SEMICONDUCTOR DEVICES AND TECHNOLOGY

David W. Greve
Department of Electrical and Computer Engineering
Carnegie Mellon University

Second printing.

Contents

PREFACE

It is customary for a book to have a preface, wherein the author explains why he wrote it and how it differs from all the other books on a similar subject.

This book came about because I was asked to update one of our sophomore courses, in part to add some material on semiconductor devices and technology and in part to increase the breadth of the course and improve its links to other parts of the ECE curriculum. Doing all this in a single semester requires a very careful choice of topics and the depth for each of the topics. That is, to a large degree one needs to make choices about what is to be left out, more so than what is to be added in. This book embodies my own personal opinions about these choices. I believe this book is written in a similar spirit to the "blue book" series (the Modular Series on Solid State Devices) by Neudeck and Pierret, now regrettably out of print but still valuable as a compact introduction to semiconductor devices. This book has broader aims, and thus a different set of topics.

In Chapter 1 have chosen to discuss only two semiconductor devices, the junction diode and the fully-depleted silicon-on-insulator field effect transistor. The SOI-FET is the easiest of the FETs to understand, both physically and mathematically. It may become the mainstream FET technology in the near future. By limiting the discussion to this one transistor type I intend to provide a useful introduction to active devices that includes the most essential device physics. Chapter 2 introduces the basic processes of semiconductor device fabrication and describes the process flow of an SOI CMOS process. This chapter also introduces the basic concepts of layout and relates device cross sections to the layout. Chapters 3-8 concern linear circuit theory and applications. Linear circuit theory remains an essential part of the ECE vocabulary- ECEs are apt to use circuits in the solution of all sorts of problems, in electromagnetic, mechanical, fluid, thermal, etc. domains. Chapter 3 is a summary of linear circuit analysis concepts, including the analysis of circuits with dependent sources. Chapter 4 concerns (low frequency) op amp circuit analysis, providing a both a link to real circuit applications and also a good example of the application of dependent sources to model practical circuits. Chapter 5 introduces energy storage elements and the transient analysis of first-order systems. Chapter 6 uses the switching power converter as an example of a system that requires transient analysis, providing another link to engineering practice. Chapter 7 addresses transient analysis in second-order systems, and sinusoidal steady state analysis is presented in Chapter 8. By placing sinusoidal steady state after the discussion of transient analysis, it is possible to view sinusoidal steady state analysis as a way to efficiently determine the forced response of a system for the special case of sinusoidal excitation. Finally, Chapter 9 introduces the (exponential) Fourier series and the Fourier transform and followed by the concepts of modulation and demodulation of analog signals.

This edition corrects (I hope most) of the errors in the previous editions; revises and expands the material on Fourier series and transforms, improves the text in some places, and adds some additional problems. I have also added a summary of essential aspects of linear algebra and improved the continuity by moving mathematical summaries to appendices. I would like to thank Bruce Krogh for providing the encouragement to become involved with this course, and Carnegie Mellon for a leave in fall, 2010 during which much of the revision was done. Background music was provided by Chet Baker, Bruce Springsteen, Neil Young, Mark Hollis, F. Chopin, and of course Ludwig van B.

David W. Greve Squirrel Hill, December, 2011

CHAPTER 1

SEMICONDUCTOR DEVICES

1. Introduction

Electronics as we know it would not exist without semiconductor devices. Semiconductor devices make it possible to perform the basic functions of switching and amplification. The most important semiconductor devices are the bipolar junction transistor and the field effect transistor. We will discuss only the field effect transistor: it is the most common semiconductor device and its basic operation is the easiest to understand. We will also learn about other components that can be fabricated with semiconductors, including resistors, a magnetic field sensor, and the *pn* junction diode.

2. Semiconductors

We have an intuitive appreciation of materials that are electrical conductors and electrical insulators. Electric *conductors* are used to make wires: common examples include the wiring used within a house or building, the traces on a printed circuit board, or the interconnect on an integrated circuit. Electrical *insulators* prevent current flow between conductors; for example, the plastic insulation on a wire or the glass epoxy substrate of a printed circuit board. Both materials are characterized by their electrical conductivity, usually designated by the symbol σ and having the dimensions 1/ohm·cm.

Figure 1.1 shows a piece of material with ideal contacts. The resistance measured between the contacts is given by

$$R = \frac{1}{\sigma}\frac{l}{A} \qquad (1.1)$$

where A is the cross-sectional area and l is the length. Resistance had the dimension of ohms (Ω).

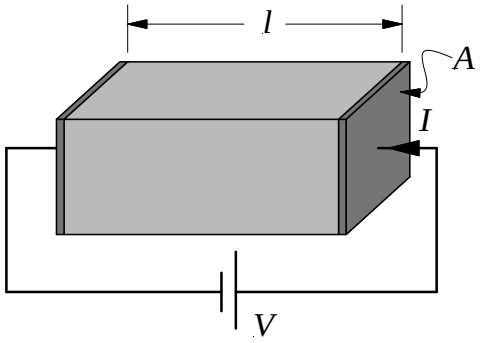


Figure 1.1. A piece of material with ideal contacts.

Figure 1.2 shows the electrical conductivity of some common materials. Note that this is a logarithmic scale- electrical conductivity varies by many orders of magnitude.

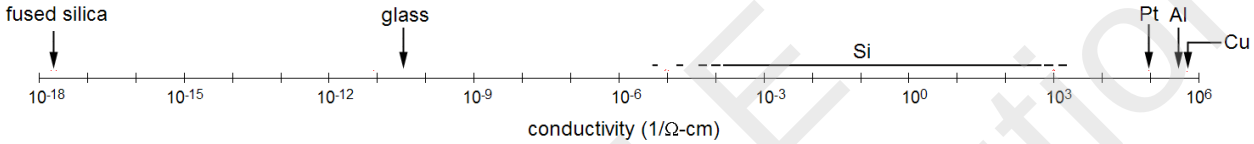


Figure 1.2. Electrical conductivity of different materials.

Semiconductors are materials that have an electrical conductivity intermediate between the electrical conductivity of good conductors (such as aluminum and copper) and good insulators (some glasses and plastics). There are a great many materials that exhibit semiconducting behavior but only a very few of them are of much interest for electronics. Silicon is the most important semiconductor and is the active material in almost all electronic devices. A few other semiconductors- for example, gallium arsenide- are essential because they can be used to make optoelectronic devices. We will focus on semiconductor silicon.

Materials are semiconductors in part because of their chemistry (the electronic structure of the constituent atoms) and in part because of their structure (the way in which atoms are organized to make the solid material). Semiconductor materials are particularly useful for electronics because the electrical conductivity of the pure material can be greatly changed by the introduction of a small number of impurities. In addition semiconductors are strongly influenced by applied fields (including electric, magnetic and electromagnetic fields). In the following sections we will describe some of the basics of the behavior of semiconductors.

The pure semiconductor at absolute zero

Semiconductors used in integrated circuits are *single crystals*; that is, they are built up by repeating a *unit cell*. Figure 1.3 shows a unit cell for silicon, where each ball represents a silicon atom and the sticks represent covalent bonds between silicon atoms. This is the same as the crystal structure for diamond. As silicon is in group IVA of the periodic table, there are four valence electrons available for bonding. In this crystal structure each atom has exactly four bonds to four nearest neighbors. Each of those bonds consists of two electrons shared between neighboring silicon atoms. This is a strong and stable crystal structure.

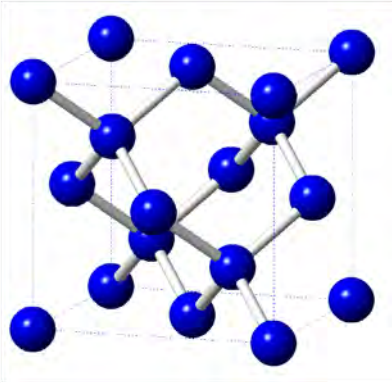


Figure 1.3. The unit cell for silicon. Each atom is bonded to four nearest neighbors. Only bonds within this unit cell are shown.

Shortly we will discuss the behavior of silicon at nonzero temperature and with deliberately introduced impurities; to do this it is convenient to have a two-dimensional representation of the bonding in the silicon single crystal. This two-dimensional representation is shown in Figure 1.4. In this diagram each line represents a valence electron that is shared between two atoms. (Two shared electrons make up a single covalent chemical bond). This figure represents a perfect single crystal (no impurities) at absolute zero. All the available electrons are in bonding states so there are no free electrons and the electrical conductivity is zero.

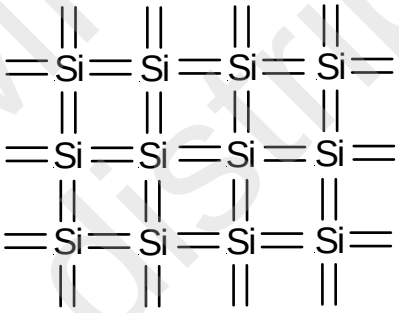


Figure 1.4. Representation of bonding in the pure silicon crystal at absolute zero.

The pure semiconductor at finite temperature

If we raise the temperature above absolute zero, a small number of electrons will be excited out of bonding states. Excitation of one electron out of a bonding state is illustrated in Figure 1.5.

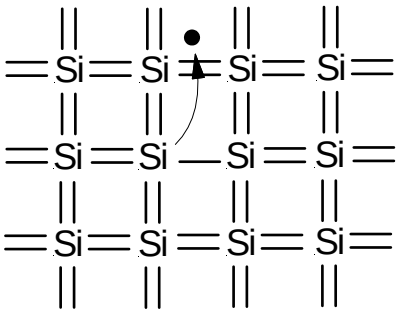


Figure 1.5. A single crystal at a nonzero temperature. Some bonding electrons are excited into higher energy states and are free to move.

where ΔQ is the charge transferred to the positive terminal (or *p* side) through the external circuit by an increment is voltage ΔV_A.

There is an important difference between a *pn* junction and a parallel-plate capacitor. In a *pn* junction the charge transferred per ΔV_A depends on the value of the applied voltage; that is,

$$\Delta Q = \Delta Q(V_A)\,. \tag{1.49}$$

The capacitance we evaluate from eq. (1.49) is actually a *differential* or *small-signal* capacitance and it is a function of the applied voltage. For a *pn* junction in reverse bias or small forward bias the capacitance is given by

$$C(V_A) = \frac{C_{j0}}{(1 - V_A / V_{bi})^n} \tag{1.50}$$

where C_{j0} is the capacitance with zero applied bias and *n* is a constant which is usually between ½ and $^1/_3$. The capacitance is important in circuits where we are concerned with changing voltages and currents.

4. Active semiconductor devices

A loose definition of an active device is one that is capable of controlling voltage or current. Implicit in this definition is the idea of an external source (to supply the current or voltage), a load (the element associated with the voltage or current being controlled) and a control terminal (to which a control signal is applied). In useful active devices the power supplied to the control signal is smaller than the power being controlled. Since there is more power dissipated in the load than supplied to the input there must be an external source, or power supply.

Almost all of the important active devices have three or more terminals. Active semiconductor devices are based either on carrier injection across a junction (bipolar junction transistors) or charge induced by an electric field (field effect transistors). Both types of devices can be constructed using *n* and *p*-type semiconductors. Figure 1.28 shows these two types of semiconductor devices along with their circuit symbols.

The bipolar transistor conists of two *pn* junctions in close proximity with a common *n* or *p* region. Figure 1.28 (top) shows an *npn* bipolar transistor. The *p* region or *base* (marked B) is the control terminal.

We concern ourselves here only with field effect transistors. In the field effect transistor (Figure 1.28, bottom) the control terminal is a *gate* electrode (marked G). The gate influences the current that flows between heavily doped *source* and *drain* regions (marked S and D). At the most basic level the operation of field effect transistors is easier to understand than that of bipolar junction transistors. In addition, field effect transistors represent the majority of semiconductor devices used today.[4]

[4] While it is reasonable to limit the discussion to field effect devices in an introductory course, this does not mean that a well-educated engineer needs to know nothing more. There are some important electronic functions that are still best performed with bipolar transistors. In addition, understanding field effect devices at an advanced level requires an understanding of concepts found in the bipolar junction transistor.

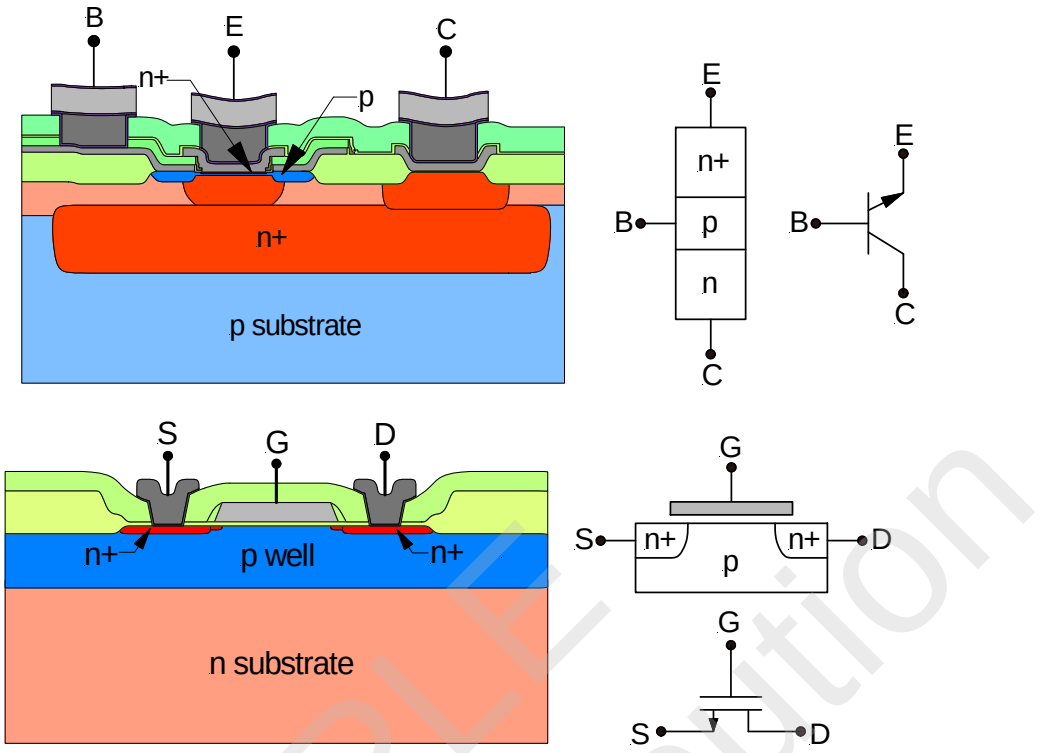


Figure 1.28. Active semiconductor devices: (top) bipolar junction transistor and (bottom) field effect transistor.

Figure 1.29 shows two different types of insulated gate field effect transistors. These are termed “insulated gate” devices because an insulating layer is placed between the gate (or control) electrode and the rest of the transistor. As a result ideally there is no DC electric current required to control the load. Most commonly, these devices are referred to as MOSFETs, which is an abbreviation for Metal Oxide Semiconductor Field Effect Transistors. Here Metal refers to the gate material (which is either a metal or a material nearly as conductive as a metal) and the Oxide refers to the insulating layer, which most often is silicon dioxide.

The top diagram in Figure 1.29 shows a *bulk* MOSFET, that is, a MOSFET that is fabricated in a silicon substrate. Many bulk MOSFETs can be fabricated in a single substrate because the substrate (*p* type in the figure) can be connected to the most negative potential in the circuit. When this is done the *pn* junction between the substrate and *n+* source and drain regions will be reverse-biased. This guarantees that essentially no current flows from the substrate.

The bottom diagram shows a *silicon- on- insulator* or *SOI* MOSFET. Here the transistor is fabricated in a thin silicon layer that is isolated from the substrate by a thick insulating layer. We will discuss in detail the operation of the SOI MOSFET, which is simpler because we do not need to be concerned with a semiconductor substrate as in the case of the bulk MOSFET. We will describe the differences between the SOI and bulk MOSFETs later.

Most MOSFETs fabricated at present are of the bulk type, although some advanced integrated circuits use SOI MOSFETs. SOI MOSFETs offer many advantages, including a substantial reduction of the influence of the substrate during circuit operation. At pre-

sent industry is anticipating a transition to SOI MOSFETs for high-performance digital logic circuits beginning approximately 2013.[5]

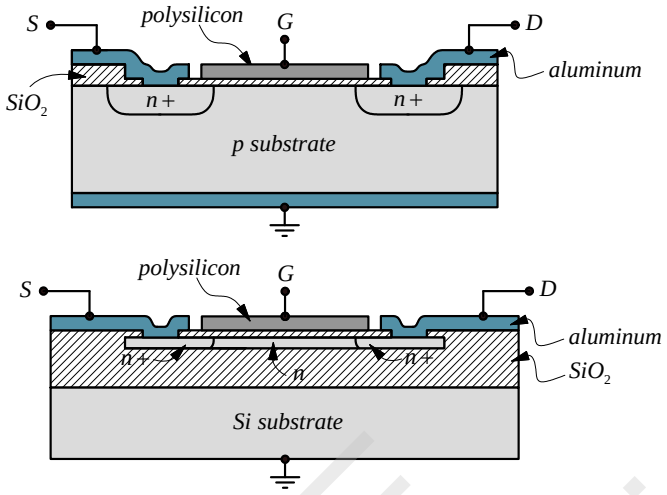


Figure 1.29. Two different types of insulated gate field effect transistors: (top) a bulk MOSFET and (bottom) an SOI MOSFET.

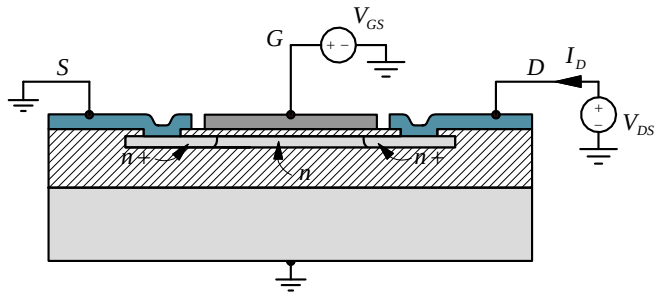
5. The SOI MOSFET

We will first discuss the qualitative operation of the SOI MOSFET. This will be followed by a derivation of its current-voltage characteristics. Finally, we will comment on some additional phenomena of importance.

Qualitative operation of the SOI MOSFET

Figure 1.30 shows a cross section of an SOI MOSFET with bias voltage sources connected. We will specify all voltages using a notation with two subscripts: for example, $V_{GS} = V_G - V_S$ is the voltage of the gate relative to the source. In order to control current at the drain terminal, we will modulate the charge carriers in a *channel* that extends from source to drain. Conduction in this transistor will be by electrons because the $n+$ source and drain regions provide good contacts to an electron channel. So we call this an n-channel transistor; when the transistor is turned on the conducting channel will consist of electrons flowing from source to drain.

Note that there is no difference physically between the source and drain regions. Negative electrons will move opposite to the direction of the electric field in the channel, that is, toward the more positive terminal (drain). The less positive terminal is the source terminal in an n-channel device.



[5] ITRS 2009 Roadmap, Process, Integration, Devices, and Structures, available from www.itrs.net.

Figure 1.30. SOI MOSFET with bias voltage sources. Electrons move toward the more positive (drain) terminal.

We will now discuss qualitatively the behavior of the drain current I_D, defined as positive flowing into the drain terminal. We will use a simplified drawing of the SOI MOSFET as shown in Figure 1.31. We can omit the substrate because the thickness of the insulator under the channel is great enough to minimize the influence of the substrate. We begin by assuming that the drain voltage is a small positive voltage, and that there is zero bias on the gate.

With the drain voltage nearly equal to zero, (Figure 1.31, top) the thin semiconductor is nearly uniform in potential and we have nearly the same situation as in a parallel plate capacitor, where the gate forms one electrode and the thin semiconductor the other electrode. When a capacitor has zero voltage applied there is no net charge on either electrode. Since the thin semiconductor is *n*-doped, with no net charge on the semiconductor there are still some mobile electrons present. The semiconductor acts like a resistor with resistance determined by the doping concentration and dimensions. There is a small drain current, which increases linearly for small values of V_{DS}.

Figure 1.31 (middle) shows what happens if the gate voltage becomes positive. A positive gate voltage means that there is a positive surface charge on the bottom of the gate and there must be an equal and opposite charge present in the semiconductor. That charge consists of additional mobile electrons. Since there are now more mobile electrons, the "resistance" of the channel decreases. The drain current is larger, and depends on the magnitude of the gate voltage (Figure 1.32).

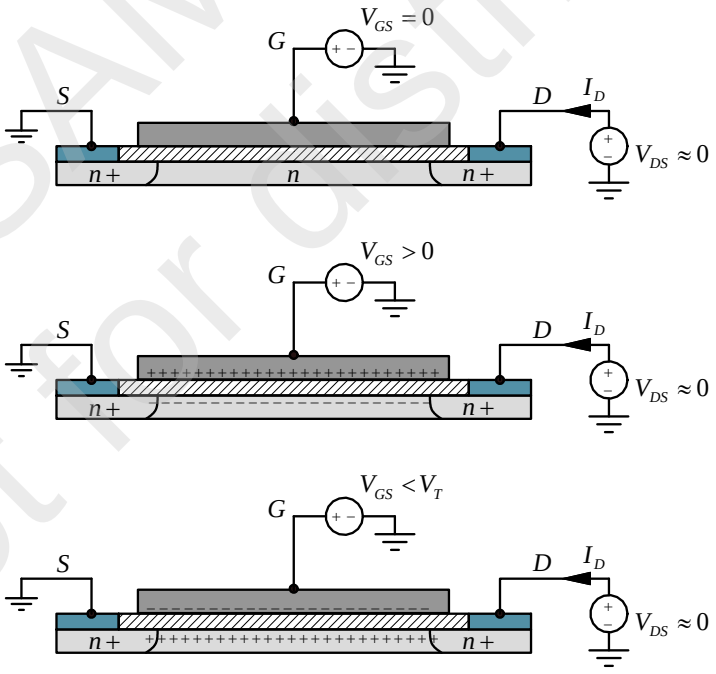


Figure 1.31. Qualitative operation of the SOI MOSFET: (top) $V_{GS} = 0$ (middle) $V_{GS} > 0$ and (bottom) $V_{GS} < 0$.

Finally, we consider the effect of a negative gate voltage (Figure 1.31, bottom). Now the gate charge is negative and must be compensated by a positive charge in the semiconductor. At first this positive charge is produced by driving electrons away from the

CHAPTER 2

SEMICONDUCTOR TECHNOLOGY

1. Introduction

Semiconductor technology refers to the sequence of process steps used to fabricate semiconductor devices. In this chapter we will introduce some of the basic semiconductor fabrication processes and we will show how they are used to make a complete integrated circuit.

2. Overview of integrated circuit design and fabrication

Before discussing fabrication it is appropriate the describe the process of designing and fabricating an integrated circuit. The integrated circuit contains many individual active devices (a microprocessor may contain hundreds of millions of individual field effect transistors, **Error! Reference source not found.**) together with possibly other parts (resistors, capacitors, and sometimes inductors may be found in an analog integrated circuit).

Design of the integrated circuit begins with a functional description of the circuit. Often the function is broken down into more or less independent units- for example, a microprocessor might be broken down into memory, control logic, and arithmetic logic units. Each of these might be broken down further into modules or individual gates. A circuit design is then developed for each unit or sub-unit. The circuit design consists of components, component values and/ or geometries, and the way in which those components are interconnected.

The next step is the *layout*, that is, the design of the masks that are used to define each of the layers in the integrated circuit fabrication process. Layout must conform to *design rules*, which describe the minimum dimensions, separations, etc. that are required in order to guarantee a manufacturable circuit.

The layout file describes the patterns to be formed on the wafer during particular steps in the fabrication process. There may be 20-30 individual patterning steps in the fabrication of a complex integrated circuit. A *mask* with transparent and non-transparent regions is created from the data in the layout file for each of these patterning steps.

Many individual integrated circuits are fabricated at the same time on a single silicon wafer substrate. Wafer fabrication involves a great many exacting process steps performed sequentially on the wafer. Because a defect formed during any of the process steps may result in a non-functional circuit, wafer processing is performed under ultra-clean conditions in special factories (commonly known as *fabs*). Each completed wafer contains hundreds or thousands of integrated circuits.

After testing, chips are separated from each other by sawing or laser scribing. Each functional chip is packaged and the package pins are connected to pads on the chip. A large, advanced integrated circuit may occupy 1 cm^2 of silicon and may cost hundreds of dollars or more. Small integrated circuits fabricated using a simple process may be a few mm^2 and in packaged form may sell for less that $0.10. This chapter provides a basic introduction to semiconductor integrated circuit fabrication. Similar process steps are used to fabricate other important products, including screens for flat panel displays; optoelectronic devices; micromechanical sensors and actuators; hard disk heads; etc. etc.

3. Individual process steps

A semiconductor process consists of many individual steps that are repeated in order to build up the integrated circuit. In this section we describe some of the essential process steps. Later we will see how they are combined in order to fabricate a complete integrated circuit.

Deposition of thin films

Thin films of metals or insulators provide for electrical interconnection or isolation between electric conductors. These layers are deposited uniformly over the entire wafer. Patterned layers are created by selectively etching portions of a deposited layer.

Figure 2.1 shows a silicon substrate after deposition of a thin film insulator followed by deposition of a thin film metal. Common insulators are silicon nitride and silicon dioxide. Conductive layers include aluminum, tungsten, and doped polycrystalline silicon.

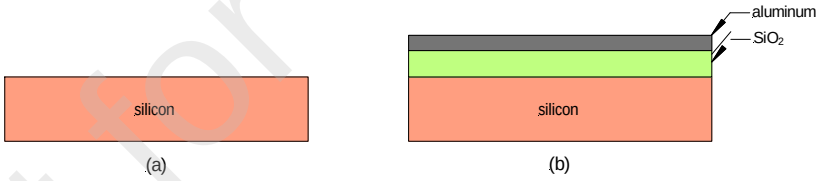


Figure 2.1. Deposition of thin films: (a) silicon substrate and (b) after deposition of a silicon dioxide layer followed by an aluminum layer.

Photolithography

Photolithography is the process used to pattern regions on the wafer. (Photo- refers to light; -litho- to stone and -graphy to writing. Literally, photolithography is “writing on stone.”)

Figure 2.2 shows the steps involved in patterning a metal layer. First a layer of *photoresist* (a light-sensitive polymer) is spread on the wafer. Then some regions of the photoresist are exposed by focusing the light that passes through a mask onto the wafer. The photoresist is developed by flooding it with a liquid developer. The developer removes the exposed regions and leaves the unexposed regions. After completion of the photo-

lithographic process the remaining photoresist masks some of the regions from subsequent etching steps.

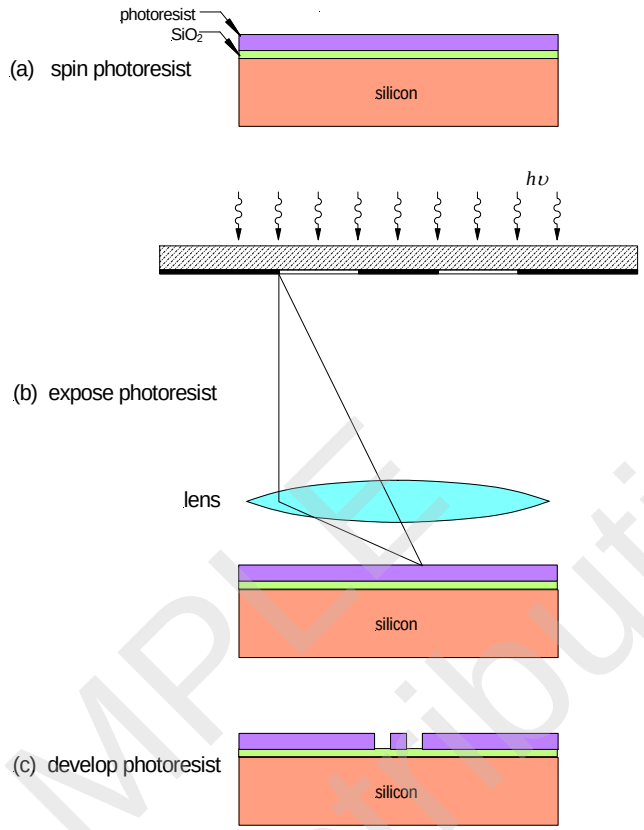


Figure 2.2. Patterning using photolithography: (a) wafer after coating with photoresist; (b) exposure of selected regions with ultraviolet light; and (c) the wafer after development of the photoresist.

Etching

Etching is the controlled removal of material. Figure 2.3 shows the etching of a metal layer. In this case the etchant is selective; that is, it etches metal only and not the underlying insulator. Etching is performed using either liquid chemicals or the excited molecular species created in a plasma discharge.

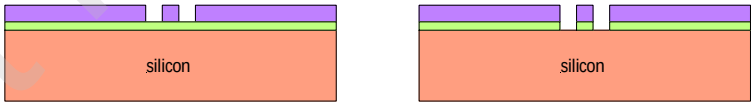


Figure 2.3. Etching of a silicon dioxide layer: (left) patterned photoresist after photolithography and (right) after etching the silicon dioxide.

Implantation and annealing

Dopants are commonly introduced by *ion implantation* (Figure 2.4). Dopant atoms are ionized and accelerated to energies high enough to penetrate a short distance into exposed regions of the semiconductor. With appropriate choice of the implantation energy ions can be blocked by a layer of insulator or other material. Implantation is followed by *annealing*, that is, heating the wafer to a temperature high enough to cause controlled diffusion of the implanted dopant atoms.

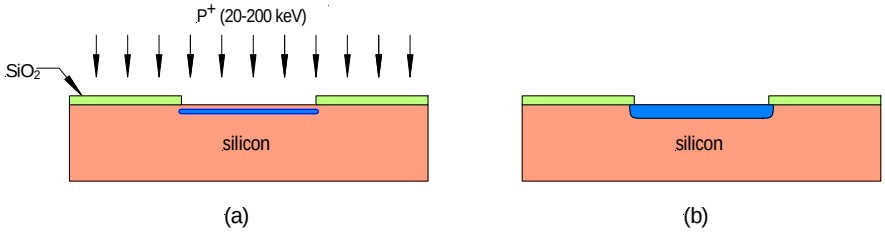


Figure 2.4. Doping by ion implantation: (a) bombardment with phosphorus atoms and (b) after annealing at an elevated temperature.

Thermal oxidation

Thin insulating layers can also be formed by *thermal oxidation* (Figure 2.5). Here the silicon substrate itself reacts with oxygen or water vapor at an elevated temperature and is converted into silicon dioxide. The insulating layer formed in this way is of very high quality and is often used as the gate insulator in field effect transistors.

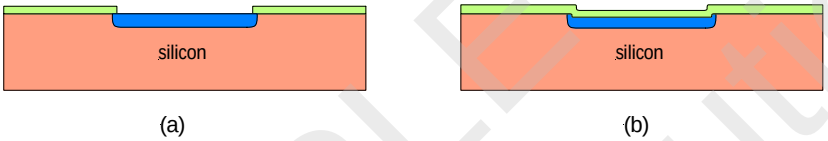


Figure 2.5. Thermal oxidation: (a) wafer before thermal oxidation and (b) after oxidation.

Chemical-mechanical polishing

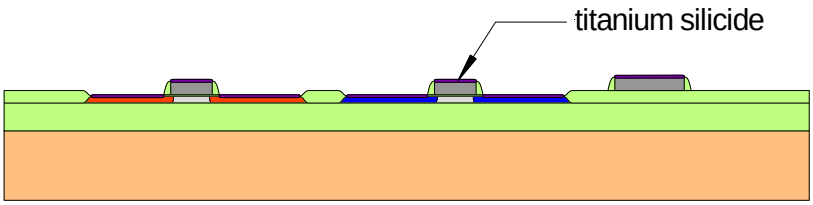
Material can also be removed from the surface by polishing or grinding. This process removes the protruding regions and leaves the recessed regions unchanged. Chemical-mechanical polishing leaves a planar surface; planar surfaces are easier to pattern and easier to cover with deposited films without thinning over steps.

Figure 2.6. Chemical mechanical polishing: (a) before polishing and (b) after polishing.

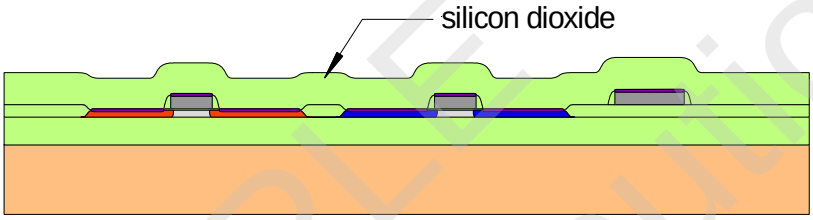
4. A complete process

In the following we will follow a wafer through a complete silicon-on-insulator MOSFET process. The result will be two transistors, one *n* channel and one *p* channel. The process described is realistic although some details are omitted. Also, the drawings are not exactly to scale; in an exact scale drawing some of the layers are so thin they are difficult to see.

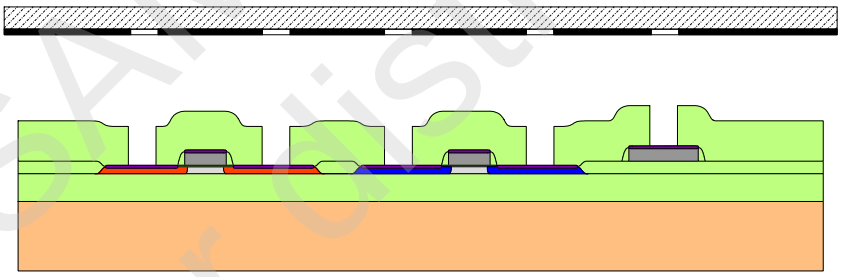
The process begins with a wafer with a thin single crystal silicon layer on top of a thick silicon dioxide insulating layer. The mask layers (layout) used for the various photo-lithographic steps are shown in Figure 2.7. In our process description we will show a cross section along the line a-a' in the layout.



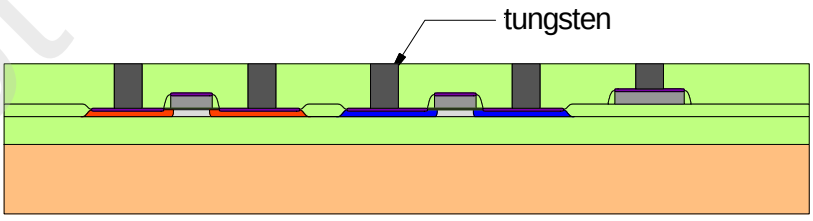
The wafer is annealed, resulting in the conversion of titanium to titanium silicide everywhere it is in contact with silicon. Then the unreacted titanium is etched away.



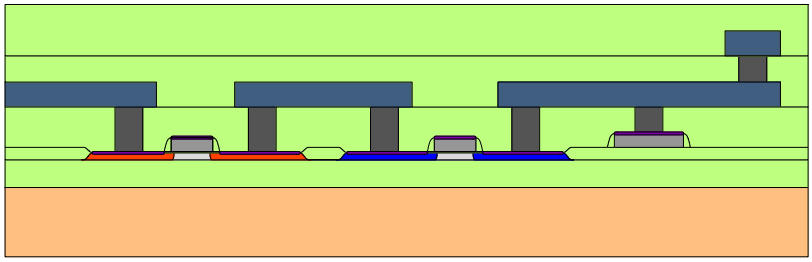
A layer is silicon dioxide is deposited.



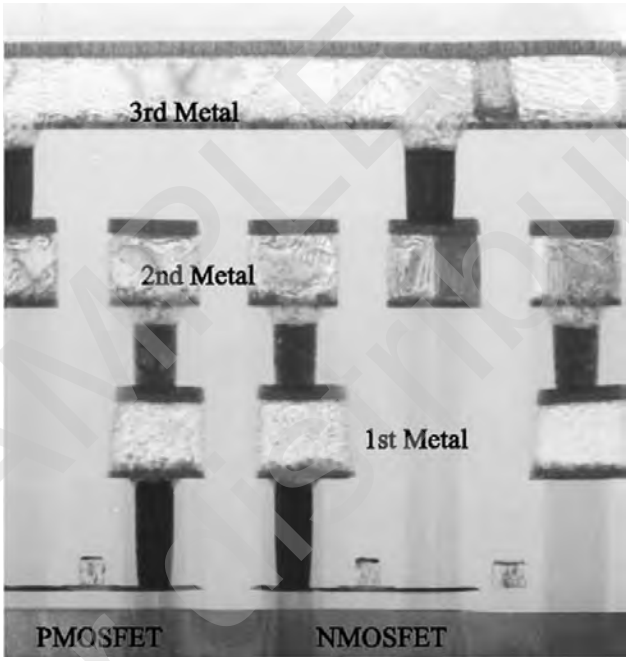
Photolithography is performed using mask #5 (contact). Silicon dioxide is etched to make contacts and then the photoresist is removed.



A thick layer of tungsten is deposited on the wafer. Then a polishing step is performed to remove excess tungsten and to make the wafer surface flat.



An additional photolithographic step is performed using mask #8 (metal 2). There may be additional layers of metallization before the wafer is finished.



Transmission electron microscope photograph of SOI MOSFETs. The transistor gatelength is 0.2 μm and the silicon channel is about 50 nm thick [Reprinted from *Solid State Electronics* **48**, issue 6, 999-1006 (2004), F. Ichikawa, Y. Nagatomo, Y. Katakura, M. Itoh, S. Itoh, H. Matsuhashi, T. Ichimori, N. Hirashita, and S. Baba, "Fully depleted SOI process and device technology for digital and RF applications,", with permission from Elsevier]. This process has three levels of interconnect.

Color figures from Chapter 8

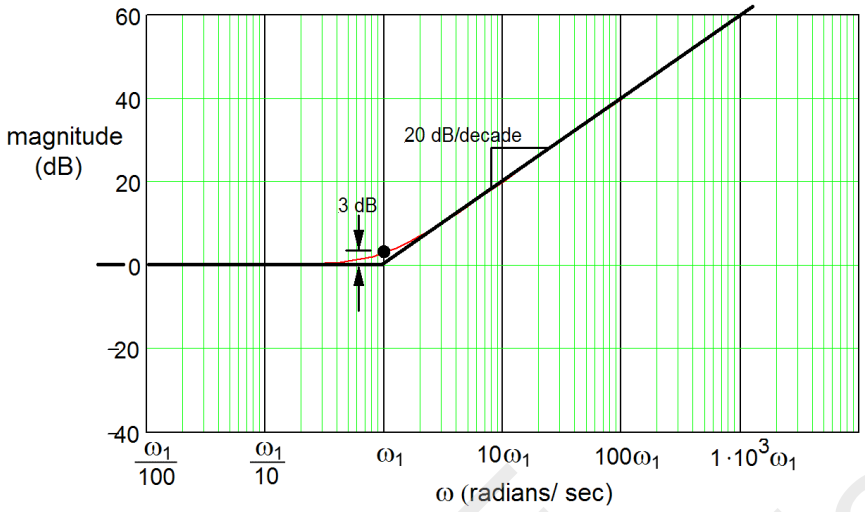


Figure 14. Asymptotic magnitude response.

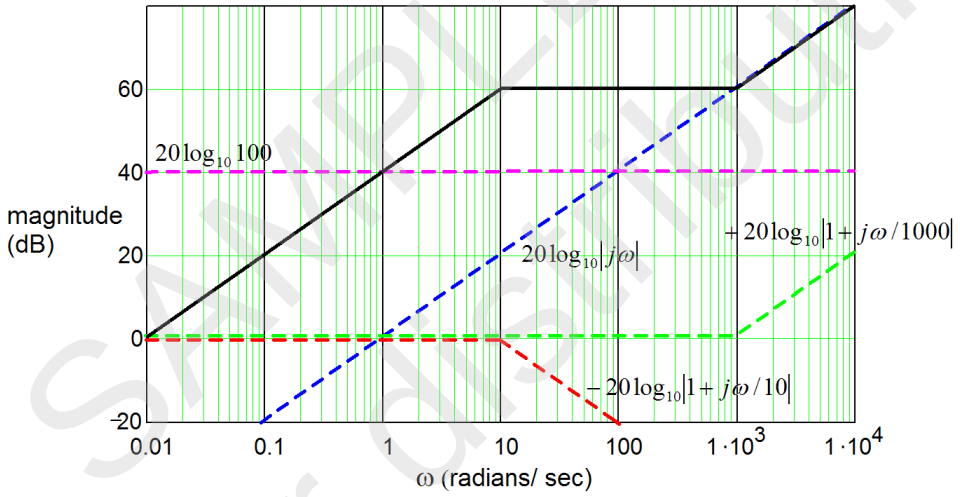


Fig. E 16. Magnitude response for the transfer function of Example 10.

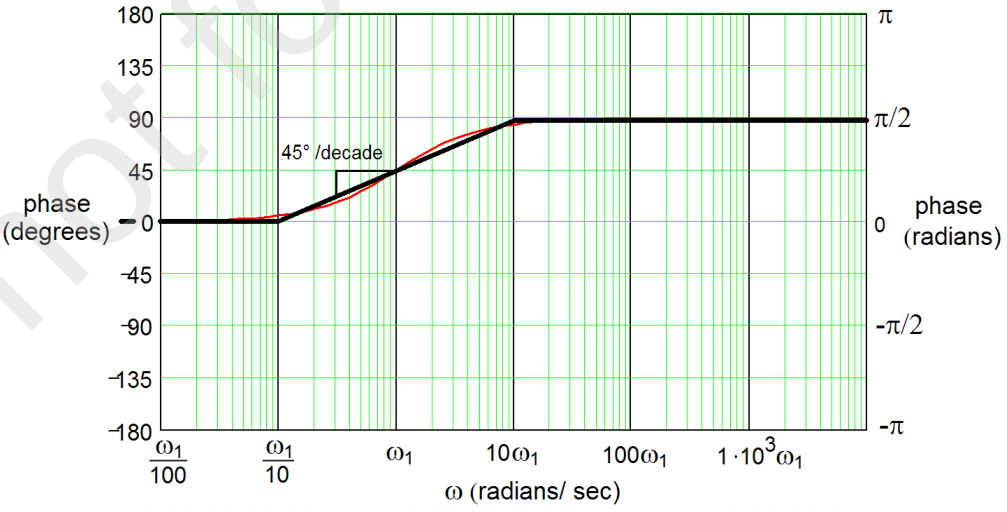


Figure 15. Phase response for the transfer function of Example 10.

CIRCUIT ANALYSIS AND APPLICATIONS

David W. Greve
Department of Electrical and Computer Engineering
Carnegie Mellon University

Contents

CHAPTER 6

POWER SUPPLIES AND ENERGY STORAGE

1. Introduction

Power supplies, whether for conversion of AC to DC or transformation of one DC voltage to another, are an essential part of electronic systems. Useful power supplies cannot be made without the use of energy storage elements. In this chapter, we study the application of energy storage elements to the transformation of electrical power.

2. Electrical power in electronic systems

For purposes of our discussion here, a *power supply* is an electronic subsystem that transforms or controls electrical power. Power supplies for line-operated electronic systems convert AC[1] (alternating current) to the DC that is required for almost all electronics. However in addition DC to DC converters are very widely used (although almost invisible, unless you look for them carefully!) to transform one voltage level into another. Another important function of power supplies is electrical *isolation*; removing the direct electrical connection between the power source and electronics can eliminate interfering signals and/ or prevent electrical shocks.

Let's look at a laptop PC to get a sense of the power supply requirements. Laptop PCs have batteries for off-line operation. A typical laptop battery provides 20 V at a few amperes. The battery is charged with a line-operated power supply (or adapter), which converts line voltage[2] into approximately 20 V DC. Within the laptop various systems have different power requirements. The microprocessor (CPU) and other digital compo-

[1] In the US AC power is nominally a sinusoid $v(t) \approx \sqrt{2} \cdot 110 \cos(120\pi t)$; that is, the RMS value is about 110 V and frequency 60 Hz. In most European and Asian countries AC power is about 220 V RMS and 50 Hz.

[2] Most AC power adapters accept input voltages from 100 V to 240 V and from 50 Hz to 60 Hz; these power adapters can be used anywhere in the world without transformers or switches.

nents require about 1.0 V DC. CD drives and disc drives require 5 V, and LCD panels 3.3 V. If there are cold-cathode (fluorescent) lamps behind the LCD panel these require a few hundred volts AC. All of these need to be generated from the 20 V available from the battery. A further complication is the variation in battery voltage while the battery discharges and the need to provide for reduced operating voltages to extend the endurance of battery operation.

An entirely different problem is encountered in hybrid automobiles. Here a power control system must direct power to a battery for charging and from the battery during acceleration, and in addition must coordinate operation of a gas engine and an electric motor.

In these, and many other power control applications efficiency is of major concern. (Other equally important factors include physical size, weight, and cost). The efficiency for a power converter can be defined as

$$\eta = \frac{P_{out}}{P_{in}} \times 100\ \% . \qquad (6.1)$$

Many power conversion systems have efficiencies in the 90-95% range.

In the following we will study the application of transient analysis to understanding the operation of power conversion systems.

Piecewise-linear analysis

In our study of power conversion systems, we will be using semiconductor devices to control the flow of current. Essentially we will be using these semiconductor devices as voltage-controlled switches (field effect transistors) or as devices which allow current flow in only one direction (junction diodes).

Both field effect transistors and junction diodes are nonlinear devices; that is, they are not in general additive and homogeneous. If we model the full nonlinear behavior of these devices, hand circuit analysis becomes difficult or impossible. So instead we use linear approximations to the characteristics that are valid for a limited range of voltages and currents. When we reach the boundary of applicability of a model, we then switch to a different model that is valid for another portion of the analysis.

This method is known as *piece-wise linear analysis*. It allows us to perform an approximate analysis of a circuit containing nonlinear components as a series of linear circuit problems. In the following, we will describe the piece-wise linear analysis technique and apply it in some simple examples.

Piece-wise linear modeling of devices

Figure 6.1 (left) illustrates the concept of piece-wise linear modeling of a nonlinear device. We approximate a non-linear $i_D(v_D)$ characteristic by several straight-line segments. The points where these segments intersect are known as *breakpoints*. For each breakpoint we have a condition on voltage or current that determines when to switch models.

and for $t > t_2$ we have the solution

$$v_C(t) = (4.13\,V)e^{-(t-t_2)/(0.01\,\mathrm{sec})}$$

where we see that the time constant is now two orders of magnitude smaller. This solution is also plotted in Figure 6.E6.

Let's discuss the performance of the rectifier circuit of Figure 6.E4. First of all, it is clear that the output has an average (DC) value with a periodic time-varying component. The periodic time-varying component is termed *ripple* and it should be substantially less than the value of the DC component. We see that the ripple is considerably worse when the load resistance decreases, that is, as the current drawn by the load increases. In order to keep the ripple voltage small, the time constant *RC* must be long compared to the period of the input waveform. For a large output current, this may require very large values for the filter capacitor. This leads to a power supply that is large, expensive, and heavy.[5]

Another important issue is the efficiency of the power supply. We see that the diode on voltage V_{ON} is dropped across the diode when the capacitor is being charged. The power dissipation in the diode is given by

$$p_D(t) = i_D(t) \cdot V_{ON} \,. \tag{6.6}$$

At the same time the instantaneous power supplied to the resistive load and the capacitor is

$$p_L(t) = i_D(t) \cdot v_L(t) \approx i_D(t) \cdot V_L \tag{6.7}$$

where V_L is the average value of the voltage across the load. So the efficiency can be no better than

$$\eta = \frac{p_L}{p_L + p_D} \approx \frac{1}{1 + V_{ON}/V_L} \,. \tag{6.8}$$

This is particularly serious when V_L is a small multiple of V_{ON}. The diode forward drop V_{ON} is a characteristic of the material used to make the diode and cannot be readily reduced. A low-voltage power supply made in this way will have very poor efficiency.

In efficient modern power supplies, diode rectifiers together with filter capacitors of moderate size are used to generate unregulated DC, that is, DC with a substantial ripple. Then this unregulated DC is converted to higher or lower DC voltages using switching circuits of the type we will discuss next. These switching circuits use MOSFETs instead of diodes. Later we will see the advantages of this apparently more complex approach.

3. Power conversion circuits

We will consider first the generation of a lower DC voltage from an available source that produces unregulated DC at a higher voltage. A good example is the production of 1- 1.5 V DC for a microprocessor from a laptop battery or a standard PC power supply.

[5] The situation will be somewhat improved if we use a *full-wave rectifier* in which the capacitor is charged on both positive and negative half cycles. However it is still true that small ripple requires large capacitors if the current demands are high.

Let's first realize that the most obvious way to do this is a bad idea. Figure 6.5 shows a circuit to produce a regulated low voltage from a higher voltage (for example, let's suppose the high voltage is 5 V and we want 1 V to power a microprocessor). The box is a three-terminal device where one terminal is a control terminal. For example, this could be a MOSFET. We adjust the control voltage so that exactly the right amount of voltage drops across the device to give us 1 V across the load.

This is highly unattractive because (1) about 80% of the power supplied by the source is dissipated in the control device and (2) the maximum current supplied by the source is the same as the current consumed by the device.

Ideally we would like a lossless converter between the two voltage levels- sort of an ideal transformer, but one that works at DC. In the following sections we will see how to do this.

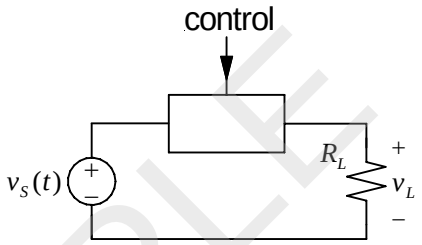


Figure 6.5. A linear regulator.

A voltage converter circuit

Let's consider the circuit in Figure 6.6 (left). The double-throw switch is an idealized representation of two MOSFETs (Figure 6.6, right). There are two positions for this switch, A and B. We will show that by appropriately controlling the switch we can maintain a nearly constant voltage across the load R_L.

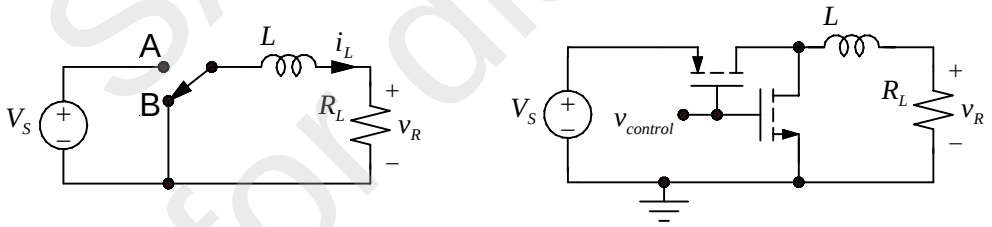


Figure 6.6. A simple voltage converter circuit (left) and implementation of the circuit using two MOSFETs as switches (right).

Suppose we begin with zero current through the inductor. Let's set the switch to position A for a time that will allow the voltage across the load resistor to increase to our desired output voltage V_1.

With the switch in position A we can use our solution for a first-order system with constant forcing function. We need the initial and final voltages for v_L. Since the inductor current is initially zero we have

$$v_R(0) = 0\,V \tag{6.9}$$

and at $t = \infty$, $di_L/dt = 0$ so $v_L = 0$ and we have

$$v_R(\infty) = V_S \tag{6.10}$$

Consequently the solution for $v_L(t)$ becomes

$$v_R(t) = V_S \cdot (1 - e^{-tR/L}) . \tag{6.11}$$

This will be equal to V_1 at time t_1 given by

$$t_1 = \frac{L}{R} \cdot \ln(\frac{V_S}{V_S - V_1}) . \tag{6.12}$$

Now suppose we throw the switch to position B at time t_1. We now have an *R-L* circuit with no source. We can use the same solution if we determine $v_R(t_1)$ and $v_R(\infty)$ for this new circuit. Now $v_R(t_1)$ is clearly still 1 V because the current through the inductor can't change instantaneously. And $v_R(\infty)$ is zero because eventually the current in the inductor will decay to zero. So we have

$$v_R(t) = V_1 e^{-R(t-t_1)/L} \tag{6.13}$$

which eventually causes the load voltage to decay to zero.

The idea for maintaining v_R nearly constant is very simple. Wait a short time and allow the voltage to decay "a little bit". Then throw the switch back to position A. As we saw before, the inductor current will now increase. As soon has increased "a little bit" then we throw the switch back to position B. So we can maintain the load voltage very close to 1 V by controlling the switch position. This operation is illustrated in Figure 6.7.

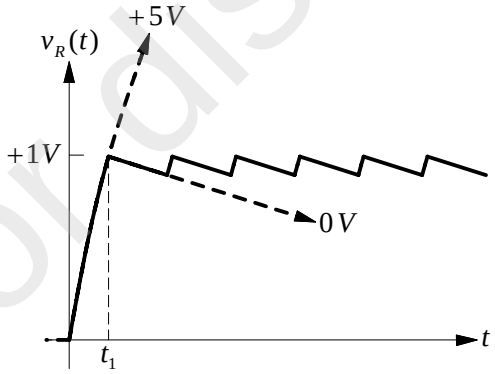


Figure 6.7. Operation of the voltage converter circuit.

So is this circuit really better? We can see that in the ideal case there is no power dissipated *except* in the load, because in our ideal circuit the switch and the inductor are lossless. For a real circuit there would be some resistance associated with the closed switch (the on resistance of the MOSFET) but with appropriate design this will be much smaller than R_L so the losses will be small.

This circuit has considerable advantages beyond being nearly lossless. We would like to switch rapidly enough that the output voltage doesn't decay significantly. We can achieve this by choosing a small inductor and switching very rapidly (say 10's to 100's

of kHz). Then the inductor can be small in value, which translates to small size and light weight.

In the following examples we further explore the operation of this circuit.

Example 3. Operation of a voltage converter.

Consider a voltage converter with R = 100 Ω , L = 10 mH, and V_S = 5 V. Determine the timing of switch positions required to drive the output voltage to 1 V and then to maintain it between 0.9 and 1 V.

Solution. Setting the switch into position A, we reach an output voltage of 1 V at time t_1 where t_1 is given by

$$t_1 = \frac{L}{R}\cdot\ln\left(\frac{V_S}{V_S - V_1}\right) = \frac{10^{-2}\ H}{100\ \Omega}\cdot\ln\left(\frac{5}{5-1}\right) = 22.3\times10^{-6}\ \text{sec}\cdot$$

With the switch in position B, the load voltage is given by

$$v_R(t) = (1V)e^{-R(t-t_1)/L}$$

which becomes equal to 0.9 V at time t_2 given by

$$0.9\,V = (1V)\cdot e^{-(t_2-t_1)/(10^{-4}\ \text{sec})}$$

or

$$t_2 - t_1 = (10^{-4}\ \text{sec})\cdot\ln\left(\frac{1}{0.9}\right) = 10.5\times10^{-6}\ \text{sec}\cdot$$

Setting the switch back to position A, we have an initial voltage of 0.9 V and the final voltage is 5 V. So we have

$$v_R(t) = (5V) - (4.1V)\cdot e^{-(t-t_2)/(10^{-4}\ \text{sec})}$$

which becomes equal to 1 V at time t_3 given by

$$t_3 - t_2 = (10^{-4}\ \text{sec})\cdot\ln(4.1/4) = 2.5\times10^{-6}\ \text{sec}\cdot$$

The control sequence is shown in Figure 6.E7.

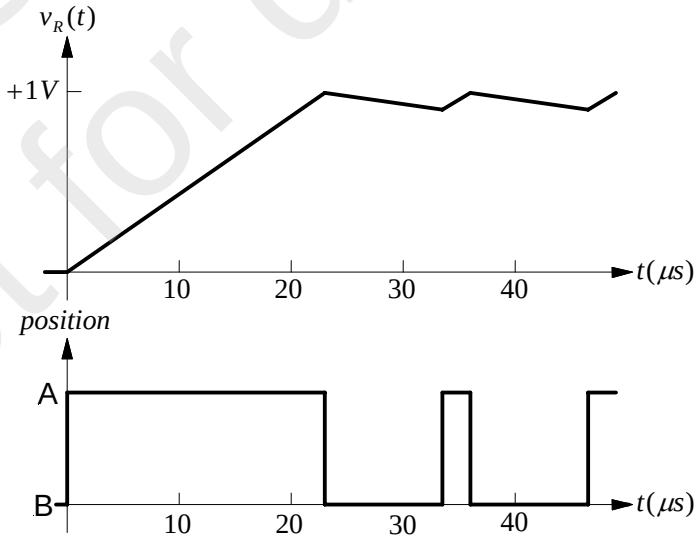


Figure 6.E7. Sequence of switch operation.

CHAPTER 8

SINUSOIDAL STEADY STATE ANALYSIS

1. Introduction

We consider here the analysis of circuits with sinusoidal sources. Sinusoidal steady state analysis was first developed for the analysis of AC power systems. However the circuit applications of sinusoidal steady state analysis are far broader than AC power systems. It can be shown that *any* signal can be represented as a superposition of sinusoids with different frequencies. As a result we can obtain the response of a linear system by determining the response to each of these sinusoids and then summing up these separate responses.[1]

2. Sinusoidal steady state analysis

We will first relate sinusoidal steady state analysis to the more general problem of transient circuit analysis. Suppose we have a linear circuit with time-dependent sources and n energy-storage elements. The most general differential equation for a circuit variable y has the form

$$y^{(n)} + q_{n-1}y^{(n-1)} + \cdots + q_1 y' + q_0 y = r(t) \qquad (8.1)$$

where $r(t)$ is a function of the time-dependent sources. In general $r(t)$ may contain any or all of the time-dependent sources along with their derivatives. To completely solve this problem we need to determine (1) the natural response (or solution to the homogeneous differential equation) $y_n(t)$ and (2) the forced response $y_f(t)$ due to the time-dependent sources (the solution to the nonhomogeneous differential equation).

[1] From an 1893 paper describing the phasor method for sinusoidal steady state analysis: "The method of calculation is considerably simplified. Whereas before we had to deal with periodic functions of an independent variable 'time', now we obtain a solution through the simple addition, subtraction, etc. of constant numbers … Neither are we restricted to sine waves, since we can construct a general periodic function out of its sine wave components … With the aid of Ohm's Law in its complex form any circuit or network of circuits can be analysed in the same way, and just as easily, as for direct current, provided only that all the variables are allowed to take on complex values." -C.P. Steinmetz

We know, however, that the natural response $y_n(t)$ is a sum of exponential solutions of the form $y_n(t) = Ae^{st}$. In a great many problems[2] these solutions will either be decreasing exponentials or exponentially damped sinusoids. That is, the real part of *s* will be less than zero. So in steady state ($t \to \infty$) this component of the solution will be zero.

Now if all of the sources in the problem are sinusoidal sources of a particular frequency ω, then $r(t)$ will also be a sinusoidal function with the same frequency. When this is the case, in the method of undetermined coefficients we choose as the forced response a sum of sine and cosine terms of the frequency ω. As a result, in steady state *all* of the voltages and currents involved in the problem will be sinusoidal with the *same* frequency. Sinusoidal currents and voltages are completely specified by their magnitude and phase. *Phasors* provide a particularly compact and efficient way to describe and manipulate sinusoids of a single frequency.

Representation of a sinusoid as a phasor

Suppose we have a sinusoidal voltage or current $y(t) = Y_m\cos(\omega t + \theta)$. We define the *phasor* $\hat{Y}$ representing $y(t)$ through the equation

$$y(t) = \mathrm{Re}(\hat{Y}e^{j\omega t}).$$

The phasor $\hat{Y}$ is a complex number with a unit (either V or A, depending upon whether it is a voltage or current). To determine $\hat{Y}$ we write this complex number in polar form $\hat{Y} = re^{j\theta}$; using the Euler relation we have

$$\begin{aligned} y(t) &= \mathrm{Re}(re^{j(\omega t+\theta)}) = \mathrm{Re}[r(\cos(\omega t+\theta) + j\sin(\omega t+\theta))] \\ &= r\cos(\omega t+\theta) \end{aligned} \qquad (8.2)$$

Clearly $\hat{Y} = Y_m e^{j\theta}$ is the phasor that represents $y(t) = Y_m\cos(\omega t + \theta)$.

Example 8.1. Representation of a signal as a phasor.

Determine the phasor that represents the signal $i(t) = 3\cdot\sin(10t + \pi/3)$ A.

Solution. Since $\sin(x) = \cos(x - \pi/2)$ we have $3\cdot\sin(10t + \pi/3) = 3\cdot\cos(10t - \pi/6)$. Consequently the phasor representing $i(t)$ is $\hat{I} = 3\cdot e^{-j\pi/6}$.

Note that the phasor does not contain any information about the frequency of the sinusoid.

Example 8.2. Plotting the voltage or current that corresponds to a phasor.

Suppose that the voltage across a circuit element is represented by the phasor $\hat{V} = 10\cdot e^{j0}\ V$ and the current by the phasor $\hat{I} = 5\cdot e^{j\pi/2}\ mA$. Plot $i(t)$ and $v(t)$ if the frequency is $\omega = 100$ rad/sec.

Solution. We have

$$v(t) = \mathrm{Re}(10\cdot e^{j0}e^{j100t}) = \mathrm{Re}[10\cdot\cos(100t) + 10\cdot j\cdot\sin(100t)] = 10\cdot\cos(100t)\ \mathrm{V}$$

and

$$i(t) = \mathrm{Re}(5\cdot e^{j\pi/2}e^{j100t}) = \mathrm{Re}(5\cdot\cos(100t + \pi/2) + 5\cdot j\cdot\sin(100t+\pi/2)) = 5\cdot\cos(100t + \pi/2)\ \mathrm{mA}.$$

The resulting signals are plotted in Figure 8.E1.

[2] The exceptions will be circuits with no loss (no resistors) and circuits that contain dependent sources and that are unstable. Examples of unstable circuits include op amp circuits with positive feedback and circuits containing active devices that are designed to be oscillators. It can be shown that circuits containing only passive components and independent sources will *always* have real parts of *s* less than or equal to zero.

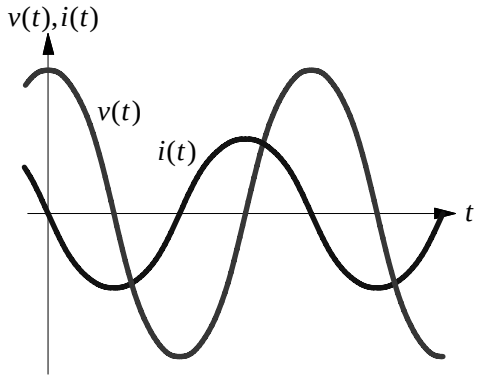


Figure 8.E1. Plots of sinusoids. At t = 0 the voltage has the value +10 V and the current is 0 A.

Plotting the sinusoidal waveforms that correspond to phasors can be simplified in the following way. Since $y(t) = \mathrm{Re}(\hat{Y}e^{j\omega t})$ the value of the signal $y(t)$ at t = 0 is given by $y(0) = \mathrm{Re}(\hat{Y}e^{j\omega 0}) = \mathrm{Re}(\hat{Y})$. That is, the real part of $\hat{Y}$ - or the projection of the complex number $\hat{Y}$ on the real axis- is the value of the signal at t = 0. Increasing t corresponds to multiplying $\hat{Y}$ by $e^{j\omega t}$ which is the same as rotating $\hat{Y}$ counterclockwise by the angle ωt. So we can imagine the phasor $\hat{Y}$ rotating counterclockwise about the origin of the complex plane. At any instant the projection on the real axis is the value of the signal y. This is illustrated in Figure 8.8.1 for the phasors in Example 8.2.

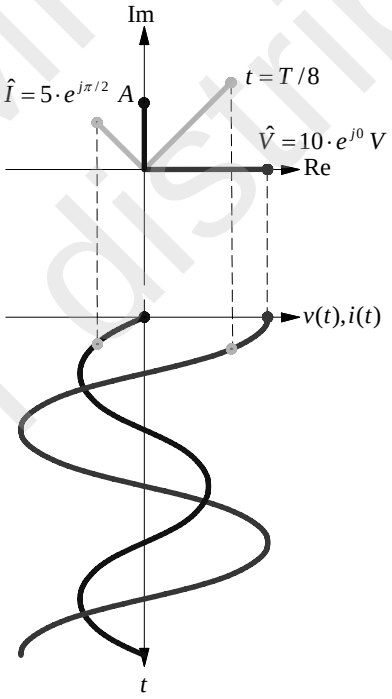


Figure 8.8.1. Visualization of the time dependence of two phasors. At t = 0 the voltage has the value +10 V and the current is 0 A. For t > 0 the two phasors rotate together and make one revolution in one period of the sinusoid ($T = 2\pi/\omega$). At any time the voltage and current can be determined from the projections of the current and voltage phasors on the real axis.

Finally, we introduce a compact notation for phasors which is commonly used in engineering. Instead of the polar form $\hat{Y} = Y_m e^{j\phi}$ we will sometimes write $\hat{Y} = Y_m \angle \phi$. In this last notation ϕ may be written either in radians or in degrees. Beware.

Element relations in phasor form

We will now derive the element relations for the inductor, capacitor, and resistor in phasor form. Recall that for an inductor

$$v_L(t) = L\frac{di_L}{dt}. \tag{8.3}$$

The signals $i_L(t)$ and $v_L(t)$ are related to the phasors $\hat{I}_L$ and $\hat{V}_L$ through

$$\begin{aligned} v_L(t) &= \mathrm{Re}[\hat{V}_L e^{j\omega t}] \\ i_L(t) &= \mathrm{Re}[\hat{I}_L e^{j\omega t}] \end{aligned} \tag{8.4}$$

so substituting into (8.3) gives

$$\mathrm{Re}[\hat{V}_L e^{j\omega t}] = L\frac{d}{dt}\mathrm{Re}[\hat{I}_L e^{j\omega t}] = \mathrm{Re}[\, j\omega L \cdot \hat{I}_L e^{j\omega t}] \tag{8.5}$$

or

$$\hat{V}_L = j\omega L \cdot \hat{I}_L. \tag{8.6}$$

Similarly from the branch relations for the capacitor and the resistor

$$\hat{I}_C = j\omega C \cdot \hat{V}_C \tag{8.7}$$

and

$$\hat{V}_R = R \cdot \hat{I}_R. \tag{8.8}$$

The relations for the capacitor and inductor can be thought of as generalizations of Ohm's law for energy-storage elements. All the current-voltage relations can be written in the general form

$$\hat{V}_Z = \boldsymbol{Z} \cdot \hat{I}_Z \tag{8.9}$$

where $\boldsymbol{Z}$ is termed the *impedance* of an element, $\hat{V}_Z$ is the phasor representing the voltage across that element, and $\hat{I}_Z$ the phasor representing the current flowing through that element. The impedances for the various elements are

$$\begin{aligned} \boldsymbol{Z_C} &= \frac{1}{j\omega C} \\ \boldsymbol{Z_L} &= j\omega L \\ \boldsymbol{Z_R} &= R \end{aligned}. \tag{8.10}$$

It is important to recognize the difference between a phasor and an impedance. Both are complex numbers but they have different units and different significance. Phasors represent a voltage or a current; they have the units of voltage or current and the corresponding time-dependent signal can be obtained by multiplying by $e^{j\omega t}$ and taking the real part. Impedances have the units of ohms and relate voltage and current phasors. The no-