think directly in terms of loop gain L

$$L = ab = -\frac{v_r}{v_t}$$

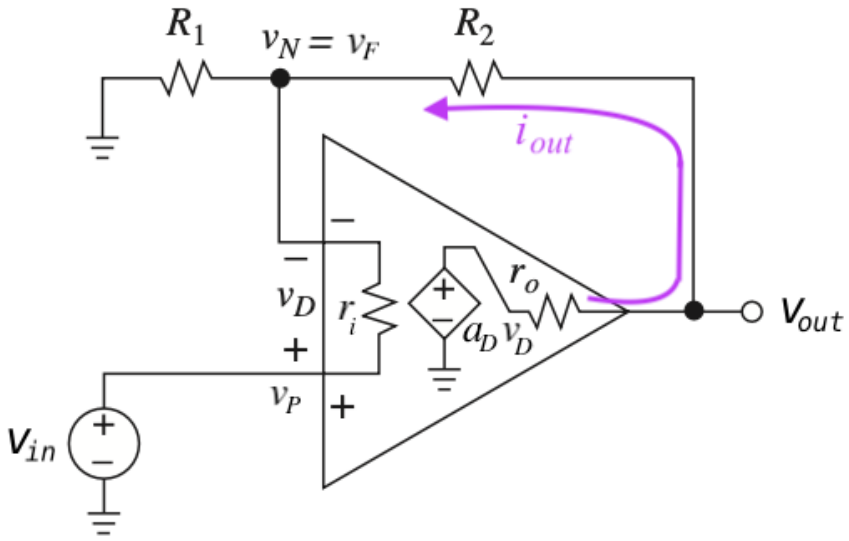


Figure 7.16.: Non inverting op amp amplifier (Franco 2015).

The minus sign is due to the − at the summing node

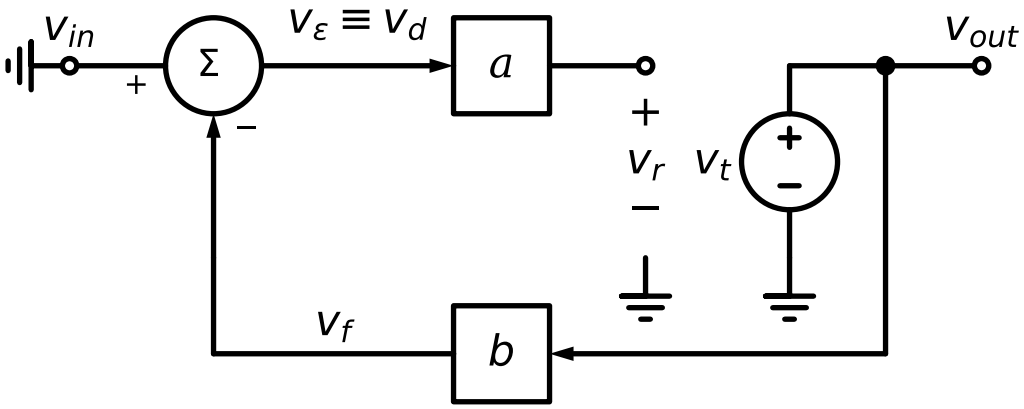


Figure 7.17.: Breaking the Loop.

Note: To find the loop gain, it is best to break the loop at the opamp's voltage controlled voltage source. This approach preserves all of the node impedances in the circuit.

$$L = -\frac{v_r}{v_t} = a_D \cdot \frac{R_1 || \cancel{r_i}}{R_1 || \cancel{r_i} + R_2 + \cancel{r_o}}$$

$$A = A_{ideal}\frac{L}{1+L} \approx A_{ideal}\left(1 - \frac{1}{L}\right)$$

$$gainerror \approx \frac{A_{ideal}\left(1 - \frac{1}{L}\right) - A_{ideal}}{A_{ideal}} = -\frac{1}{L}$$

The value of A_{ideal} is already known from the ideal op amp analysis

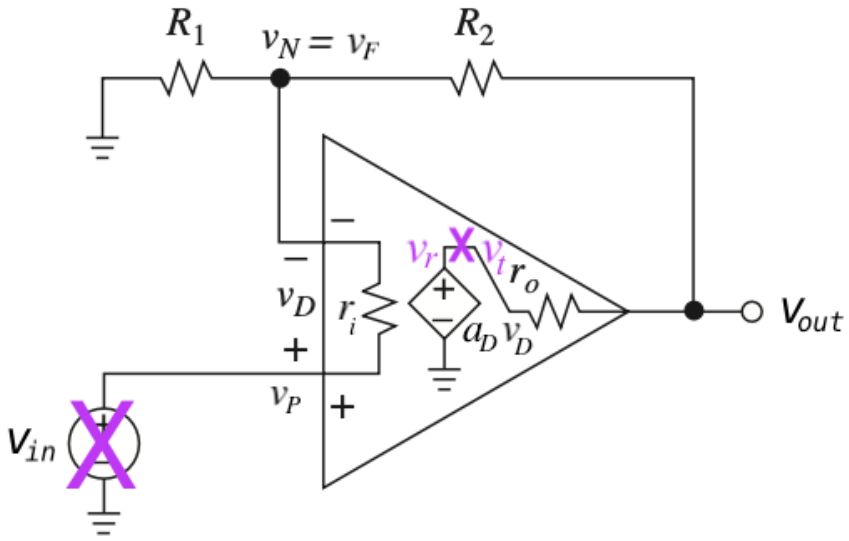


Figure 7.18.: Non inverting op amp amplifier with v_{in} nulled and the loop broken at the opamp's vcvs.

Infinite opamp gain ($a_D \to \infty$) implies infinite open-loop gain ($a \approx a_D$) and therefore infinite loop gain $L = ab$

$$A_{ideal} = \frac{1}{b} \approx 1 + \frac{R_2}{R_1}$$

Example 4 - non inverting op amp based voltage amplifier

For $a_D = 10^6$, $r_i = 10M\Omega$, $r_o = 10\Omega$, $R_1 = 20k\Omega$ and $R_2 = 80k\Omega$:

$$L = -\frac{v_r}{v_t} = a_D \cdot \frac{R_1 || \cancel{r_i}}{R_1 || \cancel{r_i} + R_2 + \cancel{r_o}} \approx 10^6 \frac{20k\Omega}{20k\Omega + 80k\Omega}$$

The exact value of L is 199600, so the above approximation has 0.2% error.

$$A = 5 \cdot \frac{200000}{1 + 200000} = 4.999975$$

$$gainerror \approx -\frac{1}{200000} = -5ppm$$

This is the same result as before, except we did not have to go through a tedious nodal analysis and plug in numbers in a "high entropy" espression.

What if the op amp gain changes ?

If the gain is cut in half :

$$A = 5 \cdot \frac{100000}{1 + 100000} = 4.999950$$

If the gain doubles:

$$A = 5 \cdot \frac{400000}{1 + 400000} = 4.999988$$

- The closed loop gain (A) is immune to large variations in the op amp gain
- The voltage gain of the overall circuit (= closed loop gain) is primarily defined by the divider ratio of the resistive feedback
  - A quantity that we can control very precisely

7.7.2. Inverting configuration

In this configuration the resistors affect both the input and the feedback path.

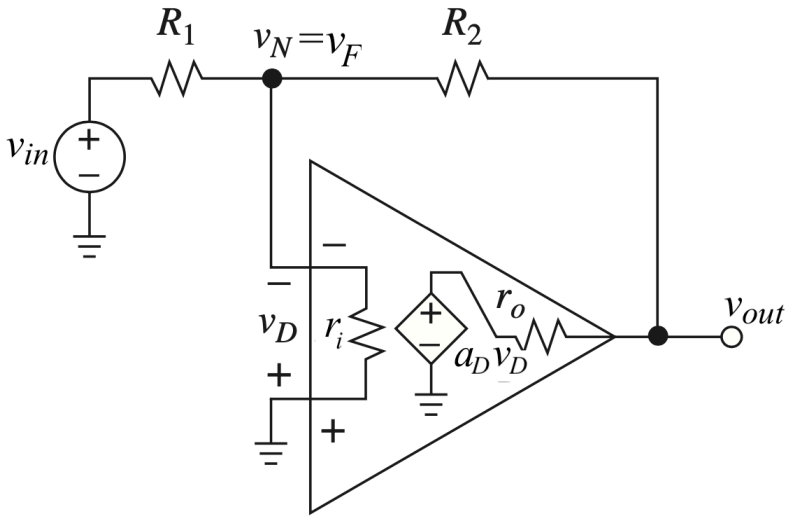


Figure 7.19.: Inverting op amp amplifier. (Franco 2015})

It is not clear how to map the circuit into the block diagram representation. The flow of the signal through the circuit elements is not unidirectional.

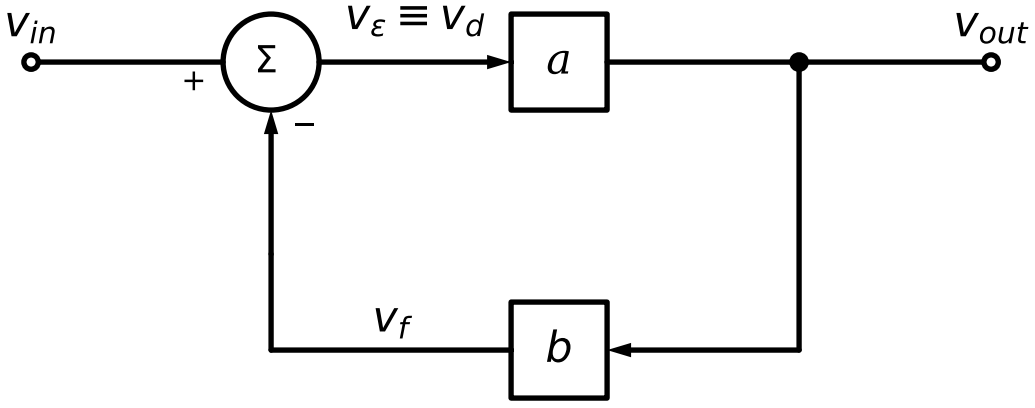


Figure 7.20.: Block diagram of negative feedback system.

We can still try to make things work using superposition.

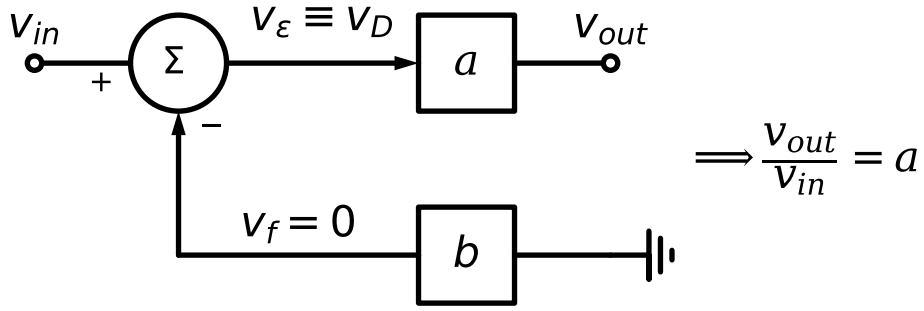


Figure 7.21.: Breaking the loop ($v_f = 0$)

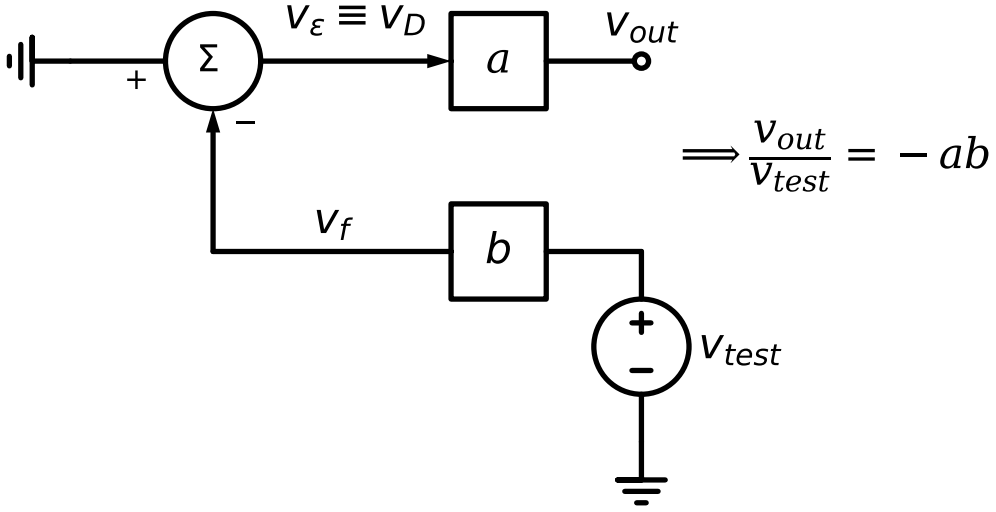


Figure 7.22.: Breaking the loop ($v_{in} = 0$)

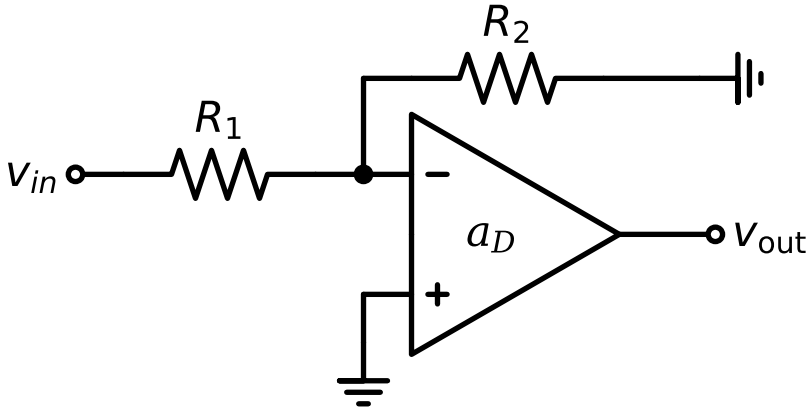


Figure 7.23.: op amp based inverting amplifier with break in the loop ($v_f = 0$)

$$a = \frac{v_{out}}{v_{in}} = -\frac{R_2}{R_1 + R_2} \cdot a_D$$

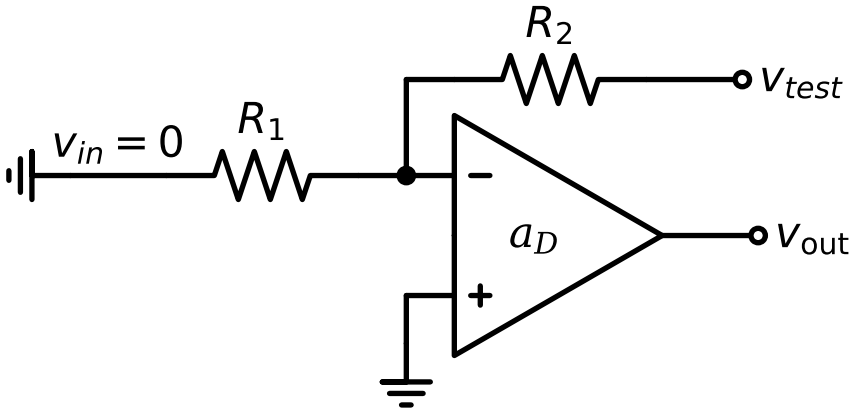


Figure 7.24.: op amp based inverting amplifier with break in the loop ($v_{in} = 0$)

$$-ab = \frac{v_{out}}{v_{test}} = -\frac{R_1}{R_1 + R_2} \cdot a_D$$

Note: the loop gain is the same as the one of the non-inverting configuration.

$$A_{ideal} = \frac{1}{b} = \frac{a}{ab} = -\frac{R_2}{R_1}$$

Beyond A_{ideal} the only other value we need to know is the loop gain $L = ab$ so we can compute the deviation from ideality

7.7.3. Comparison between non-inverting and inverting topology

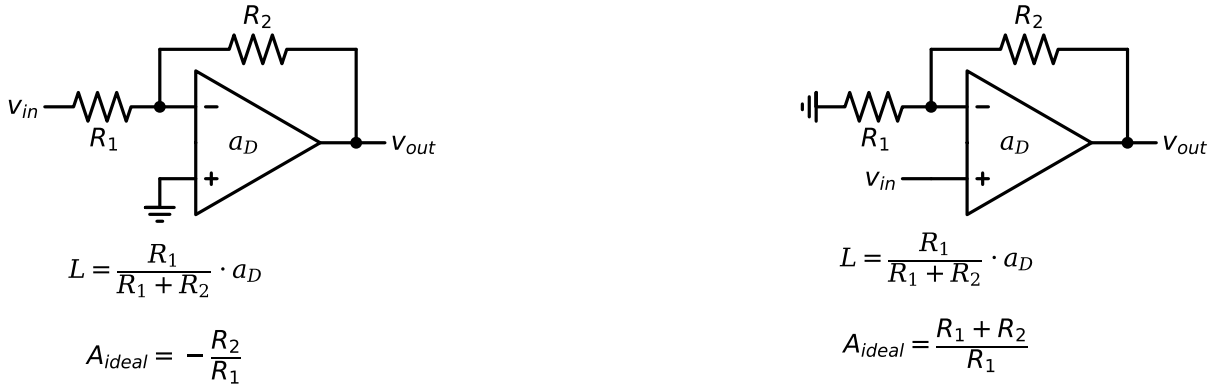


Figure 7.25.: comparison between inverting and non-inverting topology

The following model is valid for both topologies:

In summary op amp circuits can be analyzed as follows:

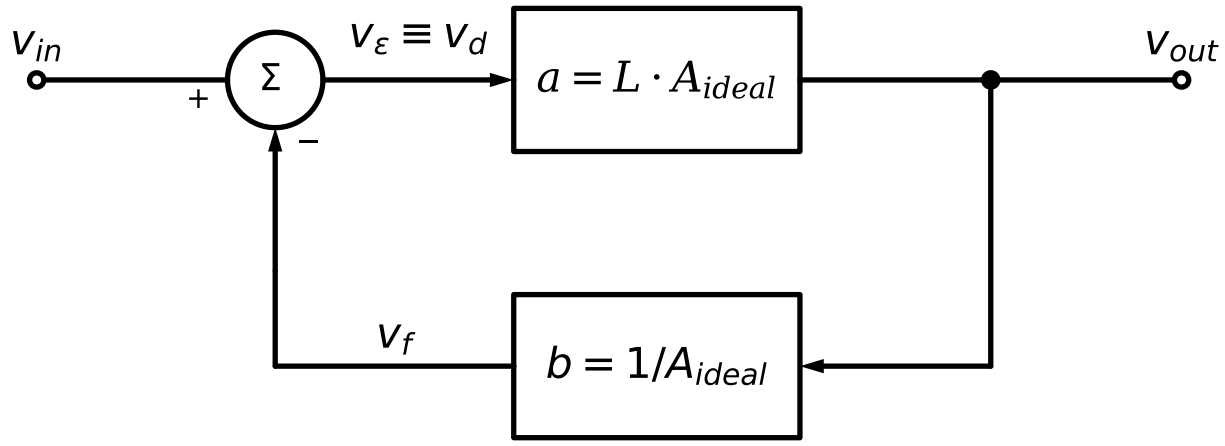


Figure 7.26.: Block diagram model valid for both topologies.

- Find A_{ideal} using nodal analysis, assuming infinite op amp gain
- Find the loop gain to compute the deviation from the ideal case
  - This is usually straight forward, especially when there are ideal breakpoints that do not alter the impedance loading around the loop
  - The best breakpoint for a voltage amplifier is right at the controlled voltage source

$$A = \frac{v_{out}}{v_{in}} = A_{ideal} \cdot \frac{L}{L+1} = A_{ideal} \cdot \frac{1}{1+\frac{1}{L}}$$

7.7.4. The four feedback configurations using op amps

Sometimes the operations of output sensing and input returning are not obvious. Even though strictly speaking the op amp is a voltage amplifier, it can be used in any of the four feedback configurations.

7.7.4.1. Series-Shunt (voltage-voltage) configuration (Voltage Amplifier)

For $a_v \to \infty$ this circuit gives:

$$A_v = \frac{v_{out}}{v_{in}} = \frac{1}{b} = 1 + \frac{R_2}{R_1}$$

$R_{in} \to \infty$

$R_{out} \to 0$

This circuit is the popular non inverting op amp based voltage amplifier.

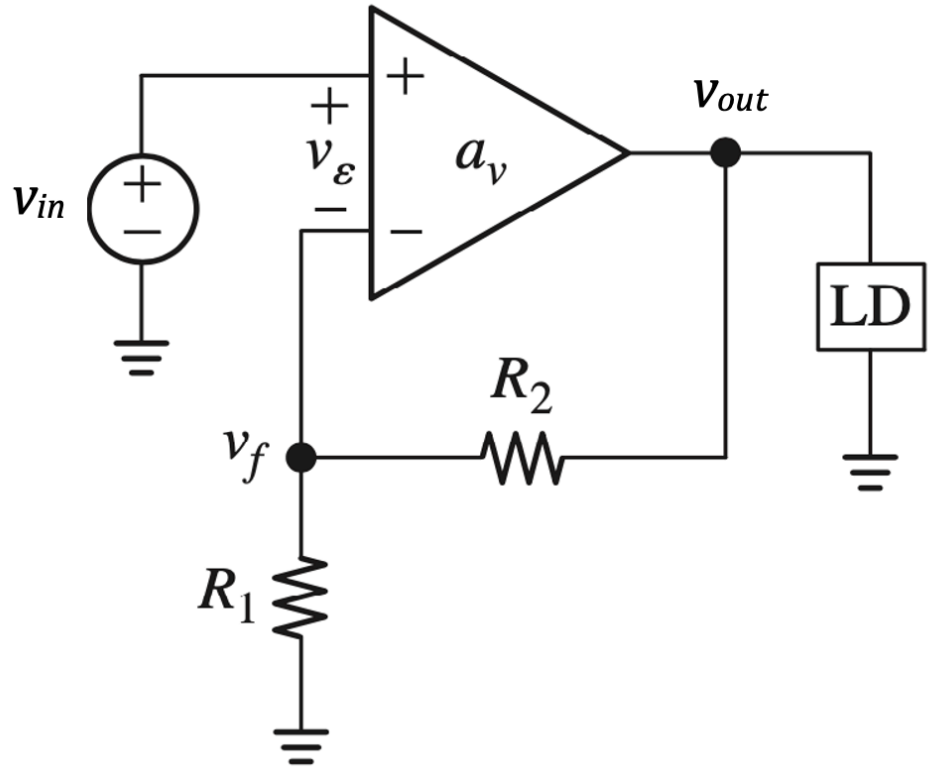


Figure 7.27.: Series-shunt configuration: $b = \dfrac{v_f}{v_{out}} = \dfrac{R_1}{R_1 + R_2}$

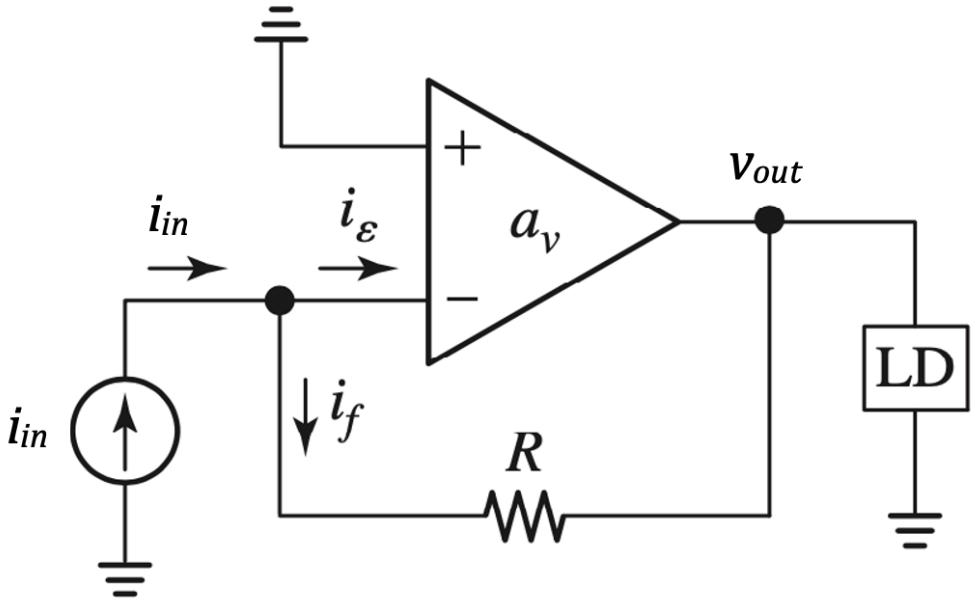


Figure 7.28.: Shunt-shunt configuration: $b = \dfrac{i_f}{v_{out}} = -\dfrac{1}{R}$

7.7.4.2. Shunt-Shunt (current-voltage) configuration (Transimpedance Amplifier)

For $a_v \to \infty$ this circuit (= TIA) gives:

$R_m = \frac{v_{out}}{i_{in}} = \frac{1}{b} = -R$

$R_{in} \to 0$

$R_{out} \to 0$

Even though the shunt-shunt configuration is a TIA, it forms the basis of the popular inverting op amp based voltage amplifier.

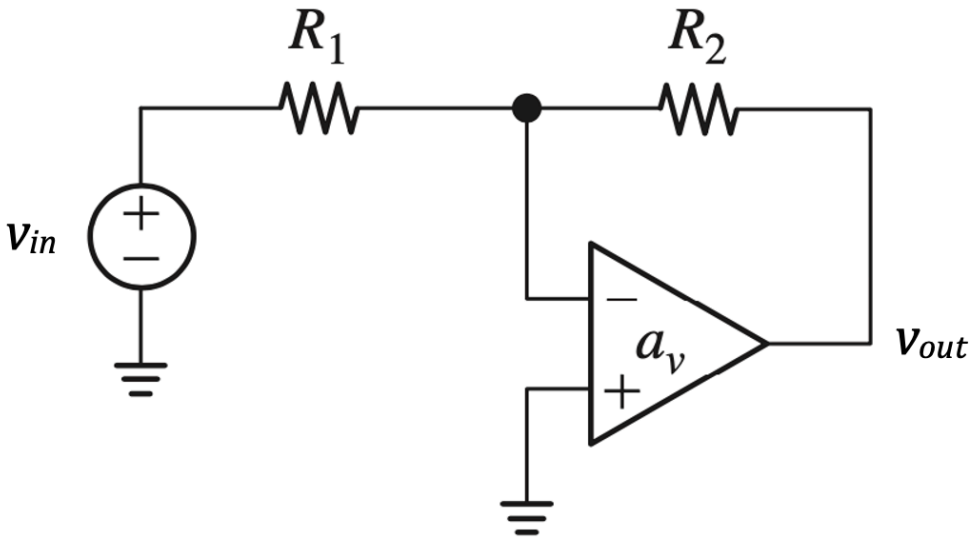


Figure 7.29.: inverting op amp based voltage amplifier

This becomes more evident if we perform a source transformation to convert the voltage source v_{in} in Figure 7.29 into a current source $i_{in} = \frac{v_{in}}{R_1}$ as shown in Figure 7.30.

For the inverting voltage amplifier as $a_v \to \infty$

$A_v = \frac{v_{out}}{v_{in}} = \frac{v_{out}}{i_{in}} \times \frac{i_{in}}{v_{in}} = -\frac{R_2}{R_1}$

$R_{in} = R_1$

$R_{out} \to 0$

7.7.4.3. Series-Series (voltage-current) configuration (Transconductance Amplifier)

For $a_v \to \infty$

$G_m = \frac{i_{out}}{v_{in}} = \frac{1}{b} = \frac{1}{R}$

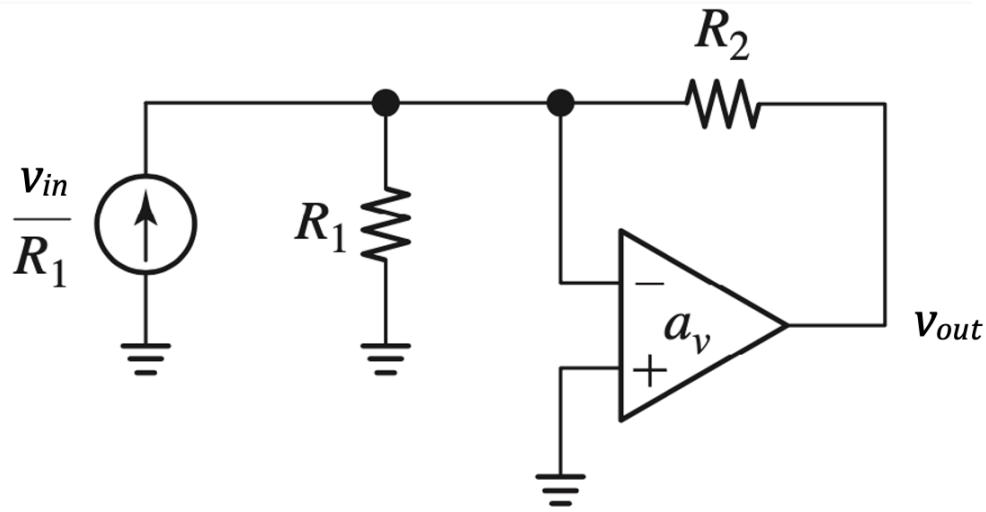


Figure 7.30.: inverting op amp based voltage amplifier with the input voltage source transformed into a current source

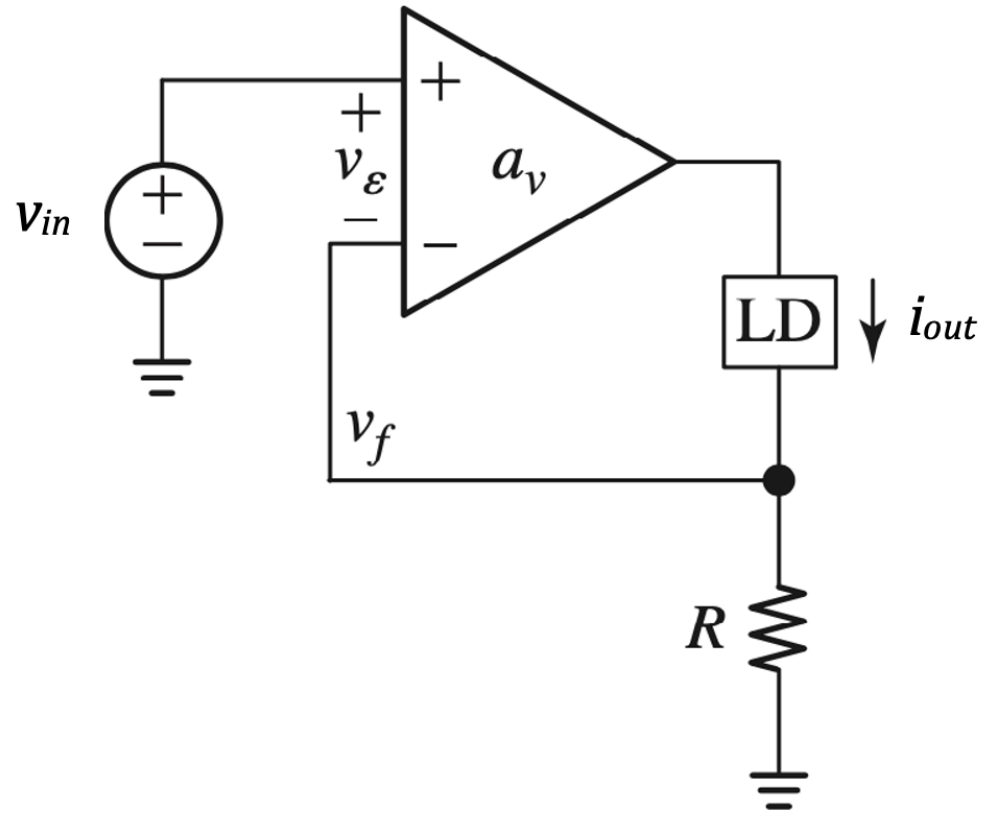


Figure 7.31.: series-series configuration: $b = \frac{v_f}{i_{out}} = R$

$R_{in} \to \infty$

$R_{out} \to \infty$

7.7.4.4. Shunt-Series (current-current) configuration (Current Amplifier)

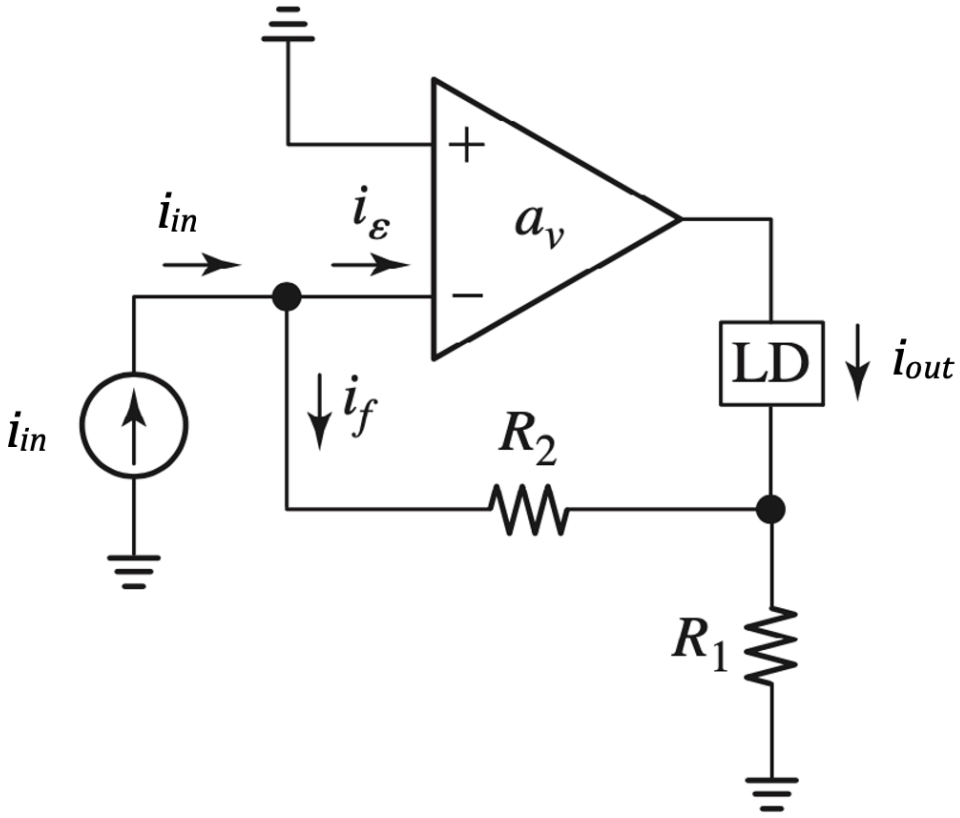


Figure 7.32.: shunt-series configuration: $b = \frac{i_f}{i_{out}} = -\frac{R_1}{R_1 + R_2} = -\frac{1}{1 + R_2/R_1}$

For $a_v \to \infty$

$A_i = \frac{i_{out}}{i_{in}} = \frac{1}{b} = -(1 + \frac{R_2}{R_1})$

$R_{in} \to 0$

$R_{out} \to \infty$

7.8. Return Ratio Analysis

7.8.1. Closed-loop gain using return ratio

Given a feedback amplifier, with a controlled source of value k (see Figure 7.33), the closed loop gain can be derived as shown in (Gray et al. 2009b):

$$A = \frac{s_{out}}{s_{in}} = A_\infty \cdot \frac{T}{1+T} + \frac{d}{1+T}$$

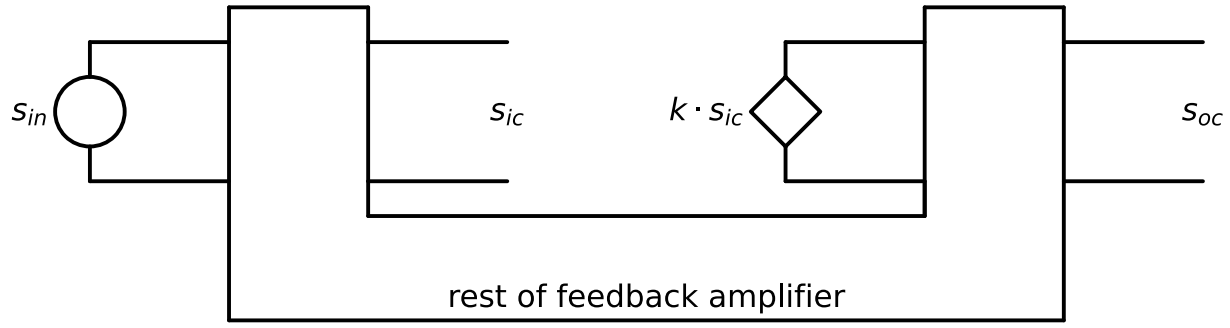


Figure 7.33.: Feedback amplifier: return ratio framework to derive the closed-loop gain

Visually, the expression for the closed loop gain can be summarized through the block diagram in Figure 7.34:

$$\left(s_{in} - \frac{s_{out}}{A_\infty}\right) \cdot TA_\infty + d \cdot s_{in} = s_{out}$$

$$A = \frac{s_{out}}{s_{in}} = A_\infty \cdot \frac{T}{1+T} + \frac{d}{1+T}$$

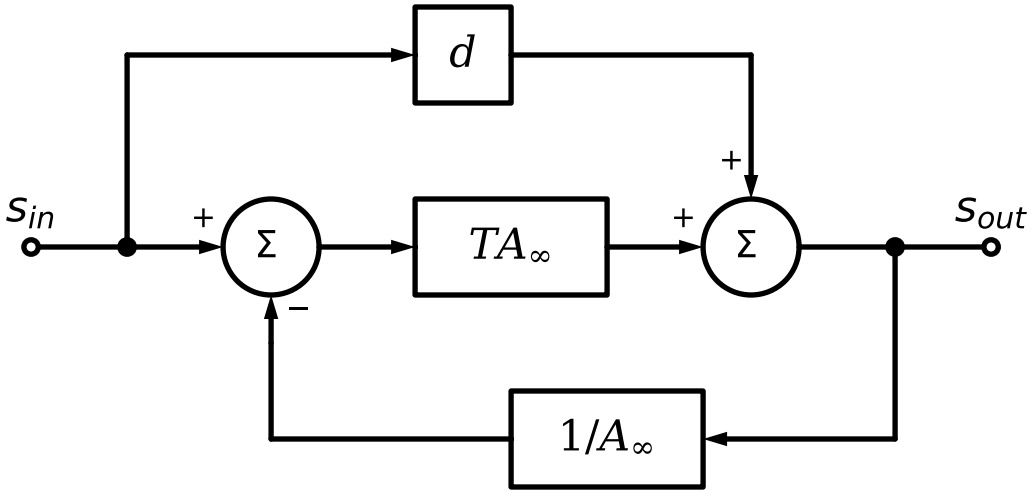


Figure 7.34.: Block diagram for closed-loop gain using return ratio framework

The three terms needed to find the closed loop gain (A_∞, T, d) are all directly computable and measurable (SPICE), and do not rely on any idealization of the feedback network.

In practical circuits the feedback network loads both the input and output of the amplifier, so idealizing its effect is not realistic. With two-port analysis incorporating the effect of loading without the use of suitable approximations/idealizations of the feedback network becomes extremely tedious.

The return ratio analysis does not try to identify the transfer function of the basic amplifier and the feedback network separately. It aims to identify directly the gain around the feedback loop. From the loop gain of a circuit, we can determine:

- stability
- closed-loop gain
- nodal impedances

The return ratio analysis can be applied to any arbitrary feedback circuit, independent of topology and port structure.

The return ratio for a dependent source in a feedback loop is found as follows:

1. Set all independent sources to zero
2. Identify a dependent source in the feedback loop that you want to analyze and break the loop by disconnecting the dependent source from the rest of the circuit. Leave the dependent source open-circuited if it is of the voltage type, or short-circuited if it is of the current-type
3. On the side of the break that is not connected to the dependent source, inject an independent test source s_t of the same sign and type as the dependent source
4. Find the return signal s_r, generated at the controlled source that was disconnected
5. The return ratio T for the dependent source is $T = -s_r/s_t$

    - Provided that we have chosen a controlled source that breaks the loop globally, the return ratio of the dependent source is equal to the loop gain of the circuit.
    - For the stabilty analysis of a feedback circuit we must have the loop gain not just any return ratio. Unless, we have a single loop feedback circuit, the return ratios computed for different dependent sources are not necessarily equal (Hurst 1991), so in general, the return ratio is not a global property of the loop (Hajimiri 2023).

The direct feedthrough d is given by:

$$d = \frac{s_{out}}{s_{in}}|_{k=0}$$

which is the transfer function from input to output evaluated for k = 0. In other words, the direct feedthrough d is the signal transfer from input to output through the passive elements, it represent a signal path from input to output that goes around rather than through the controlled source k.

The value of A_∞ is the closed loop gain when the feedback circuit is ideal (that is when $T \to \infty$). If $T \to \infty$ then $A = A_\infty$, because $\frac{T}{1+T} \to 1$ and $\frac{d}{1+T} \to 0$

Since letting $k \to \infty$ causes $T \to \infty$ (Gray et al. 2009b) the value of A_∞ can be easily found by evaluating

$$\frac{s_{out}}{s_{in}}|_{k\to\infty}$$

When $k \to \infty$, the controlling signal s_{ic} for the dependent source must be zero for the output of the controlled source to be finite. The controlled source output will be finite if the feedback is negative.

The key difference between the return ratio analysis and the two-port analysis can be seen comparing Figure 7.34 and Figure 7.35 In the two-port analysis all forward signal transmission is lumped in the block a. In the return ratio analysis, there are two forward signal paths: one path (d) for the feedback network and another path (TA_∞) for the forward gain.

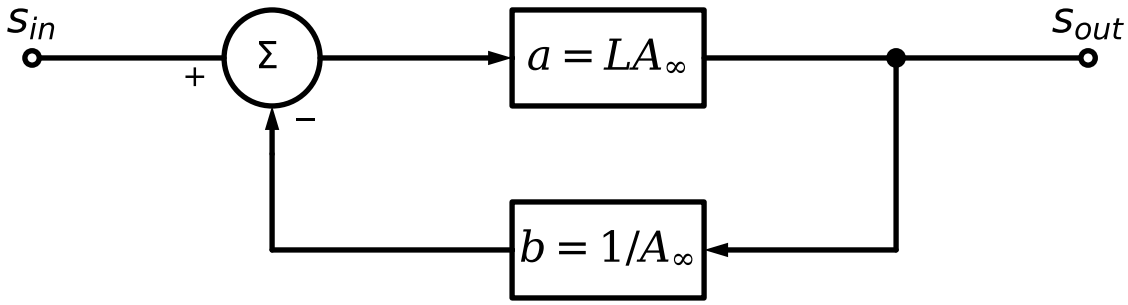


Figure 7.35.: Block diagram model for closed-loop gain using the two-port framework

Example 5 - common source stage with current source biasing

Consider the circuit in Figure 7.36 and its AC equivalent model in Figure 7.37

Current source biasing in this circuit doesn't work without feedback setting the drain voltage.

Given the finite gain and the nature of the impedances of the MOS, in this circuit is hard to decouple a(s) and b(s), so using the two-port analysis is not a good option.

$$A(s) = \frac{a(s)}{1 + a(s)b(s)}$$

In this case, since the circuit is not very complex we can use exact nodal analysis.

$$\frac{v_{in} - v_x}{R_1} + \frac{v_{out} - v_x}{R_2} = 0 \Leftrightarrow \frac{v_{in}}{R_1} - \left(\frac{1}{R_1} + \frac{1}{R_2}\right) \cdot v_x + \frac{v_{out}}{R_2} = 0 \tag{7.7}$$

$$\frac{v_x - v_{out}}{R_2} = g_m v_x \Leftrightarrow v_x = \frac{v_{out}}{1 - g_m R_2} \tag{7.8}$$

Substituting Equation 7.7 in Equation 7.8:

$$\frac{v_{out}}{v_{in}} = \frac{1 - g_m R_2}{1 + g_m R_1} = -\frac{R_2}{R_1}\left(\frac{1 - \dfrac{1}{g_m R_2}}{1 + \dfrac{1}{g_m R_1}}\right)$$

If the $g_m R$ terms are much greater than 1, the result is the same as the op amp solution.

Let's now confirm that return ratio analysis produces the same result.

Return ratio calculation

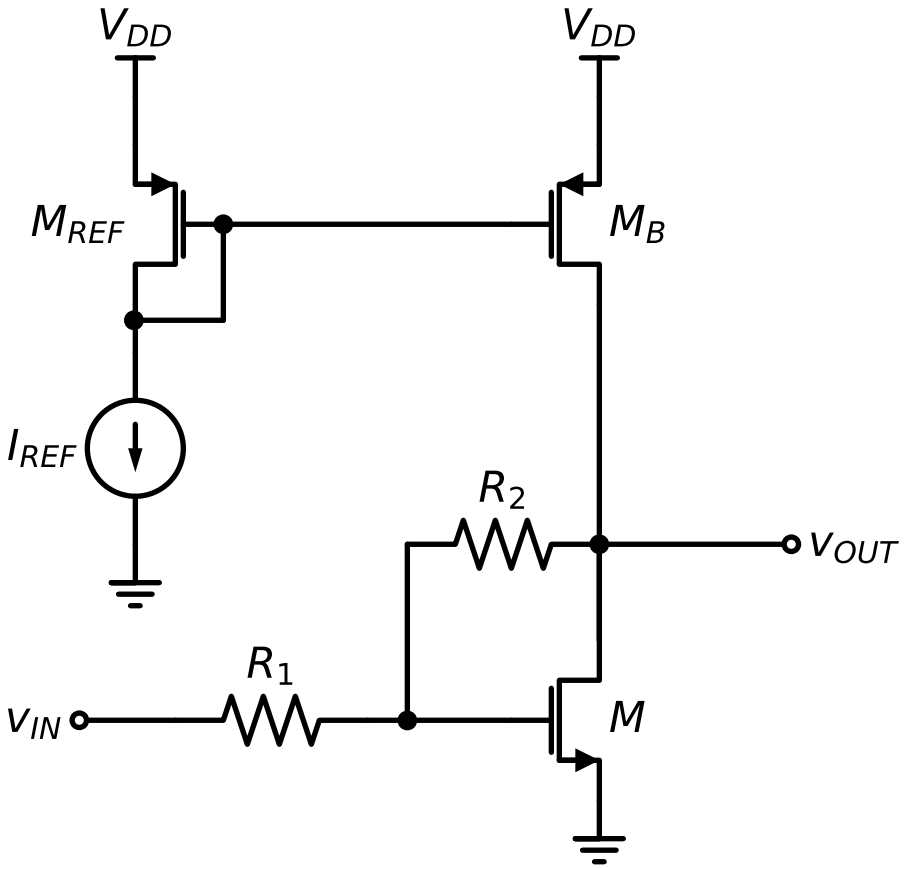


Figure 7.36.: Common source with current-source biasing

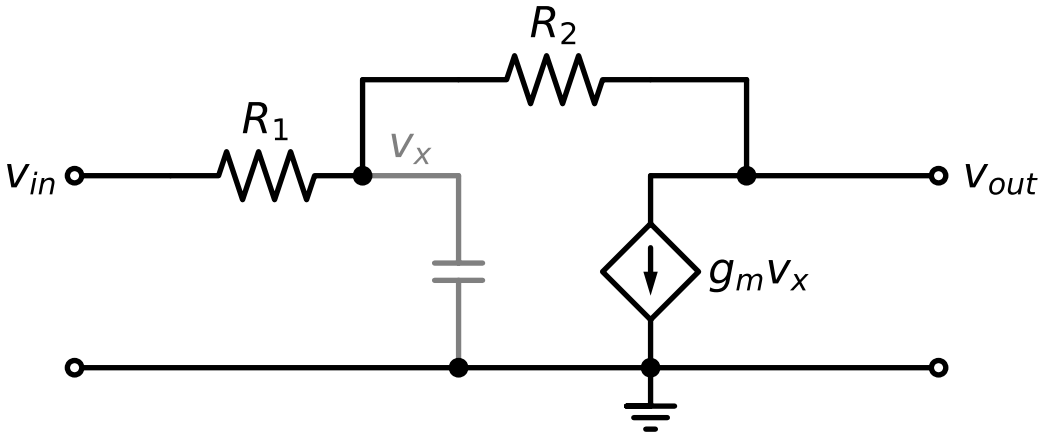


Figure 7.37.: AC equivalent model of the feedback circuit in Figure 7.36

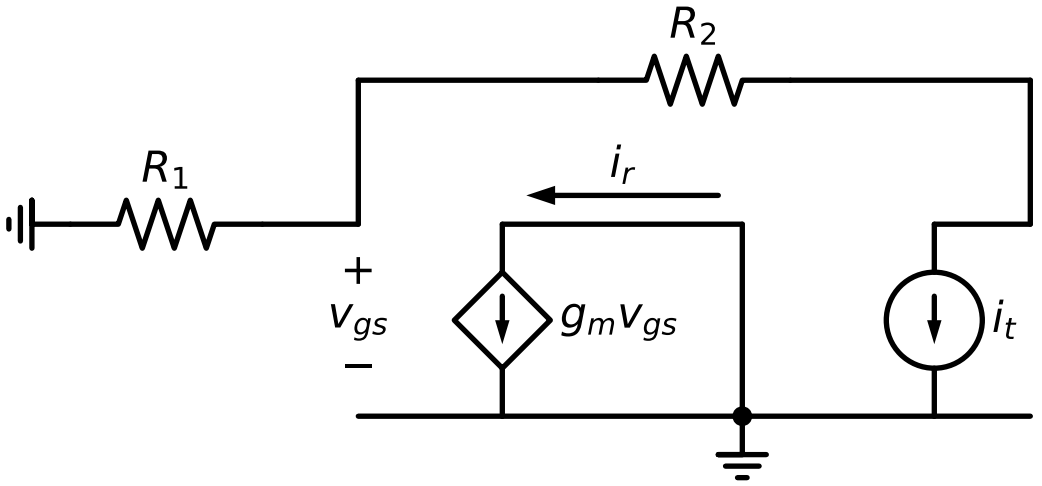


Figure 7.38.: return ratio calculation

$v_{gs} = -R_1 i_t$

$i_r = g_m v_{gs} = -g_m R_1 i_t$

$\frac{i_r}{i_t} = -g_m R_1$

$T \equiv -\frac{i_r}{i_t} = g_m R_1$

Ideal closed loop gain calculation

A_∞ is the transfer function when the gain element of the controlled source becomes ∞. In our case this corresponds to $g_m \to \infty$. If the gain element becomes ∞ the controlling signal must be zero. In our case this correspond to the input node $v_{gs} = 0$, therefore:

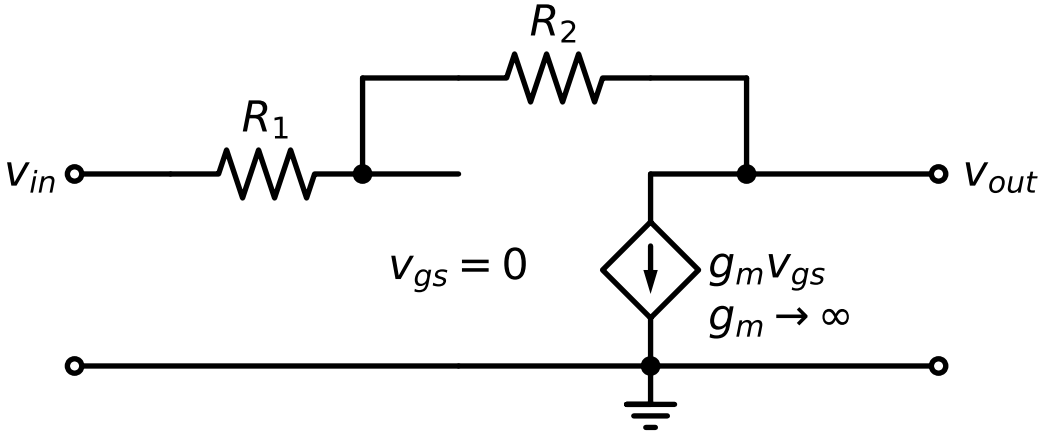


Figure 7.39.: Ideal closed loop gain calculation

$\frac{v_{in}}{R_1} = -\frac{v_{out}}{R_2} \Leftrightarrow A_\infty = \frac{v_{out}}{v_{in}} = -\frac{R_2}{R_1}$

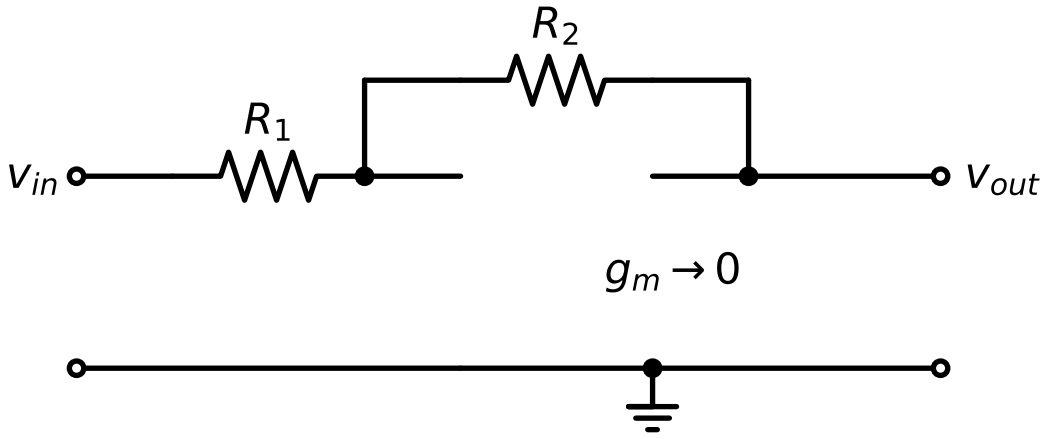


Figure 7.40.: Direct feedthrough calculation

Direct feedtrough (d) calculation

d is defined as the transfer function when the gain element becomes zero. In our case for $g_m = 0$

$v_{out} = v_{in} \Leftrightarrow d = \dfrac{v_{out}}{v_{in}} = 1$

In summary:

- $T = g_m R_1$
- $A_\infty = -\dfrac{R_2}{R_1}$
- $d = 1$

Therefore, as expected the closed loop gain computed using the return ratio framework matches with the result obtained using exact nodal analysis:

$$\begin{aligned} A &= A_\infty \cdot \frac{T}{1+T} + \frac{d}{1+T} \\ &= -\frac{R_2}{R_1} \cdot \frac{g_m R_1}{1+g_m R_1} + \frac{1}{1+g_m R_1} \\ &= \frac{1-g_m R_2}{1+g_m R_1} \end{aligned}$$

Example 6 - constant-g_m reference

In this example, we will use the return ratio framework to evaluate the stability of the bias circuit in Figure 7.41. The bias circuit in Figure 7.41 is know as constant-g_m reference or ΔV_{GS} reference.

Note: this circuit has only one-port so it cannot be analyzed using the two-ports framework.

$$\begin{aligned} I_{REF} \cdot R_2 &= V_{GS1} - V_{GS2} = V_{OV1} - V_{OV2} \\ &\approx V_{OV1}\left(1 - \frac{1}{\sqrt{m}}\right) \end{aligned}$$

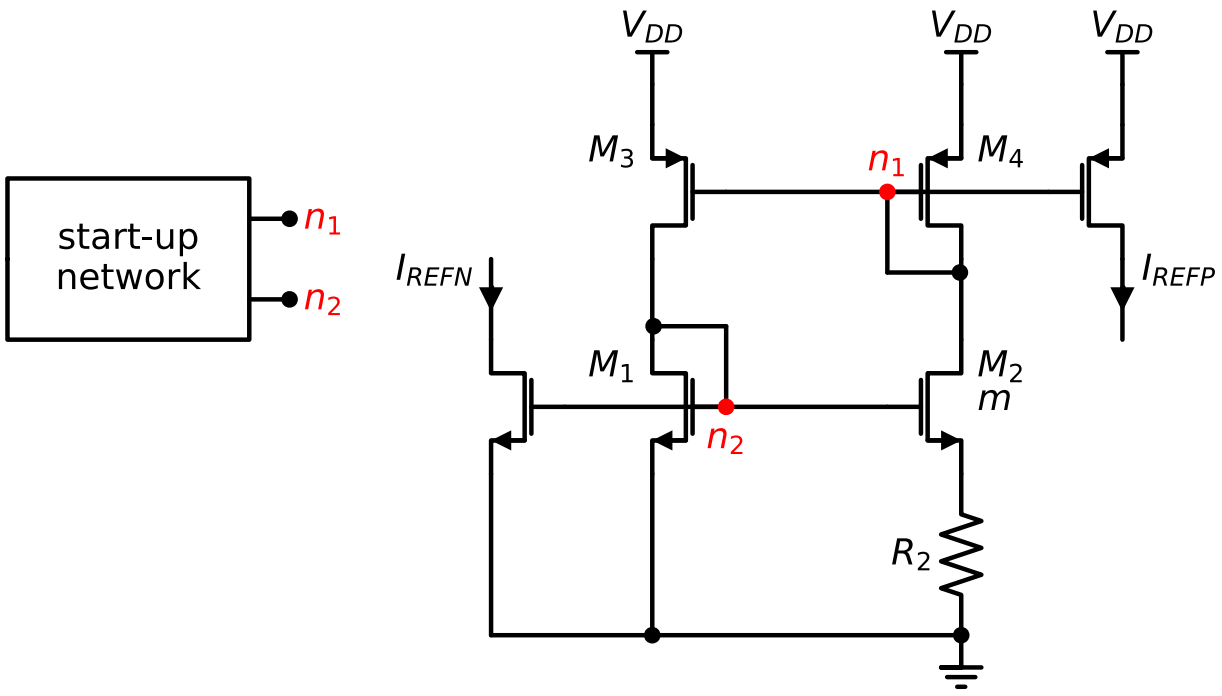


Figure 7.41.: Constant-g_m reference

$$I_{REF} \approx \frac{V_{OV1}\left(1-\frac{1}{\sqrt{m}}\right)}{R_2}$$

- This circuit has positive feedback.
- A positive feedback system is stable only if its loop gain is less than 1 ($T < 1$)

To identify which dependent source to use for computing the loop gain T of the circuit, the feedback loop we break must be global.

- M_2 and R_2 degeneration, form a local feedback loop, so M_2 cannot be used to break the main loop
- M_1 and M_4 are diode-connected; their transcondactance elements are equivalent to a resistance $1/g_m$, so cannot be used to break the main loop
- M_3 is the only *"normal"* CS gain stage in the circuit, so it is the only element that can be used to break the main loop

In general, **the return ratio of CG, CD, or degenerated CS cannot be used** to find the "global" loop gain of a circuit.

To compute the loop gain T of the circuit, it is convenient to first unroll and linearize the AC equivalent model in Figure 7.42:

and then insert the test source:

$i_r = g_{m3}v_4$

$i_t = -g_{m1}v_1$

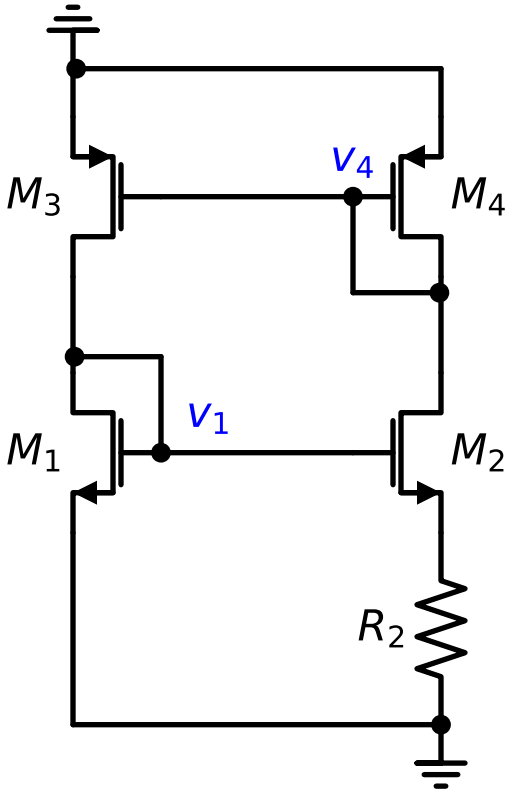


Figure 7.42.: AC equivalent model of the constant-g_m bias circuit

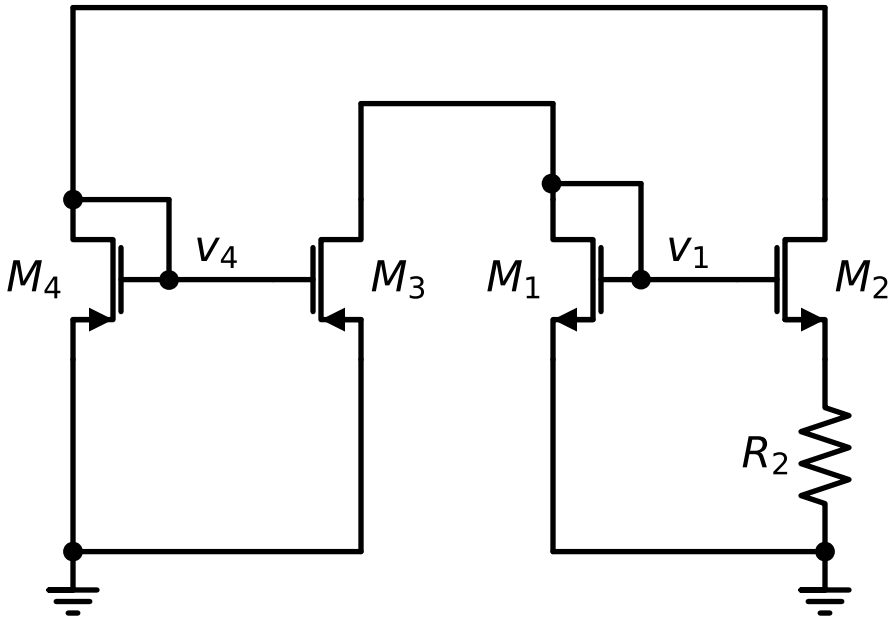


Figure 7.43.: Unrolled AC equivalent model for the constant-g_m bias circuit

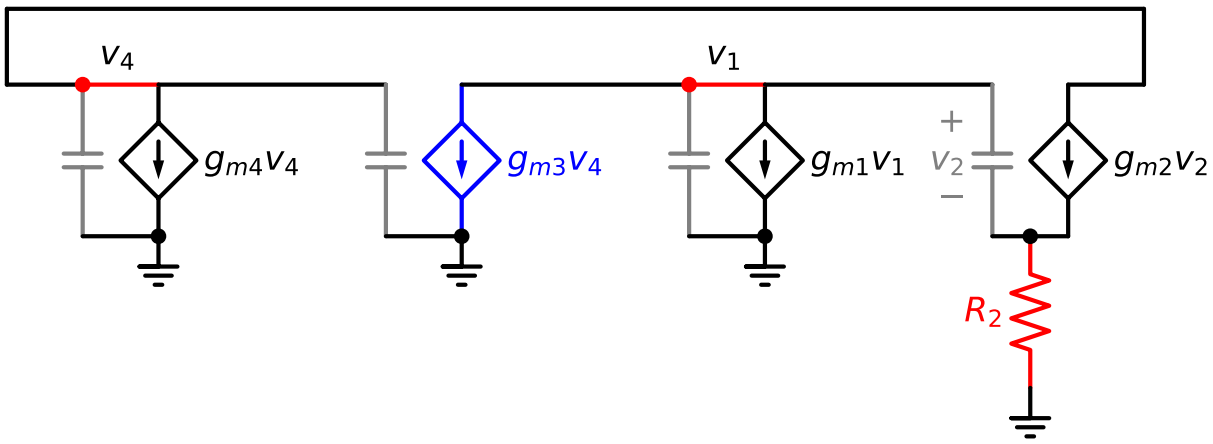


Figure 7.44.: Small signal equivalent model for the constant-g_m bias circuit, (a) red: circuit features that make s_t insertion not possible, (b) blue: best place for s_t insertion

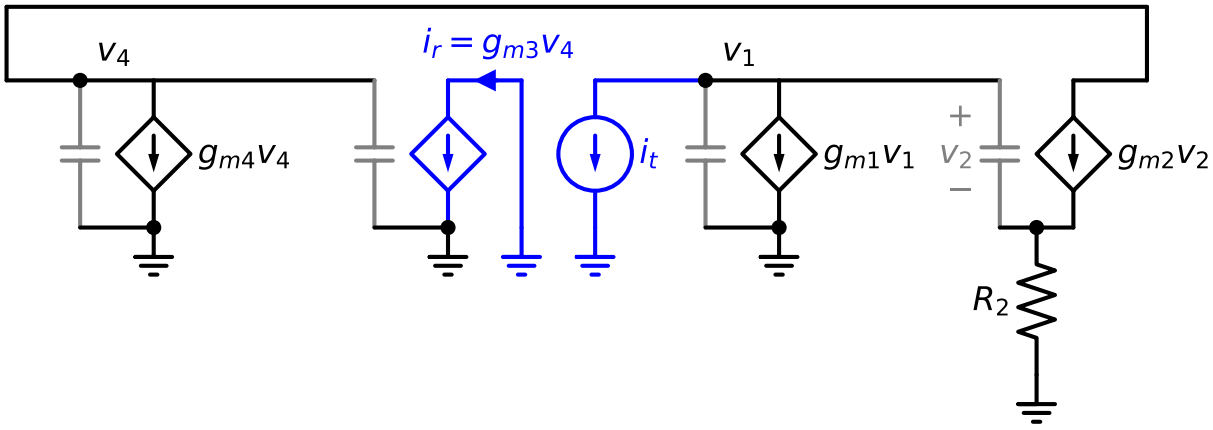


Figure 7.45.: Computing the return ratio for the constant-g_m bias circuit

$v_4 = -i_2\,(1/g_{m4})$

$i_2 = g_{m2}v_2$

$v_1 = v_2 + g_{m2}v_2R_2 \Leftrightarrow v_2 = \dfrac{v_1}{(1+g_{m2}R_2)}$

Putting the pieces together:

$$\frac{i_r}{i_t} = \left(-\frac{g_{m3}}{g_{m1}}\right)\left(-\frac{g_{m2}}{g_{m4}}\right)\left(\frac{1}{1+g_{m2}R_2}\right)$$

Therefore: $T = -\dfrac{i_r}{i_t} = -\left(\dfrac{g_{m3}}{g_{m1}}\right)\left(\dfrac{g_{m2}}{g_{m4}}\right)\left(\dfrac{1}{1+g_{m2}R_2}\right)$

The resulting return ratio is negative (i.e. the feedback is definitely not negative, even at zero frequency). The challange is to keep the magnitude of the return ratio less than unity ($|T| < 1$).

The loop has two inverting gain blocks, each loaded with a "diode connected" MOS device. The gain block formed by M_3 is loaded by the diode connected MOS M_1 and the gain block formed by M_2 and R_2 is loaded by the diode connected MOS M_4.

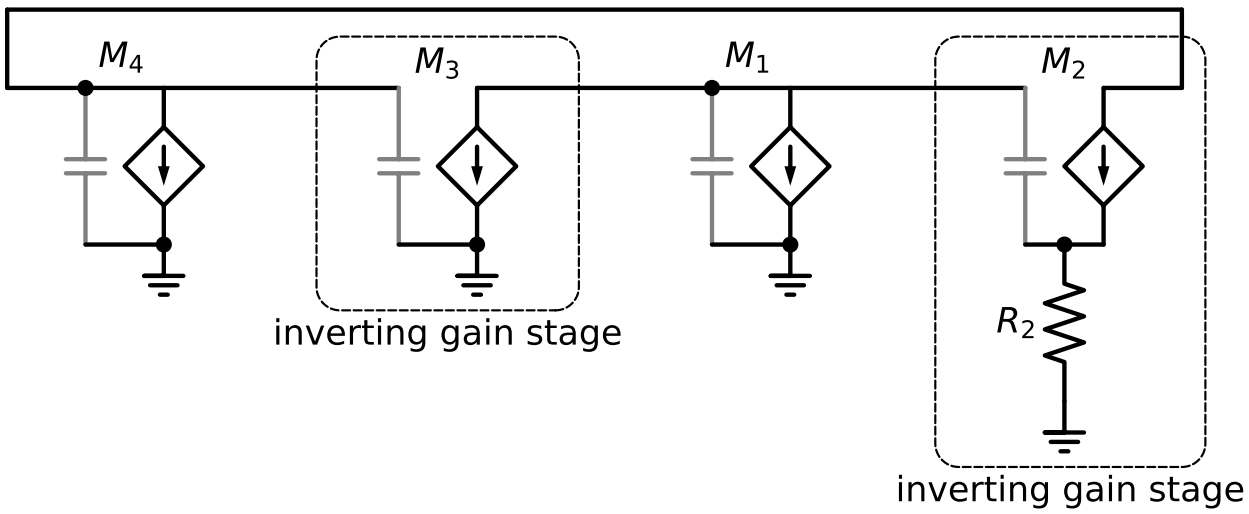


Figure 7.46.: Small signal equivalent model for the constant-g_m bias circuit

Example 7 - potential stability issue of the constant-g_m reference

Consider the design of the constant-g_m reference circuit in Figure 7.47:

Assume: $I(M_3) = I(M_4) = I_{REFN} = 100\mu A$

$V_{OV1} = 200mV$

$m = 2$

For the given circuit:

$$g_{m4} = g_{m3} = g_{m1} = \frac{2 \cdot 100\mu A}{200mV} = 1mS$$

Recalling that (assuming square law):

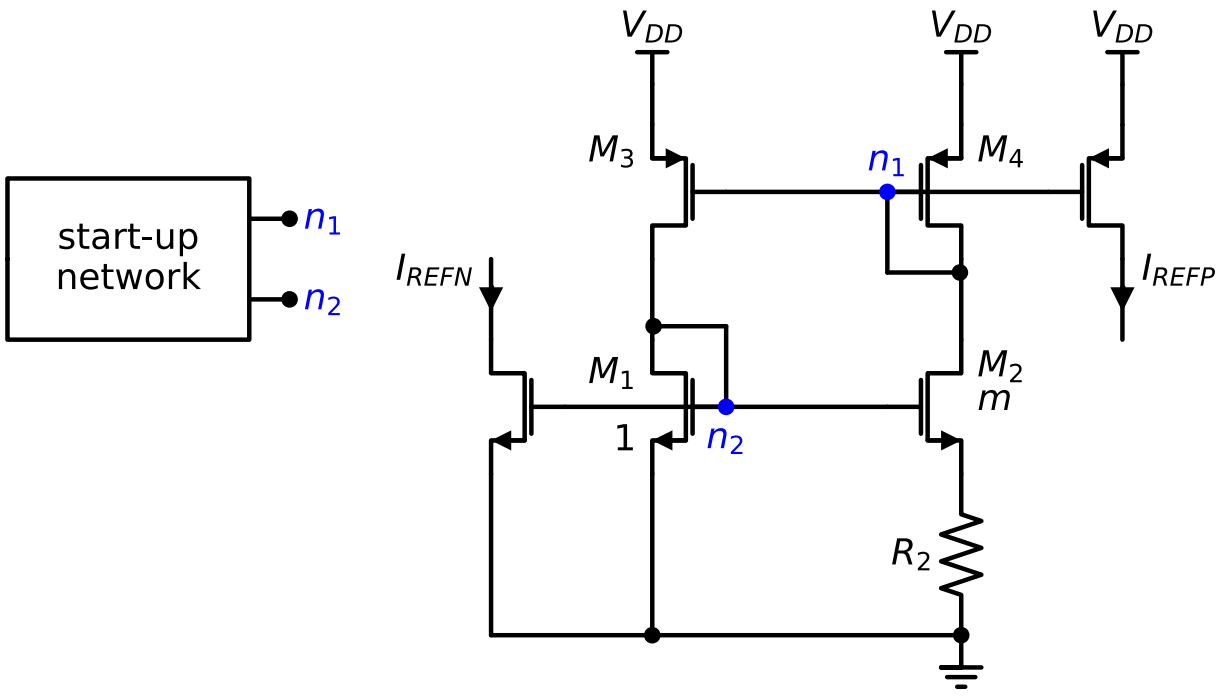


Figure 7.47.: Constant-g_m reference: design example

$$V_{OV2} = \frac{V_{OV1}}{\sqrt{m}}$$

$$I_{REF} = \frac{V_{OV1}\left(1 - \frac{1}{\sqrt{m}}\right)}{R_2}$$

for m=2:

$$V_{OV2} = 141mV$$

$$R_2 = 586\Omega$$

$$g_{m2} = \frac{2 \cdot 100\mu A}{141mV} = 1.41mS$$

$$\frac{g_{m2}}{1 + g_{m2}R_2} = 772\mu S$$

The magnitude of the return ratio is less than 1:

$$|T| = \frac{i_r}{i_t} = \left(\frac{g_{m2}}{1 + g_{m2}R_2}\right)\left(\frac{1}{g_{m4}}\right)\left(\frac{\cancel{g_{m3}}}{\cancel{g_{m1}}}\right) = 0.772(< 1)$$

Therefore, the circuit designed is stable, **but is possible to have problems**. Consider, Figure 7.48, if the parasitic capacitance C shorts R_2, the return ratio becomes larger than 1 and the circuit becomes unstable.

$$\frac{i_r}{i_t} = \left(\frac{1}{\cancel{1 + g_{m2}R_2}}\right)^{1}\left(\frac{g_{m2}}{g_{m4}}\right)\left(\frac{\cancel{g_{m3}}}{\cancel{g_{m1}}}\right) \rightarrow 1.41$$

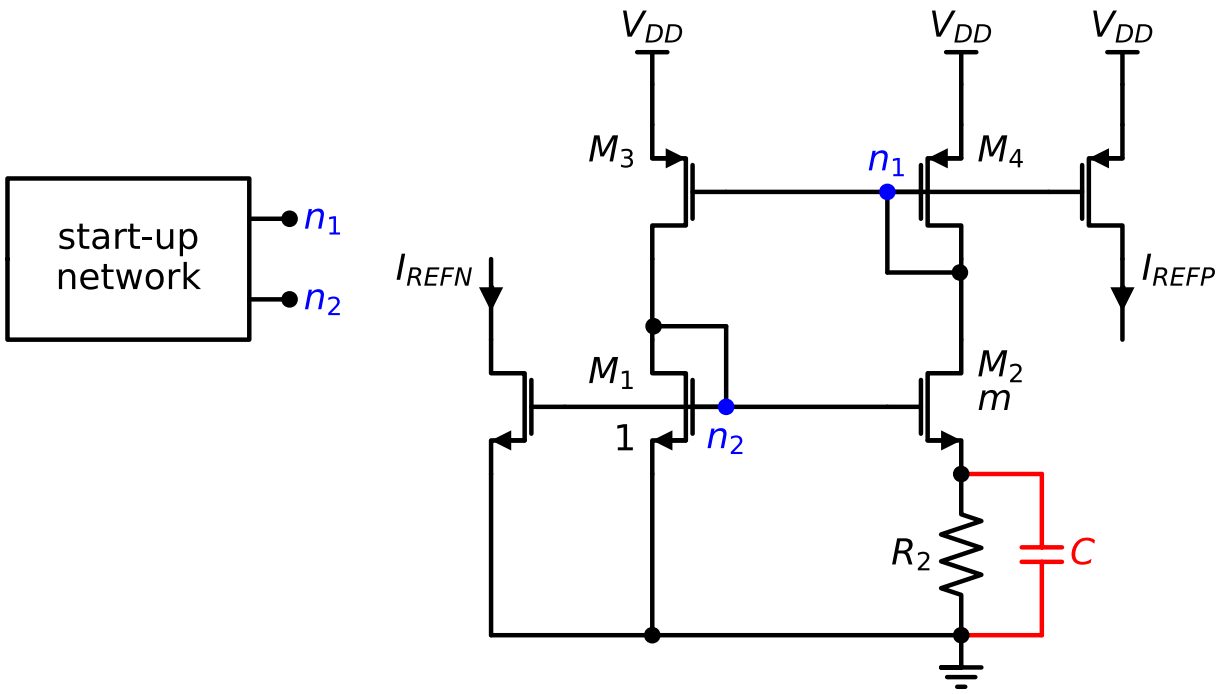


Figure 7.48.: Constant-g_m reference: possible problem

7.8.2. Closed-loop impedances using return ratio

Feedback can be used to modify the port impedances of a circuit.

Given a feedback circuit with a controlled source of value k (see Figure 7.49), the impedance at any port, including the input and output ports, can be computed (Gray et al. 2009b) using the following expression (a.k.a. **Blackman's formula**):

$$Z_{port} = Z_{port}(k=0) \cdot \left[\frac{1 + T(portshorted)}{1 + T(portopen)} \right]$$

- $Z_{port}(k=0)$ is the port impedance with the return ratio's gain element k set to 0 (either set $g_m = 0$ or $a_v = 0$)
- $T(portshorted)$ is the return ratio with the port under consideration *shorted*
- $T(portopen)$ is the return ratio with the port under consideration *open*

Often, one of the two return ratios in the formula is zero, in these cases feedback increases or decreases the impedance by a factor $(1 + T)$

In summary, Blackman's impedance formula:

- it applies to any feedback circuit, regardless of the type of feedback
- it is extremely useful and easy to use
- it is based on return ratio calculations

Example 8 - Input resistance of non-inverting op amp configuration

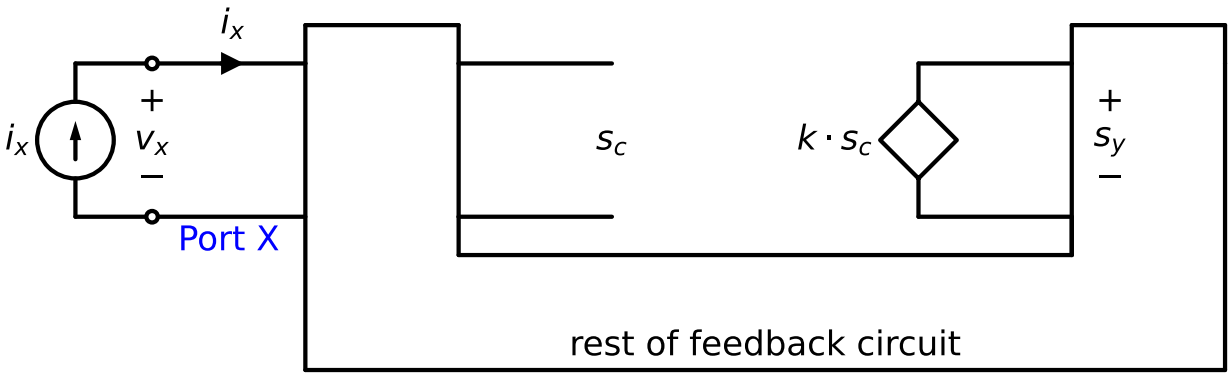


Figure 7.49.: Feedback circuit to derive Blackman's formula with respect to port x

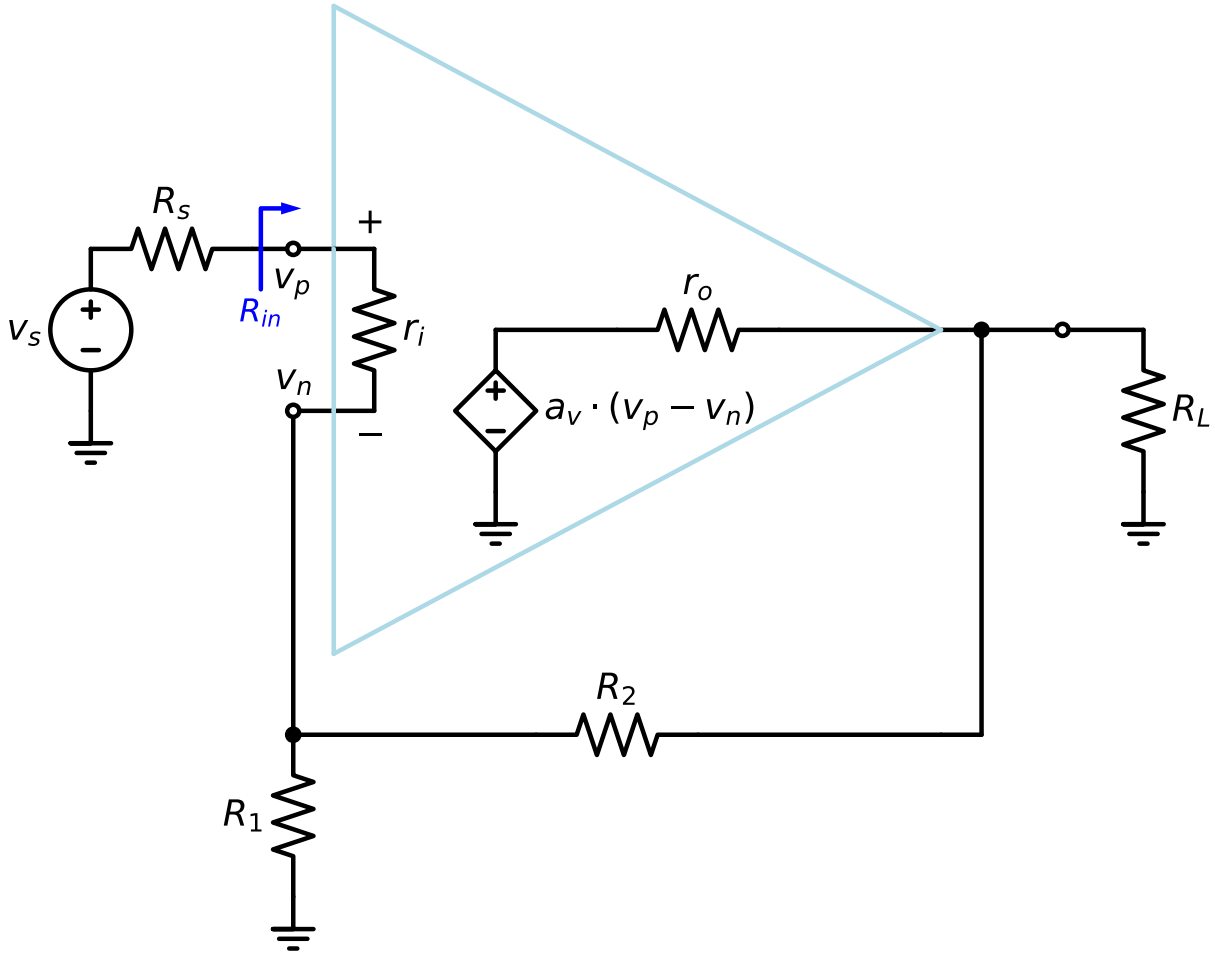


Figure 7.50.: Input resistance of non-inverting op amp circuit

$$R_{in} = R_{in0} \cdot \left[\frac{1 + T(portshorted)}{1 + T(portopen)} \right]$$

- To find R_{in0} set the op amp gain a_v to zero:

$$R_{in0} = r_i + R_1 || (R_2 + r_o || R_L) \approx r_i$$

- To compute $T(portshorted)$ short the node v_p to ground and find the return ratio:

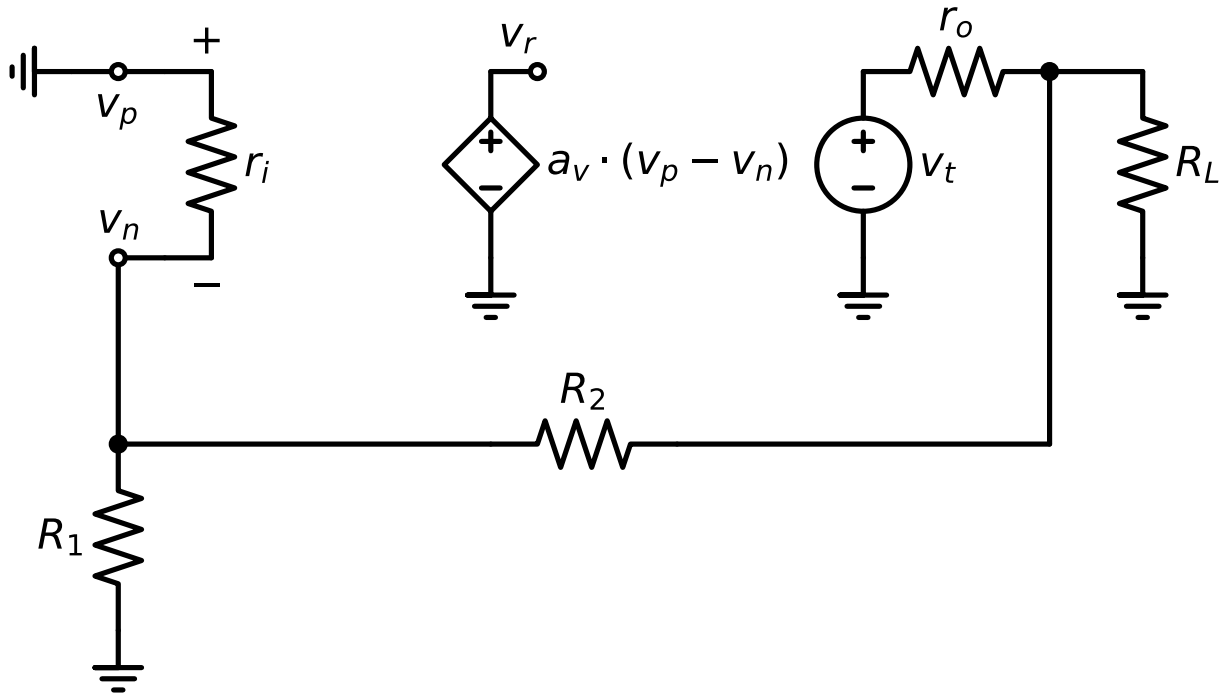


Figure 7.51.: T(with input port shorted) for non inverting configuration

$$T_{sc} = -\frac{v_r}{v_t}|_{sc} = a_v \cdot \frac{R_1}{r_o + R_L || (R_2 + r_i || R_1)} \approx a_v \cdot \frac{R_1}{R_1 + R_2}$$

- To compute $T(portopen)$ leave the node v_p floating and find the return ratio:

$$T_{open} = -\frac{v_r}{v_t}|_{open} = 0$$

Feedback increases the input impedance of the circuit significantly.

$$R_{in} \approx r_i \cdot \frac{1 + a_v \cdot \frac{R_1}{R_1 + R_2}}{1 + 0} = r_i \cdot \left(1 + a_v \cdot \frac{R_1}{R_1 + R_2} \right)$$

For $a_v = 10^6$, $R_1 = 20k\Omega$, $R_2 = 80k\Omega$:

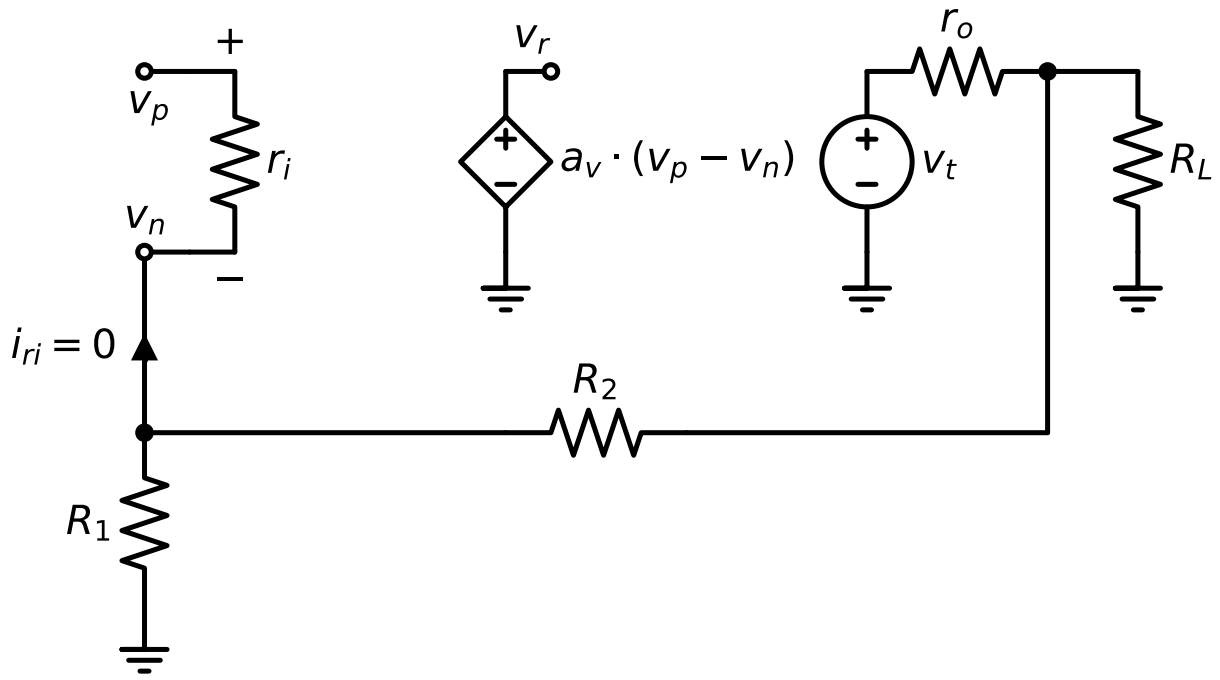


Figure 7.52.: T(with input port open) for non inverting configuration

$$R_{in} \approx r_i \cdot 200000$$

Example 9 - Output resistance of non-inverting op amp configuration

$$R_{out} = R_{out0} \cdot \left[\frac{1 + T(portshorted)}{1 + T(portopen)}\right]$$

- To find R_{out0} set the op amp gain a_v to zero:

$$R_{out0} = r_o||[R_2 + R_1||(r_i + R_s)] \approx r_o$$

- To compute $T(portshorted)$ short the node v_{out} to ground and find the return ratio:

$$T_{sc} = -\frac{v_r}{v_t}|_{sc} = 0$$

- To compute $T(portopen)$ leave the node v_{out} floating and find the return ratio:

$$v_n = \frac{v_t}{r_o + R_2 + [R_1||(r_i + R_s)]} \cdot R_1||(r_i + R_s) \approx \frac{v_t}{R_2 + R_1} \cdot R_1$$

$$v_p = v_n \frac{R_s}{r_i + R_s} \approx 0$$

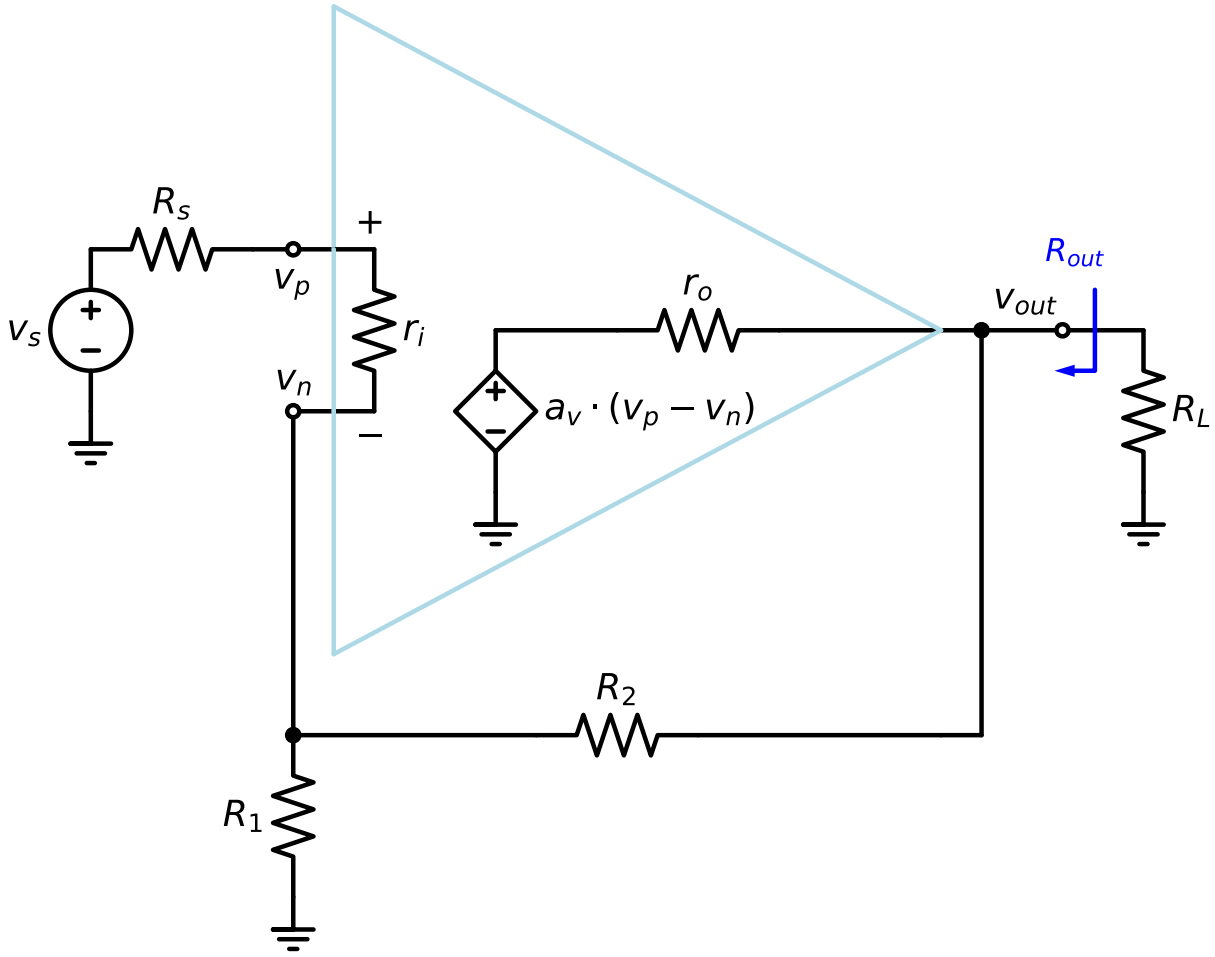


Figure 7.53.: Output resistance of non-inverting op amp circuit

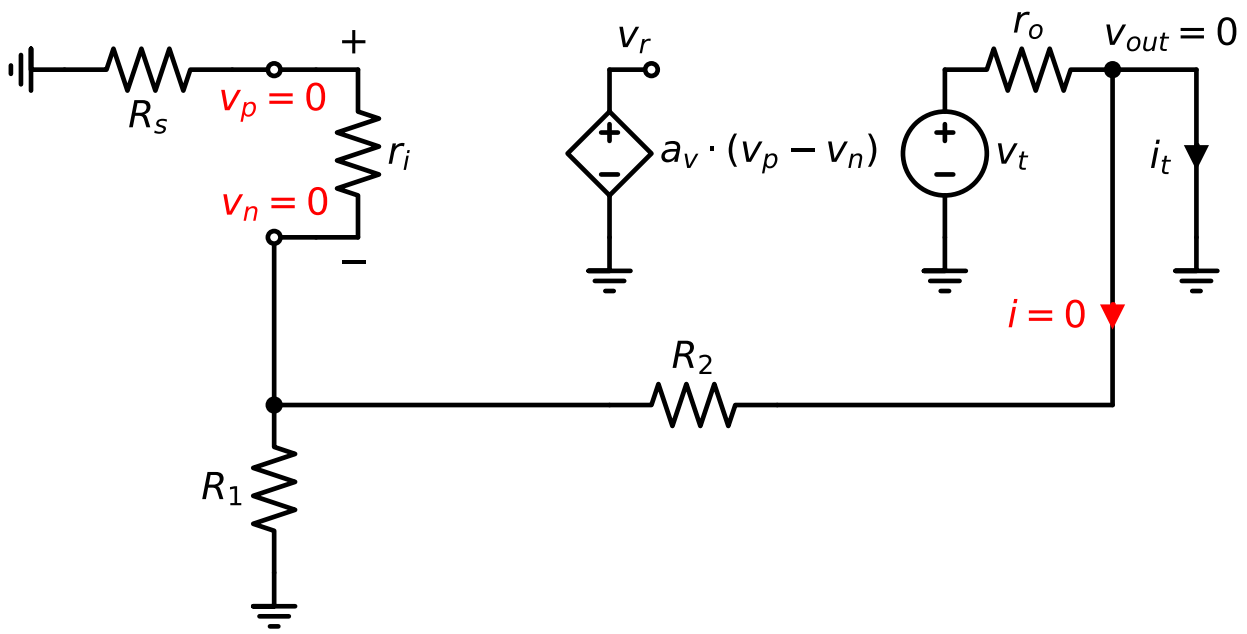


Figure 7.54.: T(with output port shorted) for non-inverting configuration

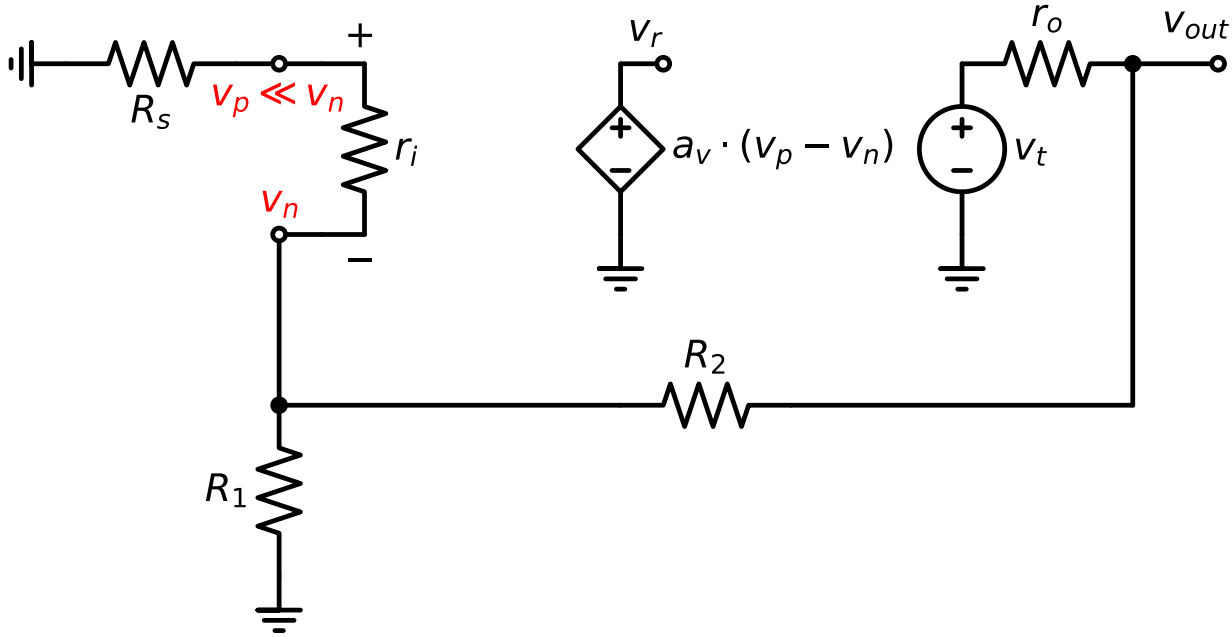


Figure 7.55.: T(with output port open) for non-inverting configuration

$$T_{open} = -\frac{v_r}{v_t}|_{open} = -\frac{a_v(v_p - v_n)}{v_t} \approx a_v \cdot \frac{R_1}{R_1 + R_2}$$

Feedback decreases the output impedance of the circuit significantly.

$$R_{out} \approx r_o \cdot \frac{1+0}{1 + a_v \frac{R_1}{R_1 + R_2}} = \frac{r_o}{1 + a_v \frac{R_1}{R_1 + R_2}}$$

For $a_v = 10^6$, $R_1 = 20k\Omega$, $R_2 = 80k\Omega$:

$$R_{out} \approx \frac{r_o}{200000}$$

Example 10 - Output resistance of bootstrapped source follower

Consider the bootstrapped source follower stage of Figure 7.56 and its equivalent small signal circuit to compute the return ratio with the output port shorted (Figure 7.57) and with the output port left floating (Figure 7.58)

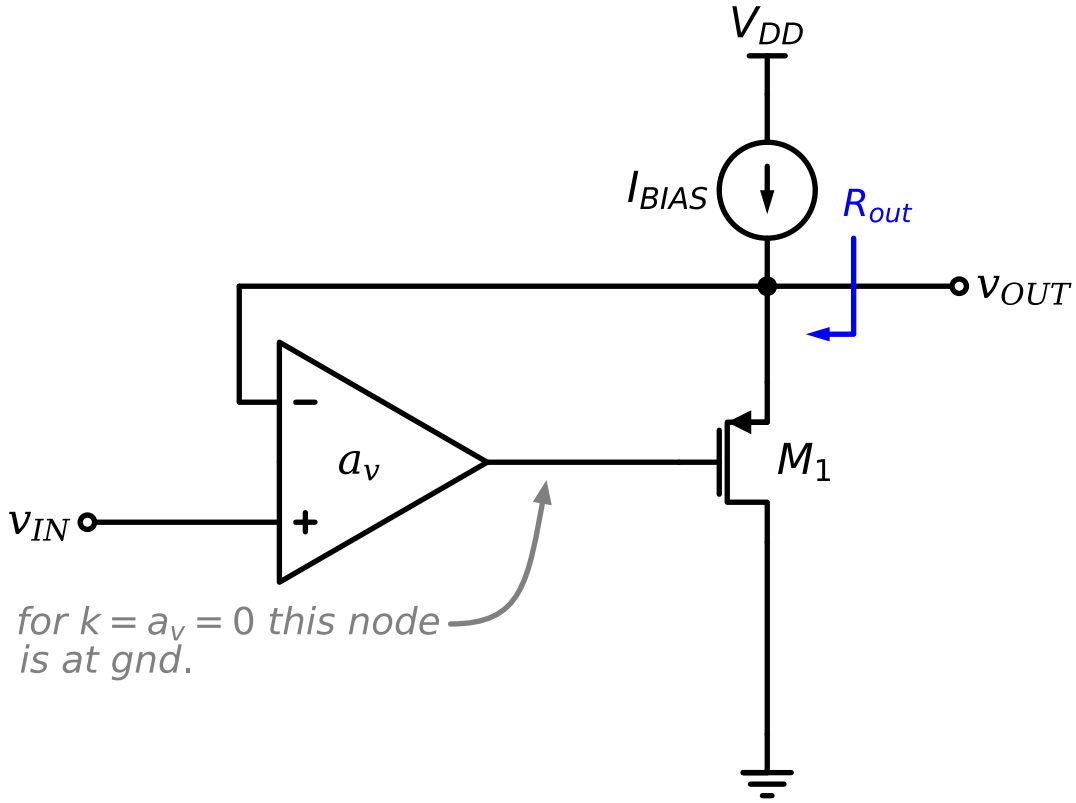


Figure 7.56.: A bootstrapped source follower stage

The output resistance for $k = a_v = 0$ is:

$$R_{out0} \equiv R_{out}(k=0) \equiv R_{out}(a_v=0) = \frac{1}{g_m}$$

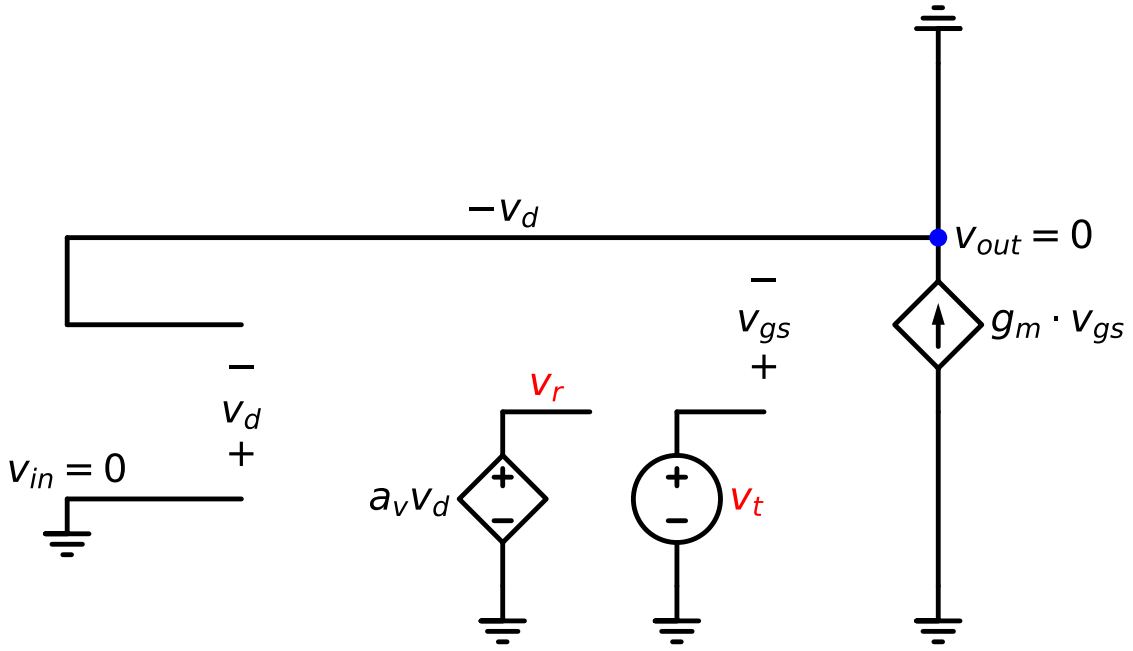


Figure 7.57.: Finding T(with output port shorted), $-v_d = v_{out} = 0$

Output port shorted implies:

$T_{sc} = -\frac{v_r}{v_t}|_{sc} = 0$

Output port open means:

$g_m \cdot v_{gs} = 0 \rightarrow v_{gs} = 0 \rightarrow -v_d = v_t \rightarrow v_r = -a_v \cdot v_t$

$T_{open} = -\frac{v_r}{v_t}|_{open} = a_v$

Therefore, the closed-loop output resistance is:

$$R_{out} = R_{out0} \cdot \frac{1+T_{sc}}{1+T_{open}} = \frac{1}{g_m} \cdot \left(\frac{1}{1+a_v}\right)$$

Example 11 - Output resistance of super source follower

The super-source follower uses feedback to reduce its output impedance.

The small signal model of the super-source follower is shown in Figure 7.60.

- To find R_{out0} we must set $v_{in} = 0$ and $k = 0$. In this circuit we can use, either g_{m1} or g_{m2} as k. We will use, $k = g_{m2}$.

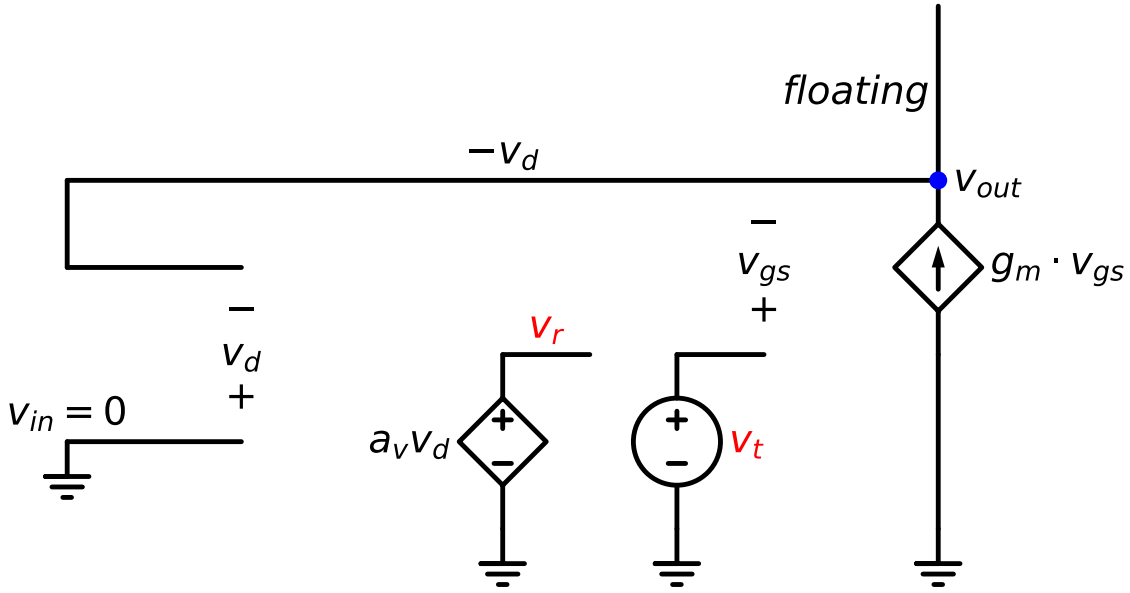


Figure 7.58.: Finding T(with output port open), $v_{out} = -v_d = -v_{gs} + v_t$

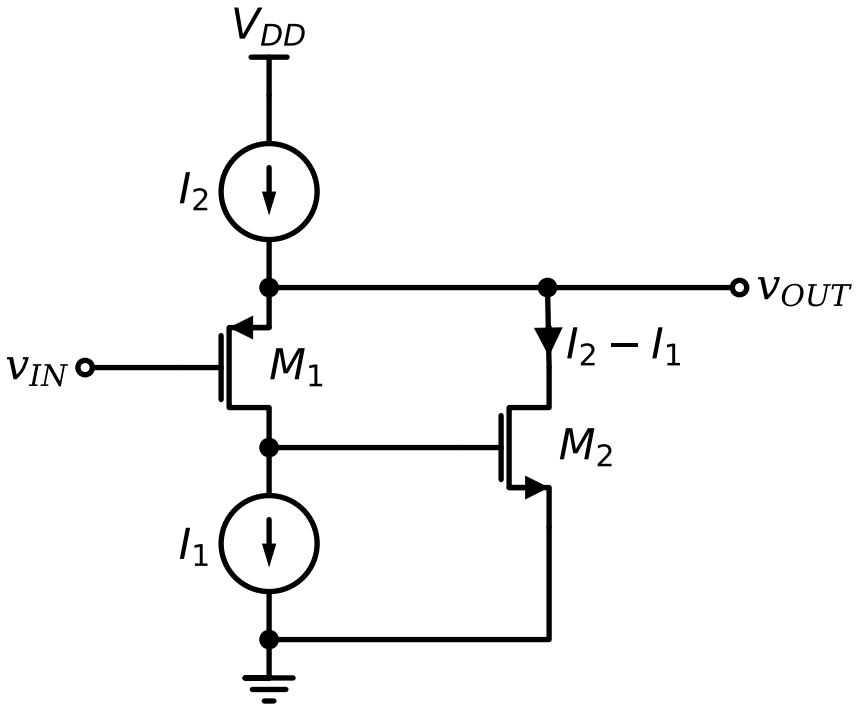


Figure 7.59.: Super-source follower stage

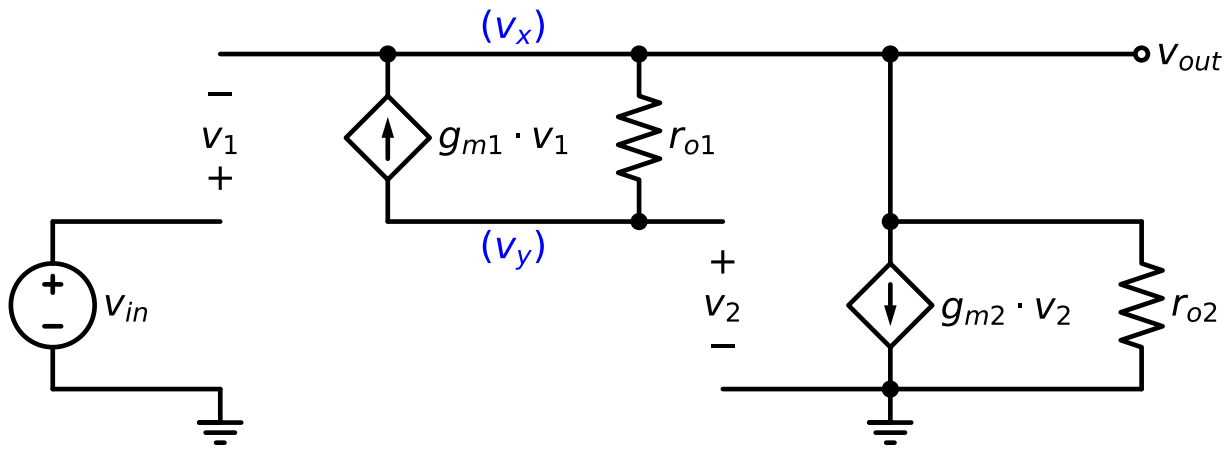


Figure 7.60.: Small-signal model of the super-source follower

The current in the g_{m1} generator flows only in r_{o1}, so M_1 has no effect on the output resistance when $g_{m2} = 0$

$$R_{out0} \equiv R_{out}(g_{m2} = 0) = r_{o2}$$

- The return ratio for g_{m2} with the output port shorted is:

$$T_{sc} = -\frac{i_r}{i_t}|_{sc} = 0$$

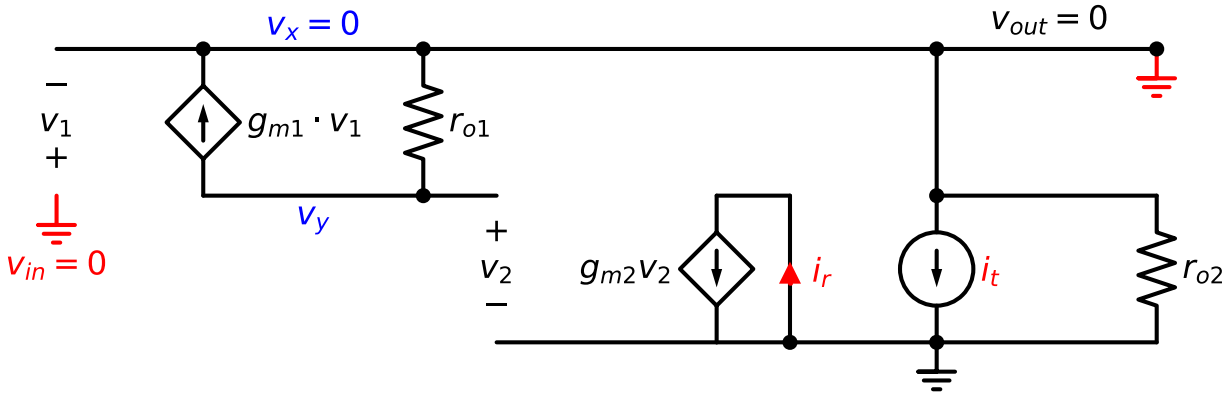


Figure 7.61.: Return Ratio for g_{m2} with the output port shorted

Shorting the output port forces $v_{out} = 0$

$v_{out} \equiv v_x \Leftrightarrow v_x = 0$

$v_x = -v_1 \leftrightarrow v_1 = 0$

$g_{m1} \cdot v_1 = 0 \Leftrightarrow v_y = 0$

$v_2 = x_x - v_y = 0 \Leftarrow g_{m2}v_2 \equiv i_r = 0$

- The return ratio for g_{m2} with the output port open is:

$$T_{open} = -\frac{i_r}{i_t}|_{open} = g_{m2} r_{o2} \left(1 + g_{m1} r_{o1}\right)$$

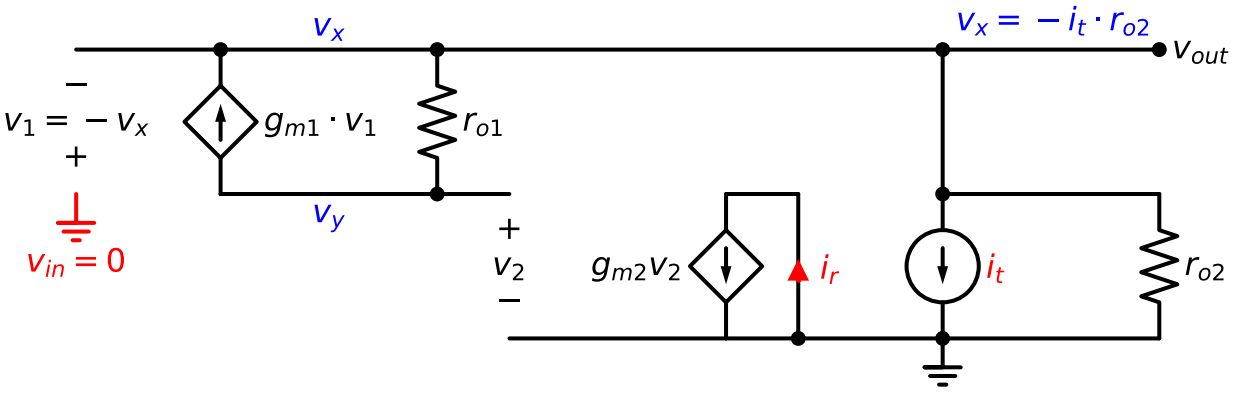


Figure 7.62.: Return Ratio for g_{m2} with the output port open

Opening the output port gives:

$v_x = -i_t \cdot r_{o2} \Leftrightarrow i_t = -\frac{v_x}{r_{o2}}$

$v_x = -v_1$

$v_y - v_x = -g_{m1} \cdot v_1 \cdot r_{o1} \Leftrightarrow v_y = v_x + g_{m1} \cdot r_{o1} \cdot v_x \Leftrightarrow v_y = v_x \left(1 + g_{m1} r_{o1}\right)$

$v_2 = v_y$

$i_r \equiv g_{m2} v_2 = g_{m2} \left(1 + g_{m1} \cdot r_{o1}\right) \cdot v_x$

$T_{open} = -\frac{i_r}{i_t} = g_{m2} r_{o2} \left(1 + g_{m1} r_{o1}\right)$

- The closed-loop output resistance is:

$$R_{out} = R_{out0} \cdot \frac{1 + T_{sc}}{1 + T_{open}} = \frac{r_{o2}}{1 + g_{m2} r_{o2} \left(1 + g_{m1} r_{o1}\right)}$$

Assuming $g_m r_o \gg 1$

$$R_{out} \approx \frac{1}{g_{m1} g_{m2} r_{o1}}$$

Example 12 - Input and output resistance of shunt-shunt stage (TIA)

Closed-loop input resistance

$R_{in}(g_m = 0) = R_F + r_o$

$T(inputshorted) = 0$

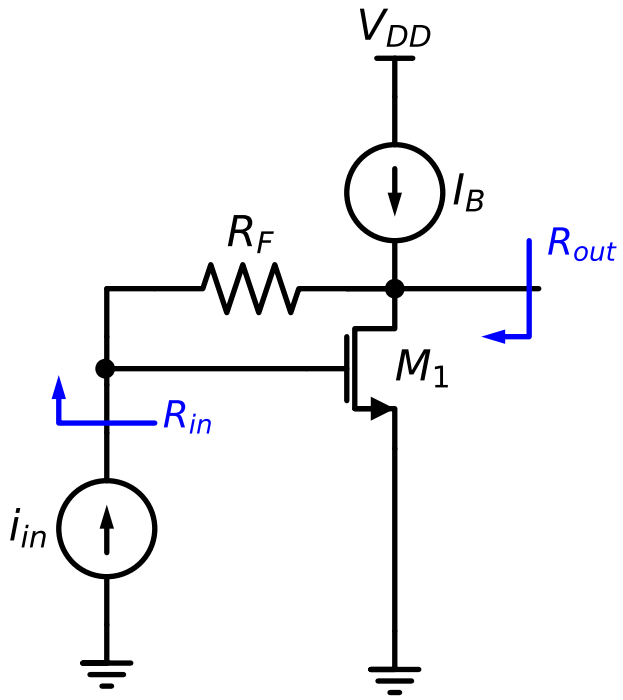


Figure 7.63.: Shunt-shunt stage (Transimpedance Amplifier)

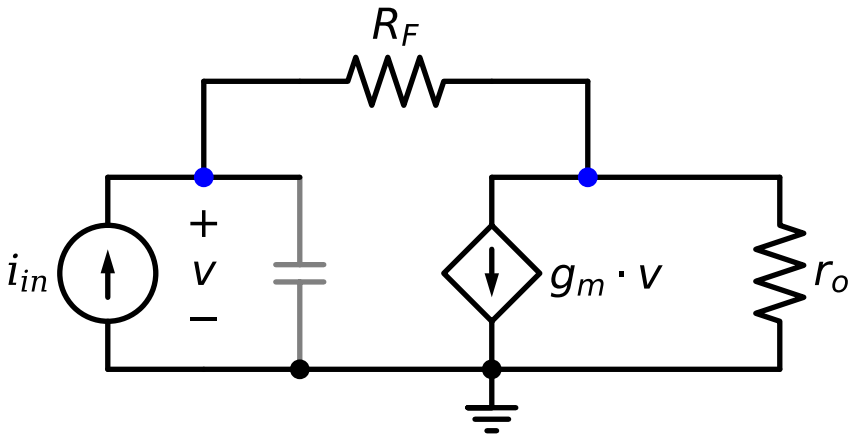


Figure 7.64.: Small signal model of the shunt-shunt stage

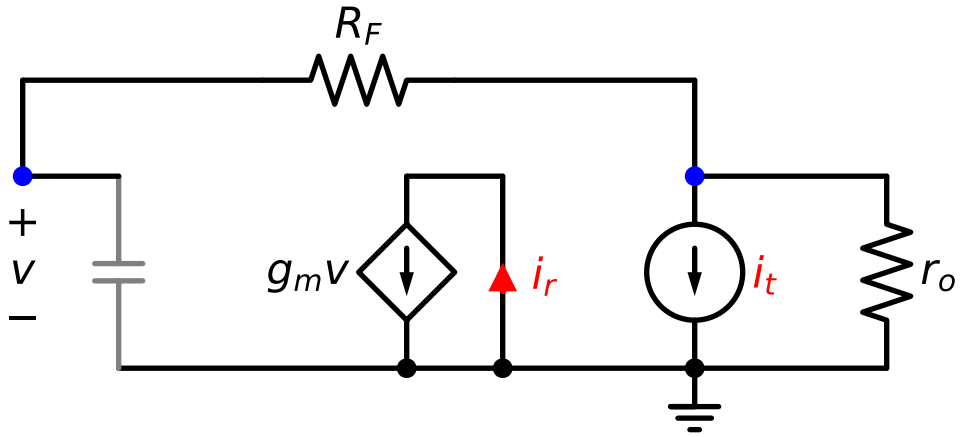


Figure 7.65.: Small signal model to find the return ratio for g_m

$T(inputopen) = g_m r_o$

Therefore, the closed-loop input resistance is:

$$R_{in} = (R_F + r_o) \cdot \frac{1}{1 + g_m r_o} \approx \frac{1}{g_m} \left(1 + \frac{R_F}{r_o}\right)$$

Closed-loop output resistance

$R_{out}(g_m = 0) = r_o$

$T(inputshorted) = 0$

$T(inputopen) = g_m r_o$

Therefore, the closed-loop output resistance is:

$$R_{out} = r_o \cdot \frac{1}{1 + g_m r_o} \approx \frac{1}{g_m}$$

Example 13 - Output resistance of active cascode

The "active cascode" circuit is also referred as "regulated cascode" or "gain-boosting" technique.

- To find R_{out0} we set $v_{in} = 0$ and $k = a_v = 0$

With $a_v = 0$ the gate of M_2 gets grounded:

$$R_{out0} \equiv R_{out}(a_v = 0) = r_{o1} + r_{o2}\,(1 + g'_{m2}) \approx r_{o1} \cdot r_{o2} \cdot g'_{m2}$$
$$with g'_{m2} = g_{m2} + g_{m2b}$$

- To find the return ratio for a_v with the output port open, we set $v_{in} = 0$ and leave the output node floating

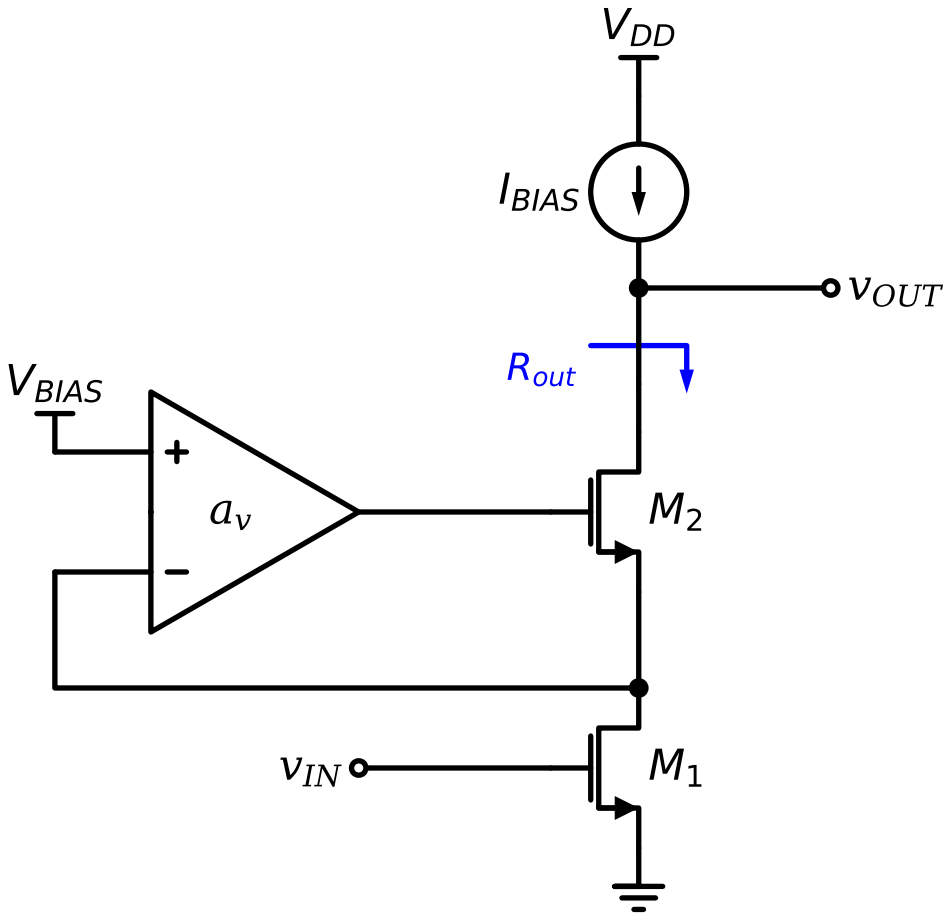


Figure 7.66.: Active cascode gain stage

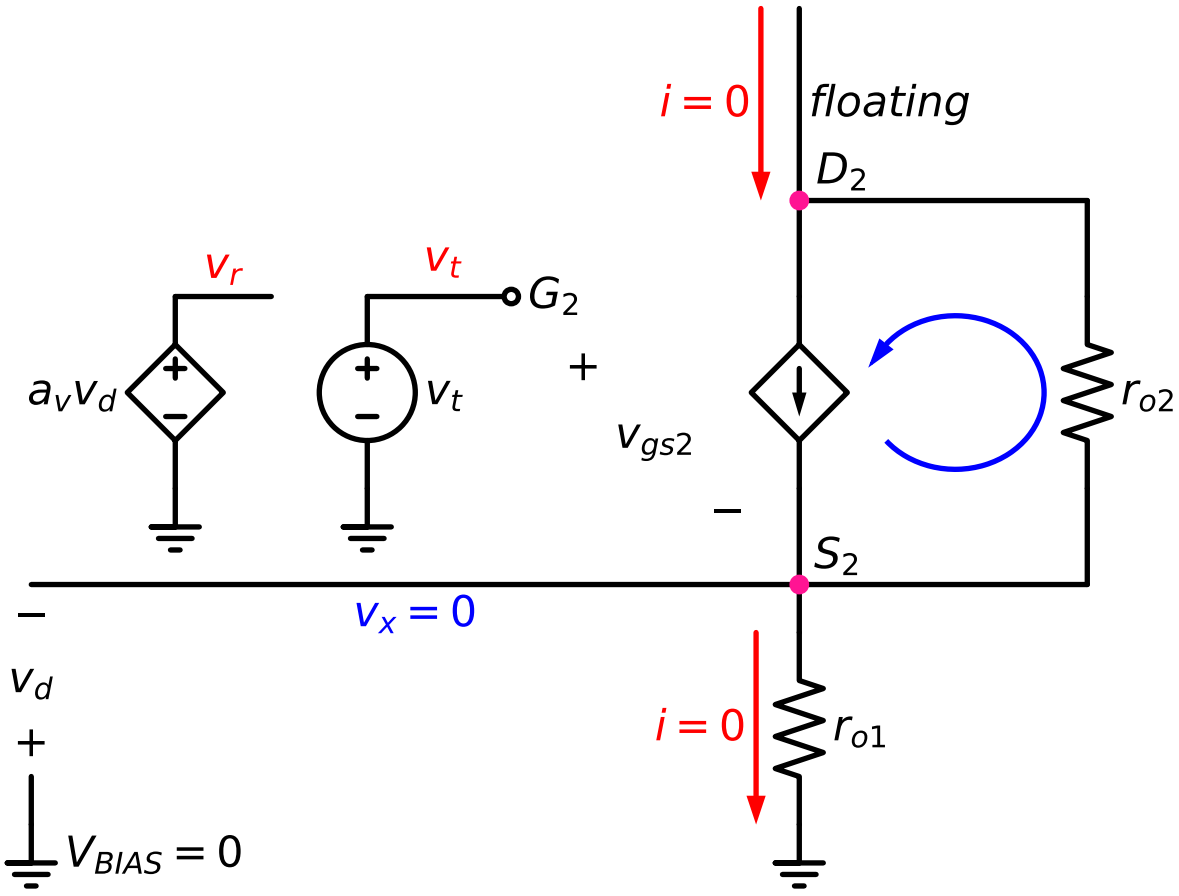


Figure 7.67.: Return ratio for a_v with the output port open

The current of the g_m-generators ($g_{m2}v_{gs2}$ and $g_{mb2}v_{bs2}$) flows only in r_{o1}, changing the drain voltage of M_2, but keeping the source voltage of M_2 to $v_x = 0 \Rightarrow v_d = -v_x = 0$. Therefore, $v_r = a_v \cdot v_d = 0$

$$T_{open} = -\frac{v_r}{v_t}|_{open} = 0$$

- To find the return ratio for a_v with the output port shorted, we set $v_{in} = 0$ and ground the output node (the output node is the drain of M_2)

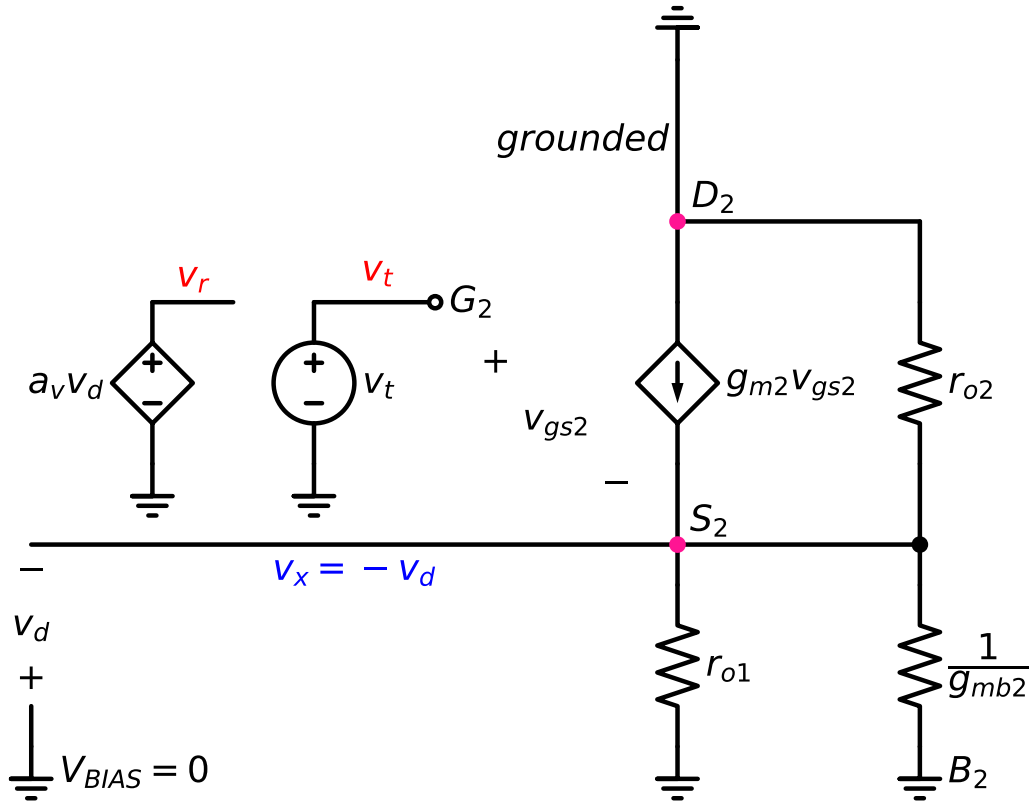


Figure 7.68.: Return ratio for a_v with the output port shorted

With M_2 drain to ground (CD), $v_x = -v_d$ is the output of a CD with v_t as input.

$v_t = v_{gs2} - v_d$

Neglecting the r_o terms w.r.t. the $1/g_{mb2}$ term:

$$-v_d \approx \frac{g_{m2}v_{gs2}}{g_{mb2}} \Rightarrow v_t \approx v_{gs2}\left(1 + \frac{g_{m2}}{g_{mb2}}\right) \Rightarrow -\frac{v_d}{v_t} \approx \frac{g_{m2}}{g_{m2} + g_{mb2}} \approx 1$$

so the return ratio is:

$$T_{sc} = -\frac{v_r}{v_t}|_{sc} \approx a_v \cdot \frac{g_{m2}}{g_{m2} + g_{mb2}} \approx a_v$$

- The closed-loop output resistance is:

$$R_{out} = R_{out0} \cdot \frac{1 + T_{sc}}{1 + T_{open}} \approx r_{o1} \cdot r_{o2} \cdot g'_{m2} \cdot a_v$$

Example 14 - Using Blackman's formula for ZVTC calculations

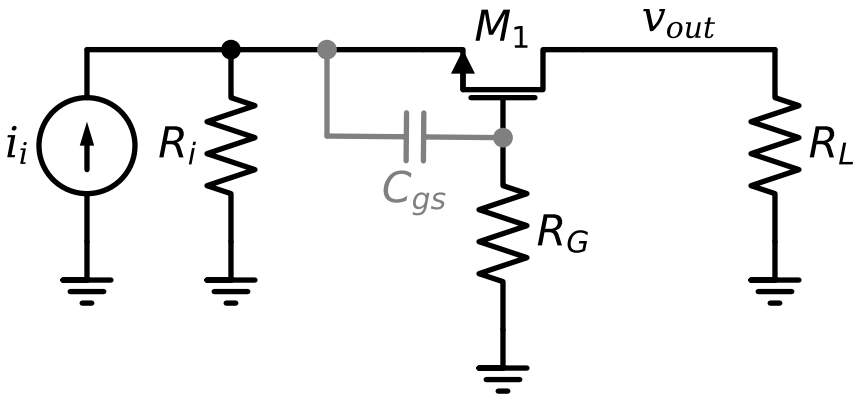


Figure 7.69.: ZVTC calculations

Using first principles (KVL and KCL) we can derive that:

$$\tau_{Cgs} = C_{gs} \cdot \frac{R_i + R_G}{1 + g_m R_i}$$

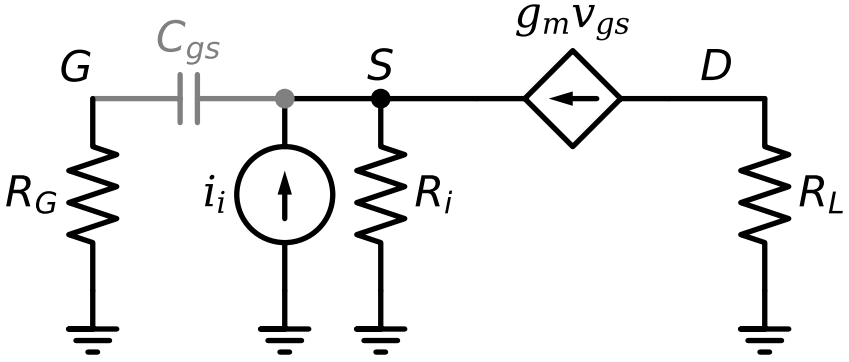


Figure 7.70.: Small signal model for ZVTC calculations, assuming r_o can be ignored

Let's look at the problem, using return ratio.

The port (gate-source) impedance with the return ratio's gain element g_m set to 0 is:

$R_{gs}(g_m = 0) = R_i + R_G$

The return ratio for the gain element g_m with port under consideration shorted is:

$T_{sc} = -\frac{i_r}{i_t}|_{sc} = 0$

Shorting the port forces $v_{gs} = 0$ so it "kills" the g_m generator.

The return ratio for the gain element g_m with port under consideration open is:

$T_{open} = -\frac{i_r}{i_t}|_{open} = g_m \cdot R_i$

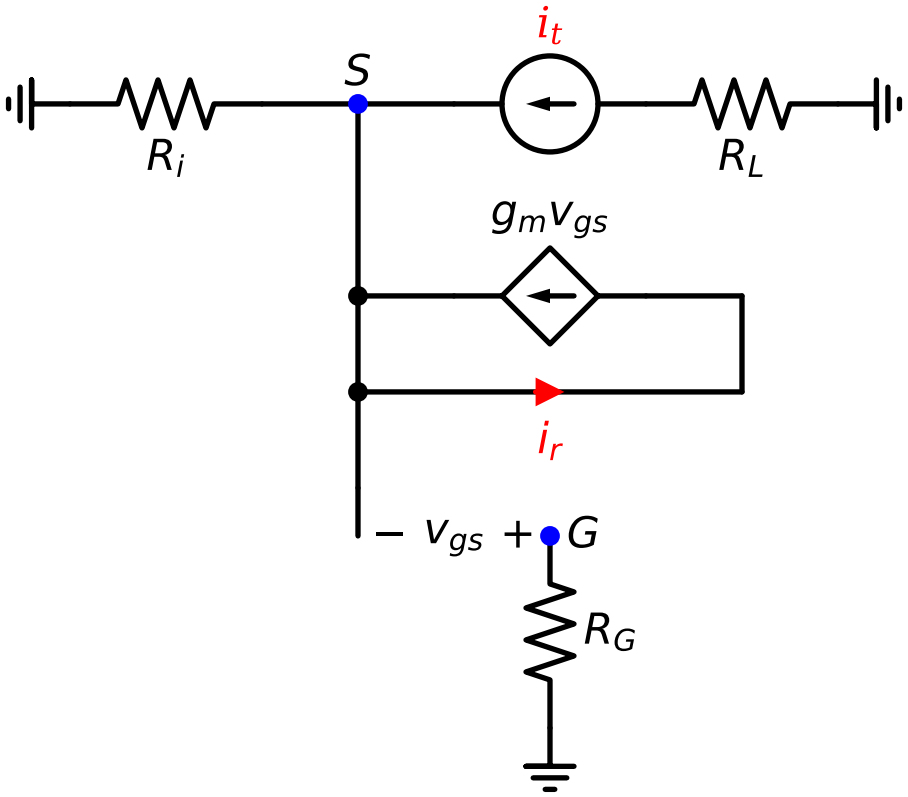


Figure 7.71.: return ratio with port G-S open

$v_s = i_t \cdot R_i$

$v_g = 0 \Rightarrow v_{gs} = -i_t \cdot R_i$

$i_r = g_m v_{gs} = -g_m R_i i_t$

The closed-loop port impedance between G and S is:

$$R_{gs} = R_{gs}(g_m = 0) \cdot \frac{1 + T_{sc}}{1 + T_{open}} = \frac{R_i + R_G}{1 + g_m R_i}$$

Example 15 - Voltage gain and output resistance of common source with degeneration using return ratio

voltage gain (assuming r_o negligible)

$$A_v = \frac{v_{out}}{v_{in}} = A_\infty \cdot \frac{T}{1+T} + \frac{d}{1+T}$$

A_∞ is the gain of the circuit when the gain element k of the controlled source tends to ∞ (note that if $k = g_m = \infty$, for the current of the dependent source to be finite, it must be $v_{gs} = 0$).

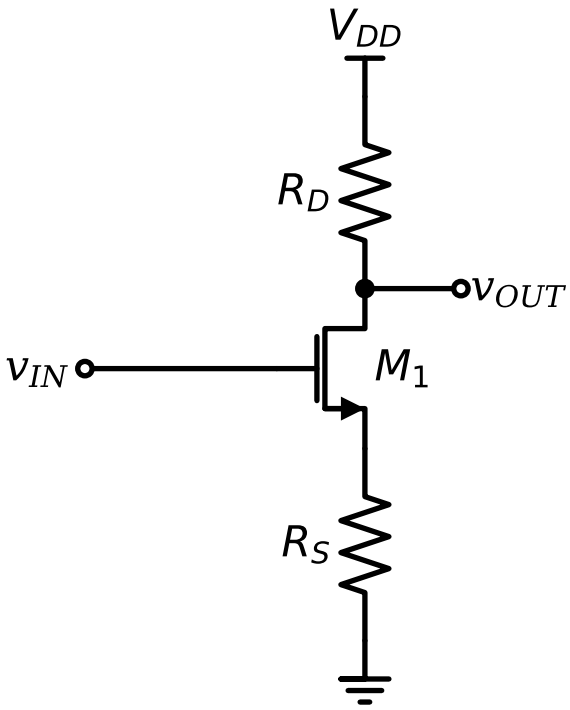


Figure 7.72.: Common source with degeneration

$$A_\infty = \frac{v_{out}}{v_{in}}|_{g_m \to \infty} = -\frac{R_D}{R_S}$$

d is the gain of the circuit when the gain element k of the controlled source equals 0.

$$d = \frac{v_{out}}{v_{in}}|_{g_m = 0} = 0$$

T is the return ratio of the dependent source.

To compute the return ratio instead of separating the dependent source from the rest of the circuit and replacing it with an independent source of the same kind as always done so far, we will follow an equivalent procedure consisting of adding an independent source s_x. In practice, this procedure is usually more easily performed, because it does not require tearing the original circuit apart. See (Middlebrook 2006) and (Hajimiri 2023).

The introduction of s_x is completely equivalent to using an independent source s_t in place of the separated dependent source and monitoring the signal s_r that returns to the separated dependent source.

$$s_t = s_x + s_r$$

$$v_{gs} = -i_t R_S$$

$$i_r = g_m v_{gs} = -g_m R_S i_t$$

$$T = -\frac{i_r}{i_t} = g_m R_S$$

Therefore, as expected:

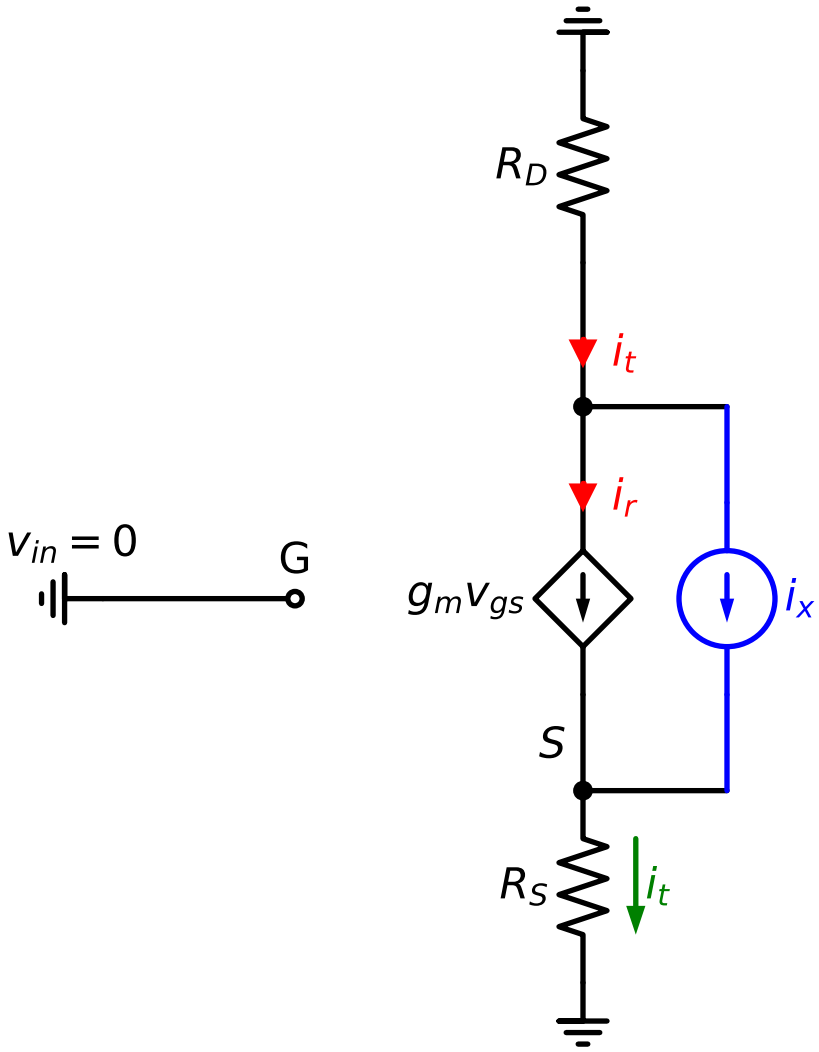


Figure 7.73.: Small signal model to find the return ratio of g_m

$$A_v = \frac{v_{out}}{v_{in}} = A_\infty \cdot \frac{T}{1+T} + \frac{d}{1+T} = \frac{-g_m R_D}{1+g_m R_S}$$

Output resistance

To find the output resistance we use Blackman's formula.

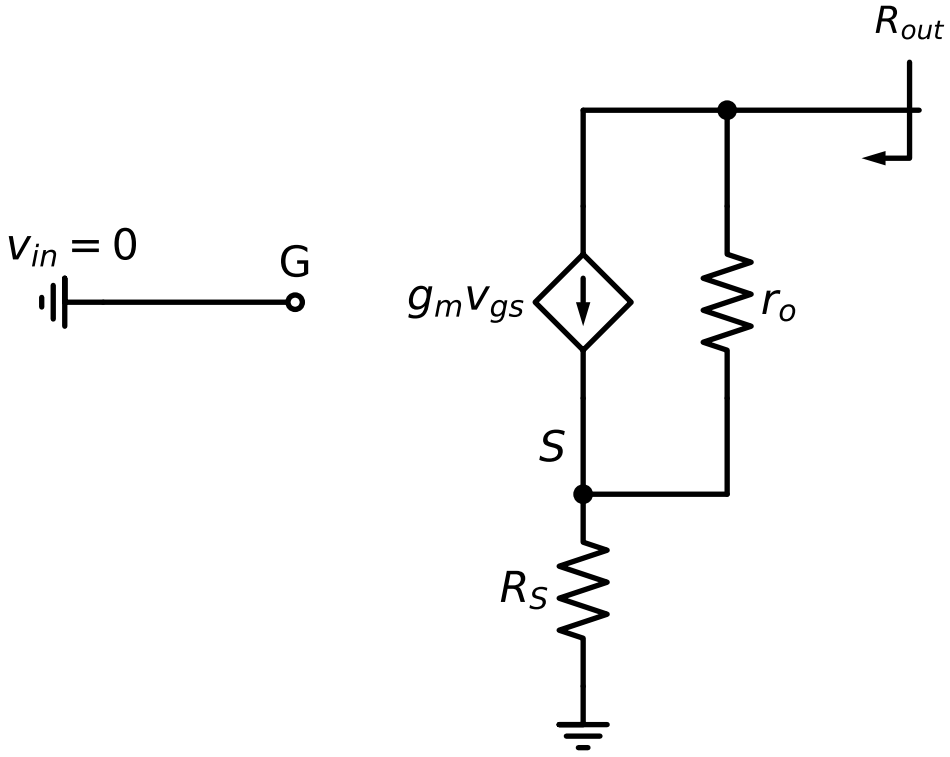


Figure 7.74.: Small signal model to find the output resistance

The output resistance with $g_m = 0$ is:

$R_{out}(g_m = 0) = R_S = r_o$

With the output port open:

$v_{gs} = -i_t R_S = 0 \Rightarrow i_r = g_m v_{gs} = 0$

So, the return ratio with the output port open is:

$T_{open} = -\frac{i_r}{i_t}|_{open} = 0$

With the output port shorted:

$v_{gs} = -(R_S||r_o) \cdot i_t \Rightarrow i_r = g_m v_{gs} = -g_m \cdot (R_S||r_o) \cdot i_t$

So, the return ratio with the output port shorted is:

$T_{sc} = -\frac{i_r}{i_t}|_{sc} = g_m(R_S||r_o)$

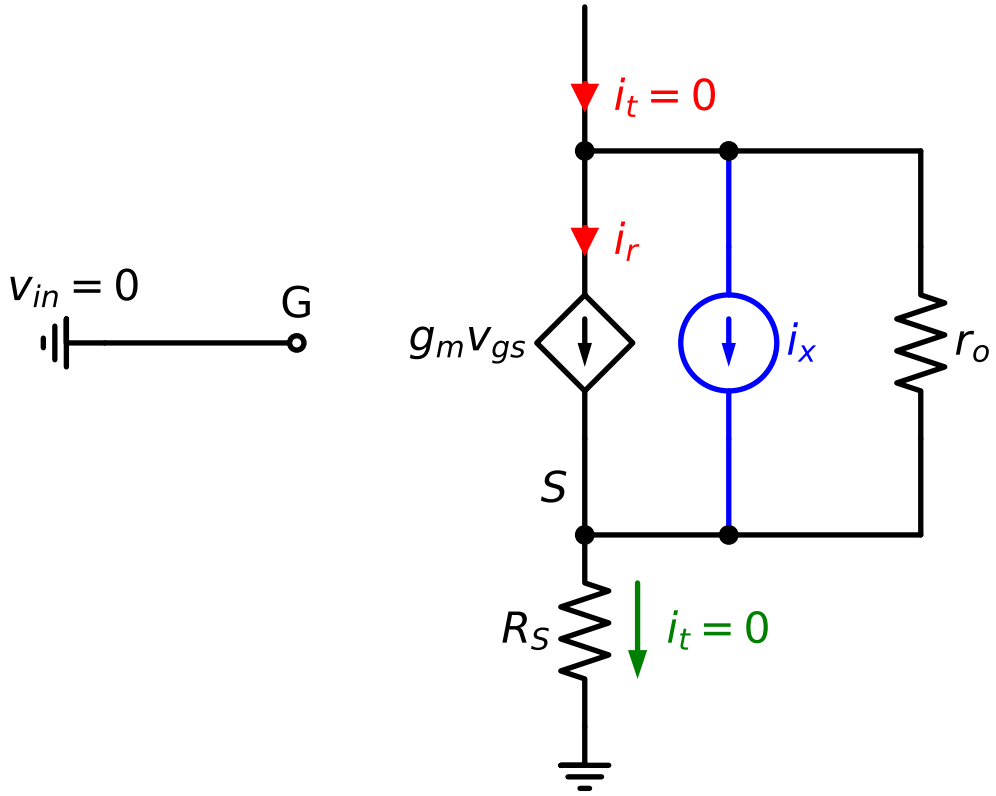


Figure 7.75.: return ratio with output port open

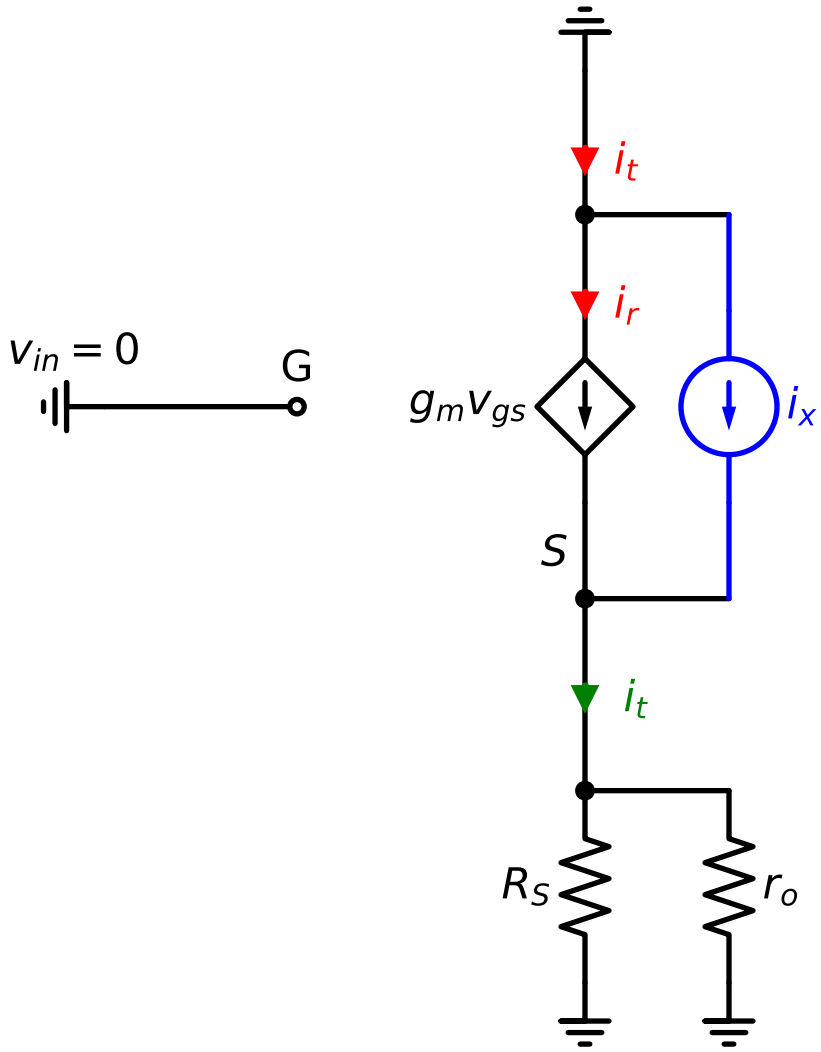


Figure 7.76.: return ratio with output port shorted

And the output resistance is:

$$R_{out} = R_{out}(g_m = 0) \cdot \frac{1 + T_{sc}}{1 + T_{open}} = (R_S + r_o)\left(1 + g_m \frac{R_S r_o}{R_S + r_o}\right) = \\ = R_S + r_o(1 + g_m R_S)$$

7.8.3. Summary - Return Ratio Analysis

- Return-ratio analysis is an alternative approach to two-port feedback analysis.
- Provided the return ratio T is computed for a dependnt source that breaks the feedback loop globally, the return ratio is a measure of loop gain. In general, the return ratio of a controlled source is a measure of how much of the signal generated by that controlled source is returned due to feedback, it is not necessarily the loop gain (Hurst 1992).
- For negative feedback circuits T > 0
- In an ideal feedback system $T \to \infty$ and the closed loop gain is A_∞
- Blackman's formula gives the closed-loop impedance in terms of two return ratios. The formula is the same for any type of feedback circuit, and it applies to any port (not only the input and output ports)
- Return ratio analysis is often simpler than two-port analayis
- Return ratio analysis uses equations that are independent of the type of feedback.
- Two-port analysis uses four-feedback configurations (series-series, series-shunt, shunt-series, and shunt-shunt), therefore, the type of feedback must be correctly identified before starting the analysis
- In practice, all physical networks are multiloop feedback structures. Unwanted local return loops exists around individual transistor through the parasitic capacitances (Tian et al. 2001) (Behmanesh and Andreani 2023). Dealing with multiple feedback mechanisms makes identifying an appropriate controlled source to use for the return ratio analysis more difficult. When dealing with multiloop feedback structures, the controlled source used to compute the return ratio, must be one that breaks the loop globally (Hurst and Lewis 1995).
- One aspect that is often cause of confusion, about return-ratio analysis and two-port analysisis is that the loop gain L computed through two-port analysis and the loop gain T computed using the return ratio, can be different. This is because they are intrinsically two different measures of transmission, however both methods lead to correct closed-loop expressions and both can be used to check stability (Chiu 2012).

Part III.

Dare to Run — Knowledge Base for State-of-the-Art Circuits

8. Introduction

TBD.

Part IV.

References

References

Behmanesh, Baktash, and Pietro Andreani. 2023. "On the Calculation and Simulation of Loop Gain in Feedback Circuits." *IEEE Transactions on Circuits and Systems II: Express Briefs* 70 (11): 4033–37. https://doi.org/10.1109/TCSII.2023.3284669.

Chiu, Yun. 2012. "Demystifying Bilateral Feedback Analysis." In *2012 IEEE 11th International Conference on Solid-State and Integrated Circuit Technology*, 1–4. https://doi.org/10.1109/ICSICT.2012.6467588.

Franco, Sergio. 2015. *Analog Circuit Design: Discrete & Integrated*. New York: McGraw-Hill Education.

Gray, Paul R, Paul J Hurst, Stephen H Lewis, and Robert G Meyer. 2009a. *Analysis and Design of Analog Integrated Circuits*. John Wiley & Sons.

———. 2009b. *Analysis and Design of Analog Integrated Circuits*. John Wiley & Sons, 4/e.

Hajimiri, Ali. 2023. "Analog. Inexact Science, Vibrant Art." 2023. https://chic.caltech.edu/wp-content/uploads/2023/01/Hajimiri_Analog_DRAFT012023.pdf.

Hurst, P. J. 1991. "Exact Simulation of Feedback Circuit Parameters." *IEEE Transactions on Circuits and Systems* 38 (11): 1382–89. https://doi.org/10.1109/31.99170.

———. 1992. "A Comparison of Two Approaches to Feedback Circuit Analysis." *IEEE Transactions on Education* 35 (3): 253–61. https://doi.org/10.1109/13.144656.

Hurst, P. J., and S. H. Lewis. 1995. "Determination of Stability Using Return Ratios in Balanced Fully Differential Feedback Circuits." *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing* 42 (12): 805–17. https://doi.org/10.1109/82.476178.

Middlebrook, R. D. 2006. "The General Feedback Theorem: A Final Solution for Feedback Systems." *IEEE Microwave Magazine* 7 (2): 50–63. https://doi.org/10.1109/MMW.2006.1634022.

Pelgrom, M. J. M., A. C. J. Duinmaijer, and A. P. G. Welbers. 1989. "Matching Properties of MOS Transistors." *IEEE Journal of Solid-State Circuits* 24 (5): 1433–39.

Pierret, R. F. 1996. *Semiconductor Device Fundamentals*. Pearson Education.

Staszewski, Robert Bogdan, Dirk Leipold, O Eliezer, Mitch Entezari, Khurram Muhammad, Imran Bashir, C-M Hung, et al. 2008. "A 24mm 2 Quad-Band Single-Chip GSM Radio with Transmitter Calibration in 90nm Digital CMOS." In *2008 IEEE International Solid-State Circuits Conference-Digest of Technical Papers*, 208–607. IEEE.

Tian, M., V. Visvanathan, J. Hantgan, and K. Kundert. 2001. "Striving for Small-Signal Stability." *IEEE Circuits and Devices Magazine* 17 (1): 31–41. https://doi.org/10.1109/101.900125.

Ulaby, Fawwaz Tayssir, and Michel M Maharbiz. 2013. *Circuit Analysis and Design*. 2nd Edition. NTS Press.