Practical Electronics
for Optical Design and Engineering

Tutorial Texts Series

- *Automatic Target Recognition*, Bruce J. Schachter, TT105
- *Design Technology Co-optimization in the Era of Sub-resolution IC Scaling*, Lars W. Liebmann, Kaushik Vaidyanathan, and Lawrence Pileggi, Vol. TT104
- *Special Functions for Optical Science and Engineering*, Vasudevan Lakshminarayanan and L. Srinivasa Varadharajan, Vol. TT103
- *Discrimination of Subsurface Unexploded Ordnance*, Kevin A. O'Neill, Vol. TT102
- *Introduction to Metrology Applications in IC Manufacturing*, Bo Su, Eric Solecky, and Alok Vaid, Vol. TT101
- *Introduction to Liquid Crystals for Optical Design and Engineering*, Sergio Restaino and Scott Teare, Vol. TT100
- *Design and Implementation of Autostereoscopic Displays*, Byoungho Lee, Soon-gi Park, Keehoon Hong, and Jisoo Hong, Vol. TT99
- *Ocean Sensing and Monitoring: Optics and Other Methods*, Weilin Hou, Vol. TT98
- *Digital Converters for Image Sensors*, Kenton T. Veeder, Vol. TT97
- *Laser Beam Quality Metrics*, T. Sean Ross, Vol. TT96
- *Military Displays: Technology and Applications*, Daniel D. Desjardins, Vol. TT95
- *Interferometry for Precision Measurement*, Peter Langenbeck, Vol. TT94
- *Aberration Theory Made Simple, Second Edition*, Virendra N. Mahajan, Vol. TT93
- *Modeling the Imaging Chain of Digital Cameras*, Robert D. Fiete, Vol. TT92
- *Bioluminescence and Fluorescence for In Vivo Imaging*, Lubov Brovko, Vol. TT91
- *Polarization of Light with Applications in Optical Fibers*, Arun Kumar and Ajoy Ghatak, Vol. TT90
- *Digital Fourier Optics: A MATLAB Tutorial*, David G. Voeltz, Vol. TT89
- *Optical Design of Microscopes*, George Seward, Vol. TT88
- *Analysis and Evaluation of Sampled Imaging Systems*, Richard H. Vollmerhausen, Donald A. Reago, and Ronald Driggers, Vol. TT87
- *Nanotechnology: A Crash Course*, Raúl J. Martin-Palma and Akhlesh Lakhtakia, Vol. TT86
- *Direct Detection LADAR Systems*, Richard Richmond and Stephen Cain, Vol. TT85
- *Optical Design: Applying the Fundamentals*, Max J. Riedl, Vol. TT84
- *Infrared Optics and Zoom Lenses, Second Edition*, Allen Mann, Vol. TT83
- *Optical Engineering Fundamentals, Second Edition*, Bruce H. Walker, Vol. TT82
- *Fundamentals of Polarimetric Remote Sensing*, John Schott, Vol. TT81
- *The Design of Plastic Optical Systems*, Michael P. Schaub, Vol. TT80
- *Fundamentals of Photonics*, Chandra Roychoudhuri, Vol. TT79
- *Radiation Thermometry: Fundamentals and Applications in the Petrochemical Industry*, Peter Saunders, Vol. TT78
- *Matrix Methods for Optical Layout*, Gerhard Kloos, Vol. TT77
- *Fundamentals of Infrared Detector Materials*, Michael A. Kinch, Vol. TT76
- *Practical Applications of Infrared Thermal Sensing and Imaging Equipment, Third Edition*, Herbert Kaplan, Vol. TT75
- *Bioluminescence for Food and Environmental Microbiological Safety*, Lubov Brovko, Vol. TT74
- *Introduction to Image Stabilization*, Scott W. Teare and Sergio R. Restaino, Vol. TT73
- *Logic-based Nonlinear Image Processing*, Stephen Marshall, Vol. TT72
- *The Physics and Engineering of Solid State Lasers*, Yehoshua Kalisky, Vol. TT71
- *Thermal Infrared Characterization of Ground Targets and Backgrounds, Second Edition*, Pieter A. Jacobs, Vol. TT70
- *Introduction to Confocal Fluorescence Microscopy*, Michiel Müller, Vol. TT69
- *Artificial Neural Networks: An Introduction*, Kevin L. Priddy and Paul E. Keller, Vol. TT68
- *Basics of Code Division Multiple Access (CDMA)*, Raghuveer Rao and Sohail Dianat, Vol. TT67
- *Optical Imaging in Projection Microlithography*, Alfred Kwok-Kit Wong, Vol. TT66
- *Metrics for High-Quality Specular Surfaces*, Lionel R. Baker, Vol. TT65
- *Field Mathematics for Electromagnetics, Photonics, and Materials Science*, Bernard Maxum, Vol. TT64
- *High-Fidelity Medical Imaging Displays*, Aldo Badano, Michael J. Flynn, and Jerzy Kanicki, Vol. TT63

(For a complete list of Tutorial Texts, see http://spie.org/tt.)

Practical Electronics
for Optical Design and Engineering

Scott W. Teare

Tutorial Texts in Optical Engineering
Volume TT107

SPIE PRESS
Bellingham, Washington USA

Library of Congress Cataloging-in-Publication Data

Names: Teare, Scott W., author.
Title: Practical electronics for optical design and engineering / Scott W. Teare.
Other titles: Tutorial texts in optical engineering ; v. TT 107.
Description: Bellingham, Washington : SPIE Press, [2016] | Series: Tutorial texts in optical engineering ; volume TT 107 | Includes bibliographical references and index.
Identifiers: LCCN 2016019252 | ISBN 9781510603622 (print) | ISBN 151060362X (print) | ISBN 9781510603639 (pdf) | ISBN 1510603638 (pdf) | ISBN 9781510603646 (epub) | ISBN 1510603646 (epub) | ISBN 9781510603653 (Kindle/mobi) | ISBN 1510603654 (Kindle/mobi)
Subjects: LCSH: Electronics. | Electronic circuit design. | Optical engineering.
Classification: LCC TK7835 .T29 2016 | DDC 621.36—dc23 LC record available at https://lccn.loc.gov/2016019252

Published by

SPIE
P.O. Box 10
Bellingham, Washington 98227-0010 USA
Phone: +1 360.676.3290
Fax: +1 360.647.1445
Email: books@spie.org
Web: http://spie.org

Printed in the United States of America.
First Printing.
For updates to this book, visit http://spie.org and type “TT107” in the search field.

Introduction to the Series

Since its inception in 1989, the Tutorial Texts (TT) series has grown to cover many diverse fields of science and engineering. The initial idea for the series was to make material presented in SPIE short courses available to those who could not attend and to provide a reference text for those who could. Thus, many of the texts in this series are generated by augmenting course notes with descriptive text that further illuminates the subject. In this way, the TT becomes an excellent stand-alone reference that finds a much wider audience than only short course attendees.

Tutorial Texts have grown in popularity and in the scope of material covered since 1989. They no longer necessarily stem from short courses; rather, they are often generated independently by experts in the field. They are popular because they provide a ready reference to those wishing to learn about emerging technologies or the latest information within their field. The topics within the series have grown from the initial areas of geometrical optics, optical detectors, and image processing to include the emerging fields of nanotechnology, biomedical optics, fiber optics, and laser technologies. Authors contributing to the TT series are instructed to provide introductory material so that those new to the field may use the book as a starting point to get a basic grasp of the material. It is hoped that some readers may develop sufficient interest to take a short course by the author or pursue further research in more advanced books to delve deeper into the subject.

The books in this series are distinguished from other technical monographs and textbooks in the way in which the material is presented. In keeping with the tutorial nature of the series, there is an emphasis on the use of graphical and illustrative material to better elucidate basic and advanced concepts. There is also heavy use of tabular reference data and numerous examples to further explain the concepts presented. The publishing time for the books is kept to a minimum so that the books will be as timely and up-to-date as possible. Furthermore, these introductory books are competitively priced compared to more traditional books on the same subject.

When a proposal for a text is received, each proposal is evaluated to determine the relevance of the proposed topic. This initial reviewing process has been very helpful to authors in identifying, early in the writing process, the need for additional material or other changes in approach that would serve to strengthen the text. Once a manuscript is completed, it is peer reviewed to ensure that chapters communicate accurately the essential ingredients of the science and technologies under discussion.

It is my goal to maintain the style and quality of books in the series and to further expand the topic areas to include new emerging fields as they become of interest to our reading audience.

James A. Harrington
Rutgers University

Contents

Preface *xi*

Acronyms and Abbreviations *xiii*

I Basic Electronics **1**

1 Introduction to Electronics **3**

- 1.1 Ohm's Law 4
- 1.2 Simple LED Circuits 5
- 1.3 Resistors 6
- 1.4 Signals 11
- 1.5 Measuring Instruments 12
- 1.6 Voltage Dividers and Regulators 15
- 1.7 Device Data Sheets 19
- 1.8 Practice Problems 19
- References 20

2 Common Electronic Circuits and Components **21**

- 2.1 Bridges and Balance Measurements 21
- 2.2 Diodes 27
- 2.3 Transformers and Inductors 29
- 2.4 Transistor Switches and Amplifiers 30
- 2.5 Power Transistors 32
- 2.6 Combining Transistors 34
- 2.7 Digital-to-Analog Conversion 35
- 2.8 Practice Problems 36
- References 37

3 Linear Amplifiers **39**

- 3.1 Inverting Operational Amplifier Configuration 40
- 3.2 Non-inverting Operational Amplifier Configuration 42
- 3.3 Operational Amplifiers: Math Circuits 44
- 3.4 Differential Amplifiers 46
- 3.5 Transfer Functions, Imaginary Numbers, and Decibels 48
- 3.6 Bode Plots 49

3.7 Practice Problems 50
References 51

4 Useful Op-Amp Circuits 53
4.1 Instrumentation Amplifiers 53
4.2 Improved Precision Rectifiers 57
4.3 Peak Detectors 58
4.4 Oscillators 59
4.5 Sample-and-Hold Circuits 59
4.6 Voltage Indicators 60
4.7 Practice Problem 61
References 62

5 Digital Electronics 63
5.1 Elements of Digital Logic 63
5.2 Binary Logic 66
5.3 Flip-Flops and Latches 69
5.4 Programmable Integrated Circuits 73
5.5 Microcontrollers 74
5.6 Field-Programmable Gate Arrays 76
5.7 Practice Problems 77
References 77

6 Instrumentation, Signal Conditioning, and Filters 79
6.1 Instrumentation 80
6.2 Passive Analog Filters 81
6.3 Active Analog Filters 87
6.4 Analog-to-Digital Conversion 88
6.5 Digital Filters and Signal Processing 89
6.6 Noise and Analysis 90
6.7 Practice Problems 92
References 92

II Optics Applications 93

7 Solar Cells and Rechargeable Batteries 95
7.1 Solar Cells 95
7.2 Batteries and Charging 96
7.3 Solar Cells and Panels 98
7.4 Solar-Panel-based Battery Charger 101
7.5 Fast Chargers 102
7.6 Solar Panels and Fast Chargers 103
7.7 Going Farther 105
7.8 Practice Problems 106
References and Notes 106

8 Photodiodes **107**
8.1 Light Detectors 107
8.2 Transimpedance Amplifiers 109
8.3 Amplifier Chaining 111
8.4 Removing Unwanted Effects 111
8.5 Photodiode Applications 112
8.6 Capacitance and Oscillation in Amplifiers 116
8.7 Practice Problems 118
References 118

9 Quad-Cells and Position-Sensitive Detectors **119**
9.1 Quad-Cells 119
9.2 Quad-Cell Mathematics 121
9.3 Quad-Cell Electronics 124
9.4 Position-Sensitive Detectors 125
9.5 Spot Location in Position-Sensitive Detectors 127
9.6 Position-Sensitive Sensor Electronics 127
9.7 Practice Problems 128
References 129

10 Proportional-Integral-Derivative Controllers **131**
10.1 Controllers 131
10.2 Proportional-Integral-Derivative Controllers 133
10.3 Bang-Bang Controllers 134
10.4 Proportional Controllers 134
10.5 Proportional-Integral Controllers 137
10.6 Proportional-Integral-Derivative Controller Redux 139
10.7 Proportional-Integral-Derivative Electronics 141
References 143

11 Strain Gauges **145**
11.1 Strain Measurement 145
11.2 Gauge Factor 147
11.3 Strain Gauge Calculations and Electronics 148
11.4 Strain Gauge Signal Processing Electronics 149
11.5 Load Cell to Microcontroller 150
11.6 Strain Gauge Applications in Optics 151
11.7 Practice Problems 154
References 154

12 Stepper Motors and Actuators **157**
12.1 Stepper Motors 158
12.2 Stepper Motor Signals 159
12.3 H-Bridge 160
12.4 Microcontroller-Driven Stepper Motor 160

12.5 Actuators 162
12.6 Feedback Sensors 163
12.7 Practice Problems 165
References 165

13 High and Higher Voltage 167
13.1 Transformers 167
13.2 Quartz Crystal Oscillators 169
13.3 Integrated-Circuit Oscillators 171
13.4 Resonant Oscillators 173
13.5 Cockcroft–Walton Voltage Multipliers 173
13.6 High-Voltage Safety Dos and Don’ts 177
13.7 Practice Problems 178
References 178

III Projects and Finishing 179

14 Circuit Boards, Connectors, and Packaging 181
14.1 Circuit Boards for Prototyping and Finished Products 181
14.2 Wall Transformers and Power 186
14.3 Cables and Connectors 188
14.4 Fuses and Circuit Breakers 189
14.5 Transient Voltage Control 190
14.6 Packaging 191
References 191

15 Tying It All Together: From Sensor to Computer 193
15.1 Instrument Overview 193
15.2 Analog Circuit 195
15.3 Microcontroller Organization 196
15.4 Error Trapping, Management, and Log Files 197
15.5 Data Storage and Analysis 198
15.6 Summary 198
References 199

Appendix: MATLAB® Introduction and Refresher 201
A.1 MATLAB Introduction 201
A.2 MATLAB Basic Commands 202
A.3 MATLAB Refresher 203
A.4 MATLAB Scripts 205
A.5 MATLAB Commands: A Closer Look 209
References 212

Bibliography 213
Index 217

Preface

Optical engineers make use of a wide range of sensors and controllers in optical systems. Photodetectors, cameras, and actuator systems all use electronics and electrical control systems to operate. There are a wide range of manufacturers of electronics systems, but sometimes all we need is a simple circuit to read a sensor or to condition a signal. Whether the needed electronics are to support a photodiode or a strain gauge, or just to make a trigger to start several instruments simultaneously, quite often the needed components are readily available on the bench. However, the how-to and confidence to start assembling the needed device may be lacking. Once a circuit has been assembled, the question arises of how to properly check its operation. This might require the use of tools that perhaps haven't been touched since that one undergraduate circuits class that still causes those exam anxiety nightmares!

Practical Electronics for Optical Design and Engineering has been organized into three parts: Basic Electronics, Optical Applications, and Projects and Finishing. The first part, Basic Electronics, focuses on a wide range of fundamental circuits important in understanding and working with electronics including making electronic measurements and techniques for moving the data from a sensor through to a computer. Optical Applications is composed of seven chapters and builds on the previous material, introducing specific electronics of interest to optical engineers. Projects and Finishing provides some ideas about how to complete projects and works through the development of an example instrument.

The book is designed so that you can get started on any chapter that catches your attention and seek more specialized information from the earlier chapters as needed. Some examples of these interesting circuits are transimpedance amplifiers and drivers for low-light photodiodes, using solar cells as a power source or battery charger, low-resolution digital-to-analog converts, analog controllers, quad-cell processing, and analog control circuits. Discussions on how practical electronics work, their design, and translating this to circuit board manufacturing, as well as the limitations of different prototyping approaches are included.

Examples throughout the book range from simple calculations to sample MATLAB® scripts. You are encouraged not only to work the examples and use the MATLAB code, but to construct and test the circuits. The aim of the MATLAB-based examples is to support an understanding of the fundamentals and relationships behind the electronics and to provide a starting point for your own code.

I hope that *Practical Electronics for Optical Design and Engineering* provides the interested reader with a functional overview of the topic of electronics and an appreciation for the way knowledge of electronics can enhance optical projects. While this book is not meant to be a complete treatise on electronics, as there are many excellent books on the topic, the aim here is to provide an introduction more closely tied to the needs of those working in optical engineering and design.

I greatly appreciate all of the colleagues and friends who have both directly and indirectly helped me in preparing and writing this book, and I am grateful for their unswerving and unselfish support. I also appreciate the feedback from the many students who over the years have helped me refine my optics and electronics lectures and laboratories. I am particularly grateful to my grandfather, Stephen Holmes Scott, who introduced me to electronics many years ago.

While I have benefited from the support of many individuals in preparing this work, any errors that remain in the text are mine to fix. I would appreciate receiving any assistance in this in the form of comments and corrections. Please direct any correspondence to the author c/o New Mexico Tech, Electrical Engineering Department, Socorro, NM 87801, USA.

I am most grateful for the support of SPIE Press for their interest in publishing this work as part of the *Tutorial Text* Series, and particularly the efforts of Senior Editor, Dara Burrows, for putting this work into its final form.

Scott W. Teare
Professor of Electrical Engineering
New Mexico Institute of Mining and Technology
Socorro, New Mexico
May 2016

Acronyms and Abbreviations

AC	alternating current
ADC	analog-to-digital converter
ADU	arithmetic data unit
AGM	absorbed glass mat
AI	analog input
BASIC	Beginner's All-purpose Symbolic Instruction Code
BJT	bipolar junction transistor
BNC	Bayonet Neill–Concelman (coaxial connector)
BW	bandwidth
DC	direct current
DIO	digital input/output
DIP	dual in-line package
DSP	digital signal processor
DUT	device under test
DVM	digital voltmeter
FPGA	field-programmable gate arrays
GBP	gain–bandwidth product
hr	hour (unit)
IC	integrated circuit
IDE	integrated development environment
LED	light-emitting diode
LSB	least significant bit
MOSFET	metal-oxide semiconductor field-effect transistor
MSB	most significant bit
op-amp	operational amplifier
PI	proportional-integral
PIC	programmable integrated circuit
PID	proportional-integral-derivative
PSD	position-sensitive detector
RCA	Radio Corporation of America (coaxial connector)
RMS	root mean square
SD	Secure Digital
SMA	sub-miniature version A (coaxial connector)

SMP	surface-mount package
S-R	set-reset
TPI	threads per inch
USB	universal serial bus
VAC	volts of alternating current
VDC	volts of direct current
VOM	volt-ohm meter

Part I: Basic Electronics

Chapters 1 through 6 provide an overview of basic electronic circuits and set the foundation for the applications to be introduced in Part II. Many readers will find the information in this section to be a review; however, those who are new to electronic circuits will find this a useful introduction to these fundamental topics.

Safety Warning
Always wear safety glasses when working on or around electronic circuits and follow all equipment and component manufacturers' safety instructions. Circuits included in these chapters may not be suitable for a particular application.

Chapter 1

Introduction to Electronics

Electronics is about manipulating electricity to accomplish a particular task and is very much a hands-on endeavor. Since the result of building electronic circuits is usually a device that performs a task, this hands-on aspect should be self-evident. What is often obscured in the study of electronics is that even the most complicated electronic device is made up of many smaller and simpler circuits and components. Once you have an appreciation for some of these smaller circuits, it is rather easy to understand more complicated circuits. It should not be very surprising then that the focus of this book is on very practical things that, when several are combined, result in electronics and electronic systems of interest and use to those involved in optical engineering.

During their university training, many future engineers often discover in their first circuits class that electrical engineering is not for them. Some of these engineers still have the recurring nightmare that they have their final circuits test the next day! In spite of the loathing that those engineering circuits classes evoke, as topics go, understanding electronics is really not that difficult and can actually be fun if it is approached in the right way.

The right way that we will be using here is to break electronics into two main parts. The stuff you *need* to know and the fun stuff you *want* to know! Of course, before we can get to the fun stuff, we have to introduce some fundamental concepts and circuits. We will start our journey into the world of electronics by focusing on signals, simple circuits, common devices, and how to make measurements. Electronics is meant to be hands-on, so constructing some of the circuits will help you understand them, and you are encouraged to do so! Example calculations and MATLAB® scripts are provided to help you appreciate how electronics problems can be solved numerically. A brief introduction and refresher to MATLAB is provided in the Appendix, along with a Bibliography of suggested reading in case you want to find some additional information.

1.1 Ohm's Law

Electronics involves the manipulation of charges, currents, and voltages. When many charges are moving in the same direction, we tend to speak of them as currents, defined as the passage of charge per unit of time. Currents can only flow if there is a complete path from once side of a power source to the other, and if the total voltage around a closed loop is zero, according to a principle known as Kirchhoff's voltage law.[1] Charges move under the influence of voltage differences, and changes in voltage can occur as currents move through resistances. This relationship is reflected in the basic rule of electronics known as Ohm's law.[1]

Ohm's law is expressed as a mathematical relationship that governs the interplay of voltage, current, and resistance, and is often stated as

$$V = I \cdot R, \tag{1.1}$$

where V is the voltage in units of volts [V], I is the current in units of amperes [A], and R is the resistance in units of ohms [Ω]. The utility of the equation will be made clear shortly. For the moment, just remember that Ohm's law is frequently used in electronics problems; fortunately, it is not very complicated.

Power describes the rate at which energy is being used. If we know the current and the voltage, we can directly calculate the power from

$$P = V \cdot I, \tag{1.2}$$

where P is the power and is measured in watts [W]. Other relationships can be created for the power by substituting in Eq. (1.1) after rearranging for either voltage or current. The value of these simple equations will become clear in just a couple paragraphs when we start to investigate a basic circuit.

The first circuit we will look at is a traditional one for introducing electronics; it is a circuit that turns on a light source using a battery and a switch. To make this circuit more interesting, we will introduce the light-emitting diode (LED) as our light source. The LED is a semiconductor optoelectronic component that is much more energy efficient than an incandescent lamp. Today, these devices are nearly as popular as the old-style "grain-of-wheat" lamps, and we will use them in many of our example circuits.

An LED is an electrical device that converts electrical power to light. The operating characteristic for an LED can be described in the simplest sense as a forward voltage and a current; in other words, the LED is sensitive to the direction of current flow. We can use this information and Eq. (1.1) to determine the resistance of the LED and its operating power, as shown in Example 1.1.

Example 1.1

The forward voltage V_f and forward current I_f for an operating red LED are 2.2 V and 20 mA, respectively. Calculate the resistance and power used.

The resistance of the LED is

$$R = \frac{V}{I} = \frac{V_f}{I_f} = \frac{2.2\ \text{V}}{20\ \text{mA}} = 110\ \Omega.$$

The power drawn by the LED is

$$P = V \cdot I = 2.2\ \text{V} \cdot 20\ \text{mA} = 44\ \text{mW}.$$

As shown in Example 1.1, we can determine very practical information about the LED using our simple relationships. So far, we have used resistance as a calculated value without stating what it means. Resistance is the ability to impede the flow of current, usually by conversion to heat, and is a very useful property. Resistors are components that provide a fixed amount of voltage drop for a given current.

1.2 Simple LED Circuits

Powering the LED requires a closed circuit path in the form of a battery that connects the power source to the LED, as well as a switch, which is a mechanical means of turning, connecting, and disconnecting the power to the LED. If we use the same LED as in Example 1.1, we will need a 2.2-V battery, which is a little hard to find. To step down the voltage from a more common battery, say a 9-V battery to provide 2.2 V, we use a resistor of the correct value. Resistors will be discussed in detail in the next section, but for now we will consider them as a means of clarifying the voltage drop relationship.

A circuit to operate the LED is shown in Fig. 1.1, where a battery is connected using wires through a switch to a resistor and an LED light source. When the switch is closed, it acts as a wire connecting both sides so that a current flows through the resistor and the LED. There is a voltage drop across the resistor that can be calculated using Eq. (1.1). If the voltage drop is too great, there might not be sufficient power for the LED to turn on, or at least the LED might not be bright enough to see. Increasing the battery voltage or

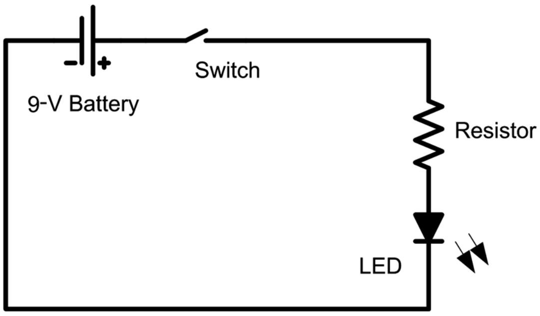


Figure 1.1 A battery wired to a switch, resistor, and LED, and relevant features.

decreasing the resistance can be tried to make the LED turn on. This calculation is performed in Example 1.2.

Example 1.2

Calculate the resistance needed so that a 9-V battery can be used to power an LED that requires 2.2 V and 20 mA to operate.

The resistance or load connected to a 9-V battery that will draw a total current of 20 mA is calculated from Ohm's law as

$$R = \frac{9\ \text{V}}{20\ \text{mA}} = 450\ \Omega.$$

The LED has an operating resistance of 110 Ω, so a resistor of 450–110 Ω will be needed to lower the voltage. The voltage drop across the resistor with a current of 20 mA will be

$$V = \frac{340\ \Omega}{20\ \text{mA}} = 6.8\ \text{V}.$$

The conclusion from Example 1.2 could also be stated as: Knowing that the current through the LED is 20 mA and that the operating voltage is 2.2 V, we need to lower the voltage by 9 – 2.2 or 6.8 V using a 20-mA current or 340 Ω.

Now many interesting details were introduced in this section as we made our decision on powering the LED. To understand this further, we need to look at how the basic electronic components are used. We will begin with a closer look at resistors.

1.3 Resistors

We have shown the nature of voltage and current in terms of driving force and moving charge provided by a battery, but we haven't said much about resistors. Resistors have many roles; in the last section, a resistor was used to drop the voltage going into the LED. Resistors do this, for the most part, by converting current to heat. Resistors are physical devices and are available in compact packages whose sizes relate to their ability to dissipate power without damage. The power-handling capability of a resistor is measured in units of watts.

Common resistors are made of carbon between two wires. It can be interesting to break one open and see inside! The physical size of the resistor relates to the amount of carbon in the resistors and to the amount of power a resistor can dissipate. The resistance value is indicated on a resistor by either a number or a set of colored bands that can be translated into the value in ohms.

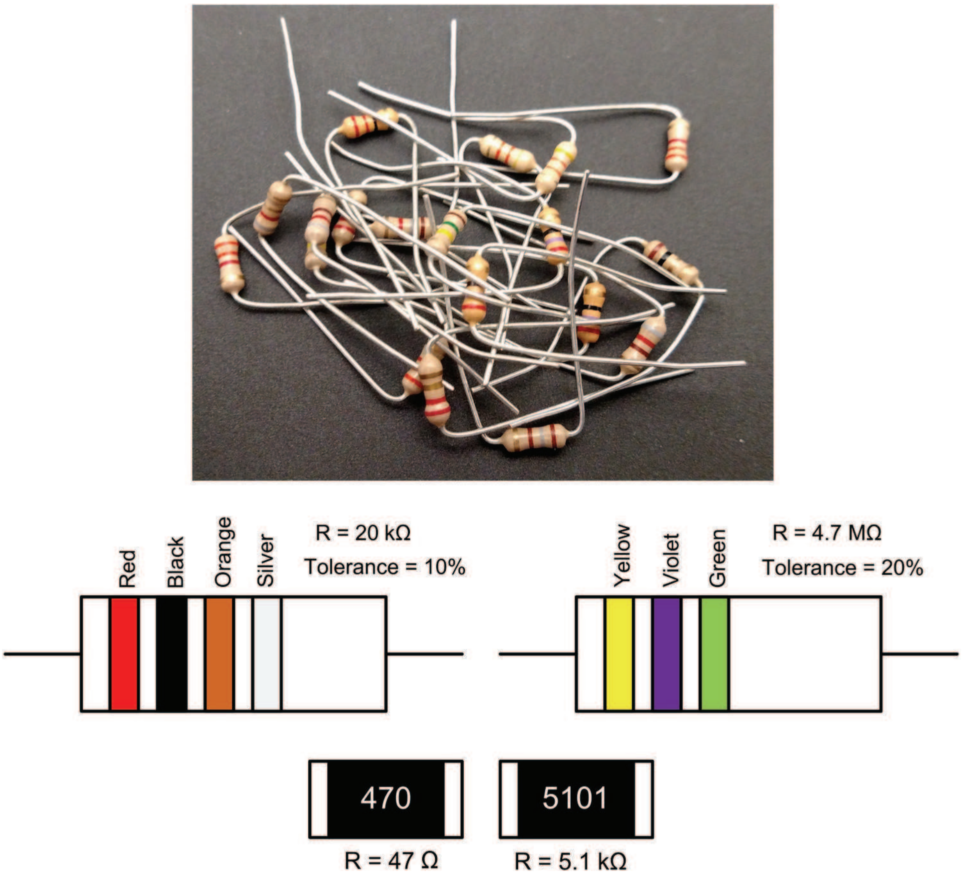


Figure 1.2 Top: A collection of various low-power resistors showing colored bands that define their resistance values. Bottom: Illustration of color bands and surface mount resistors with their values.

Colored bands, some examples of which are shown in Fig. 1.2, are the most common way to identify resistors.[2]

The colored bands, either three or four colored stripes around the body of the resistor, are located toward one end. The colored bands indicate the resistance value and the tolerance or how close the actual resistor value will be to the stated value. There are three common tolerance levels: ±5 is represented by Gold, ±10 is represented by Silver, and ±20% is represented by only three bands.

The key to translating the individual colors into numbers is shown in Table 1.1. To simplify this discussion, the first three colored bands, referred to as A, B, and C, are used to indicate the value of the resistor. The resistor value can be determined from the following equation:

$$Value = (10A + B)10^{C}. \tag{1.3}$$

Table 1.1 The four-band resistor color code decoded.

Color Band	Corresponding Number (A, B, C)	Multiplier (10^C)
Black	0	1
Brown	1	10
Red	2	100
Orange	3	1000
Yellow	4	10000
Green	5	100000
Blue	6	1000000
Violet	7	10000000
Gray	8	–
White	9	–
Gold	–	0.1
Silver	–	0.01

Notice that the third band C is used to multiply the first two digits by a factor of 10 raised to the power of C. A simple example can be used to show how this works: a three-band resistor of Red Red Orange would convert to numbers as 2 2 3 and thus be combined as 22 times 10 to the power 3, or 1000, for a resistor value of 22,000 Ω. As there are only three bands, the resistor has a tolerance of ±20%, or $R = 22000 \pm 4400\ \Omega$. If we measure the resistor's value, we would expect it to lie between 17,600 and 26,400 Ω.

So, really, the resistor color code is just shorthand for identifying the resistor value. The lower the tolerance range the more expensive the resistor is to purchase. There are also high-precision resistors that are considerably more costly to purchase so are used only for very specialized work.

The resistors with wire leads are referred to as "through hole" components; i.e., the wire ends can be poked into a hole for mounting and connection. The trend these days is toward smaller electronics, so parts are also being reduced in size. These parts mount differently and are soldered onto metal pads on the surface of an electronics board. Surface-mount resistors look like small black rectangles, as shown in the bottom of Fig. 1.2, and are often only a few millimeters in size. The numbers on these resistors refer to their resistance value, with the last number being the multiplier. There are some subtleties in the interpretation of surface-mount resistors as well as some new coding systems. For low-valued resistors, an R is used to indicate where the decimal point is located. For instance, R470 would indicate a 0.47-Ω resistor. The EIA-96 code is for 1% tolerance resistors and is a little more complicated to interpret; users should check with the manufacturer of their parts.[3]

There are occasions when a very precise resistor value is needed that is not a standard resistor value. Of course, one could purchase a large number of low-tolerance resistors and measure their individual values in the hope of finding the value needed; however, we can also construct the resistance we need by combining other resistors. We can wire resistors together in two common forms, series and parallel, as shown in Fig. 1.3.

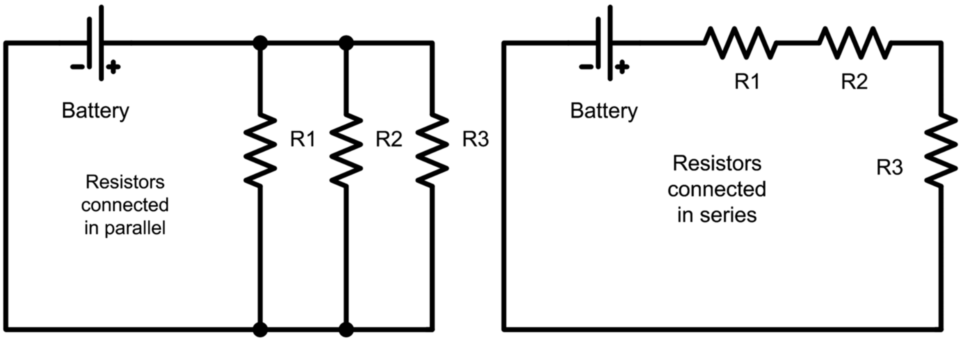


Figure 1.3 Two circuits showing resistors connected in parallel (left) and in series (right).

In this book we use the convention that straight wires that cross are not connected unless there is a dot at the crossing point.

Resistors can be combined into an effective value based on two mathematical expressions shown in Eqs. (1.4) and (1.5) for series and parallel resistors, respectively:[1]

$$R_{series} = \sum_i R_i, \tag{1.4}$$

$$\frac{1}{R_{parallel}} = \sum_i \frac{1}{R_i}, \tag{1.5}$$

where R_i show the individual resistors, and $i = 1,2,3 \ldots N$, where N is the maximum number of individual resistors to be summed.

There are some interesting limiting cases for combining resistors in series and parallel. If we have two identical resistors of value R connected in series, then the effective resistance is $2R$, while if they are connected in parallel, the effective resistance is $R/2$. If we have two resistors that are several orders of magnitude different in value, when they are connected in series, the effective resistance is approximately the larger of the two resistors; when they are connected in parallel, the effective resistance is the smaller of the two resistors. This is demonstrated in Example 1.3.

Example 1.3

Two resistors, $R1 = R$ and $R2 = 1000R$ are connected first in series and then in parallel. Using Eqs. (1.3) and (1.4), calculate the effective resistance in each case.

$$R_{series} = R1 + R2 = R + 1000R = 1001R \sim 1000R,$$

$$\frac{1}{R_{parallel}} = \frac{1}{R1} + \frac{1}{R2} = \frac{1}{R} + \frac{1}{1000R} = \frac{1000+1}{1000R} \sim \frac{1}{R},$$

$$\text{or } R_{parallel} = \text{R}.$$

Table 1.2 Common unit multiplier names and values.

Prefix	Value
k (kilo)	10^3
M (mega)	10^6
p (pico)	10^{-12}
n (nano)	10^{-9}
μ (micro)	10^{-6}
m (milli)	10^{-3}

As resistor values become larger, it is common to introduce multipiers to the unit notation based on various multipliers of 10. The letters used to represent the multipliers are taken from the metric system. As such, a resistor of 1000 Ω is commonly shown as 1 kΩ, and a resistor of 1,000,000 Ω as 1 MΩ. The common abbreviations for electronics are shown in Table 1.2. A script is shown in Example 1.4 that implements Eqs. (1.3) and (1.4) for rapid calculation of combined resistor values based on they way they are connected.

Example 1.4 is our first MATLAB script, making this a great place to get familiar with how to write a simple calculator. The goal is very modest—calculating the value of a combination of two resistors—but it also illustrates how to write a simple script. Notice that there are many lines that begin with "%" symbols.[4] These lines are called comments and are tell the reader of the script what is occurring. They are ignored by the computer and can also come after the ";" symbols, which suppress printing output. The first noncomment lines clear the variables and the command window. The next lines input the data values and provide information to the screen. Once the calculations are performed, their values are formatted and printed to the screen. Implement this script in your version of MATLAB and explore the effects of various changes.

Example 1.4

Write a MATLAB script to calculate the equivalent resistance of a 200-Ω and 400-Ω resistor connected first in series and then in parallel.

```
% Example_1_4.m
% SWT 9-4-15

% This script takes two resistor values and calculates the
% effective series and parallel resistance.

% Housekeeping
clear all; % clears the variable list and starts fresh
clc; % clears the command window

% Input values
%% Use the MATLAB input() command to request data
```

```
fprintf(‘Effective resistance calculator\n\n’);
prompt = ‘Enter the value of resistor 1 in ohms: ’;
R1 = input(prompt);

prompt = ‘Enter the value of resistor 2 in ohms: ’;
R2 = input(prompt);

%% Calculation
% Perform required calculations
Rseries = R1 + R2;
Rparallel = 1/(1/R1 + 1/R2);

%% Output
% Echo the results to the screen
fprintf(‘\nR1 is: %6.1f ohms R2 is: %6.1f ohms\n’, R1, R2)
fprintf(‘Rseries is: %6.1f ohms\n’, Rseries)
fprintf(‘Rparallel is: %6.1f ohms\n’, Rparallel)

Result:
R1 is: 200.0 ohms R2 is: 400.0 ohms
Rseries is: 600.0 ohms
Rparallel is 133.3 ohms
```

1.4 Signals

A voltage applied to a circuit can be constant or it can vary with time. Time-varying voltages can be used as power sources such as the AC voltage in a house, or can be thought of as signals containing information such as a radio transmission. The general expression for a time-varying voltage is shown as

$$V(t) = V_{DC} + v(t) = V_{DC} + \sum_{i=1}^{N} V_i \sin(\omega_i t + \delta_i), \tag{1.6}$$

where V_{DC} is a steady state voltage, and $V(t)$ is a time-varying signal. In the case of a single frequency, the time-varying signal will have an amplitude V_1 and a sine function driven by $\omega_1 = 2\pi f_1$ (where f_1 is a frequency, and t is time) and a phase δ_1 from a reference. Steady state signals are often referred to as DC for direct current, and time-varying signals as AC for alternating current.[5,6] The function is shown in Fig. 1.4.

Time-varying signals are often called periodic and can have functional forms other than sinusoidal. Ramps, steps, square functions, and many others are often encountered in electronics. When time-varying signals are involved, there can be many different frequencies. When working with time-varying signals, to account for components that have frequency dependence, the

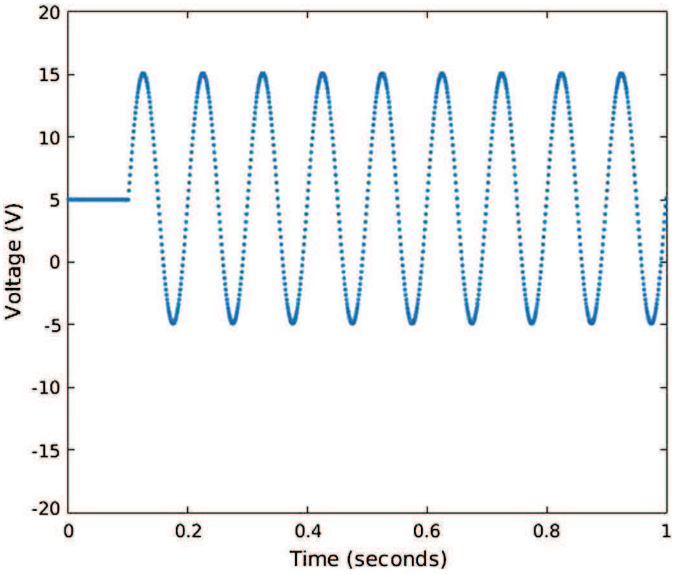


Figure 1.4 A time-varying signal based on Eq. (1.6), showing $V_{DC} = 5$ V at the start and then a superimposed sinusoidal signal $V_1 = 10$ V.

components used to control the voltage and currents are said to have impedance rather than resistance. Such frequency-dependent components include capacitors and inductors as well as many semiconductor devices.

Impedance is the broader term for talking about resistance to current flow in a circuit and is a function of frequency. Impedance includes the effects of resistance, capacitance, and inductance, the latter two of which include devices known as capacitors and inductors. Both capacitors and inductors have the ability to store energy and release it as well as to impede the flow of current. These devices will be discussed in later sections so, while introduced here, are not fully described. Don't worry. Capacitors and inductors will be included in later discussions.[5,6]

1.5 Measuring Instruments

There are many measuring instruments available for use in electronics, but for most applications the volt-ohm meter (VOM), or, as it is commonly referred to today, the digital-volt meter (DVM) is most often used. The DVM typically allows the measurement of voltage, current, and resistance at a minimum and can have manually adjustable scales or can automatically change ranges. It is necessary to understand where to place the DVM in a circuit to make measurements; otherwise, the numbers returned will not make sense. Figure 1.5 shows our earlier LED circuit with measurement devices added in their proper places. The voltmeter measures the voltage

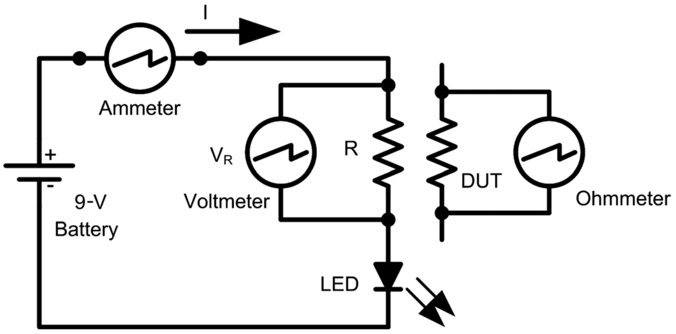


Figure 1.5 A simple circuit with measurement points and connections shown. Notice that the ohmmeter is used when the resistor is out of the circuit.

drop across the resistor, while the ammeter measures the current that is flowing in the circuit.[2] When using the ohmmeter, the resistor should not be connected to the rest of the circuit and is shown here as a separate device under test (DUT).

While a single DVM can perform each of the three measurements shown in Fig. 1.5, it can only perform the measurements one at a time. Each of the different measurement types will be considered individually:

Ammeter: To measure the current, we need the ammeter to be connected to our circuit in series, and the circuit needs to be powered. That is to say, we need to break the circuit and then reconnect the two ends with the ammeter. In this way the current flows through the ammeter. Care needs to be taken as to how much current we measure to make sure it is within the allowed range of the meter.

Voltmeter: The voltmeter measures the voltage across a device and so connects in parallel with the device. The circuit must be on in order to measure the voltage drop across a device.

Ohmmeter: The ohmmeter has its own internal power source, so the DUT must not be powered for this measurement. In general, we don't want the DUT to be in the circuit so that the rest of the circuit doesn't affect the measurement of the DUT.

One interesting observation is that when we measure the voltage drop across a resistor of known value, we are actually able to back calculate the current! In Fig. 1.5 we will expect to see that the ammeter's measurement of the current I will equal the ratio of V_R/R that we measure with the voltmeter!

Most DVMs are able to measure DC voltages using different meter settings. However, sometimes we would prefer to see the alternating signal displaced as a voltage-versus-time trace; this is where the DVM loses its usefulness. Voltage–time measurements are possible using a different

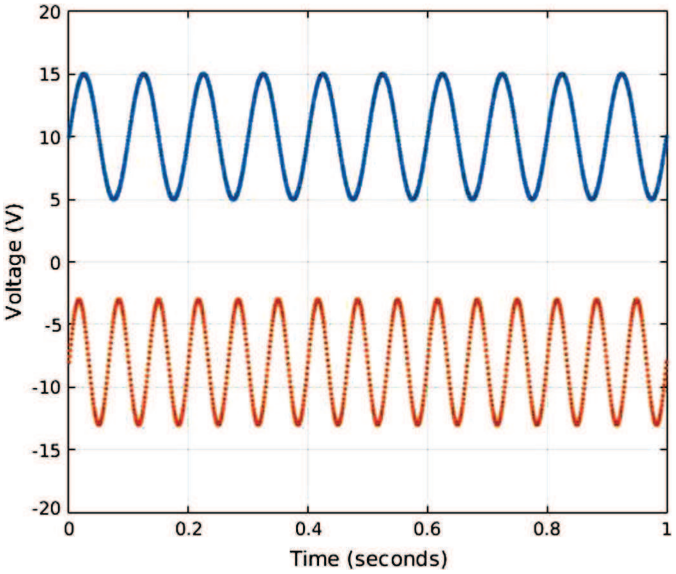


Figure 1.6 An example of an oscilloscope trace showing two signals of different frequencies, but the same amplitude. Notice the frequency difference between the upper and lower signals.

instrument known as the oscilloscope, which is a very practical tool in electronics work.

The oscilloscope is one of the most versatile tools in the electronics instrumentation arsenal. Its role is to provide the most detailed view of a signal that can be obtained, and it can display multiple signals, their relationships to each other, and specific regions of interest. Modern oscilloscopes, often referred to as digital oscilloscopes, can perform mathematical functions such as adding and differencing signals, signal averaging, Fourier transforms, and much more. Figure 1.6 shows an oscilloscope view of two signals of the same amplitude with different frequencies and offset voltages.

With these two tools—the DVM and the oscilloscope—a wide range of measurements is possible within an electronic circuit. It is very important to review the owner's manual for any electronics testing instrument you are going to use in order to avoid dangerous situations or damage to your instrument. Example 1.5 shows a MATLAB script to generate Fig. 1.6.

Example 1.5

Write a MATLAB script to plot two sinusoidal signals with offset voltages of +10 and −10 V, frequencies of 10 and 15 Hz, and a phase shift of 2.

```
% Example_1_5
% SWT 9-3-15
```

```
% Generates two sinusoidal signals for comparison

%% Housekeeping
clear all;
clc;

%% Parameters
Vo = 5; Vdc(1) = 10; Vdc(2) = -10; % in volts
freq (1) = 10; freq (2) = 15; % in Hz
delta(1) = 0; delta(2) = 0;

%% Calculations
for idx = 1:2
  omega(idx) = 2*pi*freq(idx);
end

maxCount = 1000; maxTime = 1; % seconds
for count = 0:maxCount
    time = count/maxCount*maxTime;
    for idx = 1:2
        V(idx) = Vdc(idx) + Vo*sin(omega(idx)*time) . . .
+ delta(idx);
      end
      t(count+1) = time;VV1(count+1) = V(1);VV2(count+1) = V
(2);
end

%% Output
figure(1); plot (t,VV1,t,VV2,'.');
axis([0 1 -20 20]); xlabel('Time (seconds)');
ylabel('Voltage (V)');
grid on;

Results: See Fig. 1.6.
```

1.6 Voltage Dividers and Regulators

Most circuits require a power supply, and more often than not the power supply is a battery providing a DC voltage. Batteries come in a wide range of voltages and current capacities, but usually a little more effort is required to get the voltages that we want to use for our circuit. As an example, when was the last time you found a 5-V battery in the local store? Yet many electronic devices work on 5 V. Fortunately, it is fairly easy to choose a 9-V battery as a supply and just step the voltage down to the required value.

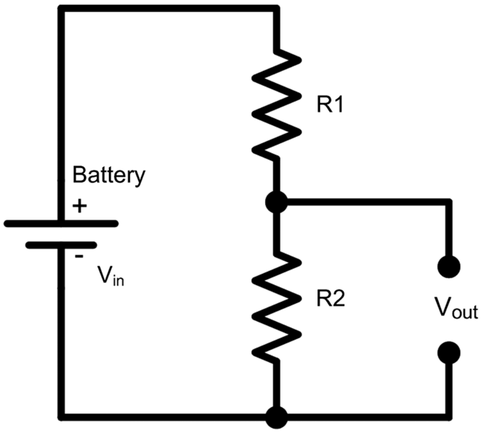


Figure 1.7 R1 and R2 are connected in series to form a voltage divider. By carefully choosing the resistor values, V_{out} can be set between the value of the ground and the battery voltage.

Let's consider a simple way of getting 5 V from a 9-V battery and introduce the voltage divider structure,[2] as shown Fig. 1.7. The voltage divider lets us use two resistors connected in series such that they will have the same current running through them.

The voltage divider rule is given as

$$V_{out} = \frac{R2}{R1 + R2} V_{in}, \tag{1.7}$$

where V_{in} is the battery voltage in the circuit, and R1 and R2 are the resistors. The quick calculations in Example 1.6 show how it works.

Example 1.6

Calculate the relationship between resistors to be used in a voltage divider to get an output voltage of 5 V using a 9-V battery as the source.

$$\frac{V_{out}}{V_{in}} = \frac{R1}{R1 + R2} = \frac{5\ \text{V}}{9\ \text{V}}.$$

Rearranging this equation yields

$$R1 = \frac{5}{4} R2.$$

If a value of $R2$ was chosen to be 400 Ω, then $R1$ would need to be 500 Ω and the total resistance in the circuit would be 900 Ω. Since we are using a 9-V battery, this circuit is drawing a current of 9/900 of 10 mA through the voltage divider to achieve the 5-V output that we wanted. Example 1.7 shows a MATLAB script that uses the input() command to take in the values of the resistors and the input voltage.

Example 1.7

Construct a voltage divider calculator using MATLAB and determine the output voltage from two 1000-Ω resistors with a load of 1 kΩ and 1 MΩ. The program takes an input voltage value in volts.

```
% Example_1_7.m
% SWT 9-4-15

% This script calculates the output of a voltage divider
% and the effect of a load resistor in parallel with the
% second resistor.

% Housekeeping
clear all; % clears the variable list and starts fresh
clc; % clears the command window

% Input values
%% Use the MATLAB input() command to request data
fprintf('Voltage divder calculator\n\n');
prompt = 'Enter the input voltage: ';
VI = input(prompt); % in volts

prompt = 'Enter the value of resistor 1 in ohms: ';
R1 = input(prompt);

prompt = 'Enter the value of resistor 2 (ohms): ';
R2 = input(prompt);

prompt = 'Enter the value of the load resistance (ohms): ';
RL = input(prompt);

% Calculation
%% Perform required calculations
REffective = 1/(1/R2 + 1/RL);
VO = VI* REffective/(R1+REffective);

% Output
%% Echo the results to the screen
fprintf('\nVout is: %6.1f V\n', VO)
```

```
fprintf('REffective is: %6.1f ohms\n', REffective)

Results 1 kohm load:
Vout is: 3.0V
REffective is: 500.0 ohms

Results 1 Mohm load:
Vout is: 4.5 V
REffective is: 999.0 ohms
```

While we are easily able to get the voltage that we want, is that all there really is to this problem? The voltage output is typically used to drive a load. This load can be represented by another resistor connected in parallel to R2. We saw earlier that there will be a change in the circuit if we add another resistor to the circuit. Let's consider a load resistor of 400 Ω. Using Eq. (1.4), the effective parallel resistance would be 200 Ω; thus, we would no longer have 5 V available as the output from the voltage divider. Of course, if the load was closer to 1 MΩ, there would not be a problem.

As such, we need a better way to create our desired voltage that is not as sensitive to the effect of adding a low resistance load. This is quite a common problem, so some very elegant solutions are available; one of the best is the voltage regulator. The voltage regulator is a rather sophisticated device, but it is so commonly used that we will introduce it now, along with a brief description (without much background detail at this point) of how to use it.

Voltage regulators are semiconductor devices designed to maintain a constant voltage level. Here, we treat them as a "black box" and demonstrate how they are used, starting with Fig. 1.8.

The 7805 voltage regulator lowers the battery voltage to 5 V. Regulators come in many voltage ranges; the number 5 in the 7805 indicates that it provides a 5-V regulated output. Often, the final capacitor C is not required and is there to act as a source of extra charge if needed. This is a very practical

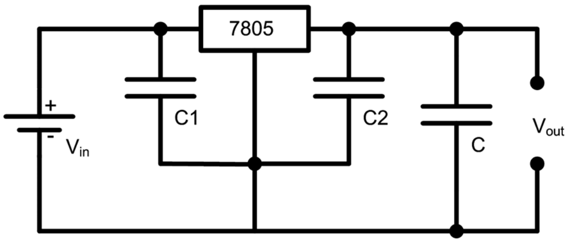


Figure 1.8 A voltage regulator used to convert a 9-V battery supply into 5 V. C1 and C2 are low-valued capacitors to support the 7805 device, while C is a larger filter capacitor. C might not always be required if the regulator is close to where the power is needed.

and simple power supply, and many different source configurations can be used in place of the battery.

When a load is not attached, the power draw of the voltage regulator is very low, making it ideal to be supplied by a battery. Voltage regulators are available in a wide range of values such as the 7812, 7815, and so forth. The major advantage of the voltage regulator is that it provides its rated voltage over wide-range current requirements. The limitations and proper operation of a voltage regulator can be found in the device data sheets.

1.7 Device Data Sheets

There is a wide range of electronics devices, and we have just begun to become familiar with a few of them. No one can remember all of the detailed specification of the myriad of electronic devices, so how can we know how to use a particular device, maybe one we have never seen before?

Not to worry. All of the information needed to work with a particular device is contained in the device data sheet. Everything from the operating specifications, performance data, size and shape information, and sometimes practical configuration diagrams are provided in the data sheet. These sheets are usually supplied by the device manufacturer and are usually available on the vendor's or the manufacturer's website.

So what does a data sheet look like? Data sheets contain a lot of information, as they try to provide all of the information important to the designer. This is where the Internet becomes a wonderful resource. Go to your favorite web search tool and look up the 7805 voltage regulator and you will see find a wide range of suppliers and many different data sheet sources. Take a look at one of the data sheets to become familiar with the type of information that they contain. There is a lot, but don't let this concern you at this point; as we move forward we will be talking more about data sheets, the symbols they use, and how to use them.

This brings up an important consideration. Many manufacturers don't want to be involved in day-to-day sales of low volumes of parts, so they enlist the help of companies that specialize in electronic part sales. This makes it somewhat easier to buy parts, as you only have to deal with one distributor for many electronics parts. This can really save on shipping costs.

1.8 Practice Problems

1. A voltage divider uses a 500- and a 900-Ω resistor in series with a 10-V battery.
    a) What is the color code you would expect to see for these resistors if they are 20% tolerance?
    b) What is the effective series resistance?

   c) What is the current through each resistor?
   d) What is the voltage drop through each resistor?
2. Three resistors are connected in parallel with values of 1000, 3000, and 1,000,000 Ω. What is the effective parallel resistance?
3. Three resistors are connected in series with values of 1000, 2000, and 3000 Ω. What is the effective series resistance?
4. What is the value of a resistor with the four-color code of yellow, purple, yellow, gold?
5. Design a circuit to power an LED that has a forward voltage of 2.5 V and a forward current of 18 mA using a 9-V battery and a voltage regulator as the supply.

Answers

1. (a) Violet Black Brown and White Black Brown; (b) 1400 Ω; (c) 7.1 mA; (d) 3.6 V and 6.4 V
2. 750 Ω
3. 6000 Ω
4. 470 kΩ ± 5%

References

1. C. K. Alexander and M. N. O. Sadiku, *Fundamentals of Electric Circuits*, Fifth Edition, McGraw-Hill, New York (2013).
2. B. Grob, *Basic Electronics*, Eighth Edition, McGraw-Hill, New York (1997).
3. Resistor Guide: www.resistorguide.com. Last accessed: 4-14-16.
4. H. Moore, *MATLAB® for Engineers*, Fourth Edition, Prentice-Hall, Upper Saddle River, New Jersey (2014).
5. P. Horowitz and W. Hill, *The Art of Electronics*, Cambridge University Press, Cambridge (2015).
6. A. S. Sedra and K. C. Smith, *Microelectronic Circuits*, Oxford University Press, Oxford (2015).

Chapter 2

Common Electronic Circuits and Components

Now that we have been introduced to some basic electronic concepts in Chapter 1, we will get to the business of making circuits to perform specific tasks. Many of these tasks involve setting and comparing voltage levels, and signal amplification. These applications will be used to introduce some additional electronic devices including semiconductor diodes and transistors as well as inductors, capacitors, and transformers. The current trend of moving to smaller electronics packages and surface mount electronics applies equally well to capacitors, diodes, and transistors, as it does to resistors. Surface mount components and their use will be discussed in a later chapter.

The major device configurations that will be introduced are bridges, rectifiers, amplifiers, and digital-to-analog converts. While each is useful for its own application, they are also useful as a means to learn about how diodes, transistors, and resistors can be used in circuits to perform specific tasks. In this chapter we will extend the idea of voltage dividers to making bridges for use in sensitive measurements, current dividers, diodes, some simple amplifiers, and power drivers, and introduce the idea of digital versus analog signals and how to convert signals from digital to analog.

2.1 Bridges and Balance Measurements

The starting point for making sensitive measurements is being able to make precise comparisons between two things. In electronics, more often than not, the comparison involves the difference between two voltage levels, with one usually being ground or zero potential. This isn't as difficult as it may seem and involves a circuit we are now familiar with, the voltage divider. If you recall from Chapter 1, the voltage divider[1,2] is just two resistors connected in series across a voltage source, as shown in Fig. 2.1.

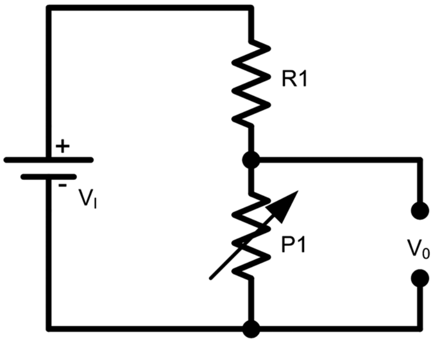


Figure 2.1 A simple voltage divider circuit showing a variable resistor or potentiometer.

The output of the voltage divider V_0 was introduced in the last chapter as part of our discussion on power supplies. It is reintroduced here with a small change; one of the resistors is now variable:

$$V_0 = V_I \cdot \frac{P1}{R1 + P1}, \tag{2.1}$$

where R1 is a fixed resistor, P1 is a variable resistor or potentiometer, and V_I is the initial or source voltage such as a battery. The potentiometer provides the resistance in the circuit to vary without needing to physically swap out one of the resistors. The effect is to have a smooth transition to a wide range of voltages at the output. Figure 2.2 shows several different potentiometers that are available and in common use.

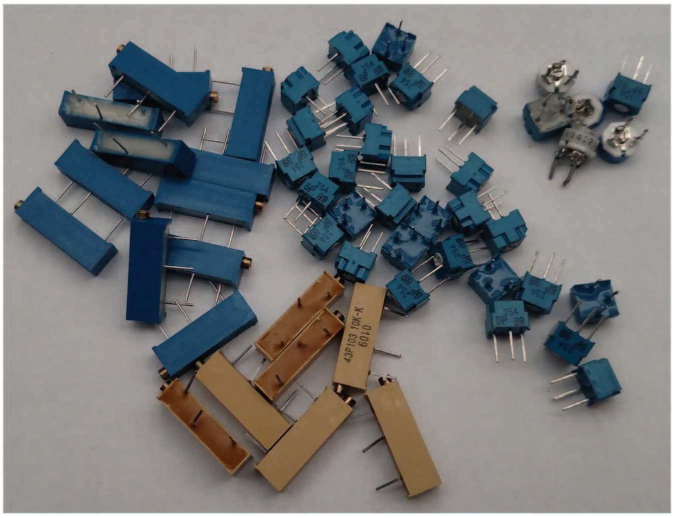


Figure 2.2 A selection of potentiometers showing some of the different shapes available. Notice that they have three connection leads!

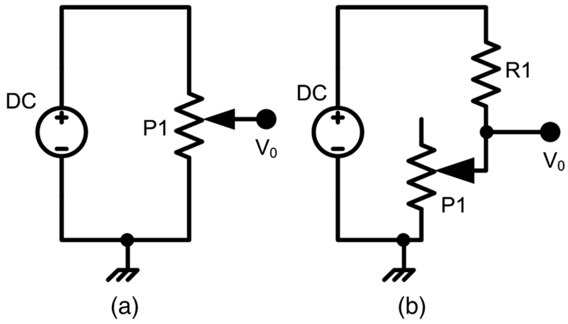


Figure 2.3 Two ways to use a potentiometer to make a voltage divider. Both work to lower the voltage, but the circuit in (b) has built-in current limiting with the resistor R1. Notice the alternative ground symbol at the bottom of the circuits.

Figure 2.1 showed a simplified symbol for a variable resistor; a more correct view would be that shown in Fig. 2.3. Some of the other symbols in Fig. 2.3 have been changed to the more commonly used symbols. The source is no longer shown as a battery, indicating that the DC voltage can come from a power supply or a battery. A case ground symbol is shown at the bottom; this is the second connection for the V_0. The symbol for the potentiometer[1] used here is much more descriptive of the operation of a variable resistor, shows all three connections to the device, and makes clear one of the issues involved in using only a single potentiometer as the voltage divider.

It may seem strange that there is a third leg to the potentiometer; the reason is easy to understand when one considers the physical construction of a potentiometer, particularly old-style wire-wound resistors. These resistors were constructed by winding a long wire around a core. If the entire length of the wire was used as a resistor, you got a single value, but you could also make contact anywhere along the length of the wire to get a different resistance based on the amount of wire between the input and the take-off location. If the contact was a wiper that could be moved up and down the wire length, the result would be a smoothly variable resistance, which is know as a potentiometer.

The symbol for a potentiometer is a resistor with three legs, one of which, the variable or wiper connection, is indicated by an arrow. As can be seen in Fig. 2.3(a), a voltage divider can be made with just one potentiometer! However, the single-potentiometer voltage divider has one big drawback. If you raise the wiper arm to the top, there is no resistance between the source and the load connected to V_0. If the potentiometer wiper is connected to ground, the source connection might short to ground, which can result in a high-current condition and potentially damage the voltage supply. Figure 2.3(b) shows a slightly better approach for connecting the potentiometer. Here, the top resistor, R1, acts as a current-limiting resistor, while the potentiometer P1 provides the voltage control. The effect of a current-limiting resistor is shown in Example 2.1.

Example 2.1

A voltage divider is constructed as in Fig. 2.3(b). The resistor R1 is 1 kΩ and the potentiometer P1 has a maximum resistance of 10 kΩ. What are the maximum and minimum voltages that can appear at V_0, and what are the maximum currents?

Using our voltage divider equation and the maximum resistance of P1,

$$V_0 = V_I \frac{P1}{R1 + P1} = V_I \frac{10\ \text{k}\Omega}{1\ \text{k}\Omega + 10\ \text{k}\Omega} = V_I \frac{10}{11}\,\text{V}.$$

Using our voltage divider equation and the minimum resistance of P1,

$$V_0 = V_I \frac{P1}{R1 + P1} = V_I \frac{0\ \text{k}\Omega}{1\ \text{k}\Omega + 0\ \text{k}\Omega} = 0\ \text{V}.$$

So, having the resistor R1 in place means that we can't get the maximum voltage out; however, even when there is a short circuit to ground, the value of R1 is still present and the maximum current that can be drawn from the supply is

$$I = \frac{V_I}{R1 + P1} = \frac{V_I}{1\ \text{k}\Omega + 0\ \text{k}\Omega} = V_I\ \text{mA}.$$

If V_I is 10 V, then the maximum current would be 10 mA through the resistor.

This example shows the value of having a 1-kΩ resistor in the circuit to limit the current: even in the worst-case scenario of a short across the voltage output, only V_I / 1000 A can be drawn from the source. Most power supplies will be able to handle such a load. While a lower resistor value will increase the amount of the voltage available at the load, the current will also rise.

The voltage divider circuit has many uses beyond providing power to a load; it can also be used to support many sensitive measurements when combined into a bridge circuit. The most common of these bridge circuits is the Wheatstone bridge.[1] The Wheatstone bridge is made from two voltage dividers fed from the same power supply. A common form of the Wheatstone bridge is the resistor checker shown in Fig. 2.4.

The resistor checker in Fig. 2.4 is a device designed to quickly sort resistors. It works because when the ratios of R1/R2 and R3/DUT are the same, the difference voltage of the bridge will be zero. This provides a quick way of identifying resistors that are a perfect match, or lie within some small variance of the desired value. These ideas are shown in Example 2.2.

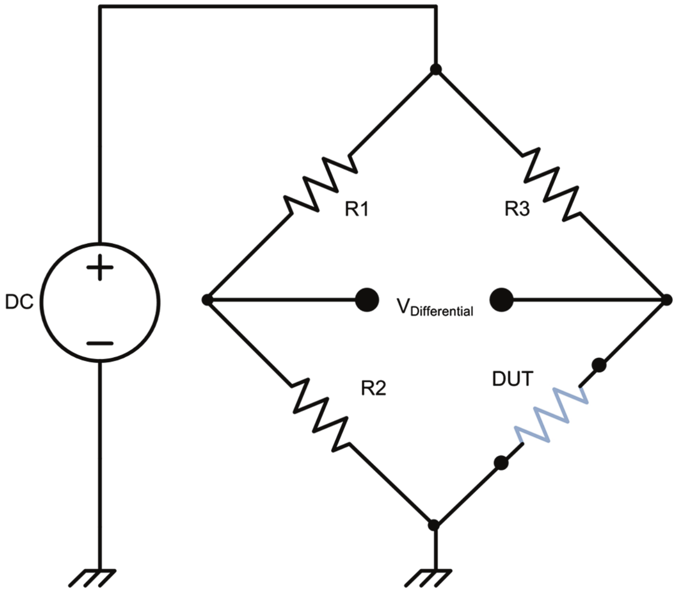


Figure 2.4 A Wheatstone bridge configured to check resistor values against a standard. The DUT is where the test resistor is placed.

Example 2.2

Write a MATLAB® script to show that a Wheatstone bridge can be considered as two voltage dividers in parallel. The voltage difference between the two sides of the bridge biased to 10 V and the current in each side is shown in Fig. 2.4, where R1 = 100R3 and R2 = 100RDUT.

```
% Example_2_2.m
% SWT 11-28-15

% This program calculates the voltages from a Wheatstone
% bridge where VL, IL are the voltages and currents on the
% left side of the bridge VR, IR are for the right side.

% Housekeeping
clear all; clc;

% Resistors
R3 = 1000; % Ohms
RDUT = 1000; % Ohms

R1 = 100*R3;
R2 = 100*RDUT;
```

```
Vbias = 10; % Volts

% Eqns
VL = R2/(R1+R2); VR = RDUT/(R3+RDUT); % voltage dividers
Vdiff = VL-VR; % voltage difference between dividers
IL = Vbias/(R1+R2); IR = Vbias/(R3+RDUT); % I in each side

% Output
fprintf(‘Bridge voltages and currents\n’);
fprintf(‘VL = %3.1fV; VR = %3.1fV\n’,VL, VR);
fprintf(‘Vdiff = %3.1f V\n’,Vdiff);
fprintf(‘IL = %3.1fmA; IR = %3.1fmA\n’,IL*1000, IR*1000);

Results:
Bridge voltages and currents
VL = 0.5 V; VR = 0.5 V
Vdiff = 0.0 V
IL = 0.1 mA; IR = 5.0 mA
```

When the two voltage dividers have the same resistance values, the voltage difference in the bridge is zero, and no current can flow even if a wire connects the terminals. If that connection was made with a meter movement, such as an old-style needle readout, a zero or null value will be shown when the two voltage dividers are the same. Such a meter movement is shown in Fig. 2.5. In the past, before everything went digital, all meters were based on magnets and coils of wire. When a current passed through the wire, the needle moved in relation to the amount of current. These meter movements with a center ‘zero’

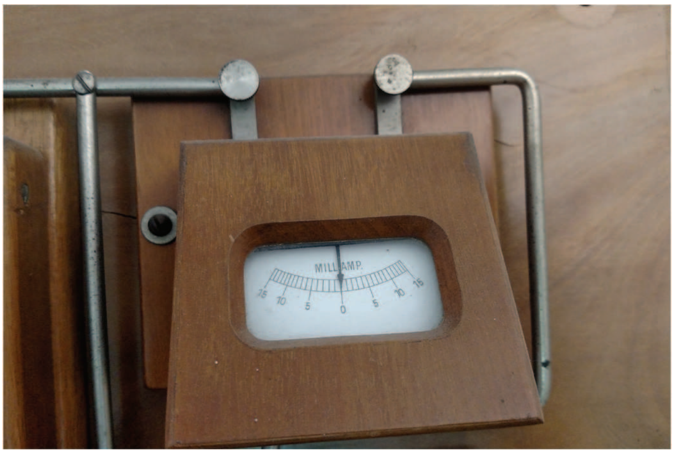


Figure 2.5 The differential voltage can be displayed on an old-style meter movement. Notice that this meter can swing to the left or right, depending on the current direction.

value are a sensitive current indicator in a Wheatstone bridge, and the direction of deflection indicates the direction of the current!

The Wheatstone bridge provides a very simple way to compare two voltages, as any voltage difference between the two sides of the bridge will cause a current to flow that can be used as an indicator of the imbalance.

2.2 Diodes

Diodes are just about the simplest of the semiconductor devices used in electronics. Diodes look a lot like resistors—little cylinders with wires sticking out the end—but they don't have a group of colored lines; most often they have just one stripe located close to one end. There are several types of diodes in use, but the diode rectifier is the most common and has the interesting property of conducting current in only one direction and blocking it in the reverse direction. Remember that stripe on the end? That is the indicator of the direction of the current flow. Figure 2.6 shows the common symbols used for rectifying a Zener diode when shown in a schematic. Notice that the "bars" in the schematic and the "band" on the actual diode correspond.

There are different types of diodes, two of which are the rectifying diode and the Zener diode. A common or rectifying diode passes current in only one direction. A Zener diode conducts current just like the rectifying diode, but once the voltage is reversed, it will conduct again after a specific voltage (the Zener voltage) has been reached.[1–3]

Understanding how the rectifying diode operates is best illustrated by considering its effects on a sinusoidal voltage signal. A sinusoidal voltage varies about a zero reference point and has an average voltage of zero. This is just a statement that because there is as much signal above as below the line, when looked at over several complete cycles, on average the voltage has a value of zero. We know from our previous calculations of power that if the

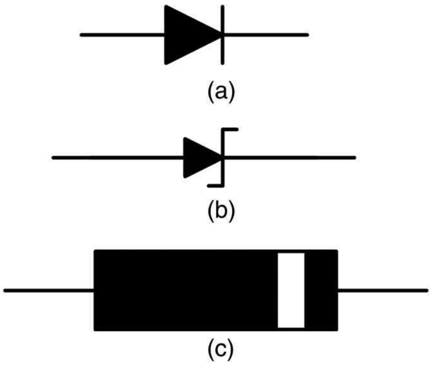


Figure 2.6 Schematic symbols for (a) a rectifying diode, (b) a Zener diode, and (c) a general diode showing the location of the stripe. Notice that the arrows in (a) and (b) point toward the bar and indicate the direction of current flow.

voltage is zero then the power is zero! But we know this isn't true, so how can we get a good measurement? Since the diode only conducts in one direction, putting a diode in the circuit will remove either the top or the bottom half of the signal. If we average the signal now, it will be either positive or negative, and we can get a measure of the magnitude!

Another excellent use for the diode is to convert AC voltages to DC voltages. This most often involves a transformer, which uses inductance between two wire coils to either raise or lower the voltage of an AC source. We will look at transformers in the next section, but here we will just say that they are useful tools for changing the amplitude of a voltage signal. Of course, if one diode is good, could more be better? Figure 2.7 shows how a sinusoidal

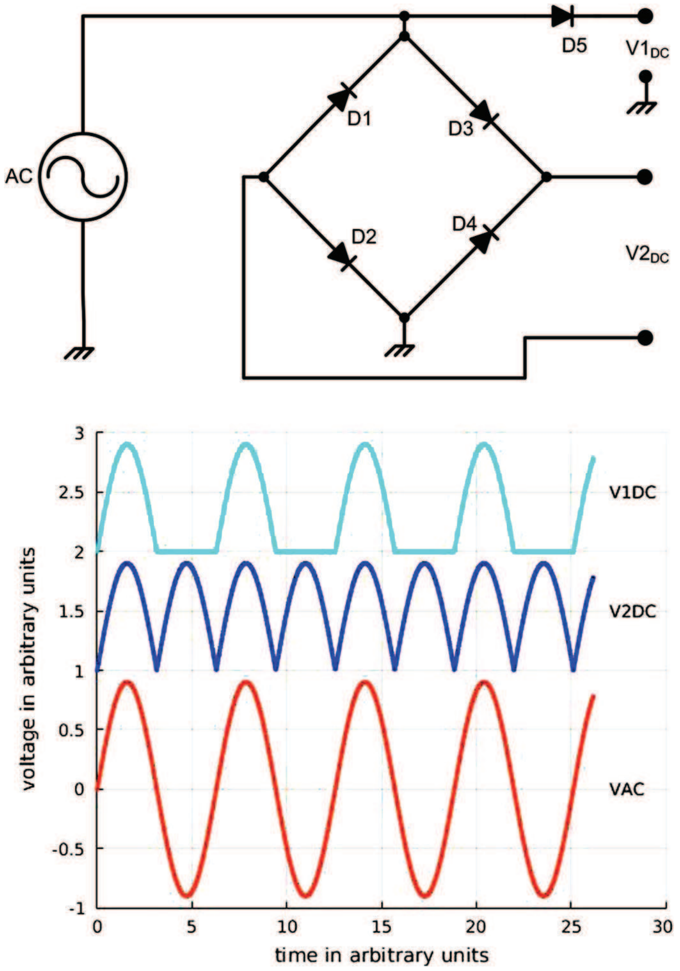


Figure 2.7 Top: Two different diode rectifier types in one circuit. Bottom: Plot showing the sinusoidal input signal (VAC is volts of alternating current).

signal can be changed in the presence of one and four diodes. Notice that the full-wave bridge rectifier looks a lot like our previous Wheatstone bridge.

Notice that in Fig. 2.7 a vertical offset of 1 V has been introduced between plots so that each waveform can be clearly seen. Output $V1_{DC}$ was generated using a half-wave rectifier, and $V2_{DC}$ was generated by the full-wave rectifier; there are some differences in their outputs.

You may have noticed that the rectified signal does not look much like a steady DC voltage as there is still a lot of ripple in it. This is easily remedied by placing a sufficiently large capacitor across the output terminals. The capacitor accumulates charge as the voltage rises, and as the voltage level drops, the amount of charge slowly decreases based on the load. Thus, the output voltage waveform will be smoother than that of the input voltage. If the fall in the voltage level is small over the period between cycles, a DC voltage has been achieved with a small ripple voltage superimposed.

Diodes have another interesting quality when they are conducting: there will be a voltage drop of about 0.7 V across the diode.[3] If several diodes are used in series, this voltage drop can easily be a few volts. While this is not always an apparent plus, it does prevent a current from getting through the forward-biased diodes unless it is above a certain threshold. We will revisit this point later.

2.3 Transformers and Inductors

A simple transformer is made from two different wire coils wound around a magnetic core and uses a changing current passing through the wire of one coil to generate a magnetic field.[1] The magnetic field passes energy to the second coil where it is converted to a voltage. The basic form of a transformer shown schematically in Fig. 2.8 consists of independent primary and secondary windings and a special core. The coil on the input or primary side of the transformer is shown with fewer turns than the output or secondary side of the transformer, indicating that the voltage out will be larger than the input. This identifies the transformer as a step-up transformer. The two coils surround a core to enhance the coupling between the two coils.

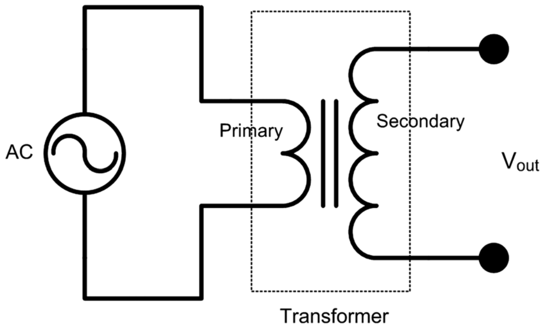


Figure 2.8 Simple transformer connected to an AC input.

The transformer operates by a varying current in one winding creating a magnetic field that induces a current in the other winding. The result is that the voltage into the first core is transformed to a new voltage based on the ratio of the number of turns or windings—the turns ratio—between the primary and secondary coils. The transformer works because the magnetic field builds and collapses in the primary, and the energy transfer to the secondary is through the magnetic field.[1]

In the case where there are more turns on the secondary than the primary, the voltage at the secondary is increased, making it seem as though an amplification is occurring; however, we are still subject to the laws of conservation of energy, so if the voltage increases, the amount of current available decreases, and there will be losses due to heating.

Transformers are commonly used in electronics and are often seen in power supplies. Quite often, household devices are connected to a wall outlet using a cable that ends at a box that is actually a transformer. These cables can be identified by the oversized plug end and can provide a wide range of voltages and currents in what appear to be similar packages, so always read these values from the plug end before you use them.

Another name for one coil of the transformer is an inductor. Inductors have many uses, particularly in electronic filters, which will be introduced in Chapter 6.

2.4 Transistor Switches and Amplifiers

Transistors come in two types that relate to their use: the general purpose and power transistor. Both of these types are available as bipolar junction transistors (BJTs) and metal-oxide semiconductor field-effect transistors (MOSFETs). There are significant differences between BJT and MOSFET devices; once we have a handle on transistors, we will introduce these differences.[2,3]

The transistor is another semiconductor device, similar to the diode, but having three connections instead of two. These connections are known as the base (B), emitter (E), and collector (C). A simple way to think of a transistor is that the current flow between the collector and emitter is controlled by the voltage applied to the base. While there are similarities between the diode and transistor devices, the transistor is where things get interesting. A simple circuit diagram with a transistor is shown in Fig. 2.9. There are two things to notice in the circuit: the first is the symbol for transistor T1 and second is all of the resistors that are included. The resistors keep the voltages on each lead of the transistor in the proper voltage ranges, a technique known as biasing.

The transistor in Fig. 2.9 is being used as an amplifier, and two different voltage sources are being used in the circuit. The DC source supplies power to the transistor, while the AC source provides a signal that is amplified by the transistor to a higher voltage signal that will appear at outputs V_C and V_E.

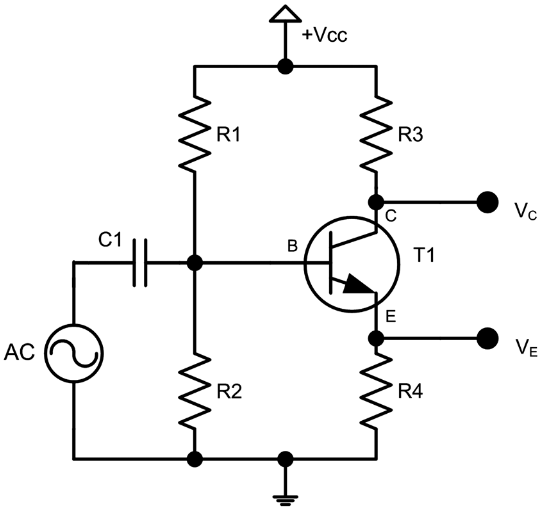


Figure 2.9 A biased transistor circuit showing the arrangement of resistors used to bias the device. Notice that both AC and DC sources are included in the circuit, requiring capacitor C1 to isolate them. Note also that voltage +Vcc is the input voltage and, like V_C and V_E, is measured with respect to ground.

This is a great place in our discussion to look carefully at what is meant by the word *amplification*. If we just look at the output voltage signal compared to the input signal, it is definitely bigger, so we say that it is amplified. To appreciate what the transistor is really doing in this circuit, we need to notice a few things. First, the transistor is connected to a power supply that is different from the input signal. This is important because it means that we don't have to conserve energy between the input and the output signals. They are separate sources but are related in shape. They track each other, but they are different.

In the transistor, the small input signal controls how much of the secondary DC source gets to the output and contributes only a very small amount to the actual process. So, while the output signal tracks the input signal, the output signal is tied to a larger voltage source that now follows the signal.

When the transistor is operating as an amplifier, the base of the transistor is said to be biased or has a DC voltage applied. Additionally, a small voltage signal to be amplified is superimposed on the bias. The transistor can also be operated in a different fashion, that of a switch. A transistor switching circuit is shown in Fig. 2.10, where the base of the transistor is driven between two voltages that turn the conduction of the transistor on and off. The effect is to have a voltage available across the resistor going to the load, or shorted to ground. This circuit is an illustration of what is possible with the transistor. Switch S1 can be replaced by many different devices such as a photocell, acoustic sensor, or a thermal sensor to turn the transistor on. These devices will require additional circuitry to accomplish the task, but the concepts remain the same.

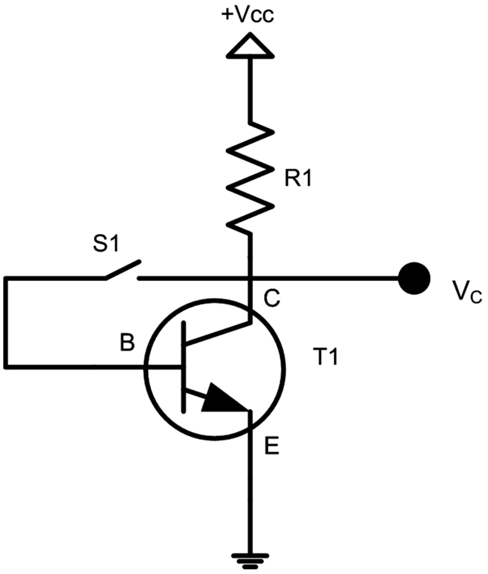


Figure 2.10 A transistor circuit configured as a switch.

In the switch mode, the transistor can be required to pass a large current to ground, so transistors used in this mode usually have a heat sink attached to help dissipate some of the heat that will be generated. It is also possible that a general-purpose transistor might not be up to the task, and an upgrade to a power transistor might be needed.

Let's return for a moment to the transistor itself and how it is used, at least under some simplified and somewhat ideal conditions. Consider an input signal generated by a sensor that does not have much power. In fact, if we draw a current from the sensor, the signal will be degraded and unusable. One of the key differences between the BJT and MOSFET is that the gate of the MOSFET does not require a current to control the drain-source current! There are other differences as well, but in many cases, this is the useful difference. The BJT and the MOSFET symbols and a physical model[2,3] are shown in Fig. 2.11.

The gate (G) of the MOSFET, which is similar to the base of the BJT, controls the current from the source (S) to the drain (D) of the MOSFET. But it does it by a field effect, i.e., an electrostatic effect. It does not draw a current to operate. Both BJTs and MOSFETs are available for general-purpose use as well as for power transistors, which will be explored next.

2.5 Power Transistors

While there is no clear distinction between general-purpose transistors and power transistors, it is usual to think of power transistors as being able to handle larger currents, say, more than one amp, or that a heat sink is needed

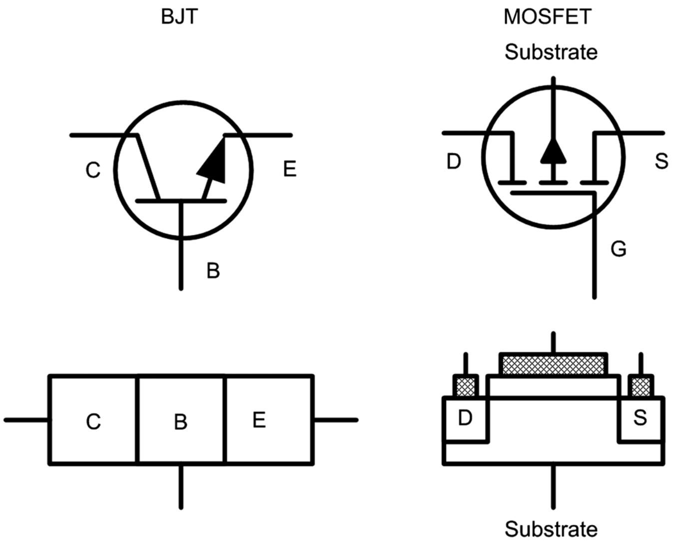


Figure 2.11 Schematic and physical models for BJT and MOSFET devices. Note that the MOSFET can have four terminals, but usually the substrate and the source terminals are connected.

to keep the transistor cool. Although not the best practice, I must confess that I usually find that I need a power transistor when the general-purpose transistor I originally selected overheats. Often this results in a puff of the "magic smoke" escaping from the transistor, and we all know that it is very hard to put the smoke back in, so a new transistor is needed. This approach can be used when you are first developing an electronic circuit, but can be catastrophic if left unattended. Transistors can get very hot!

The design paradigm for power transistors is to support the ability to handle large currents and voltages. This means that the output resistance is low, and that the transistor junctions are sized larger to better dissipate heat and are insulated to work with higher voltages. This also usually means that the transistor cannot operate at very high frequencies due to the capacitance in the junctions. It is the actual junction temperature that is important for the transistor, not the case temperature that we can feel. Sometimes, by the time we notice that a part is becoming hot, the junction is too hot and the device can be damaged, or its service lifetime can be greatly reduced.

Power transistors[3] can often be identified by being larger than regular transistors and often have either metal packaging or metal connectors on them. It is true that not all transistors look alike and not all transistors have three leads. Figure 2.12 shows three common transistor configurations, two of which are commonly used for general-purpose transistors, and the TO-220 configuration, which is is often used for power transistors. The power transistor is designed to easily mount a heat sink using the hole in the metal tab.

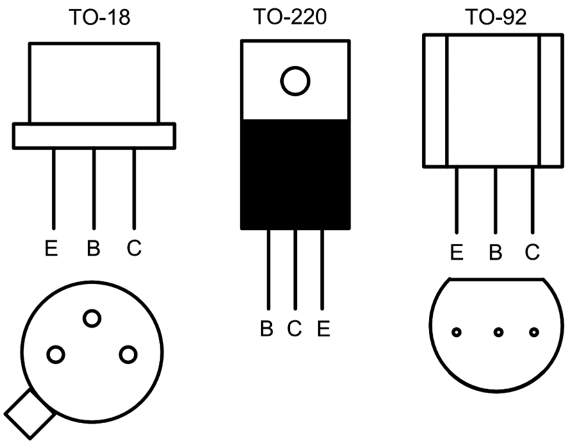


Figure 2.12 Some different through-hole transistor packages and their associated pinouts. This information is always provided on the transistor data sheets.

There are many packages for transistors, and in some ways these are a study in themselves. In addition, the pinouts often vary between devices, so you will always need to consult a data sheet for your particular device.

2.6 Combining Transistors

Sometimes one transistor is not enough, especially when working with power transistors. While it is desirable to have one transistor to do the complete job, often using more than one can provide higher heat dissipation, which translates into better power handling. One way combining transistors is shown in Fig. 2.13, where a number of power transistors are connected in parallel.

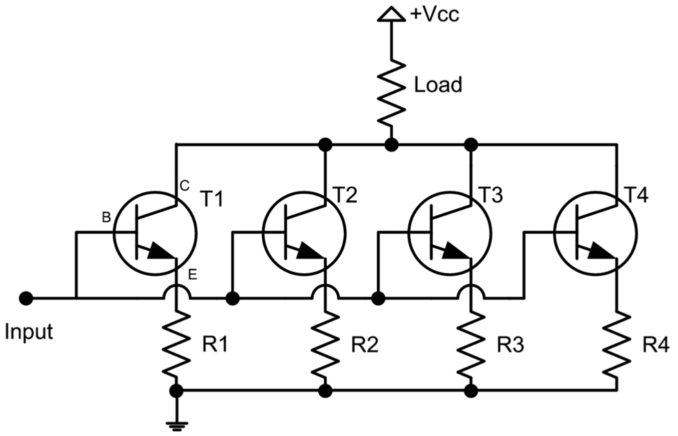


Figure 2.13 Power transistors in parallel. A low-power signal applied to the input can be used to toggle a high-power load at power greater than that of the single resistor, or can be used to reduce heating effects. This is an example of a current divider.

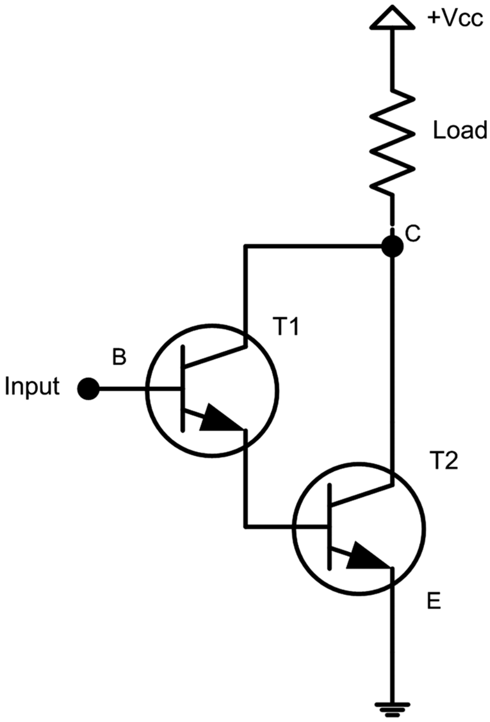


Figure 2.14 Transistors connected to form a Darlington pair, which can be constructed from two separate transistors or be a single commercial transistor.

Resistors R1 through R4 are commonly the same value or can be tuned for variations in the individual transistors.

Another technique is to connect two BJTs in a configuration known as a Darlington pair.[3] This configuration, shown in Fig. 2.14, uses the two transistors to provide much higher current gain than each of the individual transistors. Darlington pair transistors can be made from two separate transistors or commercially as a single three-terminal device.

2.7 Digital-to-Analog Conversion

Numbers can be represented in many different ways, and more often than not in electronics, you will have to consider binary representations of numbers.[2] Simply put, we can convert a number into a code that consists of a series of on (1) and off (0) states. For instance the number 3 can be written in binary as 011, interpreted not as the value 011 but rather as the sequence zero-one-one. The individual numbers are referred to as bits, so a "three count" requires a minimum of two bits, and the bits are ordered from most significant to least significant. The most-significant bit is usually labeled as MSB, and the least-significant bit as LSB. The three-bit numbers from 0 to 3 then can be

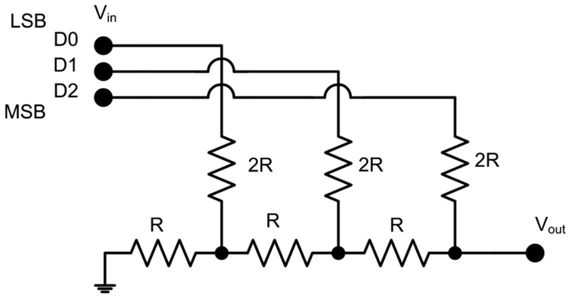


Figure 2.15 A three-bit analog-to-digital converter based on a resistor network.

represented as 0-0-0; 0-0-1; 0-1-0; 0-1-1. There is space in our three-bit number system to count higher, but we will hold here for now.

A three-bit number can correspond to three wires (okay, a ground will be needed, too, so really four wires) that hold the information. Each line can be toggled in a counting sequence to represent numbers. But what if we want the number represented as a voltage? We can combine all of the things we now know to generate a single voltage using only resistors. The design of a simple three-bit digital-to-analog converter is shown in Fig. 2.15.

The approach shown here is often called the R-2R resistor network,[4] as the circuit contains resistors of value R and 2 × R. Since we are only interested in the voltage values and are not trying to drive a load, resistor values are normally chosen to be in the range of 100 kΩ and up. This circuit can be used to drive a transistor, either a BJT or a MOSFET, to provide some buffering. There is, however, a better tool that provides both buffering and amplification. That tool will be the subject of the next chapter.

2.8 Practice Problems

1. A voltage divider uses a 500-kΩ and a 1-MΩ resistor in series with a 1-MΩ voltmeter connected across the larger resistor. If the input voltage is 12 V, what is the measured voltage?
2. A Wheatstone bridge is constructed from four resistors, two in series are 10 kΩ and 20 kΩ, the other two are 1 MΩ and 2 MΩ. What is the current flow between the center tap of the two legs?
3. A single diode is used as a rectifier in an AC circuit. Approximately how much of the input power will be lost?
4. A single general-purpose transistor is operating in a circuit such that it is on the verge of overheating. How can this circuit be improved?
5. Convert the binary number 1-1-1 to a decimal number and an analog voltage.

Answers

1. 6 V
2. 0 A
3. 50%
4. Use a power transistor of higher rating or use multiple general-purpose devices in parallel.
5. 7 decimal, the maximum voltage value

References

1. B. Grob, *Basic Electronics*, Eighth Edition, McGraw-Hill, New York (1997).
2. P. Horowitz and W. Hill, *The Art of Electronics*, Third Edition, Cambridge University Press, Cambridge (2015).
3. A. S. Sedra and K. C. Smith, *Microelectronic Circuits*, Oxford University Press, Oxford (2015).
4. T. L. Floyd, *Digital Fundamentals*, 10th Edition, Prentice Hall, Upper Saddle River, New Jersey (2008).

Chapter 3

Linear Amplifiers

Electronic devices have come a long way since the invention of the transistor in 1947 by Bardeen, Brattain, and Shockley.[1] Before the transistor was available, vacuum-tube-driven devices, now rarely seen, were the fundamental tools of signal amplification and rectification. The importance of the transistor to electronics can be exemplified in the history of the popular electronics device, the radio. Within a decade of its invention, radio designs began to switch to using transistors, and by the 1970s, billions of these new radios were manufactured, pushing out the older vacuum tube systems. The original transistors were somewhat less than reliable devices, and their failure rate kept the price of transistor radios rather high, even though only a handful of transistors was used in the radio. While these modern radios were highly inferior to the tube radios of the day, they did have a few advantages, such as being smaller, portable, and able to turn on instantly! These attributes made the transistor radio very popular, even if its audio quality was not so great.

Today, we rarely talk about the number of transistors in an electronic device, and in many commercial electronics, you might not even find a discrete transistor. As manufacturers got better at making transistors, the transistors got smaller, and the techniques of combining transistors to perform complex tasks became commonplace. Current electronics can have hundreds and even millions of transistors all contained in a single component, in what is known as an integrated circuit (IC), or more commonly, a chip.

In fact, the more common, primitive (lowest-level) electronic device is often an operational amplifier. An operational amplifier has many transistors within it and provides high-performance amplification that is linear over its operating range. Said another way, if we input a signal, we will get an output signal that is linearly related to the input signal, just with different amplitude!

In this chapter we will expand on the idea of amplification and use operational amplifiers, or op-amps, to develop circuits that are critically important to working with sensors and transducers. These are all very important tasks for optical engineers.

3.1 Inverting Operational Amplifier Configuration

Operational amplifiers or op-amps are the electronic components of choice when it comes to making amplifiers and are designed to be linear amplifiers. This might not seem so important right now, but as we move forward it will become clear just how desirable linearity is in an amplifier. The op-amp itself is an IC made from transistors and is available in a small package. A typical op-amp is shown in Fig. 3.1 in the form of dual in-line packages (DIPs). These op-amps can have as few as 8 connections and commonly have 16 or even 24. The different package designs determine how to use the different devices.

DIP devices are also referred to as through-hole devices and have pins that can be pushed into specialized breadboards or printed circuit boards. Through-hole formats are being phased out in favor of surface-mount components due to the smaller package size of the latter. The surface-mount packages are more easily handled by automated part-placement machinery but are more difficult to use for prototyping. Surface-mount components will be discussed in Chapter 14.

We are going to use a direct approach to appreciate how op-amps work, and begin by introducing the most common amplifier, the inverting amplifier. Figure 3.2 shows the classical schematic diagram for the inverting op-amp circuit.[1,2]

The inverting amplifier is so named because the output (V_{out}) and the input (V_{in}) have opposite signs. This is shown in the following equation, which defines the gain G for the inverting amplifier:

$$G = \frac{V_{out}}{V_{in}} = -\frac{R_F}{R_I}. \tag{3.1}$$

The ratio of the output voltage to the input voltage is also known as the transfer function. In this case, the negative sign indicates the inverting nature of the amplifier. R_F is the feedback resistor, and R_I is the input resistor.

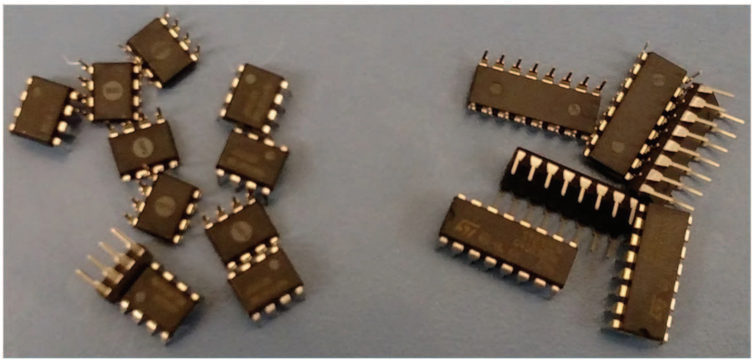


Figure 3.1 Two different DIP operational amplifiers, one with 8 pins (left) and the other with 16 pins (right).

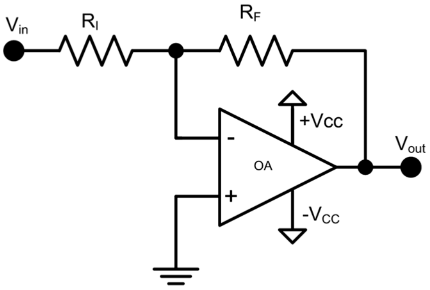


Figure 3.2 The inverting amplifier circuit using an operational amplifier (OA). The power connections are shown here, although these are normally not shown on schematics.

The gain of the inverting amplifier is totally controlled by the external resistors, making a very versatile device.

The power for many operational amplifiers requires both positive and negative voltages as well as a common reference level, known as ground. Many standard power supplies have this capability, often referred to as bipolar supplies; however, before we go too far talking about exotic power supplies, note that the power for many op-amps can be made from two connected 9-V batteries, as shown in Fig. 3.3.

The operating voltage for many op-amps is rather wide, typically from ±5 to ±15 V (direct current) and often wider ranges, so it pays to check the data sheet for your device. Operational amplifiers can also be obtained that will work with a "one-sided" power supply, and while it can be tempting to

Figure 3.3 A simple op-amp power supply made from 2 9-V batteries. This is not the most elegant solution, but it works! Connect one terminal from each battery to each other to create a common terminal between both batteries. Then the leads off of the remaining three terminals will be +9 V and −9 V referenced to the common terminal.

try to avoid the bipolar supplies, they really don't cause all that much trouble. It is also possible to make any op-amp work with a single power supply; all that is required is to change the potential used for ground, which is basically what we did by using two batteries. Operational amplifiers that are designed to work with lower voltages and use a single, one-sided supply are becoming more common, but we will use the dual-supply approach to introduce op-amps.

Our purpose for using an op-amp is to linearly amplify a small signal into a larger one; a very similar situation was seen in the last chapter on transistors. Quite often this signal is time varying, which brings up one of the important limitations of op-amps, the gain–bandwidth product. The gain–bandwidth product (GBP) is a property of the device and is given in the data sheet as a number that describes how much gain that particular device can use as a function of the frequency of the signal being applied.[1] The frequency is usually referred to as a bandwidth, which is just the range of frequencies contained in a signal. In our case, we are paying the most attention to the maximum frequency in the signal. Typically, applications with smaller bandwidth (BW) can use more gain. This relationship is shown as

$$GBP = G \times BW. \tag{3.2}$$

The GBP identifies the bandwidth that can be passed by a device at a gain of 1, or unity gain. This can be most easily understood if we consider our inverting amplifier with the resistors chosen to be identical so that the amplifier gain is -1. If the amplifier has a GBP of 1 MHz, then the working bandwidth will be 1 MHz. A gain of -10 on the device will reduce the bandwidth to just 100 kHz without causing the signal to distort.

The inverting amplifier has many excellent features, but one particular drawback: it inverts the signal. This inconvenience can be fixed by placing two inverting amplifiers back to back to restore the signal to match the input. In such a case, the gain can be applied over two amplifiers, which can help with the frequency response if it was being limited by the GBP; or one of the amplifiers can be wired for unity gain as an inverting buffer. The downside to this is that it requires space on the circuit board for two amplifiers, and more power to drive them. Fortunately, this can also be accomplished using a non-inverting amplifier configuration.

3.2 Non-inverting Operational Amplifier Configuration

There are two inputs on an operational amplifier, one marked with a negative sign, the other with a positive. The negative input is associated with the inverting amplifier, and the positive with the non-inverting amplifier,[1,3] as shown in Fig. 3.4.

The major difference between the inverting and non-inverting configuration is really which input ties to ground and which acts as the signal input. The

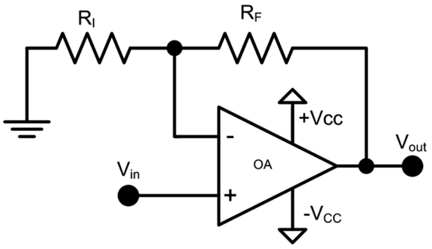


Figure 3.4 The non-inverting amplifier configuration.

most noticeable operational difference is in the gain equation, which, for the non-inverting op-amp is given as

$$G = \frac{V_{out}}{V_{in}} = \left(1 + \frac{R_F}{R_I}\right). \tag{3.3}$$

The gain relationship is not too different from the inverting amplifier, but the negative sign is gone and a 1 has appeared. Thus, if the resistors have the same value, the gain will be 2; if there are no resistors present, the gain will be 1. Example 3.1 shows a script that calculates the gain for both the inverting and non-inverting configurations.

Example 3.1

Write a MATLAB® script to calculate the operational amplifier gain for both inverting and non-inverting configurations.

```
% Example_3_1.m
% SWT 12-6-15

% This program calculates the gain for inverting and
% noninverting operational amplifiers based on the
%resistors

% Housekeeping
clear all; clc;

% Resistors
RI = 100; % Ohms
RF = 10000; % Ohms

% Equations
Rf = 10000;
Ri = 1000;
Resistor_Ratio = Rf/Ri;
Gain_inv = -Resistor_Ratio;
Gain_non = 1 + Resistor_Ratio;
```

```
% Output
fprintf('Amplifier Gain\n');
fprintf('Resistor Ratio = %6.1f\n',Resistor_Ratio);
fprintf('Gain for Inverting = %6.1f\n',Gain_inv);
fprintf('Gain for Non-Inverting = %6.1f\n',Gain_non);

Results:
Amplifier Gain
Resistor Ratio = 10.0
Gain for Inverting = -10.0
Gain for Non-Inverting = 11.0
```

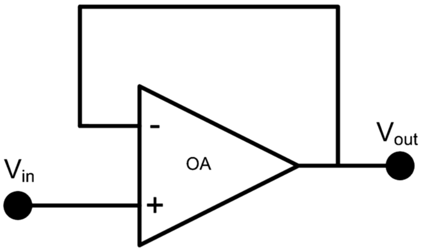


Figure 3.5 The buffer amplifier, or unity gain configuration, for an operational amplifier.

When unity gain is desired, that is, a gain of one, setting the feedback resistor to zero should work. Figure 3.5 shows the configuration of a unity gain or buffer amplifier.

The buffer amplifier[1–3] is a useful circuit to eliminate loading effects that can be caused when trying to connect a device with high impedance to one with low impedance, a situation that occurs with many sensors. Some issues might need to be addressed when using buffer amplifiers connected to capacitive loads. These issues are usually discussed in the device data sheet.

3.3 Operational Amplifiers: Math Circuits

Operational amplifiers have a wide range of uses and applications beyond amplification and are used to do analog mathematical functions. Some common circuits are used to sum or combine several signals, or to integrate and differentiate a signal. One of the most useful circuits is the differential amplifier, but it is sufficiently important that it will have its own section. Figure 3.6 shows the summing amplifier.[1–3]

The voltage out from the summing amplifier is given as

$$V_{out} = -R_F\left(\frac{V_1}{R_1} + \frac{V_2}{R_2} + \frac{V_3}{R_3}\right). \tag{3.4}$$

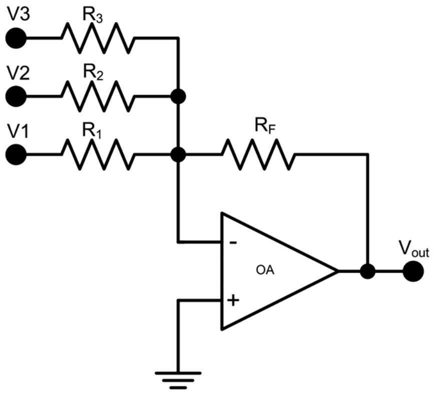


Figure 3.6 Operational amplifier configured as an inverting summing amplifier.

In this case, there are three different voltages and three different resistors being combined to produce V_{out}, but this circuit can be altered to meet the number of signals and their relationship.

The integrating amplifier[1,2] is also an inverting amplifier and is easily identified by the presence of the capacitor (C) in the feedback loop, as seen in Fig. 3.7. If you recall, capacitors act as an open circuit for constant voltages and act as a short circuit for high frequencies. The closed-circuit gain was described earlier, but the open-circuit gain has not been mentioned. When there is no feedback, an op-amp is designed to have very high gain, on the order of 100,000 or 1,000,000. So at very low frequency, the integrator is going to show a very high gain. This is often corrected by adding a resistor in parallel to the capacitor so that this situation is avoided.

The transfer function for the inverting integrator, with both a capacitor and a resistor in the feedback path, is shown as

$$G = \frac{V_{out}}{V_{in}} = -\frac{R_f}{R_I}\left(\frac{1}{1+sCR_f}\right). \tag{3.5}$$

Notice the sudden entrance of *s* in the equation. If you are familiar with 'imaginary' numbers, then you know where this is going; if you are not

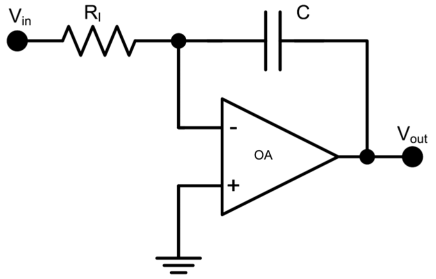


Figure 3.7 Operational amplifier configured as an integrating amplifier.

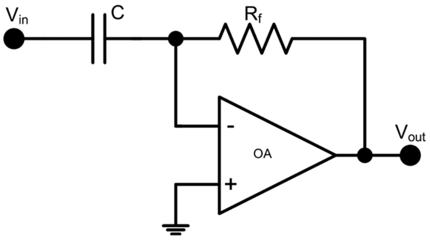


Figure 3.8 Differentiator amplifiers perform the mathematical operation of differentiation, the opposite of integration, and are also inverting amplifiers.

familiar with them, this may be a bit confusing. The definition of s is $j\omega$, where j is $(-1)^{1/2}$; $\omega = 2\pi f$, where π is a constant; f is the frequency, which brings us back to something familiar. We will touch on this again.

The circuit shown in Fig. 3.8 performs the mathematical operation for producing the derivative of the input signal over time, so it is appropriately named a differentiator. The gain for the differentiator circuit is

$$G = \frac{V_{out}}{V_{in}} = -sCR_f. \tag{3.6}$$

3.4 Differential Amplifiers

The differential amplifier[1] configuration is commonly used in noisy environments for its ability to reject signals that are common to both inputs. The circuit diagram for the differential amplifier is shown in Fig. 3.9. The differential amplifier provides linear amplification of the difference between the two input signals while suppressing the voltages that are common between the two signals. This is referred to as common-mode rejection and is expressed as a ratio in decibels.

Differential amplifiers are commonly used in many electronic applications but most obviously for the mathematical operation of subtraction. This is a key use for the instrumentation amplifier, which will be discussed in the next chapter.

The differential amplifier is an inverting amplifier combined with a voltage divider on the non-inverting input. One of the more useful properties of an

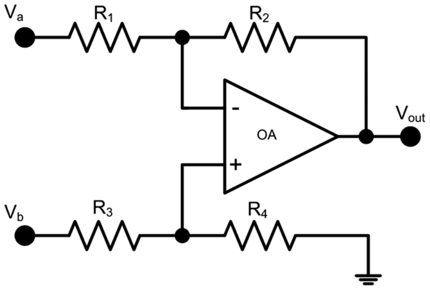


Figure 3.9 Differential amplifier showing the two inputs, V_a and V_b.

operational amplifier is that input signals that are common on its input don't pass through to the output, such that the output is driven by the difference between the two inputs. Thus, the output for the differential amplifier is as shown as

$$V_{out} = \frac{R_2 + R_1}{R_4 + R_3}\frac{R_4}{R_1}V_b - \frac{R_2}{R_1}V_a \xrightarrow{yields} \frac{R_2}{R_1}(V_b - V_a), \tag{3.7}$$

in the case where $R_4 = R_2$ and $R_3 = R_1$, so that the common mode voltage is zero. Any difference between these resistor relations will introduce errors. Thus, with careful choices of resistors, the difference relationship for this amplifier becomes apparent.

Example 3.2

Write a MATLAB script to calculate the voltage output for a differential amplifier based on the choice of four resistors.

```
% Example_3_2.m
% SWT 12-6-15

% This program calculates the output voltage for a
% differential amplifier and a check to
% see if output voltage is above the opamp
% power supply.

% Housekeeping
clear all; clc;

% Parameters
Va = 0.2; Vb = 0.1; Vcc = 15;
R1 = 100; R2 = 1000; % Ohms
R3 = 50; R4 = 1000; % Ohms

% Equations
Vo = (R2+R1)/(R4+R3)*R4/R1*Vb -R2/R1*Va;

if abs(Vo) > Vcc
   fprintf('Error: Gain too high\n');
end

% Output
fprintf('Vo = %6.1f\n',Vo);

Results:
Vo = 0.000
```

The calculator in Example 3.2 can be used to investigate a number of questions, such as the effect of having a difference in values between the input resistors. This is demonstrated in Example 3.3, where the resistor $R2$ is 90% of the value of $R1$.

Example 3.3

Consider a differential amplifier with $R_3 = 0.9\ R_1$ and $R_1 = 100$; $R_2 = R_4 = 1000$; and $V_a = V_b = 0.2$ V.

Running these values in the calculator shows a voltage difference of 18 mV. If we substitute the relationships into Eq. (3.7), we find that

$$V_{out} = \left(\frac{R_2 + R_1}{R_2 + 0.9R_1}\frac{R_2}{R_1} - \frac{R_2}{R_1}\right)V_a = \left[\left(\frac{R_2 + R_1}{R_2 + 0.9R_1} - 1\right)\frac{R_2}{R_1}\right]V_a,$$

$$V_{out} = \left(\frac{11}{10.9} - 1\right)10\ V_a = 0.092\ V_a = 18\,\mathrm{mV}.$$

3.5 Transfer Functions, Imaginary Numbers, and Decibels

In electronic circuits, the introduction of capacitors and inductors, often called reactance, changes the simplicity of the mathematics used. Reactance is a sensitive parameter of frequency, and the circuit will respond differently at different frequencies. The frequency response is normally shown over a wide range of frequencies, often covering many decades of change.

The introduction of frequency isn't the only change in the mathematics. A pesky little '*j*' starts to appear, usually with the frequency in the form of $\omega = 2\pi f$, where f is the frequency. What this means for us is that the mathematics gets harder, as we no longer trace only the magnitudes of signals but also the phase relationships, which returns us to the idea of the transfer function. In linear amplifiers, this is simply the gain; however, once reactance starts to appear, we pay more attention and refer to the gain as a function of frequency as the transfer function.

A low-pass filter[1,2] is shown in Fig. 3.10. For our purposes here, let's suppose this circuit to be constructed of resistors, so it is a voltage divider, and we will analyze it accordingly. The components will be considered in generic terms as the letter Z, representing the impedance, with an appropriate subscript. This lets us write the ratio of the output voltage to the input voltage as

$$\frac{V_{out}(s)}{V_{in}(s)} = \frac{Z_C}{Z_c + Z_R} = \frac{\frac{1}{sC}}{R_C + \frac{1}{sC}}. \tag{3.8}$$

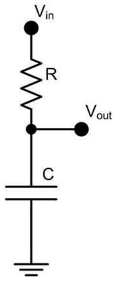


Figure 3.10 The circuit for a simple low-pass filter.

The ratio in Eq. (3.8) is not quite as simple as the resistor voltage divider but not that much different, either. This is the expression of the transfer function for the low-pass filter. If we apply some simple mathematical tools, this equation can be reduced to the form

$$H(s) = \frac{V_{out}(s)}{V_{in}(s)} = \frac{1}{sRC + 1} \Rightarrow \frac{1}{j\omega RC + 1}. \tag{3.9}$$

The transfer function provides an expression for the relationship between the output and input voltages as a function of frequency, but described as an imaginary number. To get a magnitude and a phase relationship in the form we are used to, we need to look at how to work with the imaginary numbers.

3.6 Bode Plots

The Bode plot[1] is a graphical tool for displaying the frequency response of a system and usually has two forms: a magnitude plot and a phase plot. Electronic signals can cover a wide range of frequencies, so the frequency axis is shown as a log plot and the magnitude axis is expressed in decibels, while the phase is expressed in degrees. The two plots are shown in Fig. 3.11, with the upper graph showing the magnitude plot and the lower graph showing the phase plot.

The low-pass filter is essentially just a voltage divider with the lower impedance acting as a capacitor and the output voltage measured across the capacitor. This is often called a frequency variable potential divider. The value of the Bode plot, which is a plot of the transfer function, is to provide a graphical display of the frequency response of the circuit.

The −3 dB point on the magnitude curve corresponds to the 45-deg phase frequency at 159 Hz and is shown in the figure as an asterisk. The region after the −3 dB point falls at a rate of −20 dB per decade of frequency.

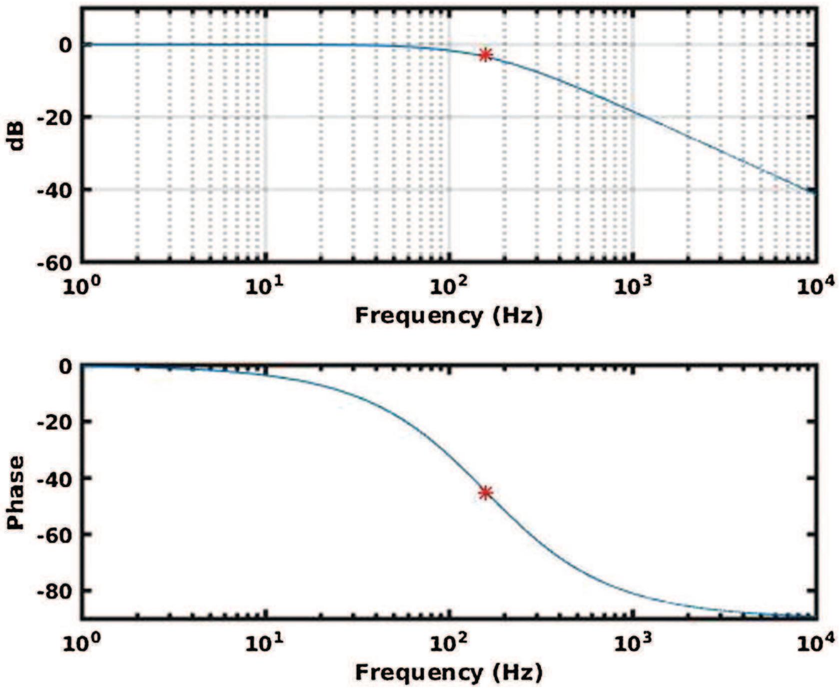


Figure 3.11 Bode plots for a low-pass filter using $R = 1$ kΩ and $C = 1$ μF. The -3 dB points are shown as asterisks on the curves.

3.7 Practice Problems

1. Calculate the gain for an inverting and non-inverting operational amplifier with $R_{in} = 1000$ and $R_{out} = 2000$.
2. A summing amplifier has three voltages, 0.1 V, 0.2 V, and 0.3 V, into its inverting amplifier. Determine the values of the input and feedback resistors so that the output of the amplifier is -10 V.
3. A differential amplifier has two input voltages of 0.1 V and 0.2 V. Determine a set of values for the resistors such that the output will be 0 V.
4. The low-pass circuit in Fig. 3.10 has an input voltage of 10 VDC, a resistor of 1 kΩ, and 300-pF capacitor. What is the output voltage V_0? What is V_0 if the input voltage is 10 VAC (volts of alternating current) at 100 kHz?

Answers

1. -2, 3
2. 100 Ω; 200 Ω; 300 Ω; 33,000 Ω
3. 200 Ω; 1400 Ω; 1000 Ω; 1000 Ω
4. 10 V; 0 V

References

1. A. S. Sedra and K. C. Smith, *Microelectronic Circuits*, Oxford University Press, Oxford (2015).
2. P. Horowitz and W. Hill, *The Art of Electronics*, Third Edition, Cambridge University Press, Cambridge (2015).
3. "Op Amp Circuit Collection," National Semiconductor Application Note AN-31, National Semiconductor Corp., Santa Clara, California (2002).

Chapter 4

Useful Op-Amp Circuits

The previous chapter introduced operational amplifiers, which have become the fundamental devices for constructing many electronic circuits. Op-amps are constructed of a number of transistors, resistors, diodes, and capacitors on a single silicon chip, resulting in their high reliability and low cost. But there is much more!

Op-amps can be used as a single stage or with several of them chained together, feeding the output of one to the input of the next. By clever selection of the external components (often referred to as lumped components), and like resistors, diodes, capacitors and inductors, operational amplifier circuits can perform a wide variety of mathematic and logic operations or act as filters. These useful circuits, made of op-amps and lumped components, are the focus of this chapter. There are many manufacturer-complied resources, often called application notes, for using op-amps. Some examples from Analog Devices, Inc. and National Semiconductor, Inc. are included in the Bibliography.

The chapter begins with arguably the most useful operational amplifier device, the instrumentation amplifier. Then we will explore several of the different op-amp-based mathematical operation circuits.

4.1 Instrumentation Amplifiers

The instrumentation amplifier[1–3] evolves from the differential amplifier that we introduced in the last chapter. The differential amplifier has many advantages over the inverting and non-inverting amplifier, most notably, the common mode rejection, but this is still not the performance promised in the "ideal" amplifier definition. The ideal amplifier is usually thought of as having the following properties:[3]

Infinite

- open loop gain
- input impedance
- output voltage range

- bandwidth
- common mode rejection ratio
- power supply rejection ratio

Zero

- input offset voltage
- output impedance
- noise

While we commonly think of the ideal amplifier as meaning idealized, or a pie-in-the-sky dream of an amplifier, what we are really saying is that this is how we would like an amplifier to behave, and as engineers, we are actually trying to achieve that level of performance.

Some of the properties listed above are more readily available than others, and if we return to the differential amplifier, we can look at some designs to improve its performance.

In Fig. 4.1, the differential amplifier can be clearly seen on the right side of the circuit connected by resistors R_1 and R_2. The two additional op-amps on the left side are configured a bit differently. If R_{gain} was infinite and the two resistors R were zero, these would look a lot like unity gain voltage followers. The great thing about the unity gain voltage follower, or buffer, is that it provides the desired high input impedance. Things get a little more interesting with the circuit when we allow the resistors in the buffers to take on more realistic values. Their values can be determined by looking at the transfer function for the instrumentation amplifier as shown:

$$H = \frac{V_0}{V_I} = \frac{V_0}{V_b - V_a} = \left(1 + \frac{2R}{R_{gain}}\right)\frac{R_2}{R_1}. \tag{4.1}$$

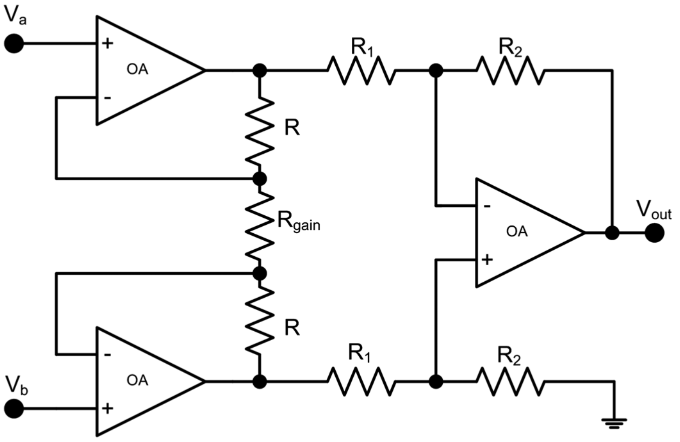


Figure 4.1 The instrumentation amplifier showing the modified input stages fed by two voltages, V_a and V_b.

Equation (4.1) shows that we would like to have the two resistors R be the same and have them contribute to the overall gain of the system. It is common to have nearly unity gain on the differential amplifier and to have the R values be large with a low-resistance R_{gain}. While there are many combinations that will give the R_2/R_1 unity, a typical value for both resistors is 10 kΩ. Precision resistors can be expensive, so we can also make the R values 10 kΩ. This leaves only the R_{gain} to select for the desired gain. Example 4.1 shows how to tune the gain of the instrumentation amplifier by adjusting the gain resistor.

Example 4.1

An instrumentation amplifier uses 10-kΩ precision resistors throughout and controls the overall gain by manipulating R_{gain}. Plot the range of gains available for this op-amp for $10\ \Omega < R_{gain} < 500\ \Omega$.

Equation 4.1 reduces to

$$G(R_{gain}) = 1 + \frac{20000}{R_{gain}}.$$

This simplified equation, or even Eq. (4.1), can be used in MATLAB® to generate the final result shown in Fig. 4.2.

The curve in Fig. 4.2 was generated using the MATLAB code[4] in Example 4.1.

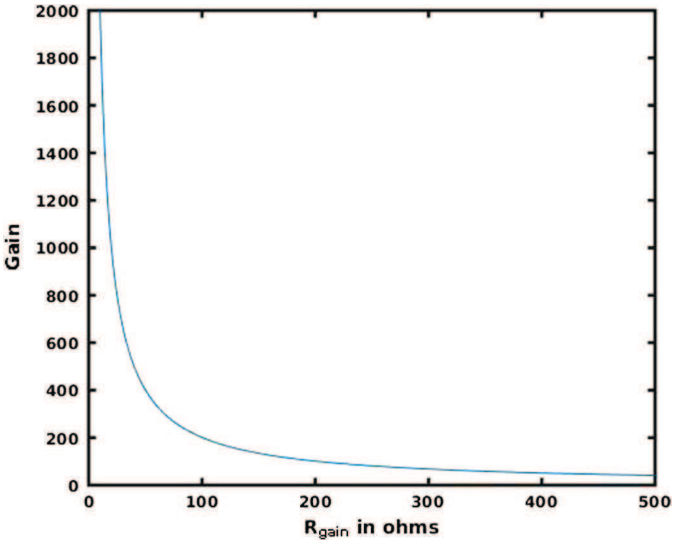


Figure 4.2 The gain for the instrumentation amplifier with R_{gain} varied and all other resistors at 10 kΩ.

Example 4.2

Plot the gain of an instrumentation amplifier as a function of R_{gain} for $R = R1 = R2 = 10000$.

```
% Example_4_2.m
% SWT 10-2-15

% How to design an instrumentation amplifier

% Housekeeping
clear all;
clc;

% Mathematical function
R = 10000; R2 = 10000; R1 = 10000;
Rgain = 10:500;

H = (1+2*R./Rgain)*R2/R1;
% Plotting
clf;
figure (1); plot(Rgain, H);
set(gca,‘FontSize’,9,‘FontWeight’,‘Bold’,‘LineWidth’,2);
axis ([0 500 0 2000]);
xlabel(‘Rgain in ohms’); ylabel(‘Gain’);

Result is shown in Fig. 4.2.
```

Figure 4.2 illustrates some very interesting dependencies in our instrumentation amplifier. To get the higher gain, we are using lower values in R_{gain}. At the lower end of the resistances, it is clear that small changes in resistance can make a big change in gain, and that as we move out past 100 Ω, the gain varies more slowly with changes in resistance.

This can be quantified if we take the derivative of H with R_{gain} as

$$\frac{dH}{dR_{gain}} = \frac{2RR_2}{R_1}\left(\frac{-1}{R^2}\right). \tag{4.2}$$

If we plot Eq. (4.2) over the same range of R_{gain} values, the dependency on the gain change can be seen and is shown in Fig. 4.3.

When selecting the resistor combinations to use for the instrumentation amplifier, it is good to consider a number of factors that might come into play. In the example here, high amplification is possible, but small changes in the gain resistance can make big changes in the gain.

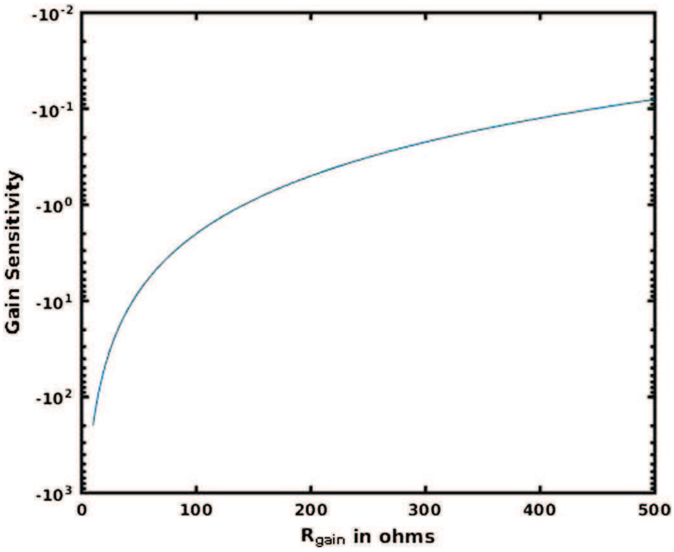


Figure 4.3 The change in gain with the value of R_{gain} for the instrumentation amplifier shown on a semilog plot.

4.2 Improved Precision Rectifiers

The precision rectifier or superdiode is an op-amp circuit that includes a diode in the negative feedback loop and has the effect of making a rectifier circuit that behaves like an ideal diode. The most commonly used method to make a superdiode is to insert a diode into the feedback loop of the op-amp configured as a buffer. This configuration does not work well if the input becomes negative, as the op-amp will have the open loop gain.

The improved precision[5] rectifier circuit avoids this issue and is shown in Fig. 4.4. The transfer function of this circuit might be unexpected when thinking about a diode but should make sense when we consider the inverting op-amp. The transfer function, i.e., the relationship between the output voltage and the input voltage, will be $-R2/R1$ when the input voltage is

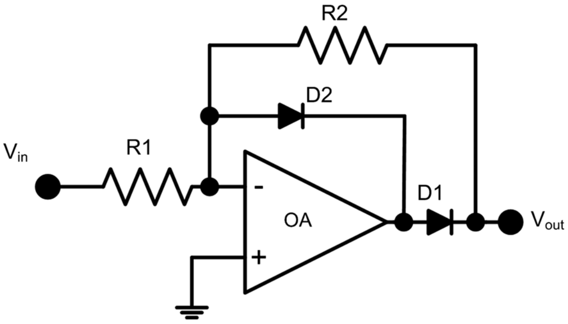


Figure 4.4 The improved precision rectifier or superdiode circuit to make an “ideal” diode.

negative and zero when the output voltage goes positive. Of course, this can be inverted by including an additional inverting op-amp stage.

4.3 Peak Detectors

A peak detector[1,3] is a circuit that samples a signal and holds the maximum voltage reached in a given period of time. The time period that the value is held can be set by the resistor–capacitor time constant or by a switch. A simple peak detector can be constructed using a diode, a resistor, and a capacitor, or a more complicated resettable circuit can be made by using the circuit for a superdiode with a transistor, a switch, and a capacitor. Both peak detectors are used to hold the peak voltage level of a signal. The circuits are shown in Fig. 4.5 with the simplest version on the left and the two reset mechanisms on the right.

The peak detector has its beginnings as a buffer or unity gain amplifier, which is then extended to a superdiode design by adding diode D1 (see Fig. 4.5). If C1 was a resistor, then this would be the simple superdiode circuit, but in this case a capacitor C1 has been included. Operationally, switch S1 is closed to bleed off any charge from the capacitor by connecting to ground. A positive signal pulse at V_{in} will create a proportional output V_{out} that will charge the capacitor through the diode, provided that the input voltage is greater than the turn-on voltage of the diode. If the next signal pulse is less than the first, the diode will not turn on, and nothing changes. If the subsequent pulse is high in voltage, it will change the voltage on the capacitor, and this value will stay until either S1 is closed or a reset signal is sent to transistor Q1 to drain the capacitor charge to ground.

While this circuit does not come up that often the peak detector is a good circuit to remember as it does fill a number of roles related to being able to capture and hold the maximum value of a signal.

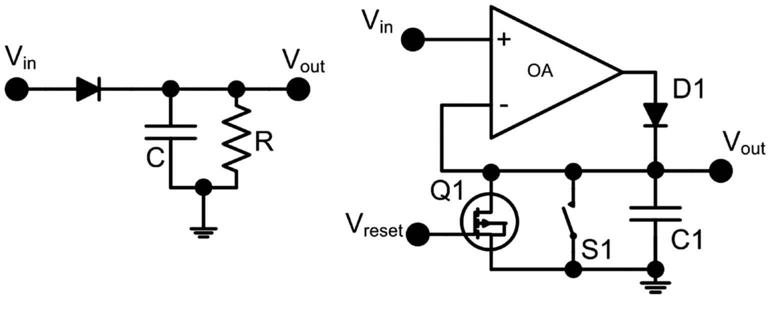


Figure 4.5 Two peak detector circuits. Left: A simple detect and hold circuit. Right: A more advanced circuit with two reset switches, one a manual switch and the other an electronic switch.

4.4 Oscillators

An electronic oscillator[1,5,6] generates a periodic signal often in the form of a square or sine wave. These signals are often used as a clock, a regularly repeating signal that can be used to mark time. If we wanted to know the maximum value of a signal every second, we could use the peak detector circuit and make it so that the peak voltage was reset every second. This could be done with the manual switch, but it would be much more practical to send an electronic signal to the transistor reset switch once every second. There are many ways to do this, but since we are working with op-amp circuits, we will use an op-amp that we will configure to provide a signal pulse out every second or have it act as an oscillator with a frequency of 1 s. An op-amp oscillator is shown in Fig. 4.6.

The oscillation frequency will be controlled by the RC time constant set up by the product of $C1$ and $R1$. A 1-s oscillation will require that this product be 1. If we choose a large resistor, say, 1 MΩ, then we will need a capacitor that is 1 μF to give the product of 1 s.

This is a very simplistic oscillator circuit; many enhancements to this circuit are possible. Oscillators are such important devices in electronics that specific devices commonly used for this purpose have been designed for easy operation. The most common of these is the 555 timer. In addition, crystal oscillators, based on the properties of a quartz crystal, can be used to provide a clock signal; these devices are available over a wide range of frequencies.

4.5 Sample-and-Hold Circuits

Other than the instrumentation amplifier, we have been using a single op-amp device to demonstrate the range of applications for operational amplifiers.

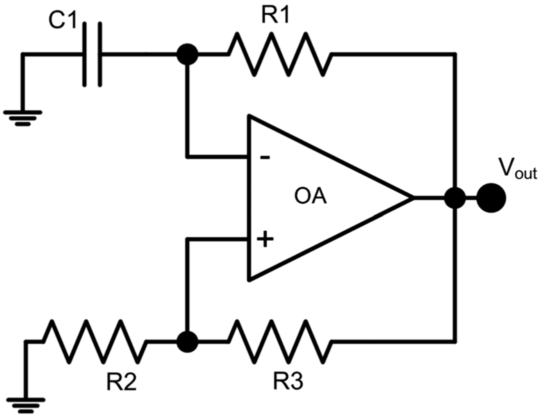


Figure 4.6 An op-amp configured as a relaxation oscillator.

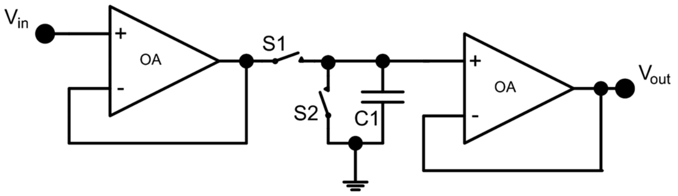


Figure 4.7 A sample-and-hold circuit constructed from two operational amplifiers.

Additional uses are possible as we begin to combine or cascade op-amps, one of the nicest being the sample-and-hold circuit.[1,3,5] This circuit is designed to have minimal effect on the signal being sampled and to ensure that the measured value can be accurately read over an extended time period. This device makes use of the nearly infinite input impedance of the op-amp and is shown in Fig. 4.7.

Some sensors will not provide an accurate voltage if they are required to supply a current. The high input impedance of the unity gain operational amplifier minimizes the current from the input signal and provides an output signal proportional to V_{in}. When the momentary switch S1 is closed, capacitor C1 is charged to V_{in} and can be read at V_{out} with the low input impedance of the second op-amp ensuring that the capacitor can hold its charge for an extended period of time. Once the sampled voltage has been read, the momentary switch S2 is closed, resetting the capacitor for the next value.

Just like the peak detector circuit we looked at earlier, replacing the manually operated switches with transistors provides electrical control of the circuit. This allows for very reliable and very fast operation of the sample-and-hold circuit and has made this a very important circuit for analog-to-digital conversion circuits. The sample-and-hold circuit is used in many different circumstances so there are many variations on the basic design. The result is that quite often the function of the circuit is obscured by the additional components.

4.6 Voltage Indicators

We have been looking at a number of very useful and versatile op-amp circuits, but let's face it—they are cool but boring. It is time for something with a little glitz! Sometimes we need a quick visual indicator of how much voltage there is at the output of a circuit. In this case, we don't necessarily want to pick up a voltmeter or an oscilloscope, but just get a good indication of what we are seeing. The circuit in Fig. 4.8 provides such an indicator.

This circuit functions by comparing the input voltage on each operational amplifier. Once this is above the reference voltage on the negative input, the LED will turn on. Variable resistors can be substituted on the reference chain to make the turn-on voltages adjustable, and differently colored LEDs, or the number of LEDs that are on, can provide a quick visual indication of the

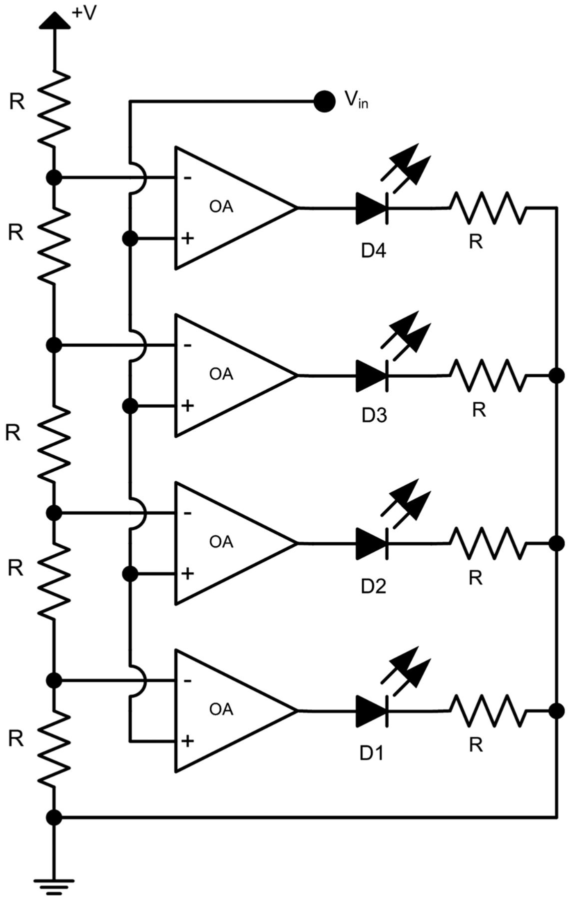


Figure 4.8 An LED voltage indicator using operational amplifier drivers.

voltage level. Depending on the desired function, a number of improvements and variations are possible when implementing this circuit, including specialty chips that have much of the functionality built in.

This type of indicator has a wide range of uses and can be constructed from operational amplifiers that come in quad packages; i.e., there are 4 individual operational amplifiers in a single DIP package with 14 or more pins. These are amazingly compact devices and come in several different mounting styles.

4.7 Practice Problem

1. Calculate the effect of 5% difference between the two input resistors to the differential amplifier in an instrumentation amplifier when the input voltages to the instrumentation amplifier are the same.

(The solution is left to the reader as this is a fairly common problem.)

References

1. P. Horowitz and W. Hill, *The Art of Electronics*, Cambridge University Press, Cambridge (2015).
2. C. Kitchin and L. Counts, *A Designer's Guide to Instrumentation Amplifiers*, Third Edition, Analog Devices, Inc., Norwood, Massachusetts (2006).
3. A. S. Sedra and K. C. Smith, *Microelectronic Circuits*, Oxford University Press, Oxford (2015).
4. H. Moore, *MATLAB® for Engineers*, Fourth Edition, Prentice-Hall, Upper Saddle River, New Jersey (2014).
5. "Op Amp Circuit Collection," National Semiconductor Application Note AN-31, National Semiconductor Corp., Santa Clara, California (2002).
6. J. Wong, "A Collection of Amp Applications," Application Note AN-106, Analog Devices, Inc., Norwood, Massachusetts (undated).

Chapter 5

Digital Electronics

Digital devices and computers dominate the information world around us, and digital electronics drives much of the consumer electronics market. Most of our displays, information sources, and communications all involve digital electronics. However, the sensors and amplifiers used to interrogate and investigate physical phenomena are based solidly in analog electronics, so moving the data collected by the sensor into a computer requires conversion into digital form.

Digital electronics is based on digital logic; i.e., it represents information digitally as specific combinations of 1s and 0s. There are three basic logic functions that can operate on binary numbers: AND, OR, and NOT. You might think that with such a limited array of variations, digital logic would not have much use, but you will quickly see that there are many constructs possible in the digital logic world, and they are all based on these three operations. Digital logic can be implemented in many ways by using a wide range of electronic devices, including discrete components, field programmable gate arrays, microcontrollers, and more.

This chapter provides a short overview of digital electronics and a somewhat fast-paced look at what can be accomplished using these diverse tools. Many books on digital electronics are available, should you need more information; here we will be looking at the high points and staying aligned to areas of interest to optical engineering.

5.1 Elements of Digital Logic

Before we enter the world of digital logic, we need to become familiar with the basics of binary numbers and Boolean operations.[1,2] This sounds like math, and it is, but there are a few rules to follow; and once we have then down, this becomes the language we need to use. Starting at the beginning with an electronics problem, we have a voltage and we want to convert it from an analog representation to a binary representation. We will start with something easy like 5 V. In our usual system of numbering, 5 V is a number that uses a base of 10, as if we are counting. There are many bases; for instance, binary numbers use base 2, and hexadecimal numbers use base 16. One thing that always causes issues is deciding what our starting number is: Is it a zero or a one? All of this will become clear in a moment.

We will first introduce a formalism for representing numbers so that we can easily change between representations. In this section we will be specific about the base we are using to represent a number. In our everyday world, we use base 10 numbers. If we show this explicitly, the number 5 in base 10 form will be written as 5_{10}. When we write 5_{10} as binary it becomes 0101_2. Here the 1s and 0s are in a specific order, and this value is shown as a four-bit number. Thus, we have two ways to represent a number, and we need a means of converting between them. When we count in base 10, we increment the numbers by 1 until we reach 9, and then we start the count again but add 10 to the value. In binary representation, there are only two values, 0 and 1, and then you add a 1 to the twos column and continue the count. A look at Table 5.1 should make this clear. We will go through the process for 16 values so that we can compare base 10, base 2, and base 16 numbers.

The basic relationship used for converting between base 2 and base 10 numbers is shown as

$$\begin{aligned} Value_{10} = {} & MSB\ x\ 2^3 + (MSB - 1)x\ 2^2 \\ & + (MSB - 2)x\ 2^1 + (LSB)x\ 2^0. \end{aligned} \tag{5.1}$$

Here, MSB is the most significant bit, the bit that is the leftmost; LSB is the least significant bit, the bit that is the rightmost. So the four-bit number 1000

Table 5.1 Comparison of base 10, 2, and 16 numbering.

Decimal (base 10)	Binary (base 2)	Hexidecimal (base 16)
0	0	0
1	1	1
2	10	2
3	11	3
4	100	4
5	101	5
6	110	6
7	111	7
8	1000	8
9	1001	9
10	1010	A
11	1011	B
12	1100	C
13	1101	D
14	1110	E
15	1111	F

in base 2 has the value in base 10 of 1×2^3 or 8. A converter for different bases is shown in Example 5.1.

Example 5.1

Write a MATLAB® script to convert the number 105 into binary, hexadecimal, and base 12.

```
% Example_5_1.m
% SWT 12-12-15

% This program converts a number into its binary, hexidecimal
% and arbitrary base value

% Housekeeping
clear all; clc;

% Parameters
myNumber = 105; % Number to be converted
myBase = 12; % The base for conversion

% Conversions using MATLAB built in commands
myBinary = dec2bin(myNumber);
myHexidecimal = dec2hex(myNumber);
myBase = dec2base(myNumber, myBase);
% Output
disp(myNumber);
disp(myBinary);
disp(myHexidecimal);
disp(myBase);

Results:
Binary: 1101001
Hexidecimal: 69
Base 12: 89
```

To convert from base 10 numbers to base 2 manually requires successive division by 2. Consider the value 14. If I divide by 2, I get 7 with no remainder; this zero remainder is the LSB. Continuing, I divide again by 2 and get 3 with a remainder of 1, the remainder puts a 1 in the next column up from the LSB. Continuing, I divide again by 2 and get 1 with a remainder of 1. Continuing, I divide again by 2 and get 0 with a remainder of 1, which is the MSB. A little clumsy and tedious, but it works. We will talk a little more about hexadecimal numbers later.

Table 5.2 Truth table for AND, OR, and NOT.

A	B	AND	OR	NOT (A)
1	1	1	1	0
0	1	0	1	1
1	0	0	1	0
0	0	0	0	1

Just like base 10 numbers, binary numbers can be used in mathematical operations as long as the rules are followed. In addition, the following are true for binary numbers: $00+00=00$; $00+01=01$; $01+00=01$; and $01+01=10$. In subtraction, the following are true: $01-00=01$; $10-01=01$; and $11-01=10$. With these basics we can solve a wide range of combinations.

It should be noted that we are applying our conversions to numbers that do not have decimal points after them. These are commonly known as integer numbers. It can take a moment to realize how to use this effectively. Consider the following idea. Our 4-bit binary number has only 16 possible discrete values. If we set the maximum value for the largest number to be 10 V, we will have created bin sizes of 10/15 V or 0.67 V steps. This is known as discretization. When finer voltage steps are needed, we will increase the number of bits in our system. The following provides a bit of a guide for how many bits correspond to the number of levels: 4 bits = 16; 8 bits = 256; 16 bits = 65,536, and so on, according to $\text{level} = 2^{\text{bits}}$.

There are also logical operations such as AND, OR, and NOT, which are best shown in a truth table; i.e., for any combination of inputs, the outputs are shown.[1,2] Table 5.2 shows a truth table for two inputs and their combinations. For two inputs, A and B, having the values of 1s and 0s or ON and OFF, the result for the application of the operations AND, OR, NOT (A) are shown. Since the NOT function applies to one input, it is shown here as applied to the input A.

From these simple operations many more complex operations can be derived, such as the XOR. The XOR (exclusive OR) has an output value of 1 only when the input values are different and zero, otherwise. This has been a short coverage of the topic of digital logic, but it is sufficient for us to start discussing digital electronics and to see how this logic is put to use.

5.2 Binary Logic

Our interest in digital logic hinges on our ability to convert the logic into electronic devices that we can use in instrumentation to determine changes in state. Let's set up a simple problem scenario where we can use an electronic circuit to differentiate between several states. The problem can be defined with the simple idea that we don't want an instrument to operate if either of two specified conditions aren't met. The two conditions are that the door to the room must be closed, and a safety gate in the room must be closed. This can

Table 5.3 Truth table for AND and OR for the Door and Gate inputs.

Door	Gate	AND	OR
1	1	1	1
0	1	0	1
1	0	0	1
0	0	0	0

be translated to a simple table (Table 5.3), in which we represent the open state as a 0 and the closed state as a 1. The table shows that of the two logic designs, the AND operator is the only one that shows a change of state on the output when only one door changes state.

Translating this to an electronic system is no too hard; we can put switches on the door and watch their status to indicate closed or open. If we put both switches in series, then both switches must be closed in order to work; however, this might require a long wire run, or introduce other problems that we prefer to avoid. As long as we can get the signal from each switch to a common location, we can use a logic device. The simplest way to make an AND is with a couple of diodes,[3,4] as shown in Fig. 5.1.

Figure 5.1 is a logic device that connects to two switches that will change state from open to ground, depending on the state of the gate and door. This is the basic idea of an interlock, a set of conditions that must be met before the system will allow operation; otherwise, the operation is “locked out.” We will consider the door first. When the door is closed, $V_{decision}$ must have a nonzero value. This requires the switch on the door to connect to ground only when the door is open and not connect to ground when the switch is closed. Thus, when the door is closed, there is no connection to ground, and $V_{decision}$ is at the high level. When the door is open, the switch connects to ground, and $V_{decision}$ is at the ground value. The gate operates in exactly the

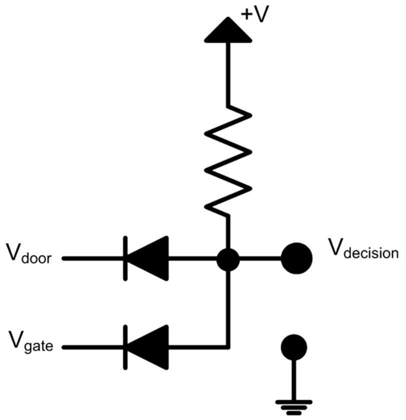


Figure 5.1 The configuration of diodes and resistors to make a logical AND device. V_{door} and V_{gate} are 0 if they are open and 1 if they are closed.

same fashion. As such, as long as either the gate or the door is open, the signal at $V_{decision}$ is 0.

The instrument that is using this logic device will see a high voltage signal at $V_{decision}$ only when both doors are closed; this will be the condition when it can operate. When either (or both) the door or gate is open, the voltage across the diode will connect to ground, resulting in a low voltage being seen at the output $V_{decision}$, and the instrument will not turn on.

It can be very clumsy to make your own gates every time they are wanted, so the electronics industry manufactures a wide range of logic devices. Discrete components are ICs that have several logic devices on them and typically have the designation 74XX. In the series, the 7400s are logic devices with the designation NAND, which stands for NOT AND. We can translate Table 5.2 by flipping the bit state in the AND column for the same inputs. The 7402 device is a NOR, 7404 is a NOT, and 7408 is an AND. Figure 5.2 shows three common electrical device logic symbols.[1,2]

The interlock example is only one of many applications for logical decision making and just begins to show the value of using electronic logic devices. This example also shows some of the connectivity of binary logic to analog electronics. When we consider that modern microcontrollers and computers are composed of billions of such logic devices, it should become clear that we are just brushing the surface. MATLAB can be used to effectively perform logical operations[5] and display the answers, as shown in Example 5.2.

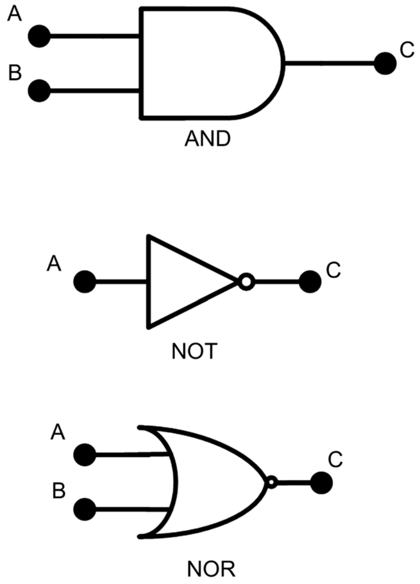


Figure 5.2 Logical symbols for AND, NOT, and NOR. The NOT symbol shows a buffer, the triangle, with a small circle at the end. The small circle indicates that the output takes the opposite value. The small circle placed on the OR symbol converts it to a NOR.

Example 5.2

Make a calculator to convert two sets of numbers $A = (1\ 0\ 1\ 0)$ and $B = (1\ 1\ 0\ 0)$ into their AND, OR, and XOR sets.

```
% Example_5_2.m
% SWT 12-12-15

% This program converts two sets of binary values A, B into
% new number sets based on logical AND, OR and XOR
% operations

% Housekeeping
clear all; clc;

% Parameters
A = [1 0 1 0];
B = [1 1 0 0];

% Conversions
C = A & B; % logical AND
D = A | B; % logical OR
E = D & ~C;% logical XOR

% Output
disp(C)
disp(D)
disp(E)

Results:
C:  1  0  0  0
D:  1  1  1  0
E:  0  1  1  0
```

5.3 Flip-Flops and Latches

Flip-flops are digital logic circuits[1,2] that have only two stable states and can be used to store information in the form of the state of the circuit. The flip-flop is the basic storage element and is an important element in digital logic. Understanding the basic principles of the flip-flop can be accomplished by looking at one of the simplest versions of the device, the set-reset latch[1] shown in Fig. 5.3.

The individual NOR gate produces a high value only when both of its inputs are low; all other combinations produce a zero output. The S-R latch

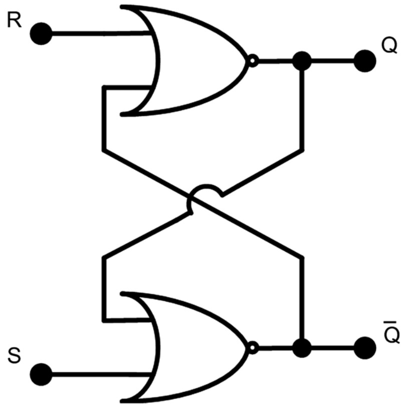


Figure 5.3 Set-reset (S-R) latch constructed from a pair of cross-connected NOR gates. The bar over the Q is another indication of the NOT condition.

operates by using the inputs, S and R, to change its outputs, Q and NOT-Q, shown in the figure with a bar over the Q as Q̄. The outputs are a complementary set of Q and Q̄ being either 1 and 0, or 0 and 1. There can be issues during the power up of the electronics such that the initial state is not known; we will address this issue later.

The operation of the S-R latch is best demonstrated by example. Consider the case where $S = R = 0$ on startup and the initial output state is $(Q, \bar{Q}) =$ (1,0). We need then to look at how the outputs cross connect to the inputs. Notice that the output Q that connects to the second input of the S latch must be high, i.e., have the logic state 1. Since the S NOR gate has a 1 and a 0 the result should be that the Q̄ output will be 1. This forces the second input on the R latch to be 1 driving the Q output to be a 0. This is consistent with the initial state of the inputs and outputs.

Changing the output state requires that the set input S be raised to a high value or 1. When $S = 1$, the S NOR gate changes its output state to 0, i.e., $\bar{Q} = 0$, which forces the R NOR gate to have both inputs be 0 and drives $Q = 1$. As this has set the second input on the S NOR gate to high, even if the S input goes to zero, the S-R latch will stay in the same state. Resetting the latch then requires momentarily applying a high value to the R input to reset the outputs.

The condition $S = R = 1$ is known as a forbidden condition, as it would require that both NOR gates output a zero state, which cannot be supported. When power is applied to the S-R latch, the state of the outputs can be difficult to predict. The starting state can depend on how signals propagate in a circuit, creating a race condition where it is unknown which signal will arrive first. This has resulted in the S-R latch being a very important structure in electronic circuits but most often combined with additional logic to ensure proper operation. This next step up in complexity is to make the S-R flip-flop shown in Fig. 5.4.

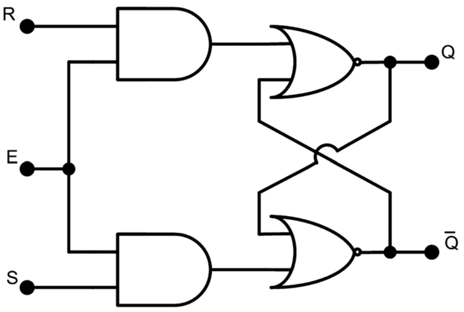


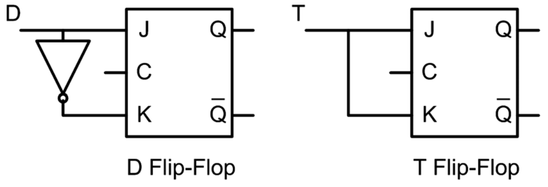
Figure 5.4 Clocked, or enabled, S-R flip-flop circuit. When the enable (E) line is high, the signals from R and S can pass to the S-R latch.

The S-R flip-flop[1] is a very simple modification to the S-R latch but is used here to introduce a very important electrical connection present on many electronic devices, the enable input, shown as E. The enable line allows inputs to a circuit to be disabled or blocked from being used by an external signal. Depending on the device, the enable line can be configured to be active when it is in the high or the low state, so it is necessary to carefully review the data sheets to know how to correctly operate it. In Fig. 5.4 the enable line must be raised high before the R and S inputs can be used.

The J-K flip-flop will be our final stop on the journey and is arguably the most usable of the basic flip-flops. It is essentially our S-R flip-flop with the $S = R = 1$ state no longer undefined; now, $S = R = 1$ has the defined role of changing the state of the output. There are a number of different types of flip-flops known as the T flip-flop, D flip-flop, and our earlier S-R flip-flop. Each of these flip-flops can be constructed from the J-K flip-flop (Fig. 5.5).

Now that we have looked at flip-flops, it is time to show them performing a task. One of the uses of a collection of flip-flops is to make a binary counter. Since we have been looking at binary numbers, this could be a useful application. A simple four-bit binary counter[1,2] can be constructed using four flip-flops connected as shown in Fig. 5.6.

There is a great deal of power in MATLAB[5] when its built-in vectorization properties are used. The ripple counter, so named because of the manner in which the values move across the components, is a means of counting in binary. The counting process can be easily implemented using a loop control statement such as a `for`[5] command to step through the various values. However, if we just want to get a collection of the binary values that result from the ripple counter, we can calculate these in a very compact way using the



D Flip-Flop T Flip-Flop

Figure 5.5 D-type and T-type flop-flops constructed from the J-K flip-flop.

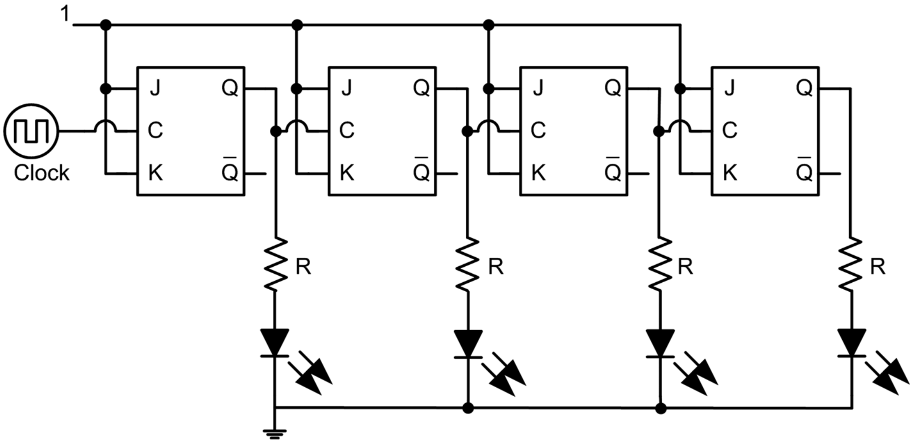


Figure 5.6 A four-bit binary ripple counter constructed from J-K flip-flops. The clock frequency appears on the leftmost LED as half of the frequency, and each consecutive LED continues to divide the signal by 2. All of the J and K inputs are tied to high (logic value 1).

`dec2bin` command. In Example 5.3 a vector of 16 values is introduced into the command and is used to generate the output of the binary counter.

Example 5.3

Construct the output of a four-bit ripple counter similar to that in Fig. 5.6.

```
% Example_5_3.m
% SWT 12-12-15

% This program shows the output of 4 bit ripple counter

% Housekeeping
clear all; clc;

% Parameters
n = 16; % Number of values to be converted
dec2bin((0:n-1))

Results:
000
001
010
011
100
101
110
111
```

5.4 Programmable Integrated Circuits

Construction of complicated logic circuits can require a large number of individual components. There are many ways to reduce the number of components by making equivalent logic circuits that make use of the rules of combinatorial logic. However, eventually, electronic designers finds themselves looking for something with a little more power, flexibility, and ease of use. This is the domain of the programmable integrated circuit, also referred to as a PIC. Originally, these chips were very limited, but they have advanced to being true microcontrollers using reduced instruction set computing that provides high-performance operation.

As with most electronics components, there are many manufacturers of PICs and many companies that have integrated the base PIC into a usable single-board microcontroller. There are various PIC chip manufacturers including Microchip Technology, Inc. and Ubicom, which require an external programmer to load a program. The usefulness of PICs led to the development of single-board PIC microcontrollers that include the programmer, extra memory, and other support devices.

The BASIC Stamp from Parallax Inc.[6] is an example of a single-board PIC microcontroller. It is is a very flexible device and can be programmed using the simple computer programing language PBASIC. While it can't do everything you can imagine, it does a pretty good job and at a very reasonable price. Parallax Inc. has been around for a long time and has a very dedicated user base.[6]

Single-board PIC microcontrollers come in a variety of sizes and capabilities. They are designed for those who are new to electronics so provide good supporting documentation and online user forums. The value of the single-board PIC microcontrollers is that the users get an operation system complete with many peripherals so that they don't have to redevelop everything. Whichever starting point is right for the user, there are some common features to be considered. These devices are designed to take input data, perform a task such as a calculation on that data, make decisions, and then output the result. Whether these calculations are digital logic, simple mathematics, or comparisons, a single device collects information and then uses it to perform a task such as turn on an LED at the right time. Essentially, if you can program it into the device, it can perform the task. This is certainly much simpler than assembling a large number of discrete logic chips.

Typically, the program needed to be used by the single-board PIC microcontroller will be generated on a computer and downloaded into the device. The device will need to have its outputs connect to the hardware or circuitry to perform the desired task, but the decision-making calculations are done in easily reprogrammable software.

Many different implementations of PICs can be used; we have just looked at the more typical single-board PIC microcontroller type. It is always possible that the problem at hand will need more computing power. In this case it is necessary to move to more advanced types of microcontrollers and/or field-programmable gate arrays.

5.5 Microcontrollers

Microcontrollers range from the simple to the extremely powerful and are now being used in commercial products everywhere from kitchen appliances through to automobiles. The reason for their popularity is twofold: (1) they have the flexibility to make changes through uploading software and (2) they are designed to easily integrate with other electronic hardware. This makes it possible for the same microcontroller to be used in many applications with only changes to the hardware. The fact that these devices are becoming very affordable is also certainly helping their popularity.

There are numerous microcontrollers on the market today with some that are popular among manufacturers and some with one-off instrument makers and the hobbyist community. One microcontroller that you might have heard of is the Arduino.[7] Arduino is an open-source computer company that creates both hardware and software products. The actual chip being used is an Atmel Corp. processor that is combined with an array of digital and analog pins and that can be accessed in a very simple fashion. The normal means of connecting these to hardware is via expansion boards that plug into the base single-board microcontroller.

The biggest challenge for many instrument builders in using microcontrollers is that you need to learn how to write programs, as microcontrollers run programs. This is not overly difficult, but if you are a beginner to programming, it can take some time to gain proficiency.

Whether you construct electronics from discrete components or move to a microcontroller depends mostly on the designers' preferences and whether the selected electronics can operate fast enough to keep up with the changing signals to make everything work. As a general statement, you can do everything in the microcontroller world that you can do in the analog world, just slower if you stay within the same cost ranges. Where the microcontroller really shines is when you want to read voltages, make decisions, store data, and control some outputs, and you want to do it fairly inexpensively.

Many books and Internet sites are dedicated to microcontrollers and Arduino products, in particular. It is sufficient for our purposes here to show that we can do some simple things with the microcontroller as well as some more complicated tasks.

Using a microcontroller is really straightforward once you know what all of the pins do, or can do. Many microcontrollers have pins to move analog voltages in, or to move digital values either in or out of the device. Once you know how to write or read to a pin, most programs are just collections of these events. The programming language is somewhat specific to the microcontroller but is usually tied very closely to the hardware.[7] A sample program is provided here.

Sample Program

Arduino program to write to an analog pin and read it through a different analog pin.

```
/*
  adcArduino
  SWTeare, NMT, 10-2015
  Read and ADC channel and report number.
*/

// Variable list
int analogPin = 3;
int valPin_3 = 0;

// Setup:
void setup() {
  Serial.begin(9600);
  analogWrite(analogPin,25);
}

// Main runtime loop
void loop() {
  valPin_3 = analogRead(3);

  Serial.println(valPin_3);
}
```

The Arduino program provided above is divided into three main parts: the definitions held in the variable list; the setup, a section of code that is read and executed just once; and a loop that runs the code in that section over and over again. This program lets the output of one pin be read by another. You will notice that there is a command `Serial.println()`. This command sends information to a second computer connected through a USB cable. In this way the data gathered at the Arduino can be logged on the computer, recorded, and analyzed as needed.[7]

The value of the microcontroller in modern electronics cannot be overemphasized. A single microcontroller can greatly reduce the number of external logic components and can be reconfigured with a software change far more easily than replacing hardware. In many ways microcontrollers such as the Basic Stamp, Arduino, and Raspberry Pi have made electronics accessible to a wider range of scientists, engineers, and hobbyists than ever before.[6,7] These low-cost devices are a simple and convenient way to introduce complex control and logic into a system with very little work required for parts assembly.

5.6 Field-Programmable Gate Arrays

Individual logic gates are typically available in IC packages that contain between one and eight gates, depending on the number of inputs and the operation. This makes it reasonably simple to incorporate logic components on a circuit board. However, some electronics require a large number of gates or an ability to reconfigure the logic between the input signals and output signals. In these circumstances, a specific type of IC that can be programmed after it is manufactured can be very valuable.

Field-programmable gate arrays (FPGAs) are semiconductor devices with configurable logic blocks that can be configured and reconfigured as needed. Originally, these devices had relatively few logic blocks available, making them somewhat comparable to discrete logic in terms of their size, and were also slower and used more energy, so their only real value involved being able to restructure logic as needed. The early limitations are now gone, and FPGAs can now provide millions of gates, more than enough for most applications, including the implementation of microcontrollers.

Today, the industry has matured, and FPGAs and microcontrollers are combined into single devices, making them even easier for electronics designers. These devices are advancing at a remarkable rate, and their reconfigurability makes fixing errors in either software or their hardware much faster and easier to implement.

An FPGA can be thought of as a collection of possibilities. Once programmed, those possibilities become specific logic operations. Field programmable gate arrays are programmed using either an electronic schematic or a hardware definition language that defines its operation. While it may seem like an advantage to be able to use a graphical system such as a schematic to program the device, it can be more efficient to program the device by implementing the code line by line. When programing, shortcuts such as being able to define connection and truth tables make the information entry easier and can speed up troubleshooting. Manufactures and suppliers of FPGAs have developed very sophisticated software packages for programming that make it easier than ever for engineers to program, analyze, and edit the device operations.

For many projects, it might seem as though FPGAs are an unneeded complexity for developing electronics. While there is some learning and upfront setup and cost, the learning curve for developing electronics in this fashion is quite manageable. Many practicing electrical engineers and electronics developers prefer these systems and will use them for a wide range of projects.

5.7 Practice Problems

1. Convert the following numbers to binary and hexidecimal: 127, 128, 255.
2. Implement the function for an ‘Exclusive Or’ (XOR), which for two binary numbers A and B is given by ‘B OR A’ AND NOT ‘A AND B.’ Use the function to calculate the XOR of ‘1010’ with ‘1100.’
3. Determine the output from a three-bit ripple counter.

Answers

1. 1111111, 7F; 10000000, 80; 11111111, FF
2. (A|B) & ∼(A&B)
3. 000; 001; 010; 011; 100; 101; 110; 111

References

1. T. L. Floyd, *Digital Fundamentals*, 10th Edition, Prentice Hall, Upper Saddle River, New Jersey (2008).
2. J. F. Wakerly, *Digital Design, Principles and Practices*, Fourth Edition, Prentice Hall, Upper Saddle River, New Jersey (2005).
3. P. Horowitz and W. Hill, *The Art of Electronics*, Third Edition, Cambridge University Press, Cambridge (2015).
4. A. S. Sedra and K. C. Smith, *Microelectronic Circuits*, Oxford University Press, Oxford (2015).
5. H. Moore, *MATLAB® for Engineers*, Fourth Edition, Prentice-Hall, Upper Saddle River, New Jersey (2014).
6. A. Lindsay, *What’s a Microntroller?* Student Guide, Version 2.2, Parallax, Inc., Rocklin, California (2004).
7. T. Karvinen, K. Karvinen, and V. Valtokari, *Make: Sensors: A Hands-On Primer for Monitoring the Real World with Arduino and Raspberry Pi*, Maker Media Inc., Sebastopol, California (2015).

Chapter 6

Instrumentation, Signal Conditioning, and Filters

Several chapters back, we began with the basic components used in electronics and then moved to exploring analog and digital electronics, and processing electrical signals. We can now begin to convert this understanding of basic electronics into an ability to design and plan more complicated instruments. Even the most complicated instruments are made up of simpler instruments and electronic circuits, so this will serve to bring more detail to our discussion of electronic instrumentation.

The instruments we will look at in this and future chapters are generally involved in performing sensing tasks and then processing that signal into the data we are interested in analyzing. In many scenarios, the input signal will be noisy or difficult to separate from other signals, meaning that a conditioning task will be needed. In optical engineering, we can make use of optical filters or gratings to separate specific light channels before the light reaches the detector. Once the light has been converted by the detector into an electrical signal, other techniques will be needed to improve the signal quality. Since our expertise is in working with the optical signals, we will focus here on the electrical side of the signals. While it is common to think of filters as being analog devices, and we will start there, the advent of digital signal processing has made it possible to use a wide range of mathematical functions available for use on signals.

In this chapter we will concentrate on electronic signal conditioning and filtering, and relating this to generic instrument design. The filters will include both analog and digital circuits, and some of the underlying concepts will be introduced as well. There is a great deal of literature on electronic filter systems, and it is common for electrical engineering programs to devote several courses to the topic.

6.1 Instrumentation

There are many ways to design and construct an instrument, depending on whether the instrument is being constructed as a one-off system for use in a controlled laboratory, or for large-scale production. Whether the instrument will be used in the field can also change how the internals of an instrument are organized. For this discussion, we will consider the signal-detection side of the instrument, as shown in Fig. 6.1, and ignore the complications of feedback controllers or actuators. Should these be required, it is usually not difficult to envision where the extra pieces would need to sit.

The division between the analog and digital sides of the signal-processing chain can become very blurred in an actual instrument, but this division is quite useful for instructional purposes. Every stage of the block diagram provides some level of signal conditioning, but we will go through the purpose of each stage one by one.

Signal of Interest: This is the information that we are interested in. Some examples for this could be an optical communication signal, an astronomical source such as a star, or so forth.

Signal Conditioning: This could be a telescope, optical filter, lens system, or fiber optic used to gather some fraction of the signal of interest. There can be many optical components in the chain, including polarizers, gratings optoelectronics, and optomechanical devices.

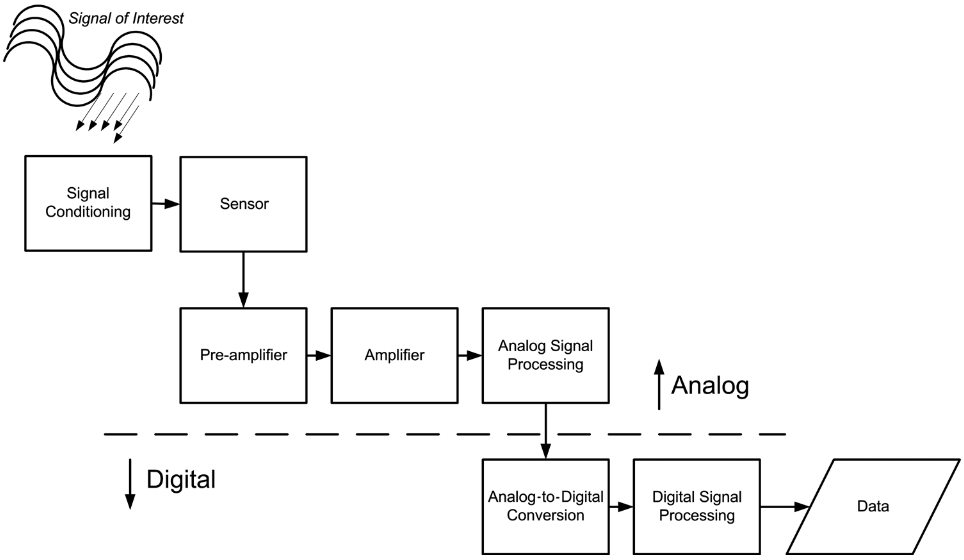


Figure 6.1 Block diagram of an instrument showing the division between the analog and digital side, illustrating the signal-processing chain from incoming signal to final data.

Sensor: The most basic optical sensors include photodiodes and cameras. The sensor can be a very intricate system, but by the end of the chain, it has become a transducer that converts optical energy to electrical energy.

Pre-amplifier: The sensor's electrical output might not be very high or might not have the ability to generate the current needed for full amplification, so an intermediate amplifier, potentially a transistor or operational amplifier, is needed to ensure that the signal is not degraded.

Amplifier: This is the more classic amplifier and will most often be a single op-amp or a combination of op-amps.

Analog signal processing: This is a stage where the analog signal is analyzed and unwanted information can be removed to isolate the signal of interest. This is commonly done with analog filters.

Analog-to-digital conversion: The signal is discretized, and often considerable information is thrown away.

Digital signal processing: This is a stage in which the digital signal is analyzed and unwanted information can be removed to isolate the signal of interest; alternatively, mathematical operations can be performed, which can include digital filters.

Data: The information that is desired has been distilled and is available for storage on a computer or other digital recording device.

Many of the components identified in Fig. 6.1 have been introduced in earlier chapters. In this chapter we will look at some of the key elements of analog signal processing, both passive (using inductors, capacitors, and resistors) and active, involving the use of operational amplifiers. This will then be leveraged to introduce some of the elements of digital filters.

6.2 Passive Analog Filters

When we discuss filters, we are talking about an electronic device that removes some portion of a signal, often related to its frequency components. We begin with two simple filter systems,[1–4] the high-pass filter and low-pass filter. The names of these filters are good indicators of what they do: a high-pass filter allows high frequencies through, and a low-pass filter passes low frequencies, while blocking higher frequencies. Circuits for each are shown in Fig. 6.2.

We will look at these two circuits first conceptually and then throw in some mathematics to see what they do. Conceptually, we invoke a very simple rule for capacitors; they pass higher-frequency signals and block lower-frequency signals. In the top circuit, the capacitor will block low frequency, such as a DC signal, from getting through to the resistor, while the high frequency can get through. The second circuit is the opposite: AC signals will tend to be shorted to the other side of the input, while the DC signals will not. Very simple conceptually, but the complete story is not really being told yet. For that, we need to look at the transfer functions of the two circuits.

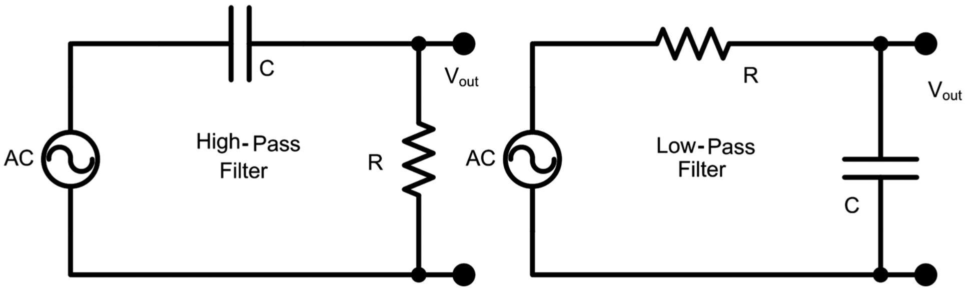


Figure 6.2 High-pass and low-pass filters connected to an alternating input signal.

If we stare at the circuits in Fig. 6.2 for a bit, it should become clear that they look very similar to the voltage divider circuits we looked at in an earlier chapter. But rather than two resistors, we have a resistor and a capacitor. If we call resistors, capacitors, and inductors by their generic name of impedances Z, we can write the voltage divider as

$$H(\omega) = \frac{V_{out}}{V_{in}} = \frac{Z_1}{Z_1 + Z_2}, \tag{6.1}$$

where ω is the angular frequency.

In the case of a resistor, $Z = R$. Capacitors have a frequency dependence so are a little more complicated, and $Z = 1/j\omega C$, where j is $(-1)^{1/2}$. This leads to the two equations

$$H_{HP}(\omega) = \frac{R}{R + \frac{1}{j\omega C}} = \frac{j\omega RC}{j\omega RC + 1}, \tag{6.2}$$

$$H_{LP}(\omega) = \frac{\frac{1}{j\omega C}}{R + \frac{1}{j\omega C}} = \frac{1}{j\omega RC + 1}. \tag{6.3}$$

In general, we are interested in the magnitudes for these values, and the magnitude of an imaginary number is the square root of the sum of the squares of the real and imaginary parts. This reduces Eqs. (6.2) and (6.3) to

$$\begin{aligned} |H_{LP}(\omega)| &= \frac{1}{\sqrt{(\omega RC)^2 + 1}}, \\ |H_{HP}(\omega)| &= \frac{\omega RC}{\sqrt{(\omega RC)^2 + 1}}. \end{aligned} \tag{6.4}$$

These equations can be best appreciated in the format of a Bode plot, as shown in Fig. 6.3.

The Bode plots shown in Fig. 6.3 clearly show the reason the filters are named as they are; the top plot shows that the low frequencies are allowed to pass until the cutoff frequency is reached, then the output starts to degrade, so this is a low-pass filter. The bottom plot shows the high-pass filter, which attenuates the

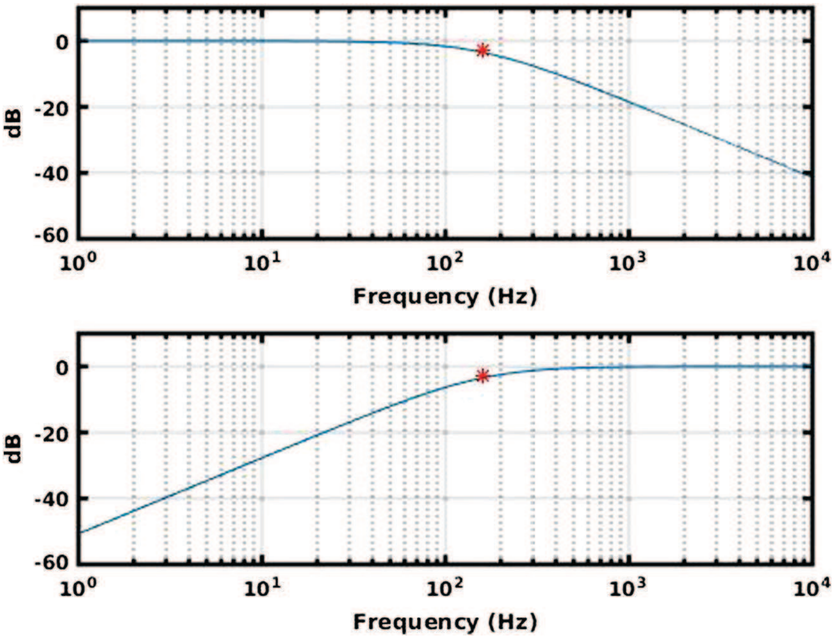


Figure 6.3 Bode plots comparing the transfer functions of the low-pass filter (top) and high-pass filter (bottom). In each case the same capacitor and resistor values are used, so the cutoff frequency, shown as a vertical solid line, is the same in both plots.

low frequencies but allows the higher frequencies through. Example 6.1 shows the calculator used to generate the Bode plots in this section. These plots are also a little more complicated than the ones we have been using, so the example provides an introduction to some useful MATLAB® commands.[5]

Example 6.1

Make a MATLAB script that will calculate the Bode plot for a first-order high- and low-pass filter using a 1-μF and 1-kΩ resistor. Show the location of the cutoff frequency.

```
% Example_6_1.m
% SWT 10-21-15
% Bode plots of high pass and low pass passive filters

%% Housekeeping
clear all; clc;

%% Input values
C = 1e-6; % capacitance in Farads
R = 1000; % resistance in Ohms
maxFreq = 1e4; % maximum frequency to scan in Hertz
```

```
%% Conversions
f = 1:maxFreq; % creates a vector of frequency values in 1 Hz
steps
omega = 2*pi*f; % calculates the angular frequency

%% Calculates the transfer function for the high and low
% pass configurations
H_lp = 1./sqrt (((omega*R*C).^2 +1).*((omega*R*C).^2 +1));
H_hp = (omega*R*C).^2./sqrt(((omega*R*C).^2   ...   +1).*
((omega*R*C).^2 +1));

%% Converts transfer functions to dB
dB_lp = 10*log(H_lp);
dB_hp = 10*log(H_hp);

cutOff= 1/((2*pi*R*C).*(2*pi*R*C));
disp(cutOff);

%% Displays Bode Plot curves
clf;
figure (1);
subplot(2,1,1);semilogx(f,dB_lp);hold on;
line([cutOff,cutOff], [-60,10],‘Color’,‘r’,‘LineWidth’,2);
axis([1 maxFreq -60 10])
grid on;
xlabel(‘Frequency (Hz)’)
ylabel(‘dB’)
title(‘Low Pass Filter’);
set(gca,‘FontSize’,9,‘FontWeight’,‘Bold’,‘LineWidth’,2;)
hold off;

subplot(2,1,2);semilogx(f,dB_hp);hold on;
line([cutOff,cutOff], [-60,10],‘Color’,‘r’,‘LineWidth’,2);
axis([1 maxFreq -60 10])
grid on;
xlabel(‘Frequency (Hz)’)
ylabel(‘dB’)
title(‘High Pass Filter’);
set(gca,‘FontSize’,9,‘FontWeight’,‘Bold’,‘LineWidth’,2);
hold off;
```

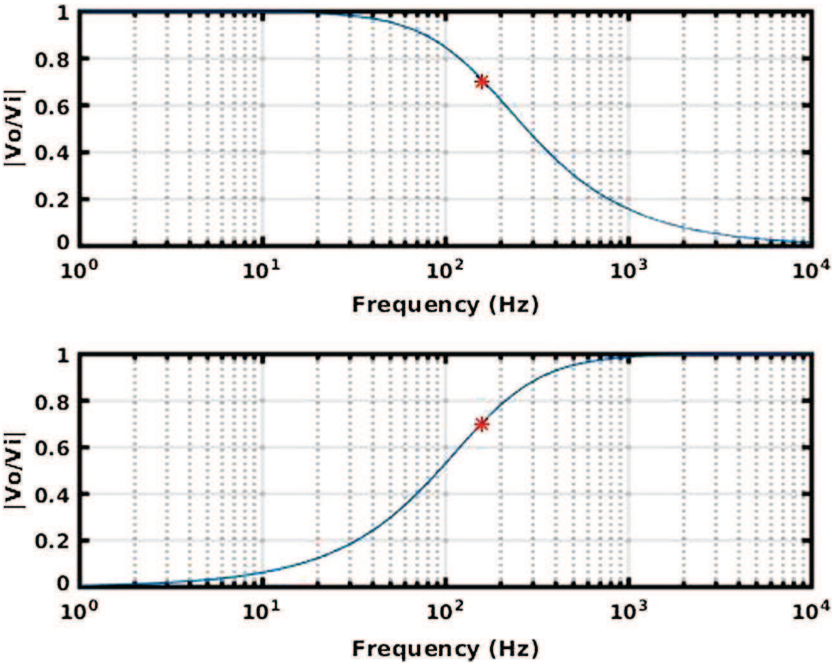


Figure 6.4 Linear plots comparing the transfer functions of the low-pass filter (top) and the high-pass filter (bottom). In each case the same capacitor and resistor values are used, so the cutoff frequency is the same in both plots.

The best way to understand the new commands shown in the above example is to use the built-in MATLAB help and the MATLAB website. It takes a little bit of thought to figure out the information using a Bode plot, so Fig. 6.4 shows how the filter outputs look in linear form.

The linear plots show that the simple first-order filters we are looking at have a very soft turn-on or roll-off with frequency. In some cases it would be desirable to have these be steeper. This can be accomplished by adding an additional stage to the filters, such as the second-order filters shown in Fig. 6.5.

The cutoff frequency of the filter is defined as the 3-dB point of the filter, i.e., the point where output drops by 3 dB from its maximum. This point will occur when the frequency is $1/2\pi RC$ for a first-order filter.

The corresponding Bode plots for these filters are shown in Fig. 6.6. Notice that the effect of having two identical stages is to improve the rate at which the filter cuts off or turns on.[4] The introduction of the second stages changes the transfer functions from those shown in Eq. (6.4) to

$$
\begin{aligned}
|H_{LP}(\omega)| &= \frac{1}{\sqrt{[(\omega RC)^2+1][(\omega RC)^2+1]}},\\
|H_{HP}(\omega)| &= \frac{(\omega RC)^2}{\sqrt{[(\omega RC)^2+1][(\omega RC)^2+1]}}.
\end{aligned}
\tag{6.5}
$$

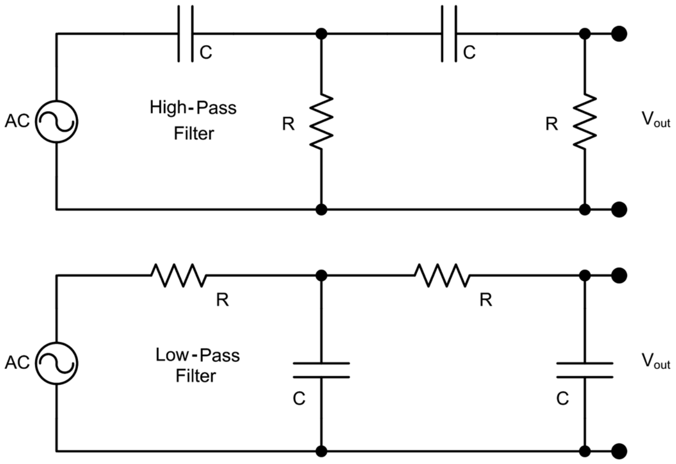


Figure 6.5 Circuits showing the second-order high- and low-pass filters.

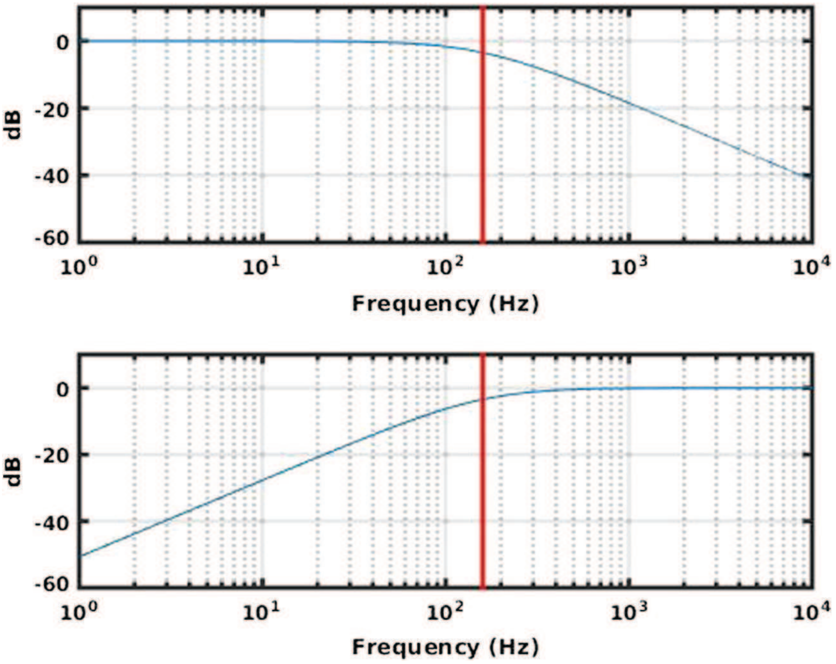


Figure 6.6 Circuits showing the second-order high- and low-pass filters.

The process of adding more resistors and capacitors can continue, but it is often more effective to move from working with passive components to active components. This is another way of saying that we return to the use of operational amplifiers.

6.3 Active Analog Filters

The idea of an active filter is something that we introduced a few chapters back when we discussed the use of integrator and derivative implementations of operational amplifiers. The integrator can actually be considered as a low-pass filter and the derivative as a high-pass filter!

We can move forward with this knowledge to discuss a filter that combines the low-pass and high-pass characteristics into a bandpass filter, and we will use this filter to develop our active analog filters. In many ways, a bandpass filter is a combination of a low-pass and high-pass filter such that a region between the two is common and allows the signal to pass. We will ignore the equations that describe the following two filters and show the two related circuits. A simple, active bandpass filter is shown in Fig. 6.7, which looks much like a combination of the integrator and differentiator circuits.[1,3,4,6]

While this circuit will result in the generation of a band pass, it is a very broad band pass. Just like when we were working with the passive high- and low-pass filters, a filter with steeper slopes is sometimes more desirable. Figure 6.8 shows a reconfiguration of the bandpass filter that uses only one additional component but improves the filter quality. This filter is often

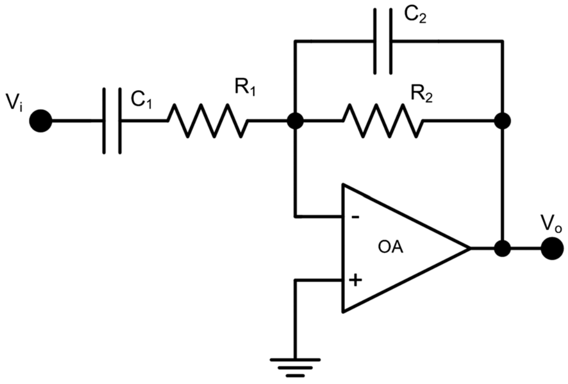


Figure 6.7 Circuit showing a bandpass filter using an operational amplifier.

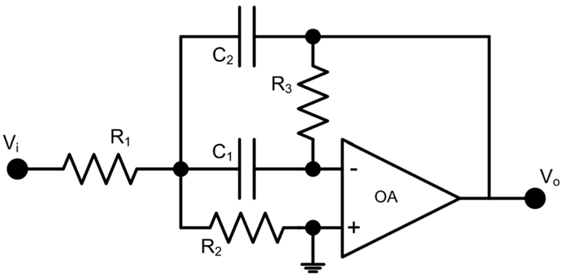


Figure 6.8 Circuit for a narrow-bandpass filter using an operational amplifier.

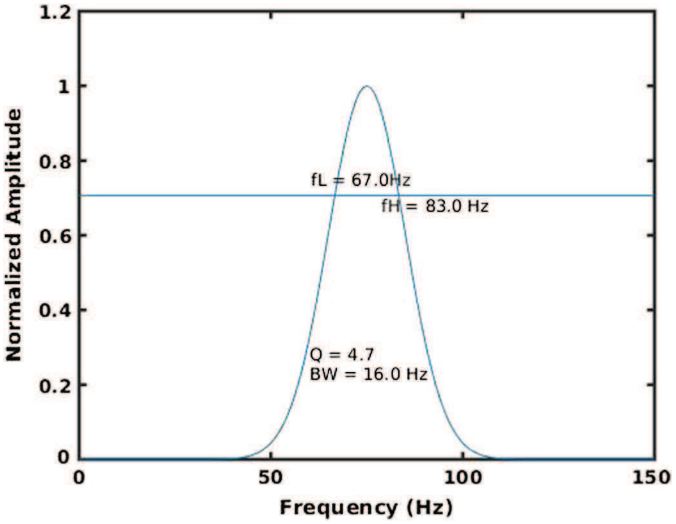


Figure 6.9 Transfer functions for a bandpass filter showing the lower and upper frequency limits to the bandpass, the bandwidth, and the quality factor value *Q*.

referred to as a multiple-feedback topology filter. The property of a filter that describes the width of the band pass is referred to as the quality factor or Q, which is defined as

$$Q = \frac{f_0}{b}, \tag{6.6}$$

where b is the bandwidth, which is the full width of the band pass at 70.7% of the maximum, also known as the -3 dB point; f_0 is the center frequency.

A comparison of the band passes of two simulated signals is shown in Fig. 6.9. In a bandpass, the bandwidth can be very much smaller than the center frequency of the filter.

6.4 Analog-to-Digital Conversion

Most often, the signals we interact with in the world around us represent some physical quantity. These signals can vary over time, distance, or other parameters, and often we are interested in knowing the change in the amplitude of the signal. The conversion of the continuous signal to a single number or an array of numbers requires that the analog signal be measured and restated as a number. The device used for this conversion process is the analog-to-digital converter (ADC).[2,4] The role of the ADC is to take a physical signal that has been converted to a voltage and determine how to best represent it as a number that a computer can use, a digital number.

The conversion process of the ADC is not perfect; it introduces a quantity of error into the measurement. This error is due to the limitation on the

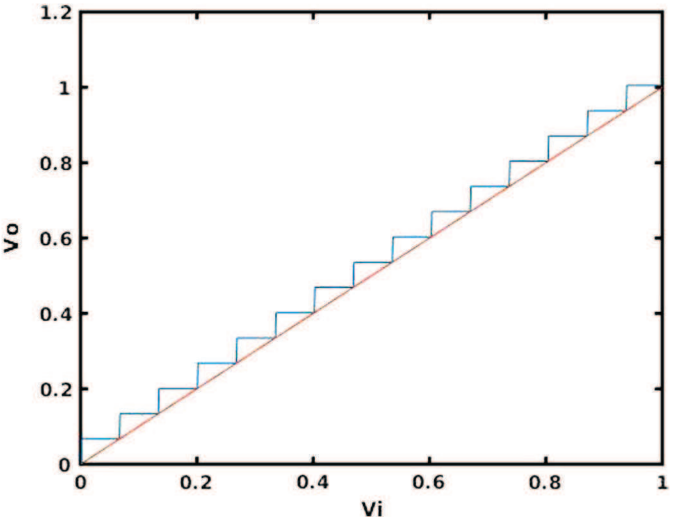


Figure 6.10 The stair-step function associated with quantization effects in ADCs. The steps represent the digitization of the continuum of the points.

number of possible values that can be obtained in the digitization process compared to the continuum of values possible. This can be made clear in an example. Consider a signal that ranges from 0 to 1 V and can take every possible value between 0 and 1. As we learned earlier, digital numbers are represented by bits, and a 4-bit number can include up to 16 numbers. If we were to consider our 1-V signal, we can break it up into 1/15 individual steps or step sizes of 0.067 V. This process of converting the continuum of possibilities into a discrete number is referred to as quantization and results in some error in the conversion process. ADC resolution is described in terms of the number of bits used in the conversion. The more bits the smaller the individual steps and the smaller the error in the digitized number compared to the actual value. This is illustrated in Fig. 6.10, which shows the conversion steps versus the continuum values.

ADCs are available as individual ICs, as parts of microcontrollers, or as specific pieces of hardware. The basic functionality has been described here, but there is significantly more involved in digitizing signals.

6.5 Digital Filters and Signal Processing

The filters that we have been discussing can also be implemented in a dedicated processor known as a digital signal processor (DSP). These devices can be used to sample continuous data, convert them into digital form, apply mathematical relationships to them, and output a result. These processors are specially designed for this purpose and can perform these tasks very fast. Many DSPs can take a signal-processing program from languages such as

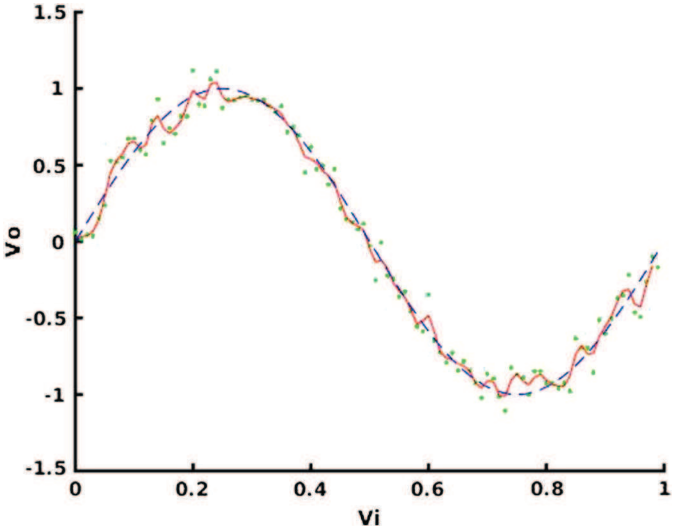


Figure 6.11 A noisy signal (shown as dots) and its postprocessed filtered form (smooth line) compared to the true signal (dashed line).

MATLAB and import them directly and use them. This greatly simplifies their use but does require some matching of software and particular boards.

While a detailed look at DSPs will require looking into another more specific book on the subject, we can demonstrate the use of digital filters on some noisy digital data to show what they can do. The simplest filter we can make is a smoothing or averaging filter as described by

$$F_i = \frac{y_{i-1} + 2y_i + y_{i+1}}{4}. \tag{6.7}$$

To understand Eq. (6.7), it is simplest to consider a dataset of numbers y. Each of the individual values in the set are numbered $y_1, y_2, \ldots, y_N$. As long as we are far enough into the dataset, we can generate the filtered data set F. By "far enough" it is meant that we have to have data for all three elements. If we start at the index 2, then $y_{2-1} = y_1$. So when we filter data this way, we have to throw out the first and the last points. The power of this filter to smooth out a noisy signal is shown in Fig. 6.11.

The result of this filter is not a perfect recovery of the original signal, but the effect of the noise is reduced. A wide array of filter types and combinations of weighting factors is available, some of which will work better than others for particular data sets.

6.6 Noise and Analysis

If you are working with signals, at some point you will end up working with noisy signals. Much of the work on this chapter has dealt with filters whose

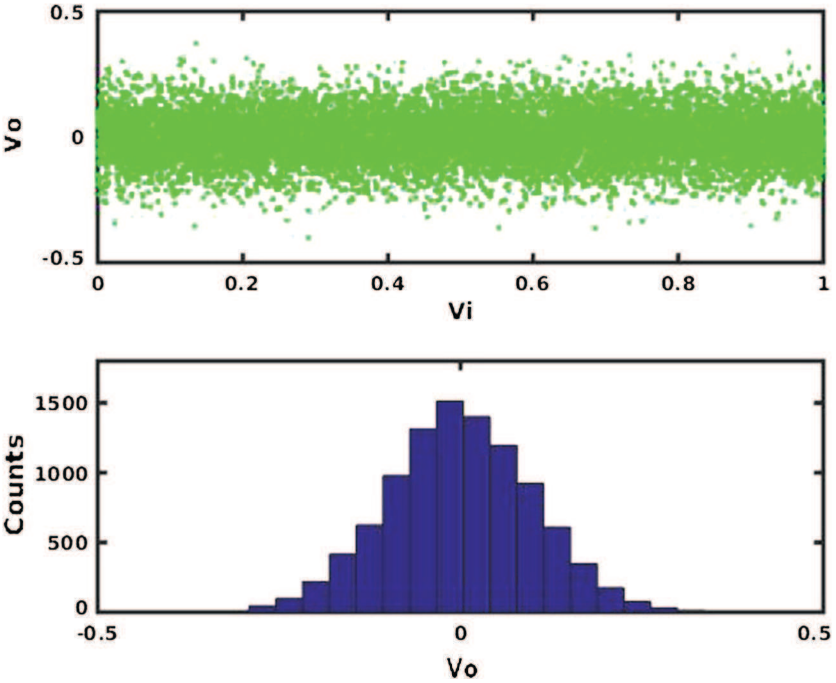


Figure 6.12 A DC signal with a random noise component superimposed (top) and the noise represented by a histogram (bottom).

purpose is to condition the signals for use. This is the point where we will talk a little bit about noise, its characteristics, and how things should look.

The typical noise source will contribute Gaussian noise to your signal. Gaussian or normally distributed noise is random noise, and we can reduce its contribution to our signal in several ways. The easiest noise signals to work with are additive noise; i.e., the signal we want has had a noise value added or superimposed onto it. This is illustrated in Fig. 6.12, which shows a DC value with a random noise signal on top of it. The original signal is just the DC value of zero, so the width of the signal is strictly due to the added noise.

A line drawn as an envelope to the histogram would take the shape of a normal distribution or a Gaussian shape. This is described as

$$I(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}. \tag{6.8}$$

The value σ describes the width of the distribution, and μ its center or average. This is the same equation as was used to generate Fig. 6.9.

Removal of the noise component from a signal is a study on its own, and there are many techniques for doing this. We will explore a couple of techniques later, but one of the simplest is to use a low-order fitting function to determine the trends in the data. Subtracting out the low-order fit from the data, if successful, will leave the data centered around zero, allowing the noise component to be analyzed, and the fit will now represent the true signal.

6.7 Practice Problems

1. What is the cutoff frequency for a first-order, low-pass filter with a 1-nF capacitor and a 1-MΩ resistor? What is the cutoff frequency if the filter is implemented as a first-order high pass?
2. A bandpass filter has a bandwidth of 100 Hz at 3 kHz. What are the low- and high-frequency limits at the 3-dB point and the quality factor of the filter?
3. What is the step size in volts of a 10-bit ADC with a maximum input voltage of 10 V?
4. Implement Eq. (6.8) in MATLAB and determine the width of the 3-dB point for $\sigma = 10$ and $\mu = 100$.

Answers

1. 159.2
2. 2950; 3050
3. 9.8 mV
4. 16 Hz

References

1. B. Carter and T. R. Brown, "Handbook of Operational Amplifier Applications," Application Report SBOA0921, Texas Instruments, Inc., Dallas, Texas (2001).
2. P. Horowitz and W. Hill, *The Art of Electronics*, Third Edition, Cambridge University Press, Cambridge (2015).
3. T. Kugelstadt, "Active Filter Design Techniques," Chapter 16 in *Op Amps for Everyone*, R. Mancini, Editor, Texas Instruments, Inc., Dallas, Texas, pp. 261–323 (2002).
4. A. S. Sedra and K. C. Smith, *Microelectronic Circuits*, Oxford University Press, Oxford (2015).
5. H. Moore, *MATLAB® for Engineers*, Fourth Edition, Prentice-Hall, Upper Saddle River, New Jersey (2014).
6. "Op Amp Circuit Collection," National Semiconductor Application Note AN-31, National Semiconductor Corp., Santa Clara, California (2002).

Part II: Optics Applications

Chapters 7 through 13 focus on electronic circuits related to specific optics applications. The electronics associated with photodiodes, quad-cells, solar cells, controllers, motors, and more are introduced.

Safety Warning
Always wear safety glasses when working on or around electronic circuits and follow all equipment and component manufacturers' safety instructions. Circuits included in these chapters may not be suitable for a particular application.

Chapter 7

Solar Cells and Rechargeable Batteries

Electronics systems rely on power, and for most portable systems this will be provided by batteries, and quite often these will be rechargeable batteries. We can take portable power systems a step farther by using solar cells to provide the operating power or to recharge the batteries. Solar energy conversion is not all that efficient and can be affected by the sun's position in the sky, shadows, and clouds; however, there is often plenty of light energy available for small solar cells used in charging small batteries, storage capacitors, or driving low-power systems.

Recharging batteries is a complicated subject, and each manufacturer and type of battery can require different charging profiles, i.e., how the voltage and current are balanced over time.[1] Most manufacturers provide very detailed information about how to optimally charge their batteries, and electronics manufacturers have responded with specific ICs that have been developed to provide the correct charging profiles.[2,3] The variations in the needs of different types of batteries make it important that you check with the manufacturers of the specific components to ensure that the batteries are not damaged when recharged. This becomes particularly important when fast charging batteries.[1]

In this chapter we will be concentrating on the generation of electrical power using solar cells, and using this to charge batteries and provide power to other electronics devices and instruments.

7.1 Solar Cells

Photovoltaic cells or solar cells[4] are devices that convert light energy into electrical energy. This process is known as the photovoltaic effect, and for our purposes we are considering devices that can change their current, voltage, or resistance when exposed to a source of sufficient light energy and wavelengths.

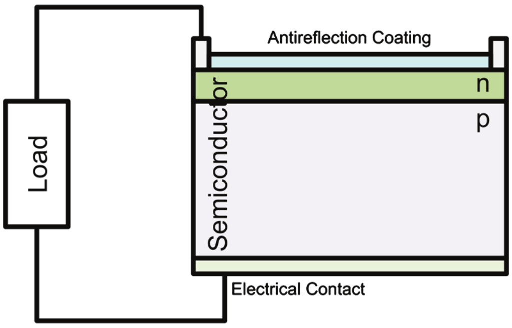


Figure 7.1 A simple description of a solar cell showing electrical and optical features. Light enters through the top antireflection coating.

When a sufficient number of solar cells are connected, a significant amount of power can be obtained, provided there is a light source.

Solar cells are not really new; the first solid state photovoltaic cells were developed in the 1880s using selenium,[4] but practical solar cells took until the mid-1950s and their eventual use on space satellites, first demonstrated on the Vanguard satellite, solidified their usefulness.[5]

Modern solar cells are based on the material silicon, just like most of our solid state electronics devices. The solar cells work by the semiconductor absorbing photons from a light source, liberating electrons that can travel through the cell to reach the contacts. The flow of electrons is a current that can be used as a source of power. This is a somewhat simple discussion of how solar cells work and, as anyone working in the semiconductor area will note, has totally ignored semiconductor hole generation and other processes.

Solar cells are somewhat more complicated in reality than we are describing here, particularly from the perspective of how light enters. There is the need for electrical contacts, protection from the elements, and reducing the effects of index of refraction, which will cause significant levels of reflection and thus light loss. A simple solar cell is shown in Fig. 7.1.

Now that we have the basic idea of what a solar cell is, we will explore how to charge batteries, describe some applications of interest, and use these applications to understand more about the operation and use of solar cells.

7.2 Batteries and Charging

It doesn't take long to come to the realization that charging batteries is not as simple as one might think.[6] Batteries and how to get charge into them is fairly complicated, and the more complex the battery the more complex the mechanics of charging. This is true for 12-V car batteries as well as for rechargeable small-voltage batteries. Since we are looking at this from the perspective of a battery charger using solar cells, we will stick with lower voltage batteries at the start.

Let's start with some basics of a typical 1.5-V alkaline, non-rechargeable battery. First, it's not really 1.5 V except when new, so we won't worry too much about the actual voltage value during its operational life other than to say it is somewhere between about 1 and 1.5 V. As stated earlier, these types of batteries are nonrechargeable, so don't try to recharge them. Just don't. Rechargeable batteries of this type are usually rated at 1.2 V and are typically nickel metal hydride (NiMH). While it may sound like rechargeable batteries are less powerful than the alkaline, this really isn't so. What makes the NiMH batteries rechargeable and the alkaline batteries not rechargeable requires a detailed discussion of chemistry that we are not going to do. Suffice it to say that there is a reversible chemical process in NiMH batteries that allows you to take charge out or move charge in and store it.

The charging or discharging rate for batteries is referred to as the C-rate. Think of C as being capacity, and the C-rating as the maximum discharge capacity for the battery. So a 1000-mAh-rated battery that has a 1-C rating can be safely discharged at 1 A of continuous discharge, i.e., without doing damage to your battery. If you needed a 2-A discharge rate, you would want a 2-C battery. Typical AA NiMH batteries have a capacity of 1500 mAh and can discharge at about 1 C.[3] To charge them and avoid overheating, we would use a rate of about 0.1 C. This means that it would take more than 10 hr to slow charge, or trickle charge, these batteries.

Charging batteries quickly and without damaging them does require an understanding of how batteries take a charge.[1] The voltage or temperature as a function of time provides a signature of the charging process and can vary for different batteries. This charging signature can be used to define an optimal way to charge the battery without overcharging or causing damage, and different battery compositions have different signatures. The simplest way to avoid damaging a battery is to charge it slowly and avoid the problem of fast charging altogether. This will be the first approach that we will consider. These chargers are known as trickle chargers, a simple system of which is shown in Fig. 7.2.

There is not too much to this charging system. An AC signal from a transformer that plugs to a main supply is passed through a full-wave bridge rectifier or diode bridge.[7,8] The capacitor C provides some filtering, the resistor serves as a current limiter, and the battery can be charged at a low rate. A system such as this can be used to provide a maintenance charge to a battery to keep it at its maximum level and can improve battery life in some types of batteries. An improved version includes a voltage regulator.[6]

Our interest is in using a solar panel to drive our *charger*[2] not our household power supply. Solar power is somewhat unreliable, and at low charging rates, it can take 10 or more hours to charge a battery. Using solar power as a charger will require that we charge in a short time period to take advantage of the time when the sun is actually shining. This will introduce a couple problems for us to consider. First is how to optimize the charger's operation, and second is how to

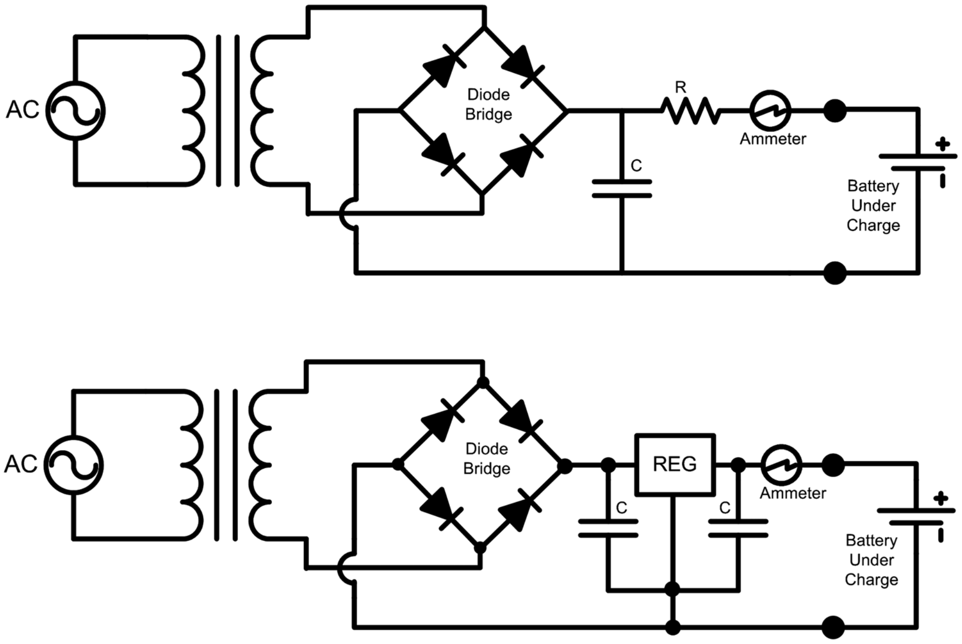


Figure 7.2 A simple charging system (top) that will provide a low-current charge rate depending on the size of the resistor R selected. An improved design (bottom) includes a voltage regulator (REG).

control the charging rate so that it does not damage the battery. The latter varies depending on the physical and chemical makeup of the battery.[1]

7.3 Solar Cells and Panels

Solar panels are comprised of several solar cells mounted together as a single unit that is often weather proofed to protect the delicate surfaces. The individual solar cells or photovoltaic cells are low-voltage devices that are often connected in series to provide a higher voltage or in parallel to provide higher currents.

The power provided from the solar cell relates directly to the amount of light that can be converted into electricity. Light losses can occur from shadows, reflectance from the surface of the cell, solar angle, the wavelengths of available light, and the season. This translates directly into the current-versus-voltage characteristic of the cell.[3,4] An ideal solar cell acts as a current source in parallel with a diode feeding a load with a current–voltage profile, as shown in Fig. 7.3.

The current–voltage profile for the ideal solar cell is described as

$$I = I_S\left(e^{\frac{qV}{kT}} - 1\right) - I_L, \tag{7.1}$$

where I_S is the diode saturation current, kT/q, the thermal voltage is 25.9 mV at room temperature, and I_L is the current through the load. This equation is

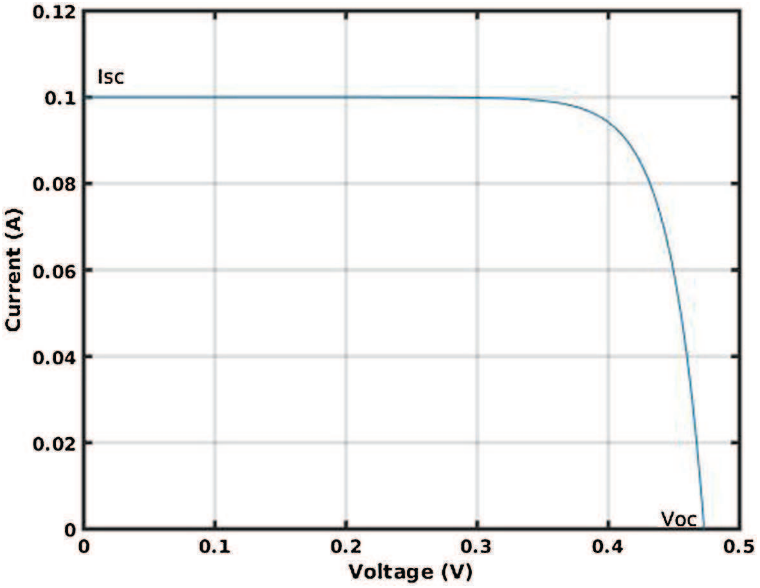


Figure 7.3 The current–voltage profile for an ideal solar cell. I_{SC} is the short-circuit current, and V_{OC} is the open-circuit voltage.

plotted in Fig. 7.3 for a load current of 100 mA and a diode saturation current of 1 nA. Two points on the curve need to be identified: V_{OC}, which is the open-circuit voltage, representing the voltage when there is no load attached, and I_{SC} the short-circuit current, where the output of the solar cell is connected together. The plot is shown as an inversion around the voltage axis, the typical representation of the information. The selection of an appropriate load will allow about 80% of the power given as the product of I_{SC} and V_{OC} to be available for use. Usually this is described by a power rectangle that can be inscribed within the curve that defines the maximum voltage and current for the cell.

The efficiency of the solar cell is defined in the equation below, where FF is given as the ratio of the areas of the $I_m V_m$ versus $I_{SC} V_{OC}$:

$$Efficiency = \frac{FF \times I_L V_{OC}}{P_{in}}, \tag{7.2}$$

where P_{in} is the power into the solar cell.

Of course this isn't the whole story; we have ignored many of the nonideal and loss characteristics, but this gives us a pretty good idea of how solar cells work. Other effects include the series resistance in the front surface of the cell. The final piece of the operation of the solar cell is the need for there to be a light source. Equation 7.1 is implemented in plotting the current–voltage curve of the solar cell in Example 7.1.

Example 7.1

Plot the current–voltage profile for a solar cell.

```
%% Example_7_1.m
% SWT 10-21-15

% Solar cell IV characteristic, ideal solar cell

%% Housekeeping
clear all; clc;

%% Parameters
Temp_C = 25;      % Operating temperature
IL = 100e-3;      % Load current
Is = 1e-9;        % Saturation current
n = 1;            % Ideality
k = 8.617e-5;     % Boltzman's constant
%% Resistances
Rs = 0;           % series resistor value in ohms
Rsh = 5;          % shunt resistor value in ohms

%% Coversions
T = 273+Temp_C;   % convert to Kelvin
kT_q = k*T;       %

%% Calculate current for a voltage input
V=linspace(-1,.5,1000);
I = Is*(exp(V./n/kT_q)-1) -IL;

%% Plotting
figure(1)
plot(V,-I)
axis([-0 .5 0 .12])
grid on;
xlabel('Voltage (V)');
ylabel('Current (A)');
text(.44,0.0025,'Voc');
text(.01,0.105,'Isc');
```

See Fig. 7.3.

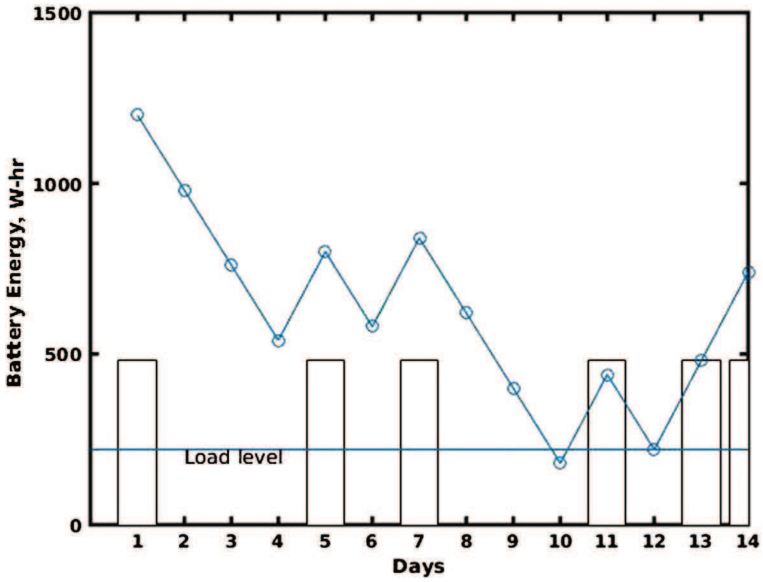


Figure E7.1 The battery voltage is depicted as the solid line with circles, and the recharging cycle is depicted as a bar chart, showing the effect of cloudy days on the battery level.

7.4 Solar-Panel-based Battery Charger

In Section 7.2 we constructed a trickle charger for low-current or maintenance charging of a battery. The same can be done using a solar panel. Even though we have solar panels that produce DC voltages, we will include a diode in our circuit shown in Fig. 7.4. The purpose of the diode is to prevent the power in the

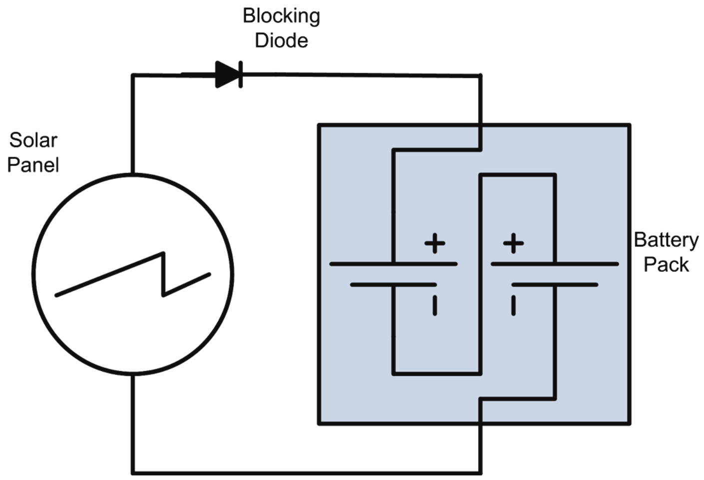


Figure 7.4 A simple two-cell battery charger using a solar panel and a blocking diode.

battery from flowing back into the solar cell when the sun is no longer shining. Without the diode, the newly charged batteries would discharge through the panel, undoing all the work of charging them. A blocking diode[7,8] is really nothing special; a regular 1N4001 diode will work just fine as long as it works within the power range you are generating. In the circuit in Fig. 4.7 we are using a solar panel that produces a fairly low amount of current at a voltage above the battery pack voltage so that we are well below the rated current of the batteries.

This particular charger would take many hours to charge even small batteries, as even in full sunlight, which has a rating of 1000 W/m^2, the small size of the solar panel, say 5 cm per side, would be illuminated by only 25 W of sunlight. Typically, there is going to be less than the maximum amount of sunlight available, maybe only 80%, so that only 20 W are available for use. With a 10% efficiency solar panel, this will give only 2 W of power for use. If we need 4 V to exceed the charging voltage of the battery, only 500 mA of current is available for charging the battery pack at the maximum. If we consider that each day there might be only 4 hr of high-quality charging time, a battery pack with 5000-mA-hr capacity will take nearly 3 days to recharge.

7.5 Fast Chargers

Fast and ultrafast battery chargers are in demand for everything from cell phones to electric vehicles.[1] After all, who wants to wait hours to "fill up" your vehicle! There is no real problem in obtaining the large amounts of current to charge batteries; the issue is that providing this current to the batteries in the wrong fashion can destroy them or limit their lifetime.[1] Battery charging and the rates at which batteries can be charged really depend on the type of battery being charged. The most common ways to damage batteries is by overcharging, which can occur by leaving batteries in unsophisticated chargers for too long. These unsophisticated chargers have limited monitoring capabilities, so they don't adjust their charging to match the battery state.

There are some basic rules to follow when charging batteries to be fast charged.[1] First, the batteries must be in good condition and, of course, specifically designed to accommodate fast charging. Second, fast charging only applies over the first 2/3 or so of the charging process, after which point, the charging must slow down. Lastly, temperatures must be maintained optimal during the charge process. These rules will allow a battery to be charged quickly, as long as the batteries themselves are in good shape and have not been charged too many times. In the future, new battery technology is expected to change these rules, pushed by the need to fast charge vehicles in minutes. But until then, how can we charge batteries at a faster pace?

Several IC manufacturers have developed battery-charging ICs designed for the rapid charging of specific batteries and their unique chemistries. As an example, Maxim Integrated provides battery charger IC products for lithium-ion, NiCd, NiMH, and lead acid batteries, to name just a few.

However, these are very specialized devices and are somewhat complicated to construct into battery chargers, but are well worth the effort to ensure that batteries are both charged fast and have a long life.

You may be getting the idea that there is a lot more to making an effective battery charger that you first thought, and this is very much the case. The circuits used for battery chargers are becoming battery management systems, and as batteries become more sophisticated in their operation and more important to their devices, it becomes more important to get the best lifespan from them as possible.

7.6 Solar Panels and Fast Chargers

A solar panel is a collection of solar cells typically wired in series. A solar panel can have as many solar cells as required to generate the total power and the required voltage level to recharge the battery. Let's consider the task of recharging a 12-V lead-acid battery that will be used to run an instrument outdoors over time so that having a solar recharging system will be useful. The first question is, what type of battery should be considered for this application? Quite often, a standard car battery will be selected, as these are relatively inexpensive and readily available. But this is not the best choice. Car batteries are used to start cars and remain fully charged or rather, be only slightly discharged during the high current burst needed to start an engine. These batteries are also wet cells requiring good ventilation, and can spill acid during transport.

A better choice is to consider a deep-cycle, sealed, absorbed-glass-mat (AGM) battery designed to allow the full capacity of the battery to be used up to the point of being nearly discharged without corrosion. In addition, these batteries require minimal maintenance and ventilation, and will not spill. These also are somewhat less expensive than other 12-V batteries with similar advantages. Somewhat more important for our proposed application is that sealed AGM batteries are less sensitive to the way they are recharged.

A typical rating of a sealed, AGM battery will be 12 V and 100 A-hr. Notice that this is considerably higher than the rating of the smaller batteries we discussed recharging earlier, so we are considering a much more robust power requirement. In a casual search through various vendors, it was easy to find recreational vehicle batteries that have improved performance, such as a UB4D 12-V 200 A-hr battery for $360 (as of the writing of this book) that weighs about 126 lbs or about 57 kg. This provides about 0.29 kg/A-hr (0.64 lbs/A-hr) and about $1.8 per A-hr. These numbers provide good comparisons when looking at different batteries and manufacturers.

Keeping the battery charged is now the challenge and, of course, the larger the solar panel we select the more power we will have available for charging, at least while the sun is shining. Solar panels come in a wide variety of ratings, sizes, and shapes. It is relatively easy to find a 100-W monocrystalline solar panel for about $150 as of the writing of this book. Such a panel is about 1.5 in. (3.8 cm) thick and about 8 ft^2 or ¾ m^2 in area. The power ratings are $V_{OC} = 22.5$ V, $I_{SC} = 5.75$ A, with the optimum values being 18.9 V and 5.29 A for a power output of 100 W, as stated earlier. Of course, if more power were desired, additional panels could be incorporated into the collecting area.

The battery capacity is 100 A-hr with a voltage of 12 V for energy of 1.2 kW-hr. If we consider that the available sunlight per day is 4 hrs and it is at 80% of the optimal intensity, then there will be $100 \times 0.8 \times 4 = 320$ W-hr of energy that can be returned to the battery each day. This is about 320/1200 or just over ¼ of the energy in the battery. If we are powering a light source that requires 10 W to operate for 10 hr per day, we need 100 W-hr of energy per day. If we can restore 320 W-hr per day, then there will be no issues in recharging the full battery each day. If we have 10 days during which there is no sunlight to recharge, the battery will be depleted by 1000 W-hr and will require close to 1000/220 or nearly 5 days to recharge under normal operating conditions. We can implement a simple model to look at the daily battery levels, as shown in Example 7.2.

Example 7.2

Determine how many cloudy days it will take to disrupt the charging cycle of a system, and how many days can be tolerated with no recharging.

```
%% Example_7_2.m
% SWT 12-14-15

% Calculate the effect of no recharging days on battery level
% for a load described with a base and peak load.

%% Housekeeping
clear all; clc;

%% Battery Parameters
batteryCapacity = 100; % A-hr
batteryVoltage = 12; % V

%% Solar Parameters
solarPanel = 100; % Watts
solarEffectiveness = 0.8;
solarHours = 6;

noCloud = [1 0 0 0 1 0 1 0 0 0 1 0 1 1]; % days no clouds
```

```
%% Load Parameters
baseLoad = 1; peakLoad = 50; % watts
peakTime = 4; % hours

%% Calculations
batteryEnergy = batteryCapacity*batteryVoltage; % watt hours
solarEnergy = solarPanel*solarEffectiveness*solarHours;
loadEnergy = peakLoad*peakTime + baseLoad*(24-peakTime);

endEnergy = batteryEnergy;
for day = 1:size(noCloud)
    netEnergy = solarEnergy.*noCloud(day) - loadEnergy;
    endEnergy = endEnergy + netEnergy;

    if endEnergy < 0
        endEnergy = 0;
    elseif endEnergy > batteryEnergy
        endEnergy = batteryEnergy;
    end

  myData(day,1) = day;
  myData(day,2) = endEnergy;

end

figure(1);bar(noCloud*solarEnergy,'w'); hold on;
plot(myData(:,1), myData(:,2),'o-');
axis([0 14 0 1500])
line([0, 14], [loadEnergy, loadEnergy]);
set(gca,'FontSize',9,'FontWeight','Bold','LineWidth',2)
xlabel('Days')
ylabel('Battery Energy W-')
text(2,200,'Load level')
```

7.7 Going Farther

The number of commercial offerings of solar cells and batteries continues to grow, and it is a study in itself to track the number of products, let alone the technical details, of these devices. Before working with any battery or charging system, it is important to fully understand the manufacturer's recommendations and operating requirements for the product and to follow these guidelines. Deviation from them can lead to shortened product life, or even dangerous conditions such as fires and energetic events.

Solar cells and batteries are of great interest to scientists and hobbyists alike, so a wide range of online resources are available to draw from, as well as manufacturers' pages, many of which provide detailed information on specific products. These online resources are some of the best ways to get specific information on products that you either have or are considering obtaining. Should you not be able to find the information you need, contact the manufacturer directly for more detailed technical documents.

7.8 Practice Problems

1. What battery rating would you use to discharge at 1A?
2. What charging rate would constitute a slow charging rate?
3. What is the effect of wiring two solar panels in series?
4. What is the effect of wiring two solar panels in parallel?

Answers

1. 1 C
2. C/100 or less
3. It increases the available voltage.
4. It increases the available current.

References and Notes

1. Check with the manufacturer of your product(s) through their webpages or manuals before working with these components or specifying charge rates and temperatures.
2. J. Drew, "Designing a solar cell battery charger," *Linear Technology Magazine*, pp. 12–15, Dec. (2009).
3. W. Li and M. Day, "Solar charging solution provides narrow-voltage DC/DC system bus for multicell-battery applications," *Analog Applications Journal* **4Q**, 8–11, Texas Instruments, Inc., Dallas, Texas (2011).
4. B. G. Streetman and S. K. Banerjee, *Solid State Electronic Devices*, Sixth Edition, Pearson Prentice Hall, Upper Saddle River, New Jersey (2006).
5. R. L. Easton and M. J. Votaw, "Vanguard I IGY satellite (1958 Beta)," *Rev. Sci. Instr.* **30**, 70–75 (1959).
6. B. Grob, *Basic Electronics*, Eighth Edition, McGraw-Hill, New York (1997).
7. P. Horowitz and W. Hill, *The Art of Electronics*, Third Edition, Cambridge University Press, Cambridge (2015).
8. A. S. Sedra and K. C. Smith, *Microelectronic Circuits*, Oxford University Press, Oxford (2015).

Chapter 8

Photodiodes

There are many circumstances in which we would want to measure the amount of light at a given location.[1] This is a common problem for engineers, scientists, optical designers, and photographers. In the case of photographers there is a need to determine how the effect of the camera's f-stop and filters will combine with the exposure time to produce a desired image. An optical engineer might want to quantify how much light is lost in its passage through an optical system, or how effective an optical coating is performing. An astronomer or optical scientist might be interested in knowing the throughput of a telescope and its optical system, and how that converts to a recognized photometric scale. These examples motivate the need for devices that can accurately quantify how much light is in a given location or system.

One of the more popular components for measuring the amount of light is the photodiode. Photodiodes are semiconductor devices whose primary role is to convert the energy contained in light into a current. These devices are rugged, reliable, and robust, and are used in a wide range of instrumentation that includes spectrometers, optical power meters, optical interlocks, encoders, and more. They are important in communications links such as fiber optics and free space optical transmission. Photodiodes are currently mass-produced commercial items that are available at very reasonable costs to the consumer.

In this chapter we will be concentrating on the electrical configuration, signal amplification, and filtering of the photodiode output and how to select and tune the performance of photodiodes. The operational amplifier plays a critical role in the operation and signal conversion from the photodiode, so we will look in some detail at the transconductance property of operational amplifiers.

8.1 Light Detectors

In the last chapter, solar cells were introduced as devices that convert light into current. In many ways, photodiodes can be thought of as small solar cells, so many of the basic concepts are now familiar. Photodiodes also have much in common with the diodes that we looked at much earlier, the difference

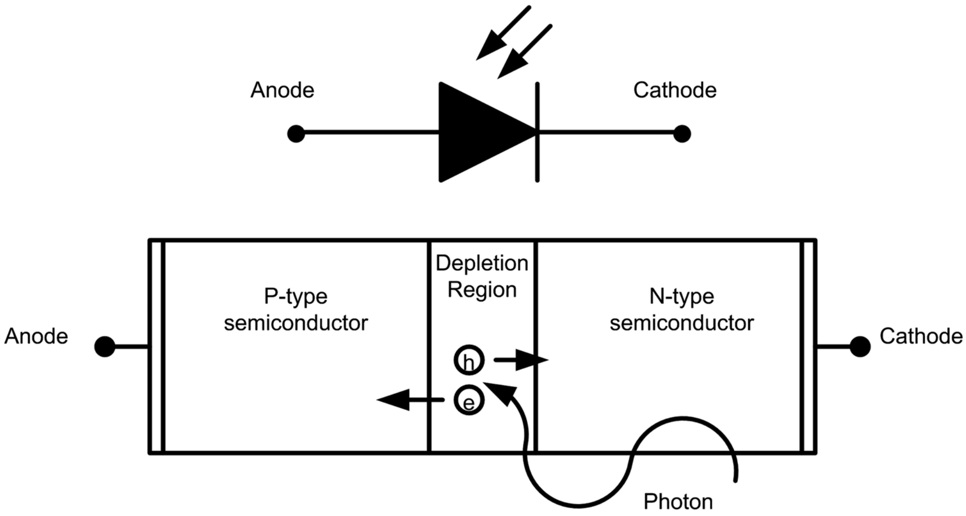


Figure 8.1 A p-n junction representation of a photodiode and the schematic symbol.

being that light can penetrate the electrically active area of the diodes, allowing them to be sensitive to light. It should be no surprise that the starting point for the photodiode is the p-n or p-i-n junction.

A photodiode operates[2,3] when a photon of sufficient energy is absorbed in or near the depletion region of the semiconductor producing electrical carriers known as electron–hole pairs. The electron is negatively charged, while the hole is positively charged so that electrons and holes can move in opposite directions and create a current called a photocurrent. This is illustrated in Fig. 8.1.

Clearly, the larger the light signal the more electron–hole pairs can be created and the larger the photocurrent produced, up to the point where the system saturates. Competing with the photocurrent is the dark current, i.e., a current that originates as a thermal effect, independent of a light source. A dark current will be present even if light is blocked from reaching the photodiode and is reduced by cooling the photodiode. It is important to appreciate that the wavelength of light also plays an important role in producing the current. Photodiodes made from silicon are able to produce signals from photons that are in the wavelength range of 190–1100 nm. Other materials are sensitive to other wavelength ranges. For example, mercury cadmium telluride can be used over a wavelength range of 400–14,000 nm, making it an important material.

Photodiodes are available in a wide range of packaging, sizes, and performance parameters. A key parameter is the responsivity, which describes the ratio of the generated photocurrent to the amount of light and is stated in units of amperes of photogenerated current per watt of light power, or A/W. The responsivity of the photodiode should be compared against the dark current for the photodiode, particularly if operation at low light levels is anticipated or accurate optical power measurements are required. A photodiode designed to quickly respond to changes in light will require a short response time that will depend on the resistance and capacitance of the photodiode, with larger photodiodes having a larger area and thus a larger capacitance. When high-speed photodiodes are needed, p-i-n diodes are more often used.

Selecting the right photodiode takes a little effort, mostly because of the wide selection and choices among many manufacturers. Some key considerations are that photodiodes are very inexpensive, making them a good choice for light detection solutions, and can come in a wide range of sizes and shapes. Photodiodes are a good choice for optical power ranges in the pico and nanowatt range, although the milliwatt range is available in larger or small collection areas. Another feature of photodiodes is that they can be custom packaged for a particular application. The process of selecting a photodiode comes down to very few decisions. First is to determine the wavelength range of interest, which will determine the semiconductor type. Second is the sensitivity and responsivity needed to support the signal rates you require. Finally, the size of the photodiode that can be physically accommodated, which will limit the light collection area, must be determined. With these parameters in mind, most manufacturers will provide sufficient information that you can start looking at their specific parts. Once the diode is selected, the next step is to determine how to get the electrical signal to be analyzed.

8.2 Transimpedance Amplifiers

The output of a photodiode is a current, but quite often it is not a very large current, particularly when a photodiode is being used to look at very low light signals or is operating close to an edge of the wavelength range. In cases where there is sufficient current, a voltage signal can be generated by passing the current through a resistor. For photodiodes, however, a transimpedance amplifier[2–6] constructed from an operational amplifier is commonly used.

A simple circuit for an operational amplifier transimpedance amplifier is shown in Fig. 8.2. The photocurrent is shown as I_{pc}, and the feedback resistor is R. In this circuit the photodiode is connected to a low-impedance load and is operating in photovoltaic mode with the output voltage as given in Eq. (8.1). This circuit is commonly used for low light operation where large gains will be required and is known as the photovoltaic transimpedance amplifier configuration. The negative output voltage is due to the use of the inverting amplifier configuration. One issue that can arise is that with the large amplifier gain, an appreciable voltage offset on the input would translate into a large output voltage. This can restrict the choice of operational amplifier used in the transimpedance amplifier.

The term transimpedance can be appreciated if we examine the equation

$$\frac{V_{out}}{I_{pc}} = -R \tag{8.1}$$

and consider that this is a representation of the amplifier gain or what we have been referring to as the transfer function of the amplifier. In this case, rather than an input voltage, we have an input current, and our transfer

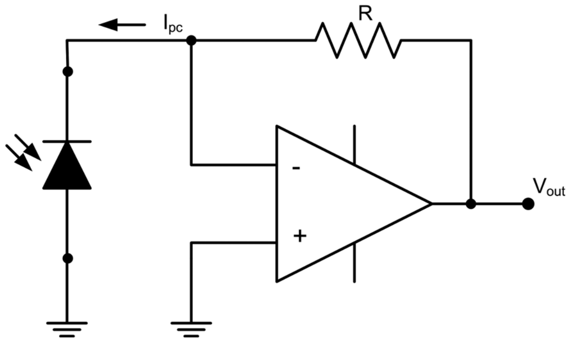


Figure 8.2 A transimpedance amplifier constructed from an operational amplifier.

function is the ratio of the voltage out to the input current, which is given as the impedance. Transimpedance is just a contraction of transfer impedance and is meant to represent the current-to-voltage conversion.

Photodiodes can be represented operationally as a current source and a capacitor in parallel. As such their operation when configured in the above circuit is often limited to DC or very low frequencies. The capacitance in the photodiode can be altered by increasing the width of the junction depletion region by reverse biasing the photodiode. This is similar to increasing the separation of a parallel-plate capacitor to lower its capacitance. Figure 8.3 shows how this can be accomplished: the photodiode is connected in reverse bias, and a second capacitor is added in the feedback loop. While this does increase the level of the dark current, it is almost always an improvement in the overall performance.

The circuit shown in Fig. 8.3 is commonly used to improve the performance of the photodiode. The bias applied to the photodiode increases the width of the depletion region, which lowers the capacitance of the photodiode. Large photodiodes with larger "zero-bias" capacitance require this approach to

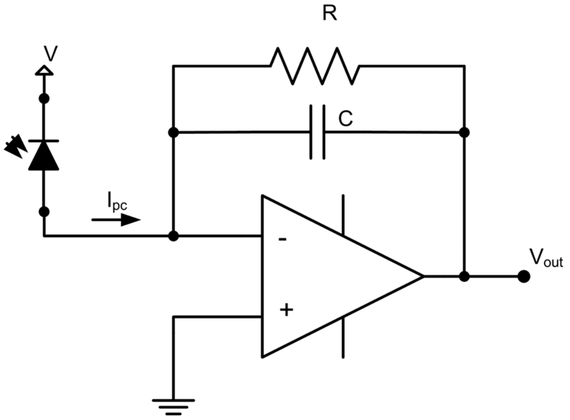


Figure 8.3 A reverse-biased photodiode circuit used to reduce the photodiode capacitance. The voltage V is used to increase the reverse bias on the photodiode.

provide a reduction in their capacitance so that the system can respond to higher-frequency signals. This circuit configuration is also known as the photoconductive transimpedance amplifier configuration. One downside to this circuit is that it is susceptible to going into oscillation and requires the addition of a parallel capacitor C in the feedback loop to maintain stable operation.

8.3 Amplifier Chaining

The output voltage from the circuits we have examined so far has been negative. The output voltage can be easily converted to a positive signal by the introduction of another inverting operational amplifier, as shown in Fig. 8.4. If all that is desired is to convert the output voltage to positive, a unity gain inverting buffer can be used; however, if additional signal gain is needed, a second-stage gain can be added in the inverter.

The two-stage amplifier system[2,4] in Fig. 8.4 is very useful for operating the photodiode; however, a little care needs to be shown in the gain of the second stage. Any noise showing on the output of the transimpedance amplifier will see the same gain as the signal of interest. We can help ourselves here by looking at the addition of filter stages in our amplifier, as was discussed in Chapter 6.

In many photodiode applications, it is desirable to have the transimpedance amplifier and the photodiode on the same chip and use a single, low-voltage supply such as a battery or the voltage from a USB computer port. These devices have been evolving for some time and are becoming more commonly available in the commercial market.

8.4 Removing Unwanted Effects

Section 8.1 introduced the idea of dark current, i.e., a current that has nothing to do with light entering the diode. In some cases, this can be a very troubling source of error. If you remember, the operational amplifier has an

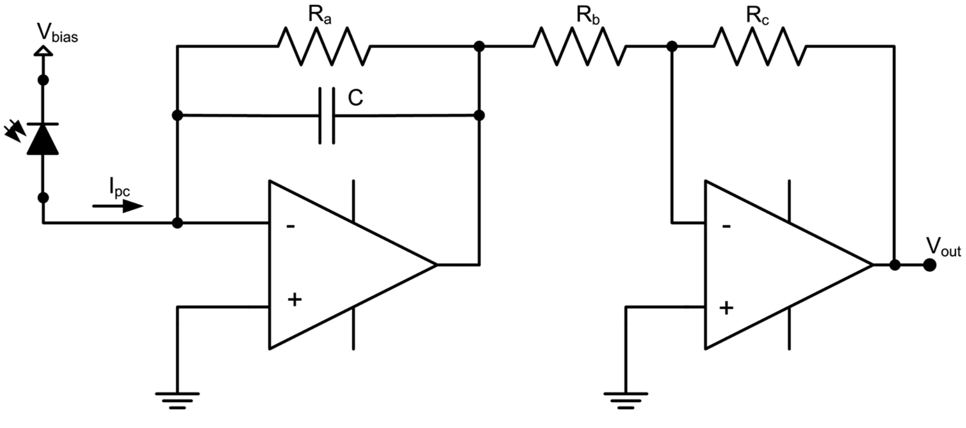


Figure 8.4 A reverse-biased photodiode feeding a transimpedance amplifier followed by an output buffer stage.

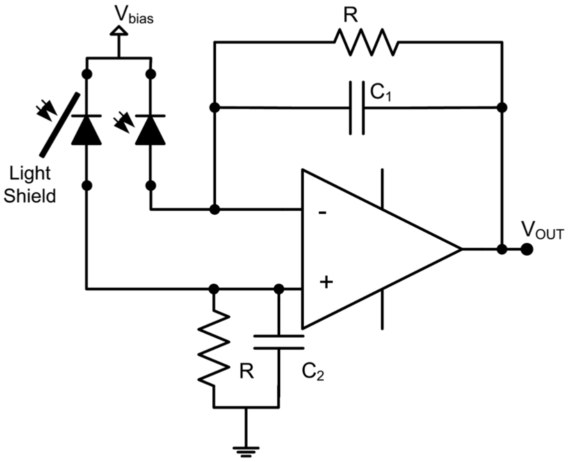


Figure 8.5 Two photodiodes connected to a difference amplifier to remove the effects of common dark current. Notice that one of the diodes is located behind a light shield to block photocurrent generation.

interesting ability to difference two signals. Figure 8.5 shows how two identical photodiodes can be connected to remove the thermally induced dark current.

This technique has been shown for the removal of the dark current and as such is significant in performing accurate light or laser power measurements. In the case of monitoring higher-power lasers, a beamsplitter or an angled window can be used to direct a small amount of the beam toward the photodiode so that the full optical power is not on the diodes. This is also a very effective means of compensating for temperature drift in the photodiodes.

An optical engineer might have a number of reasons to compare two actual light signals in this fashion such that the light shield is not needed or can be replaced with a specific optical filter with differing wavelengths or with differently sized pass bands. Similar circuits and clever optical filtering can be used to generate a wide range of composite signals that can be combined to monitor two signals compared to each other. Although we will not show it here, a summing circuit with a number of pass bands can be used to create a combined signal.

8.5 Photodiode Applications

Photodiodes have many uses, and it is easy to find even more uses once you start working with them, in particular when they are combined with a microcontroller. Two applications are illustrated in this section. The first incorporates a comparator that compares the photodiode output to a set of points defined by the microcontroller and presents a running count of these events to the microcontroller. An application for this might be an optical counter for items as they pass a specific point. The second application uses a photodiode and a sample-and-hold circuit to provide an analog value to the microcontroller to convert to a digital number. A typical application of this would be a digital power meter.

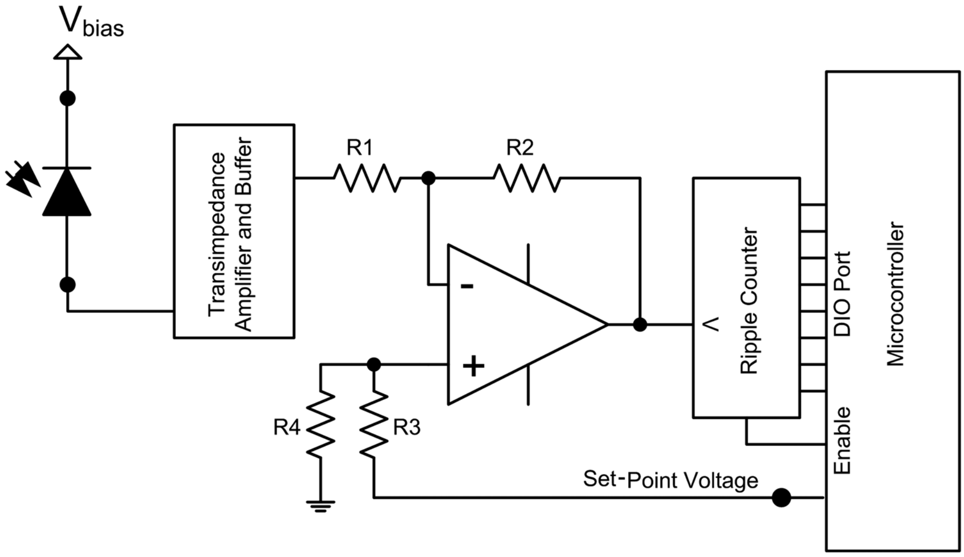


Figure 8.6 A photodiode configured to pass a digital count to a microcontroller. The operational amplifier acts as a comparator and provides a voltage output when the set point controlled by the microcontroller is exceeded. (DIO is digital input/output.)

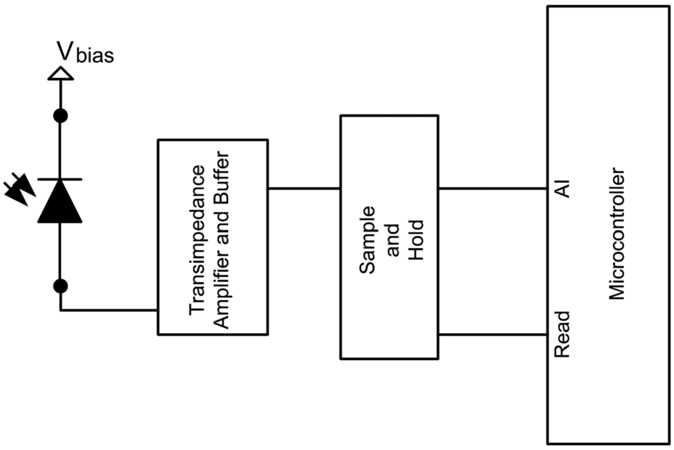


Figure 8.7 A photodiode configured to pass a digital count to a sample-and-hold circuit that is controlled by a microcontroller. The microcontroller provides the conversion to a digital number by using an analog input (AI) and its built-in analog-to-digital converter.

Figure 8.6 shows an optical counter that can be used to count events that change the intensity by a set amount. Comparator circuits were introduced in Chapter 4 as difference amplifiers.

Figure 8.7 shows an optical power meter circuit that can be used to record the intensity value seen by a photodiode. This circuit can be combined with a LED voltage level indicator to make a simple self-contained portable instrument.

Example 8.1

Design an amplifier system and photodiode with a spectral responsivity or conversion rate of 1 A/W to provide a 10-V maximum signal at 1-μW input of light centered at 600 nm with a bandwidth of ±10 nm for an efficiency of 87%.

As we are working with effectively 600-nm light, a silicon photodiode will be able to detect the light. Because the diode efficiency is 87% at the wavelength being detected, the photocurrent is

$$I_{pc} = \varepsilon(SR)L = 0.87\left(1\frac{\mathrm{A}}{\mathrm{W}}\right)1\times10^{-6}\ \mathrm{W} = 0.87\ \mu\mathrm{A},$$

where ε is the diode efficiency, *SR* is the spectral responsivity, and *L* is the input power. The bandwidth is coupled into the spectral responsivity term.

The gain for a transimpedance amplifier depends only on the resistor used in the feedback loop, as shown in Eq. (8.1). For a 1-MΩ resistor, the output voltage is

$$V_{out} = -I_{pc}R = -(0.87)1\ \mathrm{M\Omega} = -0.87\ \mathrm{V}.$$

For this output to be converted to a positive value and a 10-V maximum, we require

$$G = \frac{V_{out}}{V_{in}} = \frac{10}{0.87} = 11.5.$$

A second inverting amplifier with a gain of 11.5 will produce the 10-V output that we seek. A wide range of resistors is available to produce this gain.

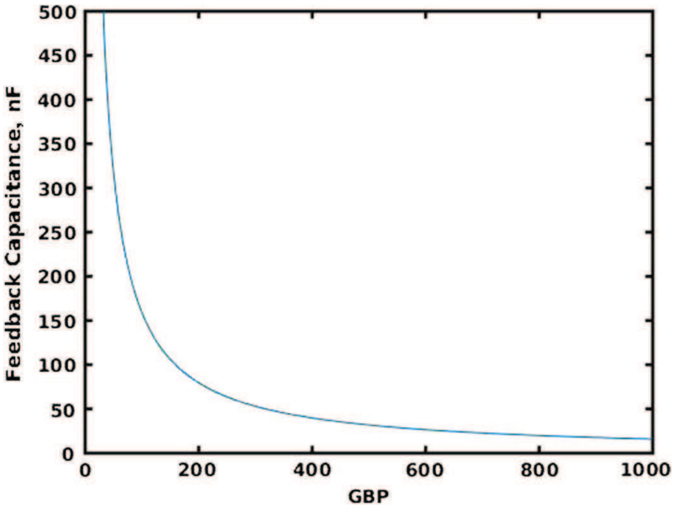


Figure E8.1 The variation in the feedback capacitance with the gain bandwidth product for a feedback resistance of 10,000 and a photodiode capacitance of 200 pF.

The equations from Example 8.1 are implemented in a MATLAB® script in Example 8.2, which tunes the resistor parameters to determine where the gain will be most effective. By knowing the coupling resistor between the two amplifiers shown in Fig. 8.4, the gain at the transimpedance and buffer amplifier can be varied as desired.

Example 8.2

```
%% Example_8_2.m

% SWT 12-16-15
% Determines the resistors to support a transimpedance
% amplifier and an inverting buffer to give a desired
% output voltage based on diode parameters.

%% Housekeeping
clear all; clc;

%% Diode Parameters
SR = 1; % A/W
Efficiency = .87;

%% Light Parameters
L = 1e-6; % W
lambda = 600; %nm

%% Buffer amplifier Parameter
Rin = 1000;
Vout = 10; % V

%% Transimpedance Parameters
RT = 1e5;

%% Photodiode output
Ipc = L*SR*Efficiency;

%% Transimpedance amplifier
VTout = -RT*Ipc;

%% Buffer amplifier
Gain = -Vout/VTout;
Rf = Gain*Rin;

%% Output
fprintf('RT = %3.3f MOhms\n', RT/1e6)
```

```
fprintf(‘Ipc = %3.3f uA\n’, Ipc*1e6)
fprintf(‘Transimpedance output = %3.3f V\n’, VTout)
fprintf(‘Buffer gain = %3.2f\n’, Gain)
fprintf(‘Buffer feedback resistor = %3.1f kOhms\n’,...
      Rf/1000)

Results:
RT = 0.010 MOhms
Ipc = 1.500 uA
Transimpedance output = -0.015 V
Buffer gain = 1000.00
Buffer feedback resistor = 1000.0 kOhms
```

8.6 Capacitance and Oscillation in Amplifiers

Accurate measurements require electronics that do not drift with temperature, have small DC errors with low noise, have fast response without a large signal overshoot, and are stable. Photodiodes, particularly large-surface-area diodes, can have significant capacitance, which can cause an operational amplifier to oscillate. Even when the photodiode capacitance is manageable, stray capacitance from the circuit board can increase the capacitive load, leading to unstable operation.[2,4,6]

The usual cure for this unwanted oscillation is to add a capacitor in the feedback loop of the transimpedance amplifier, as shown in Fig. 8.3. While it is quite common to test capacitors for an appropriate value to reduce or eliminate the ringing or oscillation, it is fairly straightforward to understand what is happening. Figure 8.8 shows a photovoltaics-driven photodiode connected to a transimpedance amplifier. Here, the photodiode is represented

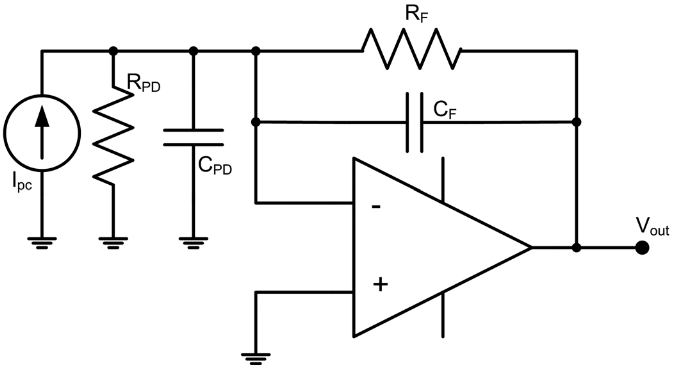


Figure 8.8 A photodiode (PD) modeled with ideal components, showing the effect of associated capacitance C_{PD}, which must be compensated with a feedback capacitor C_F.

by its ideal elements: a current source, a resistor, and a capacitor. You may recall that this circuit looks a lot like a relaxation oscillator introduced in Chapter 4, so it should not be surprising that this circuit can be unstable. In this circuit the DC response is set by the two resistors, and in general the photodiode resistance will be greater than the feedback resistance, forcing the operational amplifier DC gain to unity.

Selection of the correct feedback capacitor to suppress oscillation can be calculated as

$$C_F = \frac{1}{4\pi R_F GBP}\left(1 + \sqrt{1 + 8\pi R_F C_{\mathrm{PD}} GBP}\right), \tag{8.1}$$

where GBP is the gain bandwidth product, and the other components are shown in Fig. 8.8. The challenge is, of course, working out the gain bandwidth product and the photodiode properties. A calculator for Eq. (8.1) is shown in Example 8.3.

Example 8.3

Calculator for the feedback capacitance in a transimpedance amplifier driven by a photodiode.

```
%% Example_8_3.m
% SWT 12-17-15

% Calculate the feedback capacitance for a photodiode driven
% transimpedance amplifier

%% Housekeeping
clear all; clc;

%% Diode Parameters
CPD = 200e-12;

%% Transimpedance Amplifer
RF = 10000;
GBP = 1:1000;

%% Calculations
A = 4*pi*RF.*GBP;
CF = (1 + sqrt(1+2./A*CPD))./A;

%% Plot
figure(1); plot(GBP, CF*1e9)
axis([0 1000 0 500])
```

```
xlabel('GBP')
ylabel('Feedback Capacitance nF')
set(gca,'FontSize',9,'FontWeight','Bold','LineWidth',2)
```

8.7 Practice Problems

1. A photodiode amplifier has a transimpedance gain of $-10{,}000$ and a voltage gain of -1000. What diode spectral responsivity will be needed for 2 μW of light to be converted to 15 V if the efficiency is 100%?
2. Calculate the feedback capacitance for a photodiode-driven transimpedance amplifier with capacitance of 500 pF, resistance of 100 MΩ, and a feedback resistor of 1 MΩ for a gain bandwidth product of 1000.

Answers

1. 0.75 A/W
2. 159 pF

References

1. J. W. Hardy, *Adaptive Optics for Astronomical Telescopes*, Oxford University Press, Inc., Oxford (1998).
2. A. S. Sedra and K. C. Smith, *Microelectronic Circuits*, Oxford University Press, Oxford (2015).
3. S. M. Sze and M.-K. Lee, *Semiconductor Devices: Physics and Technology*, Third Edition, John Wiley & Sons, Inc., New York (2012).
4. P. Horowitz and W. Hill, *The Art of Electronics*, Third Edition, Cambridge University Press, Cambridge (2015).
5. "Op Amp Circuit Collection," National Semiconductor Application Note AN-31, National Semiconductor Corp., Santa Clara, California (2002).
6. T. Wang and B. Erhman, "Compensate Transimpedance Amplifiers Intuitively," Texas Instruments Application Report, SBOA055A (2005).

Chapter 9

Quad-Cells and Position-Sensitive Detectors

Quad-cells and position-sensitive detectors play an important role in measuring optical beam deflection as might be encountered in image stabilization and active or adaptive optics systems.[1–3] These are deceptively simple devices that can be constructed in several ways using discrete components or camera systems. This chapter will consider several different systems but will be predominantly concerned with the discrete components for the optically active photodiode-based subsystems. This will keep the complexity to a minimum and allow us to keep our focus on the electronics rather than the processing of the information from these devices.

Position-sensitive detectors and quad-cells can be constructed from discrete semiconductor photodiodes, and both are photosensitive. Both devices are regularly used as the sensor of a feedback-controlled positioning system that can be used to position a beam or to keep a beam of light focused on a single point.

In this chapter we will be concentrating on the electrical configuration, signal amplification, and filtering of quad-cells and lateral position-sensitive detectors and how to process their outputs into a useable form using the analog electronics that we have been working with. In general, analog electronics allows us to keep as much of the signal information as possible while performing mathematical operations as fast as possible. These features are often highly desirable in high-speed, low-light optical systems, making them a good choice for many applications.

9.1 Quad-Cells

Quadrant-cell detectors or quad-cells are used to measure the position or centration of a focused spot of light. A quad-cell can be easily constructed by placing four photodiodes together to form a square, two photodiodes on edge. Often simple optics will be placed in front of the photodiodes to provide a focused spot. In the most compact form, a quad-cell can be constructed from

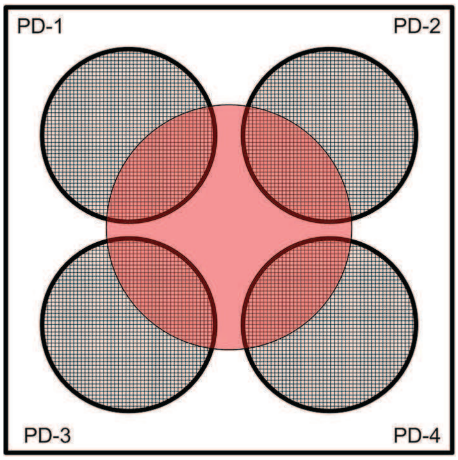


Figure 9.1 A quad-cell using four circular photodiodes (PD) and their relationship to a focused light spot.

four photodiodes placed close together, but they can also be expanded to include many more photodiodes per side, or even to have different geometries.[1,2] An illustration of a simple photodiode quad-cell using circular photodiodes is shown in Fig. 9.1.

The circular spot shown partially covering the photodiodes on the quad cells represents a focused light spot and here is evenly balanced over the four photodiodes so that each would have the same intensity of light. As the spot moves around in the area defined by the extent of the sensor, the illumination of the different photodiodes changes, indicating where the spot is centered. Notice that in this configuration, not all of the light falling on the sensor is being collected by the photodiodes. When there is sufficient light in the spot, this does not pose a problem; however, in low-light conditions, this can greatly reduce the sensitivity.[1,2]

Quad-cells are ideal for defining a location where a spot of light needs to be located. Nulling systems can use active feedback to ensure that the spot remains on the center of the quad-cell by moving a mirror in the opposite direction of the spot movement. The effect is to always keep the light balanced on all four photodiodes.

Quad-cells can be constructed from square photodiodes, or four photodiodes can be constructed on a single semiconductor wafer to reduce the amount of light that will be lost to the spaces between the individual photodiodes. This latter configuration is very popular and is used in many commercial products currently on the market. The critical issue for configuring a quad-cell is to ensure that the light spot is well matched to the size of the photodiodes. It is easy to see that if the spot were very small, it could fall on the center of the sensor and not be detected. A properly sized spot will always have light on all four photodiodes over the deflection range to be measured.

A natural extension of the quad-cell uses the pixels on an imaging camera to be grouped together through software into defined quadrant detectors. This can be done at the resolution of single pixels, or else superpixels can be defined that can contain hundreds of pixels. This approach has the advantage of being able to define a quad-cell at will, wherever the spot of light has fallen, but can also have the disadvantage of the low readout speed of large camera frames.

Powering individual photodiodes using operational amplifiers as transimpedance amplifiers[4,5] was discussed in the previous chapter and can be used to get intensity information from the photodiodes. How this information is used to determine the centration of a spot will be discussed next.

9.2 Quad-Cell Mathematics

Determining the location of the spot on the sensor requires that the intensity level of light on each photodiode be compared. This is most simply accomplished under the assumption that the light in the spot is uniformly distributed. This will not typically be the case, but will serve as the starting point. If we consider a model of the quad-cell sensor, shown in Fig. 9.2, we can calculate the displacement in the X and Y directions. Notice that the directions for positive displacements are shown, as well as a relative percentage of light that is being recorded in each quadrant. The quadrant labels A through D are used as identifiers.

The displacement of the spot is determined in the X and Y directions separately by adding up the signal in the quad-cell pairs and then comparing them mathematically.[1–3] The X displacement is calculated as

$$\Delta X = \frac{(B + C) - (A + D)}{A + B + C + D}. \tag{9.1}$$

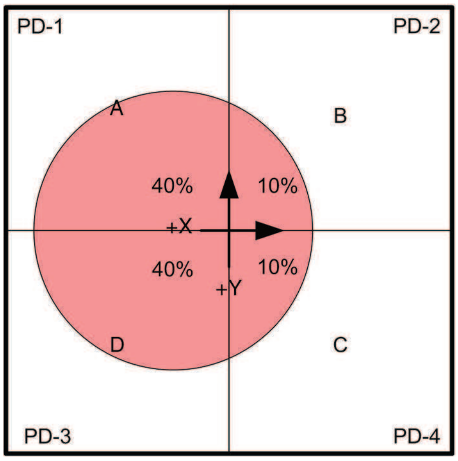


Figure 9.2 A simplified model of the quad-cell used to determine the relationships for calculating the X and Y displacement.

The Y displacement is calculated as

$$\Delta Y = \frac{(A+B)-(D+C)}{A+B+C+D}. \tag{9.2}$$

The application of Eqs. (9.1) and (9.2) to the signals measured by a quad-cell provides the displacement of the spot on the sensor. In the case of the percentages shown in Fig. 9.2, the displacements are shown in Example 9.1. Notice that the amount of light in the spot is normalized to the total amount of light measured in each photodiode. This reduces the sensitivity to light that may be lost in the "seams" between the photodiodes.

Example 9.1

Calculate the displacement of the spot shown in Fig. 9.1 using Eqs. (9.1) and (9.2).

The displacement in the X direction is

$$\Delta X = \frac{(10+10)-(40+40)}{40+40+10+10} = \frac{20-80}{100} = -0.6.$$

The displacement in the Y direction is

$$\Delta Y = \frac{(40+10)-(40+10)}{40+40+10+10} = \frac{50-50}{100} = 0.$$

The spot is displaced to the left or $-X$ direction 0.6 of one quad cell from the center and has no displacement vertically or in the Y direction.

The displacement of a spot moving across a quad-cell is shown in Fig. 9.6. Here, a light spot of 10 units in diameter is moved across the quad-cell along the X axis. The output of the quad-cell is shown on the Y axis. The quad-cell is sufficiently large that the light spot is not moved off of the light-sensitive area. This curve was calculated by comparing the displacement of the light spot centroid from the center of the quad-cell in both the X and Y directions. A simple relationship can be generated from trigonometry[7] to calculate the area on either side of the quad-cell from the displaced circle:

$$\theta_X = 2\cos^{-1}\left(\frac{\Delta X}{R}\right); \qquad D_X = \frac{\frac{A_C}{2} - \frac{R^2}{2}(\theta_X - \sin\theta_X)}{A_C}, \tag{9.3}$$

where A_C is the area of the light spot, R is the spot radius, and θ_X is the angle between the circle center and the line segment perpendicular to the displacement of the center of the quad-cell. A calculator for the curve in Fig. 9.3 is shown in Example 9.2.

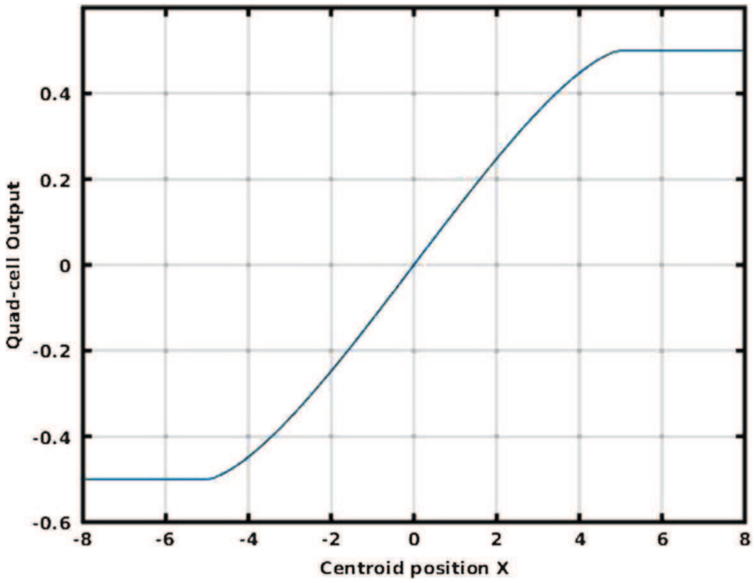


Figure 9.3 Quad-cell output for displacement of a circular light spot along the *X* axis.

Example 9.2

Calculate the quad-cell output as a light spot is moved along the *X* axis.

```
%% Example_9_2.m
% SWT 12-17-15

% Calculate the relationships of a circular spot in a
% quad cell.

%% Housekeeping
clear all; clc;

%% Circle parameters
CX=-8:0.1:8; % displacement
CY=1; % displacement
R=5; % circle radius;

%% Calculator
areaCircle=pi*R^2;

thetaX=2*acos(CX/R);
area_X=R^2*(thetaX-sin(thetaX))/2;
displacement_X=(areaCircle/2 -area_X)/areaCircle;
```

```
thetaY = 2*acos(CY/R);
area_Y = R^2*(thetaY-sin(thetaY))/2;
displacement_Y = (areaCircle/2 -area_Y)/areaCircle;

% %% Output
figure(1); plot(CX,displacement_X)
axis([-8 8 -.6. 6])
grid on;
xlabel('Centroid position X')
ylabel('Quad-cell Output')

See Fig. 9.3
```

The calculator shows results for spot movement in the *X* axis but would be very simple to modify to work in both axes. While this code does not explicitly calculate the amount of light on each element of the quad-cell, this could be easily constructed.

9.3 Quad-Cell Electronics

Quad-cells used for beam stabilization are generally operated at fairly high speed. While this can be accomplished digitally, it can also be managed in nearly real time using the analog electronics we have been working with throughout this book. This is, in fact, where operational amplifiers truly shine. The determination of the displacement in the *X* and *Y* directions can be accomplished using a single operational amplifier for each direction.[4,5] Two summing circuits can be used, one on each channel of a differential amplifier, to give the un-normalized displacement. The total intensity can be calculated in a second operational amplifier. Division of two numbers in analog is much more complicated and will be introduced next.

An electrical schematic to calculate the displacement and the total intensity is shown in Fig. 9.4. With this circuit, all that needs to be done to obtain a single voltage representation for the displacement in either *X* or *Y* is to divide the displacement and total voltages.[7]

To obtain a single number to represent the displacement in either *X* or *Y* requires that the calculated displacement be normalized. It turns out that this requires a fairly complicated bit of electronics,[7] and device manufacturers have responded by making integrated circuits that will do this operation. These devices are called analog multipliers and dividers, which can be thought of as a black box that when correctly powered will take the input signals and deliver a voltage representing the division of the two signals.

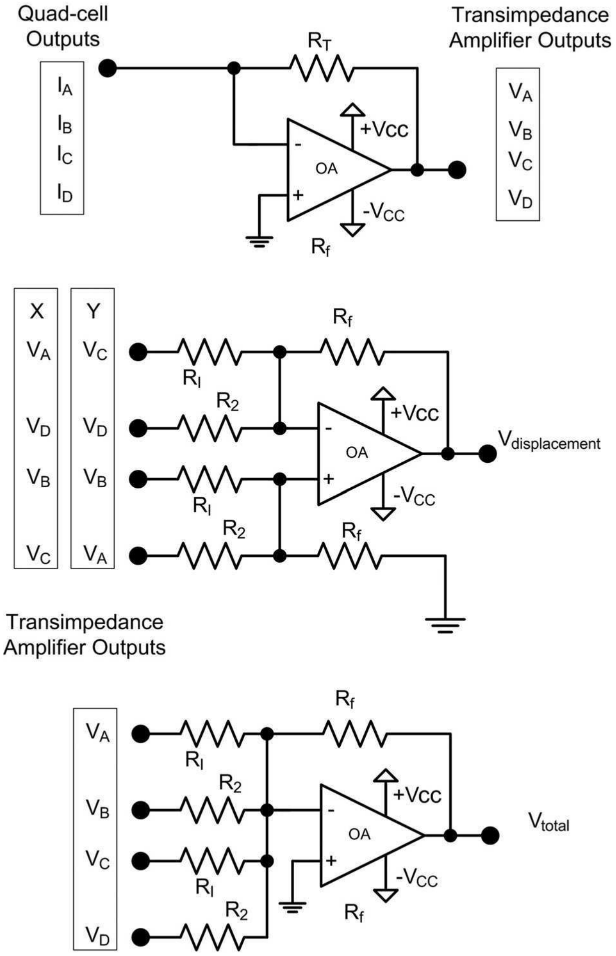


Figure 9.4 Three operational amplifier circuits that combine the quad-cell output into usable numbers for determining beam displacement. The top circuit is one of four transimpedance amplifiers for converting the photodiode current into a voltage; the middle circuit provides the sums and differences; the bottom circuit provides the sum of the light in a quad-cell as a voltage to be used in normalization.

9.4 Position-Sensitive Detectors

The position-sensitive detector (PSD) is another light-sensitive device for measuring the position of a spot on its surface. There are two types of these devices: one has a uniform surface that provides continuous surface location information, and the other has a large number of individual photodiodes on the surface and provides discrete information on position.[3] Our focus will be on the isotropic sensor that provides continuous information across the surface, referred to as a lateral position sensitive detector. These devices work

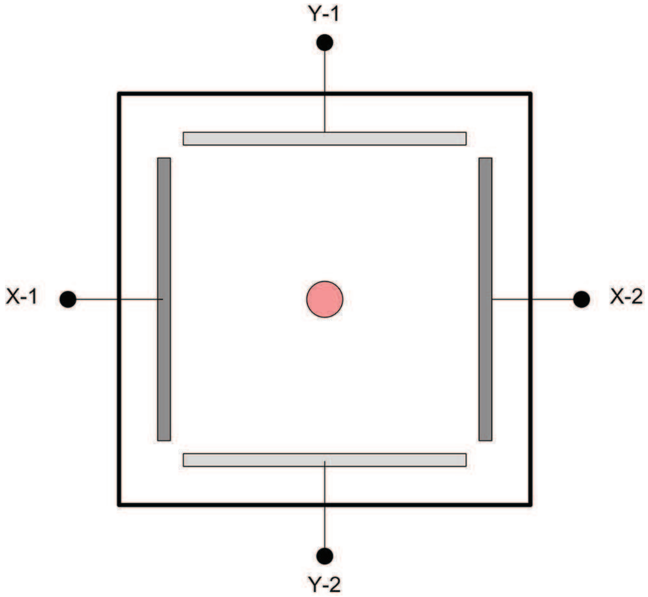


Figure 9.5 Illustration of the PSD showing the electrodes and a focused light spot in the center.

as follows: a spot of light on the surface causes a change in the local resistance at the point of contact, which causes a current to flow into the electrodes. This current is related to the position of the spot on the surface. The device configuration is illustrated in Fig. 9.5.

The PSD is a silicon photodiode with electrodes placed near the edges to provide signals that are related to the position of the light spot on its surface. One advantage of the PSD is that a much smaller focused spot can be used, providing more area in which the spot can move.

The ideal PSD will be a low-cost, low-noise, high-precision, fast-response device, and in many cases will be able to operate at low light levels. Current devices have remarkable performance, and as industrial demand for these devices increases so will their performance; however, some issues of nonlinear performance remain to be addressed in these devices.

One application of a PSD is in tracking a cooperative beacon some distance from the observing site. A guide telescope co-aligned with an imaging telescope at shorter focal length can be kept on target using a PSD. In this case small movements in the longer focal length guide scope can be seen and used to provide a correction to the tracking mount. The effect will to be to reduce the amount of jitter-induced blurring in the image. This approach can also be used to improve the tracking ability of a mount by providing continual position corrections.

9.5 Spot Location in Position-Sensitive Detectors

The light spot on a lateral PSD is referenced to the location of the electrodes. The differences in the current between electrodes on the opposite side of the detector are used to map out the spot location. As such, for a four-electrode detector, determination of the position of the spot on the detector requires differencing the currents that are generated and available at each of the four contacts and providing a means to make them comparable. This leads to the position of the spot being proportional to the differences in the currents in the same axis,[3] as shown in the following two equations:

$$\Delta X = P_X \frac{I_{1X} - I_{2X}}{I_{1X} + I_{2X}}, \tag{9.4}$$

where P_X is a gain constant, and I_{1X} and I_{2X} are currents in the X direction;

$$\Delta Y = P_Y \frac{I_{1Y} - I_{2Y}}{I_{1Y} + I_{2Y}}, \tag{9.5}$$

where P_Y is a gain constant, and I_{1Y} and I_{2Y} are currents in the Y direction.

The gain constants are used to transform the currents into positions appropriate for the desired measurement. These values will usually be very precise when the spot is in the more central region of the detector; however, outside of this region, increased position errors can occur.

The nonlinearity of the PSD usually becomes apparent as the light spot moves out of the central region of the detector toward the corners of the detector. This nonlinearity is similar to optical distortion in shape and can take forms such as pin cushion or barrel shapes. The effect is that the position of the spot and its displacement will have some error. This can be alleviated by mapping out the actual position of the light spot for several readout values and generating a "rubber sheet" mapping to provide correction.

9.6 Position-Sensitive Sensor Electronics

The current output of the PSD uses circuitry that is very similar to that of the quad-cell, shown in Fig. 9.4. Figure 9.6 shows a more complete schematic that runs from one axis of the PSD through to a single displacement value. The PSD circuit is very simple, consisting of an op-amp transimpedance amplifier, a summer, and differencing circuits.[7] The final output however does require that a ratio of the differences and sums be created.

Figure 9.6 shows that the final stage of the electronics requires an analog division. Many approaches are available for dividing the two values, including the use of a specific stand-alone integrated circuit, such as the AD534. In cases where the PSD is being used with a microcontroller, it can be very convenient

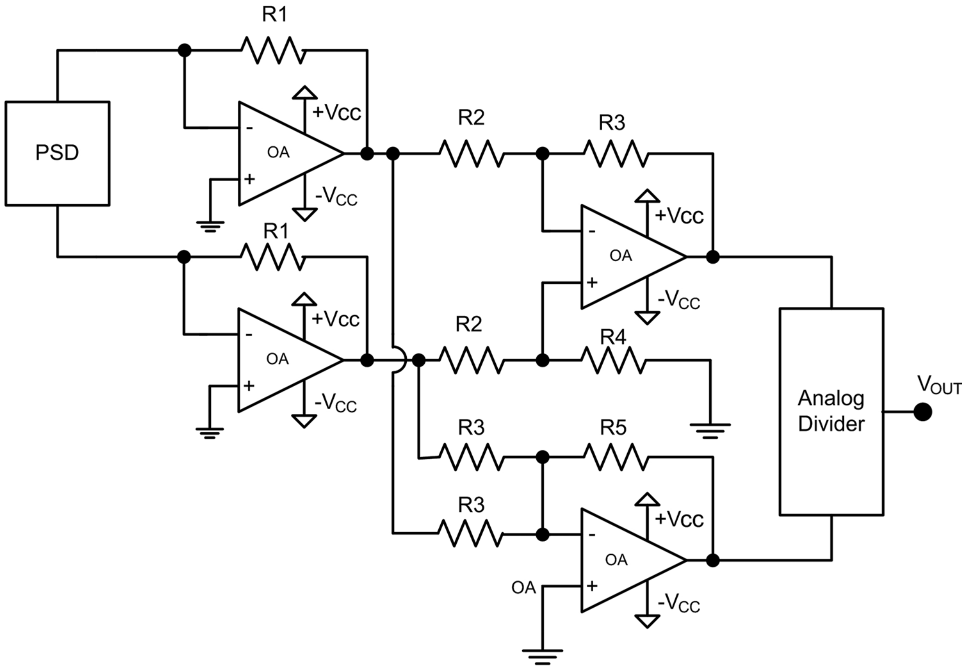


Figure 9.6 A possible implementation of PSD electronics for a single axis. The output voltage is very sensitive to the denominator normalization, so care must be taken in how the division is performed.

to connect the sum and difference signal to the microcontroller directly and perform the division process as a digital division. The circuit shown in Fig. 9.6 is rather minimal and could be improved by having additional buffer amplifiers or filters to reduce any noise in the system.

9.7 Practice Problems

1. A quad-cell is constructed from four square photodiodes, each one being 5 mm on a side. Estimate the minimum and maximum size of the light spot such that it can be displaced 1 mm from the center without losing light on any one of the photodiodes.
2. Assuming a uniform light spot, calculate the value for each of the cases in Problem 1.
3. How can the operation of division be accomplished using operational amplifiers and discrete components?
4. If the resolution of the PSD is 0.1 mm, which corresponds to 0.2 V, what is the resolution of the required analog-to-digital converter?

Answers

1. The minimum spot diameter is just greater than 2.82 mm; the maximum size will be 4 mm.

2. This problem is left to the reader to solve graphically or numerically.
3. An AD534 can be used to perform multiplication and division functions.
4. Ideally, we would measure at half the resolution value or better, so the resolution would be a minimum of 0.1 V per voltage bin of the analog-to-digital converter.

References

1. J. W. Hardy, *Adaptive Optics for Astronomical Telescopes*, Oxford University Press, Inc., Oxford (1998).
2. F. Roddier, *Adaptive Optics in Astronomy*, Cambridge University Press, Cambridge (1999).
3. S. W. Teare and S. R. Restaino, *Introduction to Image Stabilization*, SPIE Press, Bellingham, Washington (2006) [doi: 10.1117/3.685011].
4. P. Horowitz and W. Hill, *The Art of Electronics*, Third Edition, Cambridge University Press, Cambridge (2015).
5. A. S. Sedra and K. C. Smith, *Microelectronic Circuits*, Oxford University Press, Oxford (2015).
6. D. Zwillinger, Ed., *CRC Standard Math Tables and Formulae*, 32nd Edition, CRC Press, Boca Raton, Florida (2011).
7. "Op Amp Circuit Collection," National Semiconductor Application Note AN-31, National Semiconductor Corp., Santa Clara, California (2002).