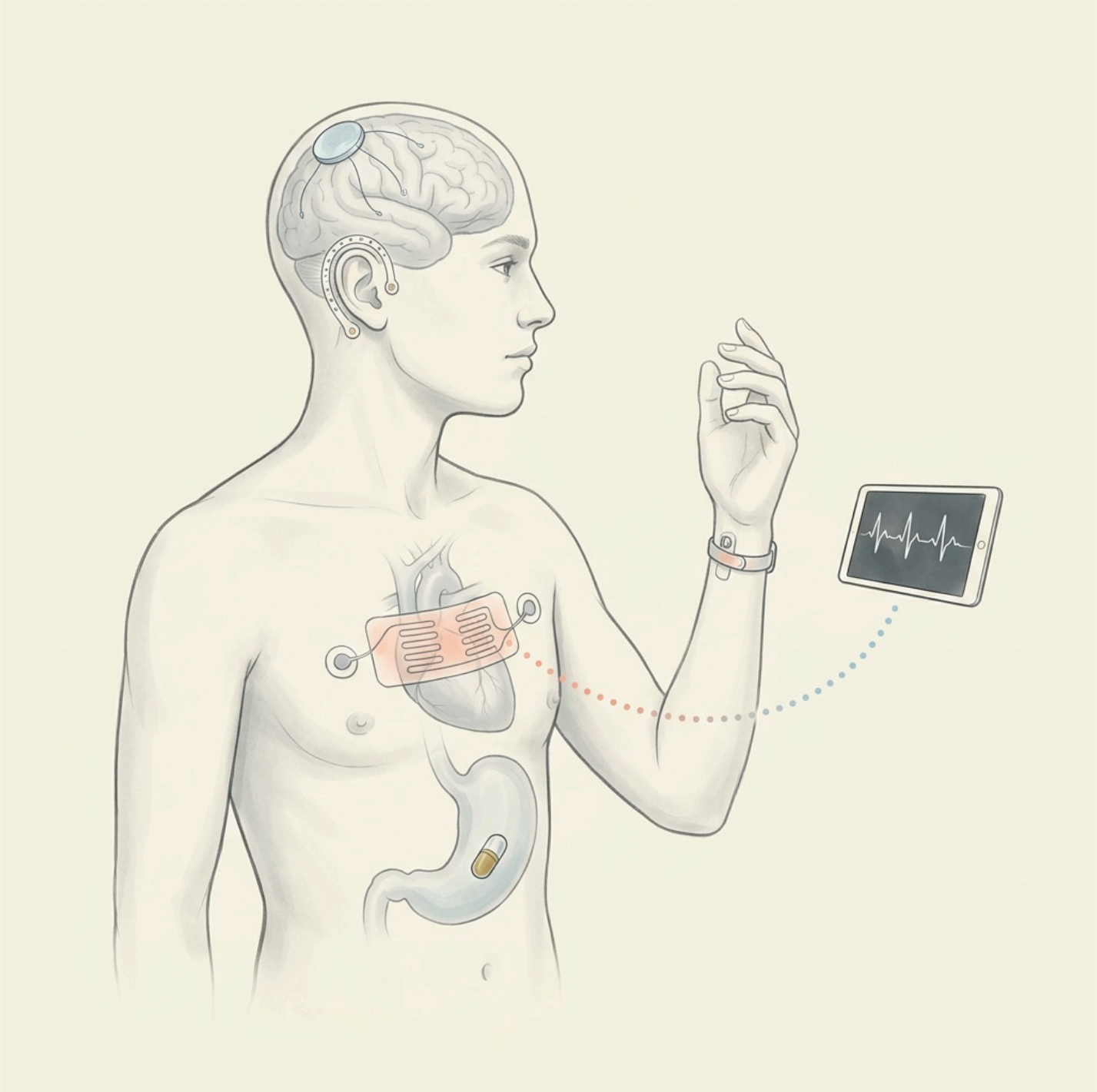
Bioelectronics

Wearable, Implantable, and Ingestible Systems

FIRST EDITION

YASSER KHAN



YASSER KHAN

BIOELECTRONICS

WEARABLE, IMPLANTABLE, AND INGESTIBLE SYSTEMS

FIRST EDITION

Bioelectronics: Wearable, Implantable, and Ingestible Systems
First Edition

Copyright © 2026 Yasser Khan. All rights reserved.

No part of this book may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the author, except for brief quotations in reviews and scholarly works.

ISBN 979-8-9826941-0-2

The author has made every effort to make the information in this book accurate, but it is provided as is, without warranty of any kind. The author is not liable for any loss, damage, or injury that results from its use. Nothing in this book is medical advice. The circuits that the laboratory exercises connect to a person are low-voltage and battery powered; never connect a person to a circuit powered from the mains, follow the instructions of your instructor and your institution's rules for human-subject testing, and follow the documentation of each part, chemical, and instrument.

Arduino, KiCad, Neuralink, Texas Instruments, and the other product and company names mentioned in this book are trademarks of their respective owners. They are used only to identify products, with no endorsement implied.

Preface

A PATCH THE SIZE OF A BANDAGE records the electrical activity of the heart for two weeks while its wearer sleeps, runs, and showers. A thread thinner than a hair carries signals from neurons in the motor cortex to a computer that moves a cursor for a person who cannot move a hand. A capsule swallowed with a glass of water measures pressure, pH, and temperature as it travels through the gut and reports them to a phone. Each of these devices sits where electronics was never meant to go: on skin that stretches and sweats, inside tissue that attacks anything foreign, or in a stomach full of acid. Building them is the subject of this book.

Bioelectronics is the meeting of biology and electronics, and its oldest experiments are older than the battery: Luigi Galvani made frog legs twitch with metal wires in the 1780s, and Alessandro Volta's explanation of those twitches led him to the voltaic pile. For most of the two centuries since, the electronics side moved much faster than the interface to the body. Silicon chips became small, cheap, and fast, but they stayed hard, flat, and rigid, while the body is soft, curved, wet, and always moving. The last three decades changed that. Thin films, printed inks, conducting polymers, stretchable conductors, and resorbable metals now let us make sensors and stimulators that bend with skin, match the softness of the brain, or dissolve when their job is done. Integrated circuits shrank the analog front end, the converter, the processor, and the radio onto one chip that runs from a coin cell.

Designing a bioelectronic device draws on all of this at once. The materials set how the device touches the body and how long it survives there. The fabrication method sets what geometries and costs are possible. The sensing principle (electrical, electrochemical, mechanical, or optical) sets how small the signal is and what can corrupt it. The instrumentation (amplifiers, filters, converters, and potentiostats) must pull a signal of a few microvolts or nanoamperes out of much larger interference. The biology decides what the signal means and what the body will do to the device in return. Power and regulation decide whether the device can be worn, implanted, or sold at all. A designer

has to see how these choices constrain one another: a softer electrode may raise the impedance, which raises the noise, which demands a better amplifier, which costs power, which shortens the battery life a patient will accept.

That kind of understanding comes from building, and it starts from two directions at once. Engineers arrive knowing circuits but not cells; biologists and chemists arrive knowing cells but not circuits. This book is an introduction for both. It grew out of Bioelectronics, a graduate course at the University of Southern California built around a team design project, and it assumes no biology and no circuits beyond Ohm's law. Each idea is explained from the start the first time it is needed, with a picture and a typical number before the equation, and four or five worked examples per chapter show the main ideas at work. Four labs at the end of the book walk through the steps of building the project hardware: laying out a printed circuit board, printing electrodes and a sensor array, designing an analog front-end board around an integrated biopotential chip, and streaming the signals to a computer over Bluetooth.

How the book is organized

The book has four parts and a set of labs (Figure P.1). Part I, *Foundations*, introduces bioelectronic devices and the circuit ideas the rest of the book uses (Chapter 1), then the cells and biomolecules that sensors detect (Chapter 2) and the neurons whose electrical signals they record (Chapter 3). Part II, *Sensing and Measurement*, builds the instrumentation: what makes a good sensor and the op amps that condition its signal (Chapter 4), physical and field-effect transducers (Chapter 5), electrochemical biosensors (Chapter 6), optical sensing (Chapter 7), and the recording of biopotentials such as the ECG and EEG (Chapter 8). Part III, *Interfacing with the Nervous System*, records from and stimulates neurons (Chapter 9) and restores hearing and sight (Chapter 10). Part IV, *Building Devices*, covers the soft materials and thin films devices are made of (Chapter 11), printing them (Chapter 12), powering them (Chapter 13), and taking a design from a class project to a medical device (Chapter 14). Part V holds the four labs.

How each chapter is organized

Every chapter opens with a short list of what you will be able to do by the end, then builds its ideas one at a time. Along the way you will find:

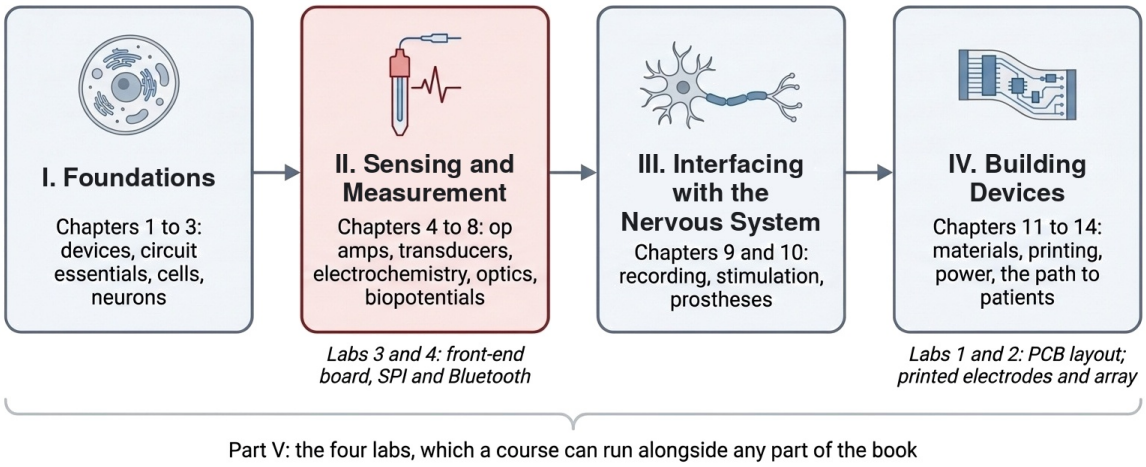


Figure P.1: The parts of the book and the labs that draw on them. Part I gives the background that engineering and life-science readers each lack; Part II is where signals become measurements; Parts III and IV apply both to neural devices and to building real hardware.

Worked examples
: marked with a red rule in the left margin, four or five per chapter, each showing one main idea with realistic numbers. Try each one before reading the solution.

Margin notes
: in the wide right margin hold definitions, key equations, historical asides, references to the research literature, and figure captions, so the main text can stay focused.

Problems
: at the end of each chapter, four or five that cover its main skills, marked by difficulty: ○ for a direct application, ● for a multi-step problem, and ●● for a design or open-ended problem. Answers to all of them are in Appendix A.

Labs
: in Part V. Each starts with a summary box and a list of what you need, then gives the procedure in labeled steps with checkpoints, the results you should expect, what to do when something goes wrong, and analysis questions.

What you need

You need algebra, the basics of calculus (derivatives, exponentials and logarithms), and high-school physics and chemistry. No circuits beyond Ohm's law are assumed: Chapter 1 gathers the circuit ideas the book uses, and Chapter 4 builds the op amp from the beginning. No biology is assumed either: Chapters 2 and 3 build the biology an engineer needs, and the electrochemistry of Chapter 6 starts from oxidation and reduction. Some programming in Python or C helps in the

labs, and Lab 1 starts the board design tool from the beginning. For the labs you need a laptop with KiCad and the Arduino IDE, access to a printing or stencil setup and a curing oven, an LCR meter or a capacitance-to-digital converter board, an integrated biopotential front end (the Texas Instruments ADS1293 for ECG or ADS1299 for EEG) on a board of your design, and a Bluetooth microcontroller board.

Companion website

The book has a companion website, https://bioelectron.org. It holds the lecture slides for every chapter, the lab guides, the design files and firmware used in the labs, and corrections to the text.

A note on the writing

Large language models (LLMs) were used to improve the clarity and grammar of the text. The author takes full responsibility for the content of this book, including any errors.

Acknowledgments

I am grateful to the teaching assistants and students of every semester of the Bioelectronics course at USC, who tested the tutorials and labs, built the first versions of the projects described in Chapter 14, and found the problems first.

I dedicate this book to my wife, Sifat Muin, and our children, Inan Khan and Ilyn Khan.

Yasser Khan
Los Angeles

Contents

Preface 3

I Foundations 15

1 *Introduction to Bioelectronics* 16

1.1 *An Ingestible Sensor, Taken Apart* 17

1.2 *Three Devices, One Architecture* 20

1.3 *What Is Bioelectronics?* 21

1.4 *A Short History* 22

1.5 *Why Bioelectronics Must Be Soft* 25

1.6 *Biosignals at a Glance* 28

1.7 *Chemical Sensing on the Body* 32

1.8 *Implanted Electronics* 37

1.9 *Ingestible Electronics* 39

1.10 *Circuit Essentials* 42

1.11 *The Anatomy of a Bioelectronic System* 49

1.12 *Three Devices in Numbers* 51

1.13 *How This Book Is Organized* 55

1.14 *Problems* 58

2 *Cells and Biomolecules* 60

2.1 *The Cell* 61

2.2 *The Cell Membrane* 63
2.3 *Diffusion and Osmosis* 65
2.4 *Sugars and Glucose* 68
2.5 *Proteins and Enzymes* 69
2.6 *Nucleic Acids and Hybridization* 72
2.7 *Blood and Body Fluids* 75
2.8 *Bacteria and Viruses* 77
2.9 *Cells in Culture and on Electrodes* 80
2.10 *Cell Signaling* 82
2.11 *Problems* 83

3 *Neurons and the Action Potential* 85
3.1 *The Nervous System and the Neuron* 86
3.2 *Ions and the Nernst Equation* 87
3.3 *The Resting Potential* 89
3.4 *Ion Channels and Stimuli* 91
3.5 *Voltage Clamp and Patch Clamp* 93
3.6 *The Action Potential* 94
3.7 *Graded Potentials and Summation* 96
3.8 *The Axon as a Cable* 96
3.9 *Myelin and Saltatory Conduction* 99
3.10 *Nerves and Reflexes* 100
3.11 *Synapses and Tweaking the Body* 102
3.12 *Problems* 105

II *Sensing and Measurement* 107

4 *Sensors and Signal Conditioning* 108
4.1 *What Is a Biosensor?* 109
4.2 *Sensitivity and Selectivity* 110
4.3 *Range, Linearity, and Response Time* 112

4.4 *Noise and Signal-to-Noise Ratio* 114
4.5 *Accuracy, Precision, and Calibration* 116
4.6 *The Detection Limit* 117
4.7 *Drift and Differential Measurement* 119
4.8 *The Operational Amplifier* 120
4.9 *Bias Current, Offset, and the Electrometer Buffer* 124
4.10 *Measuring Small Currents* 127
4.11 *Problems* 131

5 *Physical and Field-Effect Transducers* 133
5.1 *The Wheatstone Bridge* 134
5.2 *Temperature Sensors* 135
5.3 *Strain Gauges* 137
5.4 *Pressure and Mass Sensors* 139
5.5 *Capacitive Sensing* 140
5.6 *Bioimpedance* 142
5.7 *The MOSFET* 144
5.8 *The Ion-Sensitive FET* 146
5.9 *Reading an ISFET* 148
5.10 *Problems* 151

6 *Electrochemical Biosensors* 153
6.1 *Redox Reactions and Electrode Potentials* 154
6.2 *The Nernst Equation at an Electrode* 156
6.3 *The Electrode-Electrolyte Interface* 158
6.4 *The Three-Electrode Cell and Reference Electrodes* 161
6.5 *Voltammetry and Amperometry* 163
6.6 *Electrochemical Impedance Spectroscopy* 166
6.7 *Enzymatic Glucose Sensors* 167
6.8 *Potentiometric Sensors* 170
6.9 *Affinity Biosensors* 172

6.10 *Potentiostat Circuits* 174
6.11 *Microfluidics for Wearable Sweat Sensing* 177
6.12 *Problems* 179

7 *Optical Biosensing* 181
7.1 *Light and Molecular Energy Levels* 182
7.2 *Absorption and the Beer-Lambert Law* 184
7.3 *Fluorescence and the Jablonski Diagram* 186
7.4 *The Fluorometer* 189
7.5 *Total Internal Reflection Fluorescence* 189
7.6 *Surface Plasmon Resonance* 191
7.7 *Pulse Oximetry* 193
7.8 *Light Sources, Photodiodes, and the Front End* 197
7.9 *Problems* 201

8 *Biopotential Recording* 202
8.1 *Where Biopotentials Come From* 204
8.2 *The Electrode-Skin Interface* 209
8.3 *Interference: Mains, Common Mode, and Mismatch* 212
8.4 *The Instrumentation Amplifier* 215
8.5 *Right-Leg Drive and the Wilson Central Terminal* 217
8.6 *Filtering Biopotentials* 218
8.7 *Digitizing: Delta-Sigma Converters and Integrated Front Ends* 221
8.8 *Getting the Data Out: SPI and Bluetooth Low Energy* 223
8.9 *From Samples to Heart Rate and Band Power* 225
8.10 *Electrical Safety* 227
8.11 *Problems* 230

III Interfacing with the Nervous System 232

9 *Neural Recording and Stimulation* 233
9.1 *Neural Interfaces in the Clinic* 234
9.2 *Recording by Depth* 236
9.3 *Electrodes Inside the Brain* 237
9.4 *Scaling Up: Channels and Data* 241
9.5 *The Neural Recording Amplifier* 242
9.6 *Limits of Miniaturization* 244
9.7 *The Electrode-Tissue Interface in Stimulation* 245
9.8 *Safe Stimulation Waveforms* 247
9.9 *Stimulator Circuits* 250
9.10 *Problems* 254

10 *Sensory Prostheses* 255
10.1 *How a Sensory Prosthesis Works* 256
10.2 *Hearing and the Cochlea* 257
10.3 *The Cochlear Implant* 258
10.4 *The Eye and the Retina* 261
10.5 *Estimating Spatial Acuity* 264
10.6 *Retinal Disease and Where to Stimulate* 265
10.7 *The Argus II Epiretinal Prosthesis* 267
10.8 *Subretinal and Photovoltaic Implants* 269
10.9 *Trends in Visual Prostheses* 271
10.10 *Problems* 272

IV Building Devices 273

11 *Materials and Mechanics of Soft Electronics* 274
11.1 *Why Soft and Thin* 275
11.2 *Conductors, Semiconductors, and Insulators* 276
11.3 *Substrates: From Plastic to Hydrogels* 278

11.4 *Making Thin Films* 280
11.5 *Bending and the Neutral Plane* 283
11.6 *How Thin Films Fail* 286
11.7 *Stretchable Conductors* 289
11.8 *Conductive Inks and Sintering* 291
11.9 *Organic Semiconductors and Conductors* 295
11.10 *Organic Transistors: The OTFT and the OECT* 297
11.11 *Encapsulation and Stability* 301
11.12 *Problems* 303

12 *Printed Bioelectronics* 305
12.1 *Why Print?* 306
12.2 *Wetting and Surface Energy* 307
12.3 *Ink Rheology* 310
12.4 *Drying and the Coffee-Ring Effect* 312
12.5 *Inkjet Printing* 313
12.6 *Screen and Stencil Printing* 316
12.7 *Roll-Based Printing and Coating* 318
12.8 *Direct Writing: Dispensers, Aerosol Jet, and Liquid Metal* 319
12.9 *Choosing a Printing Method* 320
12.10 *Edges, Variation, and Registration* 321
12.11 *Printed Electrodes and Sensors on the Body* 322
12.12 *Problems* 325

13 *Powering Bioelectronics* 327
13.1 *Energy, Power, and Efficiency* 328
13.2 *The Power Budget of a Wearable* 329
13.3 *Energy Harvesting for Wearables* 331
13.4 *Batteries and the Lithium-Ion Cell* 333
13.5 *Batteries for Implants* 334
13.6 *Inductive Power Transfer* 336

13.7 Tissue Heating and the Specific Absorption Rate 340
13.8 Ultrasound Power 341
13.9 Powering Ingestibles 343
13.10 Problems 345

14 From Prototype to Patient, and the Design Project 346
14.1 From an Idea to a Specification 347
14.2 Planning the Work 349
14.3 Budgets: Checking the Design on Paper 351
14.4 Building and Bringing Up the Hardware 354
14.5 Testing on People Safely 356
14.6 Data, Demonstration, and Report 358
14.7 Lessons from Past Projects 360
14.8 What Is a Medical Device? 361
14.9 Device Classes and Premarket Pathways 362
14.10 Clinical Evidence 366
14.11 Quality Systems and Standards 367
14.12 Ethics of Human Studies and Devices 370
14.13 Problems 372

V Laboratory 374

L1 PCB Layout and Gerber Files 375

L2 Printed Electrodes and a Capacitive Sensor Array 392

L3 Designing an Analog Front-End Daughter Board 404

L4 Streaming Biopotentials over SPI and Bluetooth 417

A *Answers to Chapter Problems* 434

About the Author 443

Part I

Foundations

1
Introduction to Bioelectronics

In this chapter you will learn to

- Define bioelectronics, name its sensing modalities and the three places a device can live, and trace its history from Galvani's frog to soft and dissolving devices.
- Explain why devices that touch tissue must be soft and thin, using Young's modulus and the t^3 law of bending stiffness.
- Give the typical amplitude and bandwidth of the main biosignals: the ECG, EMG, and EEG, breathing, temperature, and the optical pulse.
- Compare sweat, interstitial fluid, saliva, and breath as sources of chemical information, and compute the charge and dose of iontophoresis.
- Describe pacemakers, brain stimulators, the foreign body response, dissolving electronics, and swallowed capsules, and estimate how long a dissolving device lasts.
- Use Ohm's law, the voltage divider, the RC time constant, impedance, and decibels to analyze a sensor circuit and an RC filter.
- Choose a sampling rate and an ADC resolution for a biosignal.
- Find the system blocks in a patch, an implant, and a capsule, and estimate a data rate and a battery life.
- Follow one design process through this book and its labs.

Rest a fingertip on the crown of a smartwatch and hold still for thirty seconds. The watch draws a trace with a tall spike about once a second: an electrocardiogram (ECG), the same kind of record

a cardiologist reads in a clinic. One electrode sits on the back of the watch against your left wrist; the other is the metal crown under your right fingertip. Between them lies your heart. The watch measures a voltage of about a millivolt between your two arms, amplifies it, digitizes it, and checks the rhythm for signs of atrial fibrillation. Turn the watch over and you see green and red lights that measure your pulse and blood oxygen by a second, optical method.

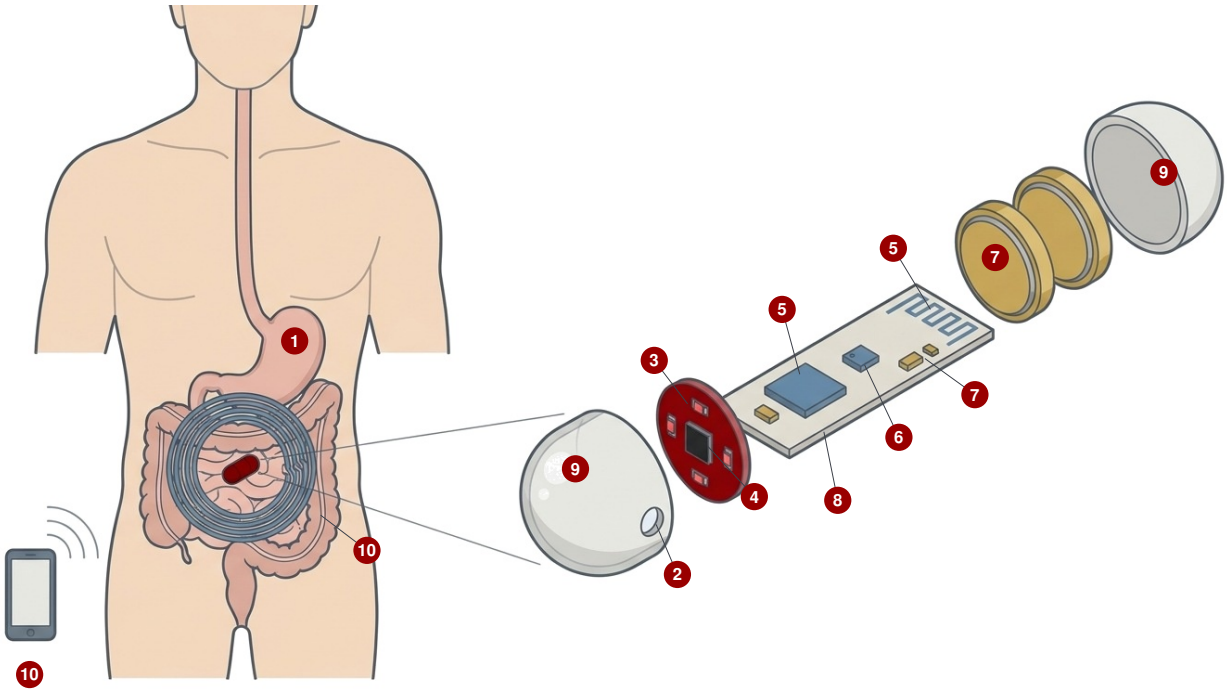
The watch is one of three kinds of device in this book. Under the collarbone of about three million people sits a pacemaker the size of a matchbox, which watches every heartbeat for years and paces the heart whenever it pauses. In a clinic, a patient swallows a capsule the size of a large vitamin pill; for the next day and a half it measures pH and temperature as it rides through the gut and radios the numbers to a receiver on the patient's belt. On the skin, in the body, or in the gut, each device reads a signal from the body, turns it into numbers, and sends them out, on a power budget of milliwatts or less.

This chapter starts by taking one device apart, a sensing capsule, and then sets it beside a patch and an implant to show that all three are built from the same blocks. That list of blocks is also the plan of this book. The chapter then introduces *bioelectronics*, its history, and the reason its devices must be soft, and tours the signals the body offers, the chemistry its fluids carry, and the devices that live inside it. A section on *circuit essentials* follows: the handful of circuit ideas (current and voltage, the divider, the capacitor, filters, decibels, sampling) that every later chapter uses. Readers who know circuits can skim it; readers from biology should work through it. The chapter ends by putting numbers on the three devices, and with a map of the book.

1.1 An Ingestible Sensor, Taken Apart

THE BEST WAY TO SEE WHAT THIS BOOK TEACHES is to take one device apart. The capsule of Fig. 1.1 is swallowed like a pill. As it rides through the stomach and the intestines it shines light on the gut wall, measures the light that comes back, works out where it is, and radios its readings to a receiver outside the body. It is a research design that Abdigazy, Khan, and coworkers lay out part by part.[1] They sort the parts of any ingestible into the blocks this book uses for every device: sensors and actuators, integrated circuits, communication, power, and packaging are essential, and localization and locomotion are optional extras.

[1] A. Abdigazy, M. Arfan, G. Lazzi, C. Sideris, A. Abramson, and Y. Khan, *Nat. Electron.* (2024).



1 The gut: acid in the stomach, then a day or two of travel. *The body, Chapters 2 and 3.*

2 Window in the shell, where the device meets the body. *Body interface, Chapters 7 and 11.*

3 Optical sensor module: four LEDs around a photodiode. *Sensors, Chapters 5 to 7.*

4 Inside the module, a front end amplifies the photodiode's current and a converter turns it into numbers. *Chapters 4 and 8; Lab 3.*

5 Microcontroller with a Bluetooth radio, and its antenna. *Processing and links, Chapters 8 and 9; Lab 4.*

6 Magnetometer: senses the field of the worn coil, to locate the capsule. *Physical sensors, Chapter 5.*

7 Two silver-oxide cells, and the regulators that set the supply voltages. *Power, Chapter 13.*

8 Circuit board that carries and connects the chips. *Labs 1 and 3; printed circuits, Chapter 12 and Lab 2.*

9 Sealed resin shell. *Materials and packaging, Chapter 11; safety and regulation, Chapter 14.*

10 Outside the body: a worn coil and a phone. *Links through tissue, Chapters 9 and 13; the design project, Chapter 14.*

Figure 1.1: A sensing capsule, taken apart. Left: the capsule (red) in the small intestine, a coil worn over the abdomen, and a phone that receives the data. Right: an exploded view, drawn after the optical-sensing capsule with magnetic localization of Abdigazy *et al.* Colors mark the blocks here and in Figs. 1.2 and 1.27: sensing and body interface cardinal red, electronics steel blue, power gold, package gray. The key ties each numbered part to the chapters and labs that teach it. Not to scale.

The dissection

Follow the signal through the capsule; the numbers are those of Fig. 1.1. Each block is a topic of this book.

The body (1). Every design starts from the place where the device must live. A capsule meets stomach acid and spends about 2 to 5 hours in the stomach, 4 to 6 hours in the small intestine, and 10 to 60 hours in the large intestine, close to the ranges of Table 1.4. It must also be small enough to swallow. The largest standard pill, size 000, is 9.91 mm across and 26 mm long; the authors use it as the reference size, because a larger capsule is more likely to get stuck. Chapters 2 and 3 describe the cells, fluids, and nerves that a device meets and measures.

Body interface and sensor (2, 3). The capsule looks at the gut through a window in its shell. Behind the window sits an optical sensor module

(the MAX86916) with four light-emitting diodes (LEDs: blue, green, red, and infrared) and a photodiode, a device that turns light into current. The LEDs light the tissue and the photodiode catches what returns. Chapter 7 explains what light reports about blood and tissue; Chapters 5 and 6 build the physical and chemical sensors that could sit behind a different window.

Front end and converter (4). The photodiode's current is far too small to use directly. An *analog front end* amplifies and filters it, and an *analog-to-digital converter* (ADC) turns the result into numbers. In this capsule both are inside the sensor module. Chapter 4 builds such circuits from the operational amplifier, Chapter 8 follows a complete front-end chip, and in Lab 3 you design a front-end board of your own.

Processor and radio (5). A microcontroller with a Bluetooth Low Energy (BLE) radio on the same chip (the nRF5340) runs the measurements and sends the results out through a small antenna. Radio through the body is costly: for a capsule in the intestine the authors give an in-body *path loss*, the whole loss between the two antennas, of about 3 dB/cm near 400 MHz and about 8 dB/cm at 2.4 GHz, the band of BLE. Absorption by tissue alone, the figure Chapter 9 uses, is about half of that; the rest comes from the spreading of the wave and from reflections where tissues meet. Through 5 cm of tissue that is 15 dB against 40 dB, a factor of about 300 in received power (decibels are explained in Section 1.10), which is why many implants use the medical band near 400 MHz. A capsule that sends a few readings a second can afford the loss when its receiver is close, on or near the body. Chapters 8 and 9 cover the links, and in Lab 4 you write firmware that streams samples over BLE.

Localization (6, 10). A reading from the gut is worth more when you know where it was taken. A tiny magnetometer (the MMC5633), 0.85 mm on a side, measures a magnetic field made by coils outside the body, and the outside system decodes the capsule's position from it. In a related study the coil is worn on a shirt, and a trained neural network turns the field readings into a three-dimensional position, so that gas readings (ammonia and oxygen) can be mapped along the capsule's path; it was tested in a model of the gut and in an intestine from a cow.[2]

[2] A. Abdigazy, M. Arfan, J. Shao, M. S. Islam, M. F. Hassan, and Y. Khan, *Cell Rep. Phys. Sci.* (2024).

Power (7). Two SR626 silver-oxide button cells in series give 3.1 V and 28 mAh. Silver-oxide cells are used because they do not go into the thermal runaway of a lithium-ion cell, and they are the only battery type approved for clinical capsule endoscopes. A step-up (boost) converter and regulators make 1.8 V for the logic and 5 V for the LEDs. With one measurement each second, the capsule draws about 1.75 mA on average, so the cells last $28\,\text{mAh}/1.75\,\text{mA} = 16\,\text{h}$: long enough for the stomach and small intestine, not for the whole colon. Chapter 13

builds such budgets and the batteries, harvesters, and wireless links that fill them.

Board and package (8, 9). The chips sit on a small circuit board, the kind you lay out in Lab 1. A 3D-printed biocompatible resin shell, 9.8 mm by 26 mm, slightly smaller than a size-000 pill, holds it all and keeps the gut's fluids out. Chapters 11 and 12 cover the materials and the printing of soft devices, and Chapter 14 the tests and approvals that stand between a prototype and a patient.

Designing your own

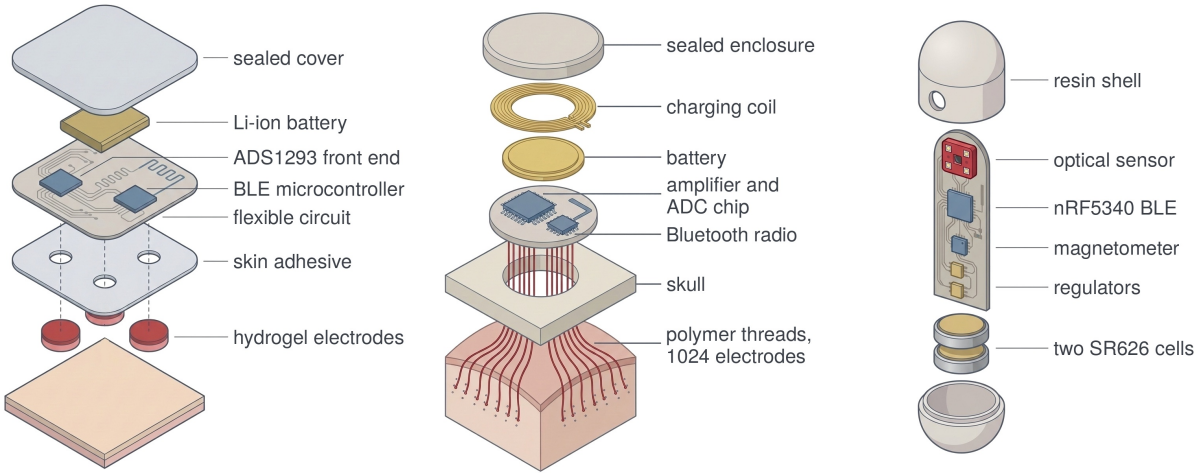
None of these choices is special to capsules. To design any bioelectronic device you answer the same questions in the same order: what signal, at what place in the body; which sensor; what front end and converter; how much data and power; which radio and battery; what package; and how the device will be tested and approved. The chapters answer them one at a time. The four labs each build one layer of hardware (a circuit board, printed electrodes, a front-end board, and firmware with a Bluetooth link), and the design project of Chapter 14 puts the layers together into a device of your own.

1.2 *Three Devices, One Architecture*

THE CAPSULE IS ONE OF THREE KINDS OF DEVICE in this book. A *wearable* sits on the skin, an *implantable* inside the body, and an *ingestible* passes through the gut. Figure 1.2 sets one of each side by side: a wireless ECG patch for the chest, a brain-computer interface set into the skull, and the capsule of the last section.

They look nothing alike and they live in very different places, yet the same colors appear in all three panels. Each device has a *body interface and sensor* (red): hydrogel electrodes on the skin, polymer threads that carry 1024 electrodes into the brain, an optical sensor behind a window. Each has *electronics* (blue) that amplify the signal, convert it to numbers, process it, and send it out by radio. Each has a *power* source (gold): a rechargeable cell, a battery charged through the skin, two button cells. Each has a *package* (gray) that keeps the body's fluids out and the device in place.

Only the choices differ, because the place differs. The patch can be large, but it must bend with the chest and survive a shower. The implant must fit in a hole in the skull, last for years, and cut its data about two hundredfold before the radio. The capsule must be swallowed, run for a day on cells that are safe in the gut, and send its signal through centimeters of tissue. Section 1.12 returns to the three devices



(a) wireless ECG patch (b) brain-computer interface (c) sensing capsule

Figure 1.2: Three devices taken apart into the same blocks (the block diagram is Fig. 1.25). (a) A wireless ECG patch: hydrogel electrodes, a biopotential front end, a Bluetooth Low Energy (BLE) microcontroller, and a battery on a flexible circuit, with a skin adhesive below and a sealed cover above. (b) A schematic of a fully implanted brain-computer interface, drawn from Neuralink's published description of its N1 implant: polymer threads carrying 1024 electrodes (fewer are drawn), a chip that amplifies and digitizes, a Bluetooth radio, and an inductively charged battery in a sealed enclosure that replaces a disk of skull. (c) The optical-sensing capsule with magnetic localization of Abdigazy *et al.*: an optical sensor, a BLE microcontroller, a magnetometer, regulators, and two silver-oxide cells in a resin shell. Colors mark the blocks in all three panels and in Figs. 1.1 and 1.27: sensing, body interface, and actuator cardinal red, electronics steel blue (including the capsule's magnetometer, which serves localization), power gold, package and structure gray. Not to scale.

with numbers, once the circuit ideas are in place.

The architecture is the plan of the book. Because every device is made of the same blocks, the book is organized by block, not by device. Part I (this chapter and Chapters 2 and 3) gives the circuit ideas and the biology. Part II (Chapters 4 to 8) builds the sensors and the front ends that read them. Part III (Chapters 9 and 10) connects electronics to the nervous system. Part IV (Chapters 11 to 14) covers the materials, printing, power, and regulation that turn a circuit into a device. Part V holds the four labs. Figure 1.27, at the end of this chapter, pins every chapter and lab to its block.

1.3 *What Is Bioelectronics?*

Bioelectronics joins biology and electronics. In practice it means building devices that exchange information or energy with living tissue. A device can *record* (an electrode on the chest picks up the heart's electrical activity), *stimulate* (a pacemaker sends a current pulse that starts a heartbeat), or *treat* (a patch releases a drug when a sensor reads a high level of a *biomarker*, a molecule whose level reports on health). The most interesting devices close the loop: they sense, decide, and act.

Where the device sits sets most of its constraints. *Wearable* devices sit on the skin, are easy to put on and take off, and see only what reaches the surface. *Implantable* devices sit inside the body, close to the organ they serve, and must survive for years in warm salt water while the immune system attacks them. *Ingestible* devices are swallowed and pass through the digestive tract in a day or two. Sections 1.7 to 1.9 look at each in turn.

What the device measures sets its *transducer*, the part that turns one kind of signal into another. The sensors in this book fall into four groups:

- *Electrical* sensors measure voltages that cells produce (biopotentials such as the ECG, EMG, and EEG) or the impedance of tissue.
- *Optical* sensors shine light into tissue and measure what comes back: the photoplethysmogram (PPG), pulse oximetry, and fluorescence.
- *Electrochemical* sensors turn the concentration of a molecule (glucose, lactate, sodium, cortisol) into a current or a voltage at an electrode.
- *Physical* sensors respond to temperature, strain, pressure, or motion, usually through a resistance or a capacitance.

Optical and electrical sensors are already in consumer products for heart rate and rhythm; chemical sensing on the skin is still mostly in research, apart from the continuous glucose monitor. Whatever the transducer, the signal then passes through the same chain: an analog front end that amplifies and filters it, an analog-to-digital converter (ADC), a processor, a radio, and a power source, all packaged so the body tolerates them (Section 1.11).

No single discipline builds such a device: materials scientists, chemists, electrical and mechanical engineers, biologists, clinicians, and data scientists all contribute. This book is written for readers from electrical engineering and from the life sciences alike, so each idea from circuits, chemistry, or biology is explained from the start the first time it is used.

1.4 A Short History

In the 1780s, in Bologna, Luigi Galvani hung frog legs by brass hooks and touched the nerves with metal. When a circuit of two different metals connected the nerve and the muscle, the legs twitched. Galvani concluded that the body holds its own "animal electricity," which the metals merely released, and published the work in 1791. Alessandro Volta disagreed. He argued that the electricity came from the contact of two dissimilar metals, with the moist tissue simply closing the circuit: "metallic electricity." To prove it, Volta stacked disks of zinc and copper separated by brine-soaked cloth and in 1800 described the first battery, the voltaic pile.

Both men were partly right, and their argument contains the two halves of this book (Fig. 1.3). Nerves and muscles do carry electrical signals: every cell holds a voltage of tens of millivolts across its

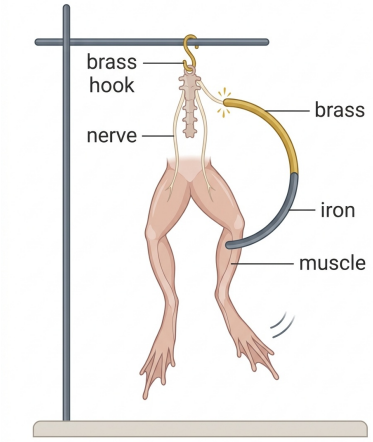


Figure 1.3: Galvani's experiment. A frog leg hangs from a brass hook; an arc of two metals touches nerve and muscle, and the leg contracts. Galvani credited the animal, Volta the two metals; both were partly right.

membrane, and excitable cells fire brief pulses called action potentials (Chapter 3). That is Galvani's animal electricity, and it is what an ECG or EEG electrode records. Two different metals in a salt solution do form a cell with a voltage between them. That is Volta's metallic electricity, and it is the electrode-electrolyte interface that every biosensor and every stimulating electrode must manage (Chapter 6). Galvani's twitching leg was also the first electrical *stimulation* of tissue.

Instruments followed the science. Willem Einthoven's string galvanometer recorded a clean human electrocardiogram in the early 1900s, and he named the waves P, Q, R, S, and T that we still use. Hans Berger recorded the first human electroencephalogram (EEG) in 1924. The electronics that makes such signals easy to handle came later: John Bardeen and Walter Brattain demonstrated the point-contact transistor at Bell Laboratories in 1947, and the integrated circuit followed in 1958 and 1959. Sixty years of steady work then took the industry from that single transistor to chips with billions of them.

Implants grew alongside. The first fully implanted cardiac pacemaker went into a patient in Stockholm in 1958, before integrated circuits were available to build one. In 1957 Djourno and Eyriès evoked sound sensations by stimulating the auditory nerve of a deaf patient, and the United States Food and Drug Administration (FDA) approved the first cochlear implant in 1984; roughly 740,000 had been placed worldwide by the end of 2019.[3] Deep brain stimulation followed in 1987.

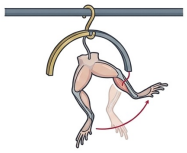
[3] US National Institute on Deafness and Other Communication Disorders.

Chemistry and light joined in. In 1962 Leland Clark and Champ Lyons trapped an enzyme at an oxygen electrode to measure glucose, the ancestor of every glucose meter (Chapter 6). In 1974 Takuo Aoyagi showed that the pulsing part of the light that passes through tissue gives the oxygen saturation of arterial blood, the principle of the pulse oximeter (Chapter 7). In 2000 a swallowed capsule with a camera filmed the small intestine from the inside.[4]

[4] Enzyme electrode: L. C. Clark and C. Lyons, *Ann. N.Y. Acad. Sci.* (1962). Capsule endoscope: G. Iddan *et al.*, *Nature* (2000).

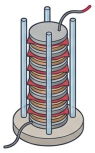
The newest part of the story is about mechanics and materials (Fig. 1.4). In 1994 a transistor made entirely of polymers was printed, showing that electronics need not be rigid crystals. In 2011 John Rogers's group laminated circuits thinner than the outer layer of skin onto the body like a temporary tattoo, and in 2012 built silicon transistors that dissolve harmlessly in body fluid after their job is done. In 2013 George Malliaras's group laid organic electrochemical transistors on the surface of the brain; because each transistor amplifies right where the signal is picked up, they recorded brain activity with a better signal-to-noise ratio than plain electrodes (Chapter 11). In 2014 a pulse oximeter was built entirely from organic light-emitting diodes and photodiodes (Section 1.6), and in 2015 Charles Lieber's group injected a mesh of nanoscale electronics into the brain through

1791



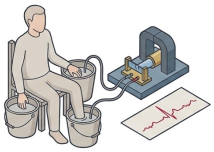
Animal electricity
Galvani makes frog legs twitch with an arc of two metals.
L. Galvani, De viribus electricitatis (1791)

1800



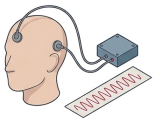
The first battery
Volta stacks zinc, copper, and wet cloth into a pile.
A. Volta, Phil. Trans. R. Soc. (1800)

1903



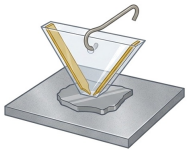
The electrocardiogram
Einthoven's string galvanometer records the heartbeat.
W. Einthoven, Pflügers Arch. (1903)

1924



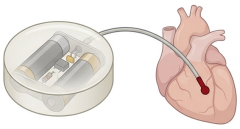
The human EEG
Berger records brain waves through the scalp.
H. Berger, Arch. Psychiatr. Nervenkr. (1929)

1947



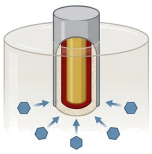
The transistor
Bardeen and Brattain amplify with point contacts on germanium.
J. Bardeen and W. H. Brattain, Phys. Rev. (1948)

1958



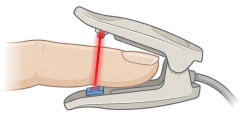
Implanted pacemaker
Elmqvist and Senning place a pacemaker fully inside the body.
R. Elmqvist and Å. Senning, Med. Electron. (1960)

1962



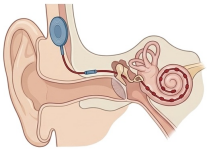
The enzyme electrode
Clark and Lyons trap an enzyme at an electrode to read glucose.
L. C. Clark and C. Lyons, Ann. N.Y. Acad. Sci. (1962)

1974



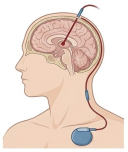
Pulse oximetry
Aoyagi reads blood oxygen from the pulsing part of the light.
T. Aoyagi et al., Jpn. Soc. Med. Electron. Biol. Eng. (1974)

1984



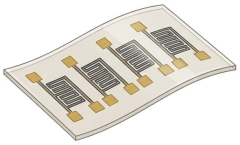
Cochlear implant
The FDA approves the first implant that restores hearing.
US Food and Drug Administration (1984)

1987



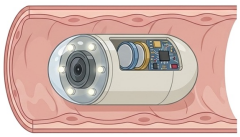
Deep brain stimulation
Benabid stops tremor with pulses deep in the brain.
A.-L. Benabid et al., Appl. Neurophysiol. (1987)

1994



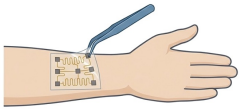
Printed transistor
Garnier prints a transistor made entirely of polymers.
F. Garnier et al., Science (1994)

2000



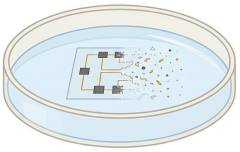
Capsule endoscope
A swallowed camera films the small intestine.
G. Iddan et al., Nature (2000)

2011



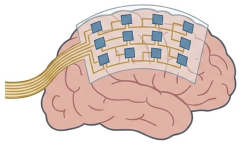
Epidermal electronics
Rogers laminates circuits as thin and soft as the skin's outer layer.
D.-H. Kim et al., Science (2011)

2012



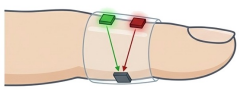
Transient electronics
Rogers builds silicon circuits that dissolve in the body.
S.-W. Hwang et al., Science (2012)

2013



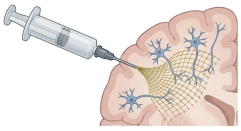
Organic transistors on the brain
Malliaras records brain activity with OECTs on the cortex.
D. Khodagholy et al., Nat. Commun. (2013)

2014



All-organic pulse oximeter
Khan and Arias read pulse and oxygen with organic LEDs and photodiodes.
C. M. Lochner, Y. Khan, et al., Nat. Commun. (2014)

2015



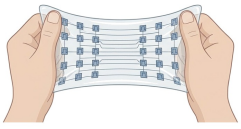
Injectable electronics
Lieber delivers an electronic mesh to the brain through a needle.
J. Liu et al., Nat. Nanotechnol. (2015)

2016



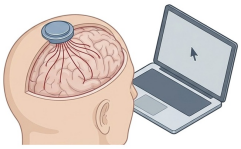
Sweat sensor array
Gao and Javey measure four sweat analytes at once on a wristband.
W. Gao et al., Nature (2016)

2018



Stretchable transistors
Bao makes arrays of rubbery transistors that work at twice their length.
S. Wang et al., Nature (2018)

2024



Wireless brain implant
A person moves a cursor with Neuralink's 1024-electrode implant.
E. Musk and Neuralink, J. Med. Internet Res. (2019)

Figure 1.4: Two centuries of bioelectronics. Each panel gives the year, the advance, and a key reference. The field moved from discovering bioelectricity, to measuring it, to electronics small enough to implant and swallow, and now to electronics as soft as tissue. The drawings are schematic.

a syringe needle.[5]

Complete systems followed. In 2016 Wei Gao, Ali Javey, and coworkers put sensors for glucose, lactate, sodium, and potassium, a skin thermometer, and a flexible circuit board into one wristband that analyzes sweat during exercise (Section 1.7). In 2018 Zhenan Bao's group made arrays of transistors from rubbery polymers, 347 to the square centimeter, that keep working when stretched to twice their length.[6] In January 2024 the first person received Neuralink's wireless brain implant, which records from 1024 electrodes on polymer threads (Section 1.12). The next section explains why softness matters so much.

[5] Printed polymer transistor: F. Garnier *et al.*, *Science* (1994). Epidermal electronics: D.-H. Kim *et al.*, *Science* (2011). Transient silicon: S.-W. Hwang *et al.*, *Science* (2012). Transistors on the brain: D. Khodagholy *et al.*, *Nat. Commun.* (2013). Injectable mesh: J. Liu *et al.*, *Nat. Nanotechnol.* (2015).

[6] Sweat wristband: W. Gao, S. Emaminejad, *et al.*, *Nature* (2016). Stretchable transistor array: S. Wang, J. Xu, *et al.*, *Nature* (2018).

1.5 *Why Bioelectronics Must Be Soft*

Skin, brain, and heart are soft, curved, and moving; silicon chips and circuit boards are stiff and flat. The mismatch hurts in three ways. A stiff device touches a curved surface at only a few points, so the electrical or optical contact is poor and changes whenever the person moves. Motion makes the stiff device slide relative to the tissue, which adds artifacts to the signal. And inside the body, a stiff implant rubs against soft tissue with every heartbeat and breath, which inflames it and grows scar tissue around the device (Section 1.8). A rigid board is also bulky and uncomfortable, which limits how long a person will wear it.

Young's modulus

The stiffness of a material is its *Young's modulus* E, the ratio of stress (force per area) to the strain (fractional stretch) it causes, $E = \sigma/\varepsilon$, in pascals ($1\,\text{Pa} = 1\,\text{N/m}^2$). Table 1.1 lists typical values. The range is enormous: brain tissue has a modulus of about 1 to 2 kPa, silicon 130 to 190 GPa, a factor of about 10^8. Skin, at roughly 0.1 to 1 MPa, is a hundred to a thousand times stiffer than brain but still five orders of magnitude softer than silicon.

You cannot make silicon as soft as brain, and you do not need to. What a tissue feels is not the modulus of the material but the *stiffness of the structure*, and that depends very strongly on thickness.

Material	E
Brain tissue	1 to 2 kPa
Heart, muscle	8 to 20 kPa
Hydrogels	1 to 1000 kPa
Skin (in plane)	0.1 to 1 MPa
PDMS	1 to 3 MPa
Polyimide	2.5 GPa
Parylene C	about 3 GPa
FR-4 board	about 20 GPa
Gold	79 GPa
Platinum	168 GPa
Silicon	130 to 190 GPa

Table 1.1: Typical Young's moduli, from tissue to the hard materials of electronics. Tissue values depend on how they are measured; treat them as orders of magnitude.

Bending stiffness and the t^3 law

Bend a thin plate of thickness t into an arc. The material on the outside stretches, the material on the inside compresses, and a *neutral plane* in the middle does neither. The bending moment needed per unit width,

KEY EQUATION

$$D = \frac{Et^3}{12(1-\nu^2)}$$

Bending stiffness per unit width. Halve the thickness and the film is eight times easier to bend.

divided by the curvature $1/R$, is the *bending stiffness* of the plate:

$$D = \frac{E\,t^3}{12\,(1-\nu^2)}, \tag{1.1}$$

where ν is Poisson's ratio (about 0.3 for most solids and 0.5 for rubber). The cube is the important part: a film ten times thinner is a thousand times easier to bend. Chapter 11 shows where Eq. (1.1) comes from and treats films of several layers; here we use it. The surface of a film bent to radius R is strained by

$$\varepsilon_{\text{surface}} = \frac{t}{2R} \tag{1.2}$$

when the neutral plane sits in the middle of the film, so thin films also break less easily. Silicon cracks at a strain of about 1 %, which a 100 µm chip reaches at $R = 5$ mm but a 1 µm membrane only at $R = 50$ µm. That is why thinned silicon can be flexible.

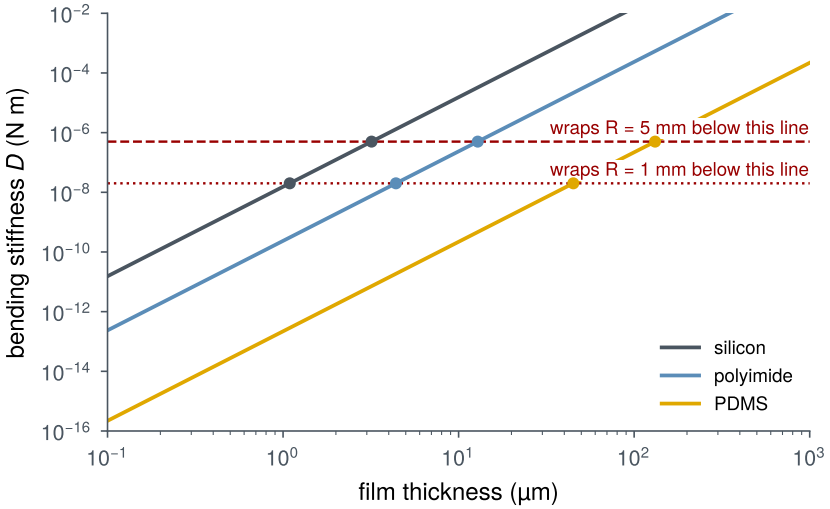


Figure 1.5: Bending stiffness per unit width versus thickness. Each tenfold drop in thickness lowers D a thousandfold. The red lines mark the largest stiffness that still wraps a body surface of radius 5 mm (dashed) or 1 mm (dotted) for an assumed work of adhesion of 10 mJ/m^2 (Eq. 1.3); the dots are the thickest films that conform.

Worked Example 1.1 *Bending stiffness of neural probes and circuit boards*

Compare the bending stiffness per unit width of (a) a silicon neural probe 15 µm thick ($E = 170$ GPa, $\nu = 0.28$), (b) a polyimide probe of the same thickness ($E = 2.5$ GPa, $\nu = 0.34$), (c) a polyimide film 5 µm thick, and (d) a standard 1.6 mm FR-4 circuit board ($E = 20$ GPa, $\nu = 0.12$).

SOLUTION. Equation (1.1) for each:

$$D_{\mathrm{Si}} = \frac{(170\times10^{9}\,\mathrm{Pa})(15\times10^{-6}\,\mathrm{m})^{3}}{12(1-0.28^{2})} = 5.19\times10^{-5}\,\mathrm{N\,m},$$

$$D_{\mathrm{PI},15} = \frac{(2.5\times10^{9}\,\mathrm{Pa})(15\times10^{-6}\,\mathrm{m})^{3}}{12(1-0.34^{2})} = 7.95\times10^{-7}\,\mathrm{N\,m},$$

$$D_{\mathrm{PI},5} = 7.95\times10^{-7}\times(5/15)^{3} = 2.94\times10^{-8}\,\mathrm{N\,m},$$

$$D_{\text{FR-4}} = \frac{(20\times10^{9}\,\mathrm{Pa})(1.6\times10^{-3}\,\mathrm{m})^{3}}{12(1-0.12^{2})} = 6.93\,\mathrm{N\,m}.$$

At equal thickness the polyimide probe is 65 times more flexible than the silicon one, close to the ratio of the moduli. Thinning the polyimide from 15 to 5 µm buys another factor of $3^{3} = 27$, so the thin polyimide film is about 1,800 times more flexible than the silicon probe. The circuit board is 2.4×10^{8} times stiffer than the 5 µm film. Thickness matters as much as material: a silicon membrane thinned to 15 µm/4 = 3.75 µm would match the 15 µm polyimide probe, because $65^{1/3} \approx 4$. ■

When does a film wrap the body?

A thin film laid on a curved surface sticks on its own if the energy it gains by adhering exceeds the energy it must store to bend. Bending a plate to radius R stores $D/(2R^{2})$ per unit area; sticking releases the *work of adhesion* γ per unit area, of order 10 mJ/m² for a wet film on tissue (an order-of-magnitude value that depends strongly on the surfaces). The film wraps when

$$\gamma \geq \frac{D}{2R^{2}} \quad\Longleftrightarrow\quad t \leq t_{\max} = \left[\frac{24\,(1-\nu^{2})\,\gamma R^{2}}{E}\right]^{1/3}. \tag{1.3}$$

For a fingertip (R = 5 mm) and γ = 10 mJ/m², the thickest films that conform are about 13 µm of polyimide, 3 µm of silicon, and 130 µm of PDMS (the dots of Fig. 1.5). The thickest polyimide film then sees a surface strain of only 0.13 %; a 1.6 mm board bent to the same radius would need 16 %, so it cannot follow the finger at all. Because $t_{\max}$ grows only as $\gamma^{1/3}$, a tenfold error in the adhesion changes it by about a factor of two. The same energy argument guided the design of silk-supported electrode arrays that wrap the folds of the brain,[7] where thin films cut into open meshes conformed better than thick, continuous sheets.

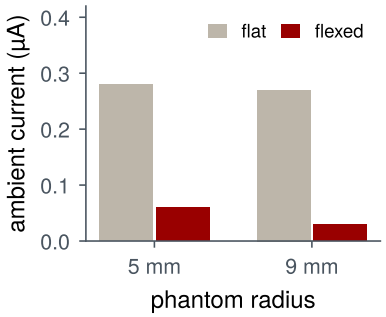


Figure 1.6: Ambient light reaching a flat and a flexed organic photodiode on finger phantoms of two radii, read from a published bar chart (Lochner, Khan, *et al.*, 2014).

Contact and signal quality

Soft sensors give cleaner signals, and the reason is contact. A rigid board rests on a few high points of the skin and leaves air gaps; a thin

[7] D.-H. Kim *et al.*, *Nat. Mater.* (2010).

film follows the surface. An optical sensor that does not sit flush lets room light leak in beside it, and the leak changes with every movement. In a direct test, an organic photodiode flexed around finger phantoms collected 5 to 9 times less ambient light than the same photodiode held flat (Fig. 1.6). An electrode's impedance scales roughly with the inverse of the area that truly touches the skin, so an electrode resting on a few high points has a large impedance that changes as the person moves, and the amplifier then picks up hum from power lines (Section 1.10 shows why).

1.6 Biosignals at a Glance

Clinicians call four quantities the *vital signs*: heart rate, blood pressure, breathing rate, and body temperature. Blood oxygen saturation is often added as a fifth. They are cheap to measure and change early when something goes wrong, and wearable devices already exist for each of them. Beyond the vital signs, the body produces electrical signals from heart, muscle, brain, and eyes, and an optical pulse in every tissue with a blood supply.

The first two questions about any biosignal are *how big* and *how fast*. Table 1.2 collects the answers. Amplitudes run from a few microvolts (EEG) to a few millivolts (ECG, EMG). The *bandwidth*, the range of frequencies a signal contains, runs from a fraction of a hertz (breathing, temperature) to several kilohertz (action potentials). These two numbers set the gain of the amplifier, the filters, the sampling rate, and the resolution of the converter (Section 1.10).

Signal (site)	Amplitude	Bandwidth	Sampled at
ECG (skin)	0.5 to 4 mV	0.05 to 150 Hz	250 to 1000 Hz
EEG (scalp)	5 to 300 µV	0.5 to 100 Hz	250 to 500 Hz
EMG (skin)	10 µV to 5 mV	20 to 500 Hz	1 to 2 kHz
EOG (beside the eyes)	50 µV to 3.5 mV	DC to 50 Hz	100 to 250 Hz
ECoG (brain surface)	10 µV to a few mV	0.5 to 500 Hz	1 to 2 kHz
Spikes (in the brain)	50 to 500 µV	0.3 to 7 kHz	20 to 30 kHz
PPG (optical)	AC/DC 0.1 to 2 %	0.5 to 10 Hz	25 to 400 Hz
Breathing		0.1 to 1 Hz	10 to 50 Hz
Temperature	35 to 38 °C (core)	below 0.01 Hz	0.1 to 1 Hz

Table 1.2: Amplitudes and bandwidths of common biosignals, with the sampling rates practical systems use (a few times the highest frequency of interest, Section 1.10). ECoG is the electrocorticogram on the surface of the brain; it and the spikes of single neurons return in Chapter 9.

Electrical signals: ECG, EMG, and EEG

Every cell in the body keeps a voltage of tens of millivolts across its membrane, inside negative (Chapter 3 explains how). Heart, muscle, and nerve cells can reverse that voltage briefly, an *action potential*, and pass the pulse to their neighbors. When many cells fire together, their currents flow through the surrounding tissue, which conducts like a

salt solution, and set up voltage differences that reach the skin. A pair of electrodes measures the difference between two points of that field.

The electrocardiogram. The ECG records the electrical activity of the heart. Each beat starts at the heart's natural pacemaker in the right atrium (the sinoatrial node), spreads across the two atria (the *P wave*), pauses at the junction between atria and ventricles, and sweeps through the ventricles (the *QRS complex*); the ventricles then recover (the *T wave*). Figure 1.7 shows one beat. Heart rate comes from the time between R peaks: peaks 0.83 s apart mean $60/0.83 = 72$ beats per minute. The shape and timing of the waves reveal problems in how the heart conducts, and an abnormal rhythm, an *arrhythmia*, shows in the spacing of the beats. The most common sustained one, *atrial fibrillation*, appears as missing P waves and irregular R-R intervals, and it raises the risk of stroke.

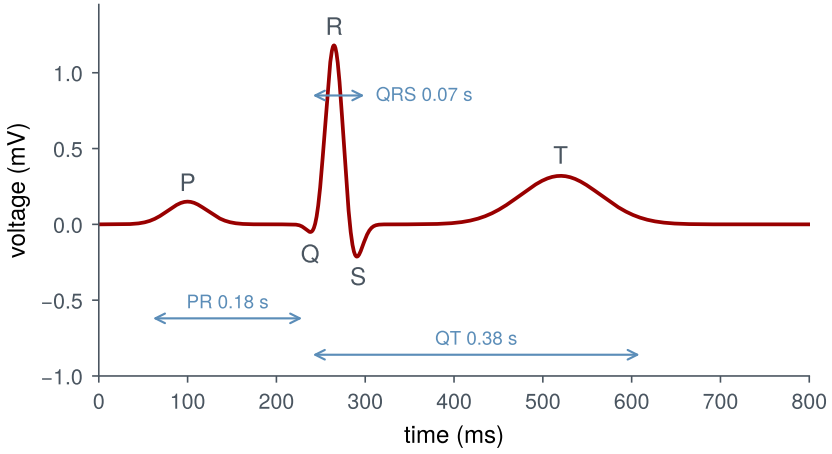


Figure 1.7: One synthetic ECG beat (illustrative, not a recording) with the waves and intervals clinicians read. Normal adult values: PR 0.12 to 0.20 s, QRS under 0.12 s, and a QT that shortens as heart rate rises.

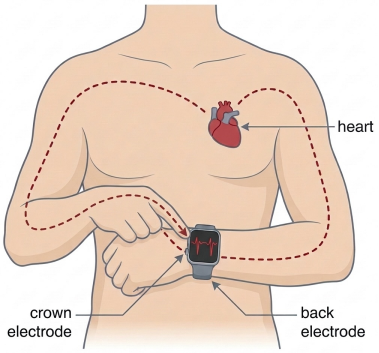


Figure 1.8: A smartwatch ECG. The back electrode touches the left wrist; a fingertip of the right hand touches the crown electrode. The loop runs from one arm through the chest to the other arm, so the heart lies between the two electrodes.

Hospital ECGs use *wet* electrodes, patches with an electrolyte gel that gives a low, stable contact impedance; watches use *dry* metal electrodes, good enough for heart rate and rhythm. A clinical recording uses twelve *leads* (voltages between chosen points); most wearables record one. The smartwatch of the chapter opening records the voltage between the two arms (Fig. 1.8). The arms act as wires, and the heart must lie between the electrodes: two contacts on one wrist would see almost no ECG, which is why the second hand has to touch the crown. Chapter 8 develops the leads and their amplifiers.

Electromyography. The EMG records the electrical activity of skeletal muscle. A muscle is driven by *motor units*, each a nerve fiber and the muscle fibers it controls. A motor unit fires only 7 to 20 times per second, but each of its events is sharp, so the surface EMG spans

about 20 to 500 Hz, and the sum of many units looks like noise whose envelope tracks the force (Fig. 1.9b). Amplitudes run from tens of microvolts at rest to several millivolts in a strong contraction. The layer of fat under the skin is a poor conductor that moves the electrode away from the muscle and shrinks the signal. Wearable EMG serves gait analysis, rehabilitation, and gesture recognition, for example to control a prosthetic hand.

Electroencephalography. The EEG records the summed activity of millions of neurons in the cortex through electrodes on the scalp. Skull and scalp spread and weaken the signals to 5 to 300 µV, and eye movements, blinks, jaw muscles, and body motion contaminate them. EEG power is described in bands: delta (0.5 to 4 Hz, deep sleep), theta (4 to 8 Hz), alpha (8 to 13 Hz, awake and relaxed with eyes closed), beta (13 to 30 Hz, alert), and gamma (above 30 Hz). The related *electrooculogram* (EOG) comes from the eye, whose front is positive relative to its back; electrodes beside the eyes read gaze direction from it.

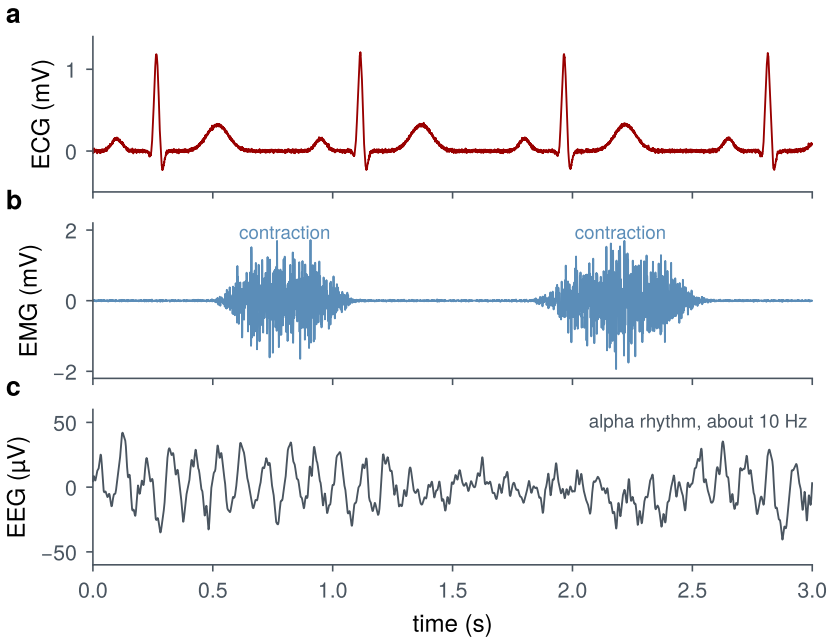


Figure 1.9: Three biopotentials over the same 3 s (synthetic illustrations). (a) ECG, about 1 mV, with a beat every 0.85 s. (b) Surface EMG, bursts of band-limited activity during two contractions. (c) Scalp EEG, tens of microvolts, with a 10 Hz alpha rhythm.

The three signals overlap in frequency. The EMG of a clenched jaw shares the upper EEG band and is up to a thousand times larger, so it can swamp an EEG; the ECG often appears in EMG recorded on the trunk. A filter cannot separate signals that share a band, so electrode placement matters as much as the electronics.

Temperature and breathing

Internal body temperature normally lies between about 35 and 38 °C, depending on where it is measured, and follows a daily rhythm with a swing of about 0.5 °C. Wearables measure skin temperature as a proxy. Skin is several degrees cooler than the core and much more variable: exercise, fever, meals, and the room all change it, and a sensor that loses contact reads the air. A *thermistor*, a resistor whose resistance falls by about 4 % per degree near skin temperature, read with a voltage divider (Section 1.10), resolves 0.1 °C easily; Chapter 5 treats temperature sensors.

Adults at rest take about 12 to 20 breaths per minute. Breathing rate and the *tidal volume* (the volume of each breath) track lung diseases such as asthma, which affect hundreds of millions of people. A *chest strap* measures the expansion of the rib cage with a stretchable strain sensor; a *microphone* on the chest hears the airflow, and wheezing sounds different from normal breathing; a *humidity sensor* under the nose sees each humid exhalation (Example 1.4); and *accelerometers* on the chest feel the motion of breathing and coughing.

The optical pulse

Shine light into a fingertip and measure what comes out: the detected light dips slightly with every heartbeat, as a pulse of arterial blood swells the vessels and absorbs a little more. That trace is the *photoplethysmogram* (PPG). It has a large, steady *DC* part from light absorbed by tissue, bone, and blood that does not pulse, and a small *AC* part, typically 0.1 to 2 % of the DC, from the pulsing arterial blood. Timing its peaks gives heart rate. Clinical finger and earlobe probes work in *transmission*, with the light source and the detector on opposite sides of the tissue; watches and forehead sensors work in *reflectance*, with the two side by side, collecting light that scatters back.

Oxygenated and deoxygenated hemoglobin absorb light differently. In the red, near 660 nm, deoxygenated hemoglobin absorbs about ten times more; in the near infrared, past about 800 nm, oxygenated hemoglobin absorbs more. A *pulse oximeter* records the PPG at two wavelengths, divides the pulse at each by its own DC part, and compares the two. Their ratio changes with the fraction of hemoglobin that carries oxygen, and a calibration turns it into the oxygen saturation, SpO_2, normally 95 to 100 %. Chapter 7 derives the method from the Beer-Lambert law.

Rigid finger clips are uncomfortable for long use, which motivates flexible, printed oximeters (Fig. 1.10). Organic light-emitting diodes are easy to print in red and green but not in the near infrared. The two forms of hemoglobin absorb green light almost equally, so a green

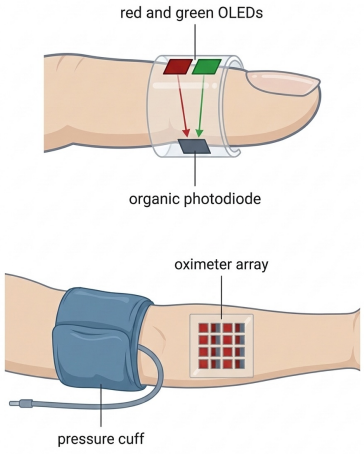


Figure 1.10: Printed organic oximeters. Top: a transmission probe of red and green organic LEDs and an organic photodiode on plastic, wrapped around a fingertip. Bottom: a reflectance array of red and near-infrared pixels laminated on the forearm; a pressure cuff on the arm stops the blood flow, and the array maps the fall in oxygenation.

pulse can serve as the reference that the infrared pulse usually provides, and a red and green pair can replace the usual red and infrared pair. An all-organic sensor built on that idea measured pulse rate and oxygenation with errors of about 1 and 2 % against a commercial probe, and a later reflectance array laminated on the forearm mapped the change in oxygenation in two dimensions.[8]

[8] C. M. Lochner, Y. Khan, A. Pierre, and A. C. Arias, *Nat. Commun.* (2014); Y. Khan, D. Han, *et al.*, *Proc. Natl. Acad. Sci. USA* (2018).

Blood pressure and pulse wave velocity

Blood pressure is reported in millimeters of mercury (1 mmHg = 133.3 Pa) as two numbers: the *systolic* pressure at the peak of each beat and the *diastolic* pressure between beats, with values below 120 over 80 mmHg counted as normal. Sustained high pressure (*hypertension*, 130/80 mmHg or above in current US guidelines) is a leading risk factor for stroke and heart attack. A cuff gives one reading every few minutes at best; cuffless sensors follow the pulse beat by beat. The pressure pulse travels along an artery at the *pulse wave velocity* (PWV), faster in a stiffer artery. Two PPG sensors 30 cm apart on one arm whose pulses arrive 40 ms apart give PWV = 0.30 m / 0.040 s = 7.5 m/s. A PWV between the neck and the groin above about 10 m/s marks a stiff aorta, a risk factor for heart attack and stroke in its own right.

1.7 Chemical Sensing on the Body

Blood carries nearly everything a physician wants to know, but getting it means a needle. Every other body fluid is a compromise between access and information (Table 1.3). *Interstitial fluid* (ISF), which bathes the cells and exchanges with the blood through the walls of the capillaries, tracks blood closely for small molecules such as glucose, but it lies under the outer layer of the skin. *Sweat* reaches the surface on its own and is rich in ions and small molecules, but its composition depends on how fast it flows. *Saliva* is easy to collect but changes with eating and drinking. *Breath* needs no contact at all, but its trace gases sit at parts per million or less.

Every chemical sensor has two parts. A *recognition element* binds or reacts with the target and ignores the rest: an ion-selective membrane for sodium or potassium, an *enzyme* (a protein that speeds one reaction) for glucose or lactate, or an antibody or an *aptamer* (a short, folded strand of DNA or RNA) for a protein or a hormone. A *transducer* turns that event into an electrical signal: a voltage across a membrane (*potentiometry*), a current from a reaction at an electrode (*amperometry*), a change in the current of a transistor, or a change in light. Chapter 2

Fluid	Access	Typical targets	Relation to blood
Blood	needle or finger stick	everything; the reference	reference
ISF	microneedle, filament under the skin	glucose, lactate, alcohol, drugs	small molecules track blood with a lag of minutes
Sweat	skin surface; stimulated or exercise	Na^+, Cl^-, K^+, lactate, urea, cortisol	ions depend on sweat rate; glucose about 1/100 of blood
Saliva	mouthguard, swab	lactate, glucose, cortisol	strong but imperfect correlation
Breath	mask, nose clip	acetone, isoprene, NO, H_2O_2	gases equilibrate with blood in the lungs

Table 1.3: Body fluids for chemical sensing. Moving down the table, access gets easier and the link to blood gets weaker.

introduces the biomolecules, Chapter 4 the measures of a good sensor, and Chapter 6 the electrochemistry; here we look at what each fluid offers.

Sweat

A sweat gland is a coiled tube in the dermis with a duct that opens on the skin (Fig. 1.11). The coil secretes a fluid close to blood plasma in salt content; as it rises through the duct, cells in the duct wall take back sodium and chloride, so the sweat that reaches the surface is much more dilute than plasma. Chloride leaves the duct through a channel protein called CFTR. Two consequences follow. When sweat flows fast, the duct has less time to take the salt back, so sweat sodium and chloride rise with sweat rate. And when CFTR is defective, as in the genetic disease *cystic fibrosis* (CF), chloride stays in the sweat. The sweat chloride test is the standard test for CF: 60 mM or more in stimulated sweat is consistent with CF, and 29 mM or less makes it unlikely.

Sweat also carries metabolites such as lactate and urea, excreted drugs and alcohol, and lower levels of hormones and proteins.[9] Larger molecules pass poorly: glucose is present at only about a hundredth of its blood level, and cortisol, a stress hormone, at nanomolar levels. Lactate in sweat is made largely by the gland itself and is a poor proxy for blood lactate. Because a sensor works with a few microliters, it should sit directly on the skin so that fresh sweat reaches it before it evaporates.

[9] M. Bariya, H. Y. Y. Nyein, and A. Javey, *Nat. Electron.* (2018), review wearable sweat sensors and sweat composition.

Making the skin sweat. Waiting for exercise to make sweat is unreliable, especially for an infant being tested for CF. Sweat can instead be stimulated by *iontophoresis*: a small current drives a sweat-inducing drug from a gel into the skin (Fig. 1.11). The usual drug, *pilocarpine*, mimics the nerve signal that tells sweat glands to secrete. It is positively charged, so it sits under the *anode*, the positive electrode, and the

current pushes it into the skin; the water that flows with the current from anode to cathode drags more along. A few minutes of stimulation make the glands under the gel sweat for tens of minutes.

The dose follows from counting charge. A current I for a time t moves a charge $Q = It$ in coulombs (Section 1.10); dividing by the *Faraday constant* $F = 96\,485\,\mathrm{C/mol}$, the charge of a mole of electrons, gives the moles of singly charged ions moved. Only a fraction t_d of the current, the drug's *transport number*, is carried by drug ions; the rest is carried by small, mobile ions such as Na^+ and Cl^-. The moles of drug delivered are

$$n_{\mathrm{drug}} = \frac{t_d\, I\, t}{z\, F}, \tag{1.4}$$

with z the charge number of the drug ion. Comfort and the risk of irritation limit the current density to about $0.5\,\mathrm{mA/cm^2}$.

KEY EQUATION

$n_{\mathrm{drug}} = t_d It/(zF)$

Iontophoretic dose: only the fraction t_d of the charge is carried by the drug.

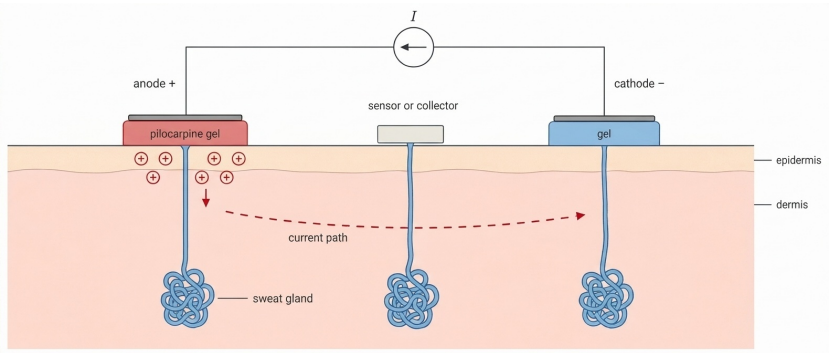


Figure 1.11: Iontophoresis for sweat stimulation. Current from the anode drives positively charged pilocarpine from the gel into the skin, where it reaches the sweat glands and makes them secrete. A sensor or collector between the electrodes picks up the sweat as it emerges.

Worked Example 1.2 *Stimulating sweat by iontophoresis*

A sweat test passes 1.5 mA for 5 min through a $3\,\mathrm{cm^2}$ pilocarpine gel. (a) What is the current density? (b) How much charge flows? (c) If 10 % of the current is carried by pilocarpine ($z = +1$, molar mass 208 g/mol), how much drug enters the skin? (d) If the gel and skin present about 10 kΩ, what voltage must the current source supply?

Solution. (a) $1.5\,\mathrm{mA}/3\,\mathrm{cm^2} = 0.50\,\mathrm{mA/cm^2}$, at the comfort limit.

(b) $Q = It = (1.5 \times 10^{-3}\,\mathrm{A})(300\,\mathrm{s}) = 0.45\,\mathrm{C}$, or $Q/F = 4.66\,\mu\mathrm{mol}$ of elementary charge.

(c) Equation (1.4): $n = 0.10 \times 4.66\,\mu\mathrm{mol} = 0.466\,\mu\mathrm{mol}$, which is $0.466 \times 10^{-6}\,\mathrm{mol} \times 208\,\mathrm{g/mol} = 97\,\mu\mathrm{g}$. The transport number is the uncertain input, because small, mobile ions carry most of the current.

(d) By Ohm's law, $V = IR = (1.5\,\mathrm{mA})(10\,\mathrm{k\Omega}) = 15\,\mathrm{V}$. A wearable must make this from a small battery with a step-up converter,

a circuit that raises the battery's voltage. It sets the current, not the voltage, for two reasons: the current fixes the dose, and skin resistance falls as the skin hydrates, so a fixed voltage would let the current, and with it the risk of irritation, rise. ■

Sweat sensors. Most sweat sensors are electrochemical. Ions are measured with an ion-selective membrane, whose voltage changes by about 59 mV for each tenfold change in concentration at room temperature (the Nernst slope of Chapter 3, which Chapter 6 applies to electrodes). Metabolites such as glucose and lactate are measured with an enzyme that converts the target and produces a current at an electrode. A landmark wristband integrated sodium, potassium, glucose, and lactate sensors with a skin thermometer and a flexible circuit board, so that each reading could be corrected for temperature on the spot.[10] Microfluidic patches collect the sweat in a channel and measure the sweat rate from how fast it fills (Fig. 1.12).

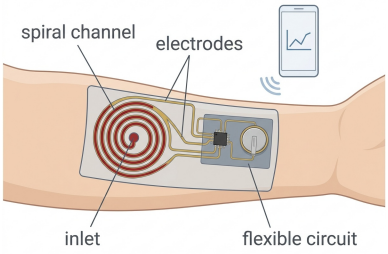


Figure 1.12: A wearable sweat patch: a spiral microfluidic channel collects sweat from an inlet on the skin; electrodes in the channel measure sweat rate and sodium, and a small flexible circuit board reads them out and transmits the results to a phone.

[10] W. Gao *et al.*, *Nature* (2016).

Four routes to chemical sensing. Four lines of work show the range of ideas (Fig. 1.13). Joseph Wang's group printed electrodes on temporary-tattoo paper and transferred them to the skin. With the enzyme lactate oxidase on the working electrode, the tattoo followed the lactate in sweat during exercise, flexing with the skin as it did so.[11] Jason Heikenfeld's group asked what sweat can and cannot report. An analyte must cross from the blood through the interstitial fluid and the wall of the sweat gland, so small molecules that pass through membranes, such as ethanol, arrive nearly undiluted, while large molecules such as proteins arrive diluted by orders of magnitude. A patch that stimulates sweat locally and reads it continuously followed the alcohol level in blood, a few minutes behind.[12]

[11] W. Jia, A. J. Bandodkar, *et al.*, *Anal. Chem.* (2013); J. Kim, A. S. Campbell, B. E.-F. de Ávila, and J. Wang, *Nat. Biotechnol.* (2019), review wearable biosensors.

[12] J. Heikenfeld *et al.*, *Nat. Biotechnol.* (2019); A. Hauke *et al.*, *Lab Chip* (2018).

Enzymes exist for only a few targets. For the rest, such as hormones, drugs, and neurotransmitters, the recognition element can be an *aptamer*, a short strand of DNA or RNA selected to fold around one target molecule. Kevin Plaxco's group tethers an aptamer to a gold electrode and puts a redox label at its free end. When the target binds, the strand folds, the label moves relative to the electrode, and the current changes within seconds; when the target leaves, the change reverses. Placed in a vein of an awake rat, such sensors followed drug levels in the blood for hours.[13] Anne Andrews's group couples aptamers to a thin-film transistor instead. When the target binds, the charged backbone of the aptamer rearranges close to the channel and changes its current, which let the transistor detect small molecules such as serotonin and dopamine in undiluted, salty fluids, and later cortisol in sweat with a sensor worn like a watch.[14] Chap-

[13] N. Arroyo-Currás, J. Somerson, P. A. Vieira, K. L. Ploense, T. E. Kippin, and K. W. Plaxco, *Proc. Natl. Acad. Sci. USA* (2017).

[14] N. Nakatsuka *et al.*, *Science* (2018); B. Wang, C. Zhao, *et al.*, *Sci. Adv.* (2022).

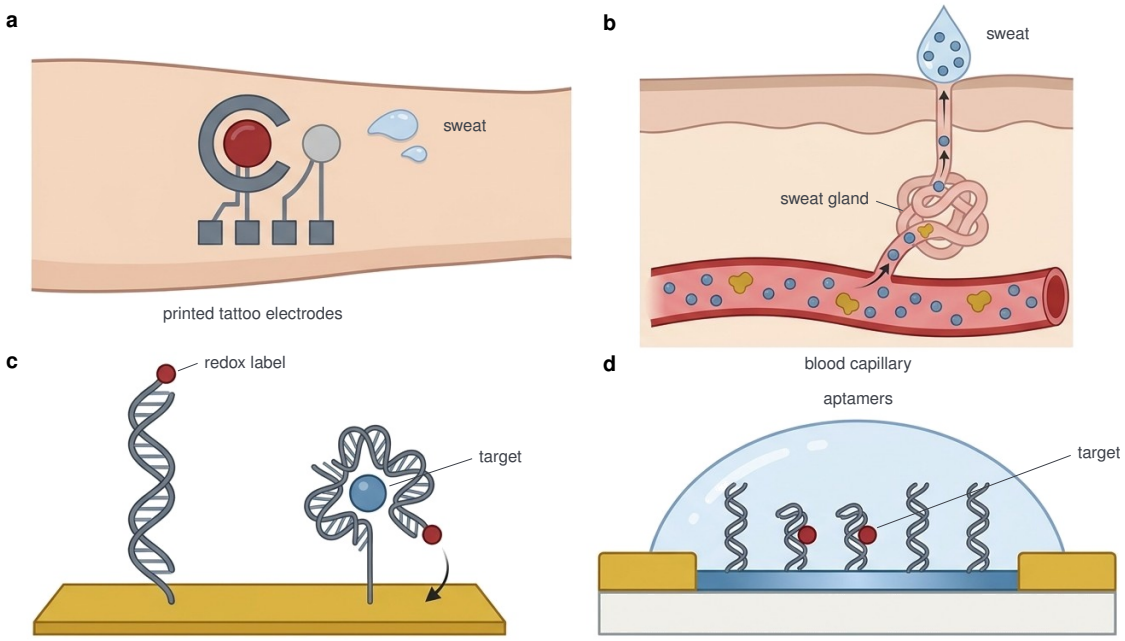


Figure 1.13: Four routes to chemical sensing on the body, drawn after the work of four groups. (a) Tattoo electrodes with an enzyme read lactate in sweat (Wang). (b) Small molecules pass from the blood into sweat; large ones mostly do not (Heikenfeld). (c) An aptamer on gold folds around its target and moves a redox label (Plaxco). (d) An aptamer on a transistor moves its own charge near the channel (Andrews).

ter 6 returns to enzymes, membranes, and aptamers, and Chapter 5 to the transistor as a sensor.

Interstitial fluid, saliva, and breath

Microneedle arrays reach the interstitial fluid in the outer skin without reaching most nerves or blood vessels. Needles a fraction of a millimeter long pierce the outer barrier layer of the skin, and electrochemical sensors on or inside the needles measure the fluid in place. A wearable microneedle patch with its own electronics measured glucose, lactate, and alcohol at once.[15] The *continuous glucose monitor* reaches the same fluid through a single flexible filament inserted under the skin for one to two weeks; it is the most widely used wearable chemical sensor (Chapter 6).

Saliva is a noninvasive window on metabolism: a mouthguard with an enzyme electrode can follow salivary lactate during exercise, and salivary glucose correlates with, but does not equal, blood glucose. *Breath* carries hundreds of volatile organic compounds at parts per trillion to parts per million: acetone rises in diabetic ketosis, and nitric oxide rises with the airway inflammation of asthma. The ideal gas law turns these fractions into concentrations: one part per million of acetone in breath at 37 °C is only about 39 nmol/L, in air saturated with water vapor. A breath sensor's selectivity and its correction for humidity, not its sensitivity alone, decide whether it works.

[15] F. Tehrani *et al.*, *Nat. Biomed. Eng.* (2022).

1.8 Implanted Electronics

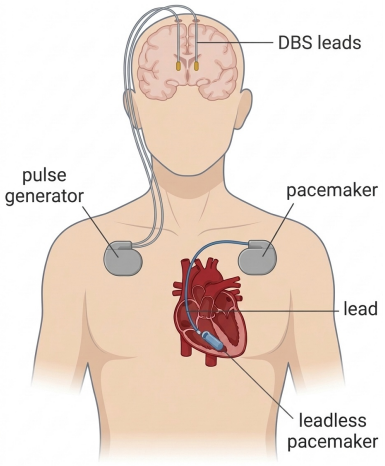


Figure 1.14: Three implants. A traditional pacemaker under the skin of the chest with a lead through a vein into the right ventricle; a leadless pacemaker, a capsule anchored inside the right ventricle; and a deep brain stimulator, with leads through small holes in the skull and a pulse generator in the chest.

Implanted electronic devices date back to the late 1950s, before integrated circuits existed (Fig. 1.4). The cardiac pacemaker came first, and it is still the model for every implant that followed (Fig. 1.14).

Pacemakers. The pacemaker of the chapter opening monitors the heart's rhythm and, when the heart pauses or beats too slowly, delivers a brief current pulse that starts a contraction: it stimulates *on demand*. Because the electrode touches the heart muscle, a pulse of a few volts lasting a fraction of a millisecond is enough. The device sits under the skin of the chest, and its leads run into the heart through a large vein. Its pulses and its sensing circuits together draw only about 10 µA on average, so a battery of about 1 Ah lasts about a decade, after which the whole generator is replaced (Chapter 13). *Leadless* pacemakers, about a tenth the size, sit entirely inside the right ventricle, anchored by small tines, and avoid the complications of the lead.

Neural implants. *Deep brain stimulation* (DBS) treats the symptoms of Parkinson's disease, essential tremor, and other movement disorders: electrodes pass through small holes in the skull to targets deep in the brain, and a pulse generator much like a pacemaker sits in the chest and sends 130 to 185 pulses per second. The *cochlear implant* stimulates the auditory nerve to restore hearing (Chapter 10). A *brain-machine* or *brain-computer interface* records the activity of neurons, decodes it, and uses it to drive a computer cursor, a robotic arm, or a paralyzed limb; its electrodes sit on the scalp, on the brain, or inside it, with larger and sharper signals the closer they get (Chapter 9).

The foreign body response

An implant lives in a hostile place: warm, salty, wet, moving, and guarded by an immune system built to destroy intruders. The tissue moves constantly. Heart muscle fibers change length by 10 to 20 percent with each beat, and at 75 beats per minute a pacemaker lead flexes about 40 million times a year, so *fatigue*, the growth of cracks under repeated loading, is a classic failure.

The body reacts to any implant in a predictable sequence (Fig. 1.15). Within seconds, proteins from blood and tissue fluid stick to the device surface and label it as foreign. Immune cells arrive over hours and days: first neutrophils, then monocytes that mature into *macrophages*, which attack the material with reactive oxygen species and enzymes.

Macrophages that cannot digest the device fuse into large *foreign body giant cells*. Over weeks, cells called fibroblasts lay down collagen and wrap the device in a *fibrous capsule*, scar tissue tens to hundreds of micrometers thick.

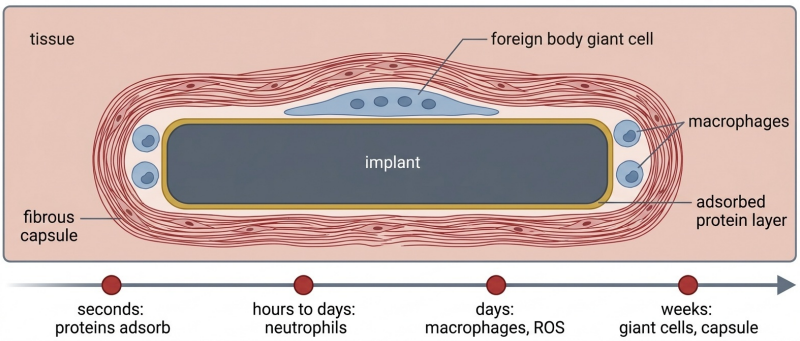


Figure 1.15: The foreign body response. Proteins adsorb on the implant within seconds; immune cells arrive over hours and days; macrophages that cannot remove the implant fuse into giant cells and release reactive oxygen species (ROS); and over weeks a collagen capsule walls it off. Not to scale.

For a sensor or an electrode, the capsule is the enemy. It is a layer of poorly conducting, poorly perfused tissue between the device and its target: it raises electrode impedance, slows the diffusion of glucose to an implanted sensor, and, in the brain, pushes neurons away from a recording site. Softer, smaller, and thinner devices that move with the tissue provoke a smaller response, and coatings that release anti-inflammatory drugs, as on the steroid-eluting pacemaker leads used since the 1980s, reduce it further.

Packaging. The electronics must also be protected from the body. Traditional implants use a *hermetic* package, a welded titanium can that keeps water vapor out for decades but is rigid and millimeters thick. Soft implants use *polymer* encapsulation (silicone, parylene, polyimide), which only slows water down (Chapter 11). Lifetime is tested by *accelerated aging*: the device soaks in warm saline, on the rule of thumb that the failure rate doubles for every 10 °C. Circuits that fail after 45 days at 87 °C, where failure comes $2^{(87-37)/10} = 32$ times faster than at body temperature, would last about 4 years in the body, provided the hot soak does not start failure modes that never occur at 37 °C.

Transient and bioresorbable electronics

Many implants are needed only for a while: a sensor that watches pressure in the skull for a week after a brain injury, or a stimulator that speeds nerve regrowth for a month. Removing them means a second surgery with its own risk of infection. *Transient* or *bioresorbable* electronics avoid it: every part is made of materials that dissolve in body fluid after the device has done its job, into products the body

absorbs or excretes. The materials that work are few. The conductors are magnesium (Mg) and zinc (Zn), which dissolve quickly, and iron (Fe), tungsten (W), and molybdenum (Mo), which dissolve slowly; the semiconductor is silicon, as membranes tens to hundreds of nanometers thick; the insulators are silicon dioxide and magnesium oxide; and the substrate is silk or a degradable polymer. They dissolve by reacting with water, for example

$$\mathrm{Mg} + 2\,\mathrm{H_2O} \longrightarrow \mathrm{Mg(OH)_2} + \mathrm{H_2}, \tag{1.5}$$

into ions and soluble species the body already handles; silicon becomes silicic acid, $\mathrm{Si(OH)_4}$, which the kidneys clear. Reported rates are 0.5 to 3 µm/h for Mg and Zn and 1×10^{-4} to 0.02 µm/h for W, Mo, and Fe, depending on the solution and the film;[16] silicon membranes dissolve at a few nanometers per day near neutral pH (Fig. 1.16). Dissolving devices have measured pressure and temperature inside the skull, recorded from the surface of the brain, and tracked blood flow around an artery after surgery, and then vanished.[17]

[16] R. Feiner and T. Dvir, *Nat. Rev. Mater.* (2017).

[17] Transient silicon: S.-W. Hwang *et al.*, *Science* (2012). Skull: S.-K. Kang *et al.*, *Nature* (2016). Brain surface: K. J. Yu *et al.*, *Nat. Mater.* (2016). Artery: C. M. Boutry *et al.*, *Nat. Biomed. Eng.* (2019).

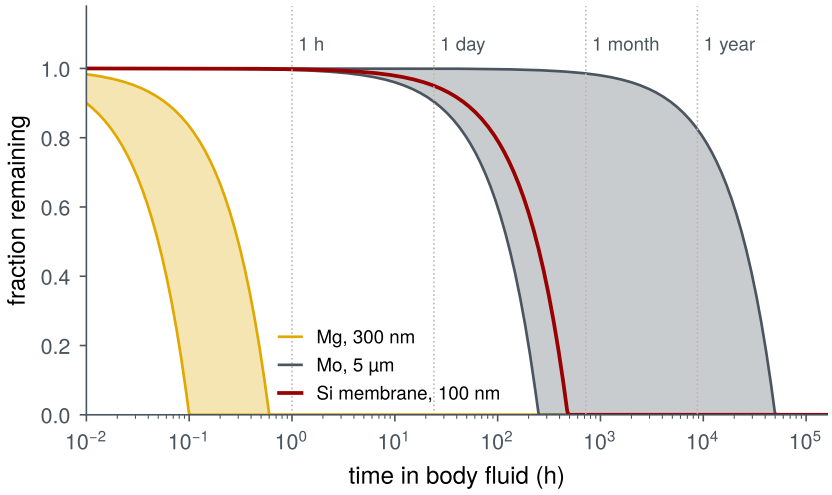


Figure 1.16: Fraction of a conductor remaining versus time in body fluid, without encapsulation. Bands span the quoted rate ranges: a 300 nm Mg trace is gone within an hour; a 5 µm Mo trace lasts from about ten days to several years; a 100 nm silicon membrane at an assumed 5 nm/day lasts about three weeks. Encapsulation adds a delay before dissolution starts.

1.9 *Ingestible Electronics*

A SWALLOWED CAPSULE VISITS places no wearable can reach and no surgeon wants to open. The device that made ingestibles routine is the wireless *capsule endoscope*, a camera in a pill about 11 mm by 26 mm with LEDs, a lens, an image sensor, batteries, and a radio. It photographs the small intestine at about two frames per second for roughly eight hours while a recorder on the patient's belt receives the images.[18] Sensing capsules followed (Fig. 1.17). They measure pH,

[18] Capsule endoscope: G. Iddan, G. Meron, A. Glukhovsky, and P. Swain, *Nature* (2000). Gas sensing: K. Kalantar-Zadeh *et al.*, *Nat. Electron.* (2018). Bacterial-electronic: M. Mimee *et al.*, *Science* (2018).

pressure, and temperature to follow how contents move through the gut, or gases that reflect the activity of gut bacteria; one carries engineered bacteria that glow in the presence of blood.

The digestive tract is a long tube whose chemistry changes along its length (Table 1.4). The stomach is strongly acidic; the small intestine is near neutral and grows slightly alkaline toward its end; the colon is slightly acidic at first, where bacteria ferment food into acids, and close to neutral afterward. A capsule's pH record shows these changes as landmarks (Fig. 1.18): a rise of more than three pH units when it leaves the stomach, and a drop of about one unit when it passes from the small intestine into the colon. A fall in temperature marks its exit from the body. In Fig. 1.18 the pH jumps at 3.2 h, drops at 8.0 h, and the temperature falls at 30 h: the stomach emptied in 3.2 h, the small intestine took $8.0 - 3.2 = 4.8$ h, and the colon $30 - 8.0 = 22$ h, all within the normal ranges. Doctors use such records to diagnose *motility* disorders, such as a stomach that empties too slowly.

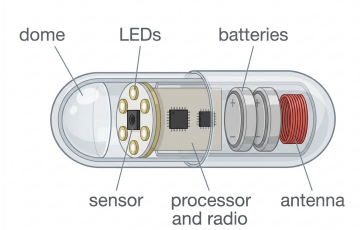


Figure 1.17: Inside an ingestible capsule: a transparent dome and LEDs at one end, a sensor (camera, optical, or pH), a circuit board with the processor and radio, button-cell batteries, and an antenna, in a sealed biocompatible shell.

Region	pH	Capsule transit time	Notes
Mouth, esophagus	6 to 7	seconds	
Stomach	1 to 3.5	2 to 5 h (normal under 5 h)	acid, enzymes, grinding
Small intestine	6 to 7.5	2.5 to 6 h	bile, enzymes; most absorption
Colon	5.5 to 7	10 to 59 h	bacteria, gases (H_2, CO_2, CH_4)
Whole gut		10 to 73 h	

Table 1.4: The gastrointestinal tract as an environment for a capsule. Transit times are typical normal ranges for a capsule that is not digested; they vary widely from person to person.

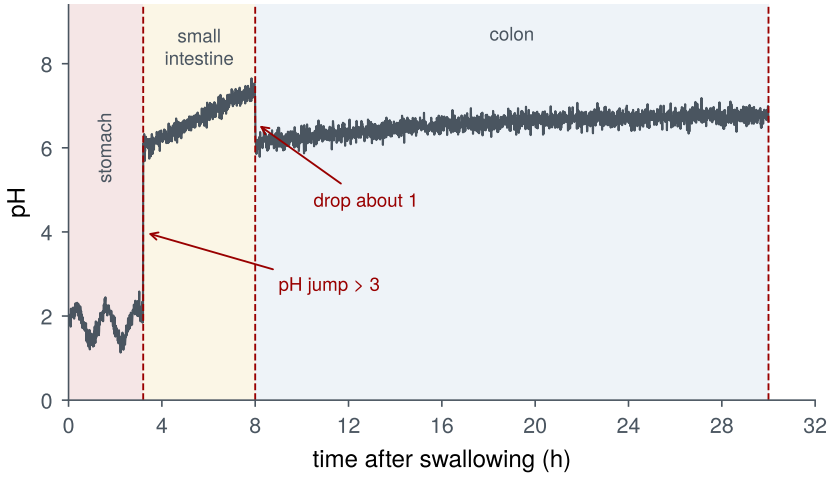


Figure 1.18: An illustrative capsule pH record (synthetic). The pH jumps when the capsule leaves the stomach, rises slowly along the small intestine, drops at the junction with the colon, and recovers slowly in the colon. The dashed lines mark the landmarks used to compute transit times; the last, at 30 h, is the exit from the body, read from the temperature record (not shown).

From cameras to laboratories. Research capsules now carry chemical sensors, living cells, and drugs (Fig. 1.19). Kourosh Kalantar-Zadeh's group built a capsule with gas sensors behind a membrane that lets

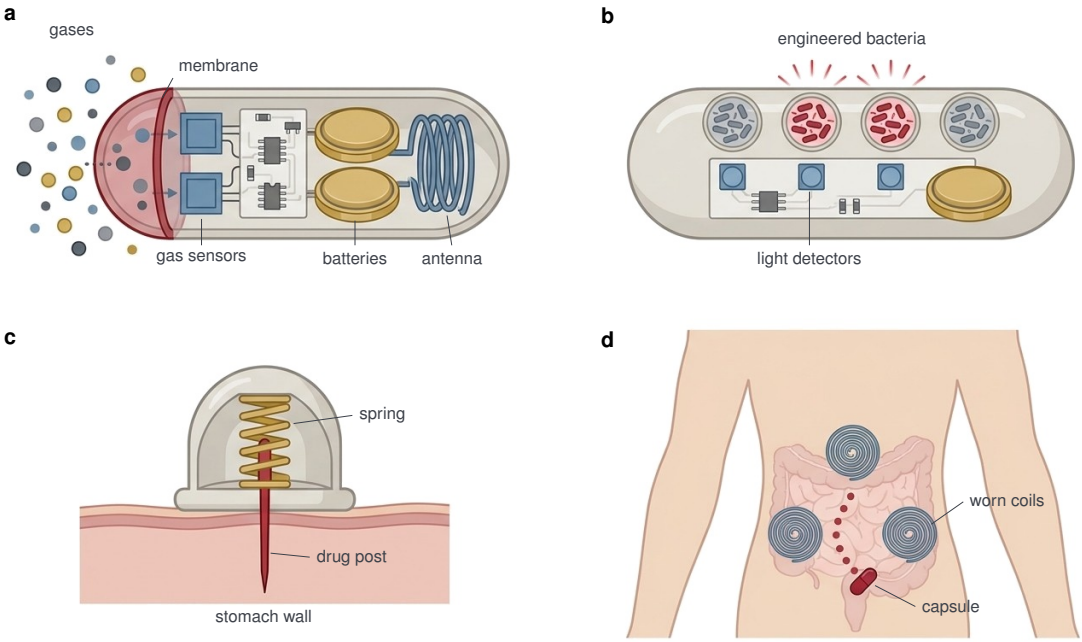


Figure 1.19: Ingestible devices that sense and treat, drawn after the work of four groups. (a) Gut gases pass a membrane and reach sensors inside (Kalantar-Zadeh). (b) Engineered bacteria glow when they meet blood, and detectors under them read the glow (Traverso and Lu). (c) A capsule rights itself on the stomach wall and injects a post of drug (Abramson and Traverso). (d) A coil worn on a shirt gives each gas reading a position (Khan).

gas through and holds liquid back. In a pilot trial in people it reported oxygen, hydrogen, and carbon dioxide along the gut, and the oxygen record showed where the capsule was. Giovanni Traverso, Timothy Lu, and coworkers paired living sensors with electronics: engineered bacteria in the capsule glow when they meet heme from blood, photodetectors under them read the glow, and a radio reports the result; the capsule detected bleeding in the stomachs of pigs. Capsules can also treat. Alex Abramson, Traverso, and coworkers shaped a capsule like the shell of a leopard tortoise, so that it rights itself on the stomach wall; a spring then pushes a tiny post of solid insulin into the tissue, which gave blood levels in pigs like those of an injection under the skin.[19] Where a reading was taken matters as much as its value. The capsule of Section 1.1 carries a magnetometer for that reason, and a related capsule from the same group maps ammonia and oxygen in three dimensions with a coil worn on a shirt.

Safety and power. A capsule must be small enough to swallow and to pass, sealed against stomach acid, and cool. Batteries are the usual power source, and they are small: a silver-oxide button cell 6.8 mm across holds about 28 mAh at 1.55 V. A capsule that senses and streams over Bluetooth continuously, drawing about 2.2 mA from such a cell, goes silent after about 13 hours, usually before it leaves the colon. Measuring for 10 s each minute and sleeping in between stretches the same cell to about 73 hours, enough for nearly every transit. Turning

[19] Self-orienting capsule: A. Abramson *et al.*, *Science* (2019). Reviews of the field: C. Steiger, A. Abramson, P. Nadeau, A. P. Chandrakasan, R. Langer, and G. Traverso, *Nat. Rev. Mater.* (2019); A. Abdigazy *et al.*, *Nat. Electron.* (2024).

circuits off when they are not needed, *duty cycling*, is the main power-saving tool of every battery-powered device (Chapter 13).

1.10 Circuit Essentials

EVERY DEVICE IN THIS CHAPTER reads its sensor through a circuit. This section collects the circuit ideas the rest of the book is built on: current and voltage, resistance and power, the voltage divider, the capacitor and its time constant, sinusoids and impedance, decibels, filters, and sampling. It is written for readers from biology and chemistry; electrical engineers can skim it. Later chapters add the operational amplifier (Chapter 4), the transistor (Chapter 5), and the amplifiers and filters of biopotential recording (Chapter 8).

Charge, current, and voltage

Electric *charge* is measured in coulombs (C). An electron carries $-e$ and a sodium ion $+e$, where $e = 1.602 \times 10^{-19}\,\mathrm{C}$; a mole of singly charged ions carries the Faraday constant, $F = 96\,485\,\mathrm{C/mol}$. *Current* is the rate at which charge flows past a point, in amperes:

$$I = \frac{\mathrm{d}q}{\mathrm{d}t}, \qquad 1\,\mathrm{A} = 1\,\mathrm{C/s}. \tag{1.6}$$

In a metal wire electrons carry the current; in body fluids and tissue, ions do. Where a wire meets the body, at an electrode, chemical reactions or charge stored at the surface hand the current from one kind of carrier to the other (Chapter 6).

Voltage, or potential difference, is the energy each unit of charge gains or loses between two points, in volts ($1\,\mathrm{V} = 1\,\mathrm{J/C}$). A voltage is always measured *between* two points. In a circuit one point is chosen as the reference and called *ground*, and other voltages are quoted against it. If current is like the flow of water in a pipe, voltage is like the pressure difference that drives it. Some typical values: a lithium cell holds about 3.7 V, a resting nerve cell about 70 mV across its membrane (Chapter 3), and the ECG between two points on the skin about 1 mV.

Resistance, Ohm's law, and power

In a resistor the current is proportional to the voltage across it:

$$V = IR, \tag{1.7}$$

which is *Ohm's law*, with the resistance R in ohms (Ω, volts per ampere). Its inverse, the conductance $G = 1/R$, is in siemens (S). A wire

or a printed trace of length L and cross section A, made of a material with conductivity σ, has the resistance

$$R = \frac{L}{\sigma A}. \tag{1.8}$$

A platinum trace (9.4×10^6 S/m) 10 cm long, 50 µm wide, and 1 µm thick has 212 Ω; the same trace in a conducting polymer at 1×10^5 S/m has 20 kΩ, which is why polymers serve as electrode coatings and metals as wires (Chapter 11).

A current through a resistor turns electrical energy into heat at the rate

$$P = VI = I^2R = \frac{V^2}{R}, \tag{1.9}$$

in watts (1 W = 1 J/s). For example, 1 mA through 1 kΩ drops 1 V and dissipates 1 mW. The same product VI gives the power a battery delivers. Battery capacity is quoted in milliampere-hours: a 150 mAh cell supplies 1.5 mA for 100 h, and at 3.7 V it stores 0.150 Ah × 3.7 V = 0.56 Wh, or 2.0 kJ. A device's battery life is its capacity divided by its average current (Section 1.11).

KEY EQUATION

$V = IR, \quad P = VI = I^2R$

Ohm's law and the power a resistor turns into heat.

Series, parallel, and the voltage divider

Two rules govern every circuit. Charge is neither created nor lost, so the currents flowing into any junction add up to the currents flowing out (*Kirchhoff's current law*). And the voltages around any closed loop add up to zero (*Kirchhoff's voltage law*), just as a walk that returns to its start climbs as much as it descends. From them follow the rules for combining resistors. Resistors in *series* carry the same current, and their resistances add, $R = R_1 + R_2$. Resistors in *parallel* share the same voltage, and their conductances add, so

$$R_1 \parallel R_2 = \frac{R_1 R_2}{R_1 + R_2}; \tag{1.10}$$

two equal resistors in parallel give half of either.

Two resistors in series across a source V_s form a *voltage divider* (Fig. 1.20). The same current, $V_s/(R_1 + R_2)$, flows through both, so the voltage across R_2 is

$$v_{\text{out}} = V_s \frac{R_2}{R_1 + R_2}. \tag{1.11}$$

The divider is the simplest way to read a resistive sensor. Make R_1 a thermistor, a strain gauge, or a humidity-sensitive film, and v_{out} follows its resistance. With $R_1 = R_2$ the output sits at $V_s/2$, and a small fractional change x in R_1 moves it by about $V_s x/4$: a 1 % change with a 3.3 V supply gives 8.3 mV. Chapter 5 improves on the divider with the Wheatstone bridge.

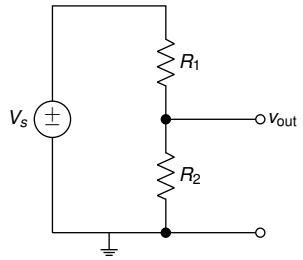


Figure 1.20: A voltage divider: the output is the fraction $R_2/(R_1 + R_2)$ of the source. Make R_1 a sensor and v_{out} follows its resistance.

The divider also describes *loading*. A real source always has some resistance of its own in series. An electrode on the skin, for example, has an impedance Z_e (think of it as a resistance for now), and the amplifier it feeds has an input impedance Z_{in}. The two form a divider, and the amplifier sees only the fraction $Z_{\text{in}}/(Z_{\text{in}} + Z_e)$ of the biopotential. A gelled electrode, at tens of kilohms (say 30 kΩ), into an amplifier of 10 MΩ delivers 99.7 % of the signal; a dry electrode resting on hair, at 1 MΩ, delivers 91 %. When the two electrodes of a pair differ like this, the hum that power lines induce on the body reaches the amplifier's two inputs in different amounts, and the difference looks like a signal: 10 mV of hum times $(0.997 - 0.909)$ leaves 0.9 mV of false signal, as large as the ECG. The cure is an amplifier whose input impedance is far larger than any electrode's, 1 GΩ or more, which receives 99.9 % of the signal through either electrode (Chapter 8).

The capacitor and the RC time constant

A *capacitor* is two conductors separated by an insulator. A voltage V across it stores a charge $+Q$ on one plate and $-Q$ on the other, in proportion:

$$Q = CV, \tag{1.12}$$

with the capacitance C in farads (F). For two parallel plates of area A a distance d apart, $C = \varepsilon_0 \varepsilon_r A/d$, where $\varepsilon_0 = 8.854 \times 10^{-12}$ F/m and ε_r is the relative permittivity of the insulator. Two electrodes 5 mm square separated by 100 µm of silicone rubber ($\varepsilon_r = 2.7$) make 6.0 pF; pressing them closer raises C, which is how the capacitive pressure sensors of Lab 2 work. Every cell membrane is a capacitor too, of about 1 µF per square centimeter (Chapter 2).

Because current is the rate of flow of charge, a capacitor passes current only while its voltage changes:

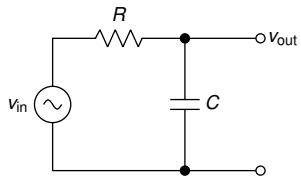
$$I = C\,\frac{\mathrm{d}V}{\mathrm{d}t}. \tag{1.13}$$

Two consequences follow. A steady (DC) voltage drives no current through a capacitor once it has charged, so a capacitor *blocks DC*. And its voltage cannot jump, because a jump would take an infinite current.

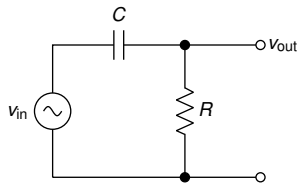
Connect a capacitor through a resistor to a source that steps from 0 to V (Fig. 1.21a). At first the capacitor holds no charge, the full step appears across R, and a current V/R flows. As the capacitor charges, its voltage rises, the voltage across R falls, and the current slows. The result is an exponential approach,

$$v_C(t) = V\left(1 - e^{-t/\tau}\right), \qquad \tau = RC, \tag{1.14}$$

with the *time constant* τ in seconds when R is in ohms and C in farads (Fig. 1.22a). After one time constant the capacitor has covered 63 % of



(a) low-pass



(b) high-pass

Figure 1.21: The two first-order RC filters. (a) Low-pass: the output is taken across the capacitor, which shorts high frequencies to the bottom rail. (b) High-pass: the series capacitor blocks DC and slow changes. Both have the corner frequency $f_c = 1/(2\pi RC)$.

the step, after three 95 %, and after five more than 99 %; a discharging capacitor decays as $e^{-t/\tau}$ in the same way. A 10 kΩ resistor and a 100 nF capacitor give $\tau = 1$ ms. The same exponential describes how a cell membrane charges (Chapter 3) and how many sensors respond to a sudden change at their input (Chapter 4).

Sinusoids, frequency, and impedance

Many signals are best described by the frequencies they contain. A sinusoid,

$$v(t) = A\sin(2\pi f t + \phi), \tag{1.15}$$

has an amplitude A, a frequency f in hertz (cycles per second), a period $T = 1/f$, and a phase ϕ that sets where the cycle starts; $\omega = 2\pi f$ is the angular frequency. Power lines run at 60 Hz in North America and 50 Hz in most of Europe and Asia, the alpha rhythm of the EEG near 10 Hz, a resting heart near 1.2 Hz. Any signal can be built from a sum of sinusoids (Fourier's theorem), and its *spectrum* says how much of each frequency it holds. Slow, smooth signals need only low frequencies; sharp events, such as the QRS complex or a neural spike, need high ones. The bandwidth of Table 1.2 is the range of frequencies that matters for each signal's use.

KEY EQUATION

$|Z_C| = \dfrac{1}{2\pi f C}$

A capacitor's impedance falls with frequency: open at DC, nearly a short at high frequency.

For sinusoids, the ratio of the voltage across a part to the current through it is its *impedance* Z, which generalizes resistance. A resistor has $Z_R = R$ at every frequency. A capacitor, by Eq. (1.13), passes more current the faster its voltage changes, so its impedance falls with frequency:

$$Z_C = \frac{1}{j2\pi f C}, \qquad |Z_C| = \frac{1}{2\pi f C}. \tag{1.16}$$

The j is the imaginary unit ($j^2 = -1$; engineers write j because i is taken by current). It records that the current through a capacitor peaks a quarter cycle before its voltage; for most of this book the magnitude is what matters. A 100 nF capacitor has an impedance of 26.5 kΩ at 60 Hz, 159 Ω at 10 kHz, and an infinite one at DC. Impedances combine by the same series, parallel, and divider rules as resistances, with complex arithmetic. A resistor in series with a capacitor, for example, has $Z = R + 1/(j2\pi f C)$: a real part R and an imaginary part whose size, $1/(2\pi f C)$, is the capacitor's *reactance*. The two voltages peak a quarter cycle apart, so their magnitudes add at right angles, like the sides of a right triangle: $|Z| = \sqrt{R^2 + (1/2\pi f C)^2}$. The electrode-skin interface, a cell membrane, and the tissue between two electrodes are all modeled by networks of resistors and capacitors (Chapters 2, 5, 6, and 8).

Decibels

Ratios in electronics span many orders of magnitude, from the gain of an amplifier to the attenuation of a filter, so engineers quote them on a logarithmic scale. A voltage ratio in *decibels* (dB) is

$$G_{\mathrm{dB}} = 20 \log_{10} \frac{V_2}{V_1}, \tag{1.17}$$

and a power ratio is $10 \log_{10}(P_2/P_1)$. The two agree because power goes as voltage squared. A few values are worth remembering: a factor of 10 in voltage is 20 dB, a factor of 2 is 6 dB, a factor of $1/\sqrt{2} = 0.707$ is -3 dB, a factor of 1000 is 60 dB, and a factor of $1/100$ is -40 dB. Decibels add when stages follow one another: an amplifier of 40 dB followed by a filter that passes -6 dB gives 34 dB overall. A signal-to-noise ratio (SNR) in decibels uses the same rule (Chapter 4).

RC filters

A *filter* passes the frequencies that carry information and removes the rest. The simplest is a resistor and a capacitor used as a voltage divider whose ratio depends on frequency (Fig. 1.21). In the *low-pass* filter the output is taken across the capacitor. At low frequency the capacitor's impedance is huge, it draws almost no current, and the output equals the input; at high frequency its impedance is small, and it shorts the output toward the bottom rail. The divider rule with Z_C gives the gain

$$|H_{\mathrm{LP}}| = \frac{1}{\sqrt{1 + (f/f_c)^2}}, \qquad f_c = \frac{1}{2\pi RC} = \frac{1}{2\pi\tau}, \tag{1.18}$$

where f_c is the *corner*, or cutoff, frequency. Below f_c the gain is close to 1 (0 dB); at f_c it is $1/\sqrt{2}$ (-3 dB); above it the gain falls in proportion to $1/f$, by 20 dB for each tenfold rise in frequency (Fig. 1.22b). The 10 kΩ and 100 nF pair has $f_c = 159$ Hz. Swap the two parts and take the output across the resistor, and you have a *high-pass* filter,

$$|H_{\mathrm{HP}}| = \frac{f/f_c}{\sqrt{1 + (f/f_c)^2}}, \tag{1.19}$$

with the same corner. It blocks DC completely, because the capacitor passes no steady current, and passes fast changes. That makes it the standard way to remove the steady offset of skin electrodes, up to a few hundred millivolts, before a biopotential is amplified; a series capacitor used this way is called *AC coupling*. Time and frequency are two views of the same circuit, tied by $f_c = 1/(2\pi\tau)$.

One RC stage makes a *first-order* filter, and 20 dB per decade is a gentle slope. Sharper filters, of second order (40 dB per decade) and higher, combine several stages or use op amps (Chapter 8).

KEY EQUATION

$f_c = \dfrac{1}{2\pi RC}$

Corner frequency of a first-order RC filter; the gain there is -3 dB.

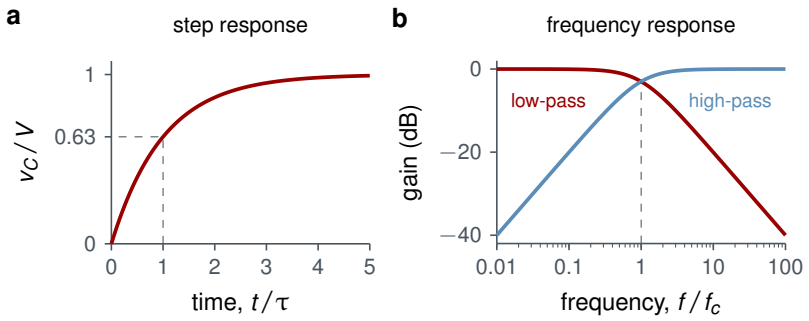


Figure 1.22: The RC filter in time and in frequency. (a) After a step of size V, the voltage on the capacitor rises as $1 - e^{-t/\tau}$, covering 63 % of the step in one time constant. (b) Gain of the low-pass (cardinal) and high-pass (steel) filters of Fig. 1.21. Both are -3 dB at the corner f_c (dashed) and change by 20 dB per decade of frequency far from it.

Worked Example 1.3 *RC filters for an ECG*

An ECG for heart-rate monitoring needs the band from about 0.5 to 150 Hz. (a) A high-pass filter uses $R = 1\,\text{M}\Omega$. Choose C for a corner at 0.5 Hz, and find the time constant. (b) A low-pass filter uses $R = 10\,\text{k}\Omega$. Choose C for a corner near 150 Hz, and find its gain in decibels at 60 Hz and at 1 kHz. (c) What do the two filters do to a 300 mV electrode offset and to 60 Hz hum?

Solution. (a) From Eq. (1.18), $C = 1/(2\pi f_c R) = 1/(2\pi \times 0.5\,\text{Hz} \times 1\,\text{M}\Omega) = 318\,\text{nF}$. The nearest standard value, 330 nF, gives $f_c = 0.48\,\text{Hz}$ and $\tau = RC = 0.33\,\text{s}$. After a sudden jump at the input, such as an electrode bump, the output settles back to its baseline in about $5\tau = 1.7\,\text{s}$.

(b) $C = 1/(2\pi \times 150\,\text{Hz} \times 10\,\text{k}\Omega) = 106\,\text{nF}$; the standard 100 nF gives $f_c = 159\,\text{Hz}$. At 60 Hz, $f/f_c = 0.377$ and $|H| = 1/\sqrt{1+0.142} = 0.936$, or -0.6 dB: the ECG band passes almost untouched. At 1 kHz, $f/f_c = 6.28$ and $|H| = 0.157$, or -16 dB: muscle noise and other fast interference fall by a factor of six.

(c) The offset is DC, so the high-pass filter removes it entirely once its capacitor has charged; the amplifier after it can then have a large gain without its output running into the supply. The hum is a different story. At 60 Hz the high-pass gain is 0.99997 and the low-pass gain 0.936, so the hum passes almost as well as the ECG, because it lies inside the ECG's band. No filter that keeps the ECG can remove it; the cure is a differential amplifier that rejects the voltage common to both electrodes (Chapter 8). ■

Sampling and ADC resolution

An analog-to-digital converter turns a voltage into numbers. It measures the input at regular instants, f_s times per second (the *sampling rate*), and rounds each measurement to one of 2^N levels, where N is its number of bits (Fig. 1.23). Each choice sets a limit.

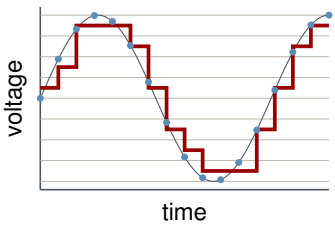


Figure 1.23: Sampling and quantization. The converter measures the signal (gray) at regular instants (dots) and rounds each value to one of 2^N levels, here $N = 3$ and 8 levels (one in each band between the faint lines), holding it until the next sample (staircase). One level spacing is one LSB.

Sampling rate. To capture a signal whose highest frequency is f_{max}, the sampling rate must exceed twice that frequency, the *Nyquist criterion*:

$$f_s > 2f_{max}. \tag{1.20}$$

A component above $f_s/2$ is not lost but disguised: it shows up at a false, lower frequency, an effect called *aliasing*. A 300 Hz component sampled at 400 Hz appears at $|300 - 400| = 100$ Hz, where nothing can tell it from a real 100 Hz signal. So an analog low-pass filter, the *anti-aliasing* filter, goes in front of every converter. Because a simple filter rolls off slowly, practical systems sample three to five times faster than f_{max}: an ECG with content to 150 Hz at about 500 Hz, and neural spikes with content to 7 kHz at 20 to 30 kHz (Table 1.2).

Resolution. An N-bit converter divides its full-scale range V_{FS} into 2^N steps. One step, the *least significant bit* (LSB), is the smallest change it can see, so to resolve a detail ΔV it needs

$$\text{LSB} = \frac{V_{FS}}{2^N} \leq \Delta V \quad \Longrightarrow \quad N \geq \log_2 \frac{V_{FS}}{\Delta V}. \tag{1.21}$$

KEY EQUATION

$N \geq \log_2(V_{FS}/\Delta V)$

Bits needed to resolve ΔV in a range V_{FS}.

A 10-bit converter with a 3.3 V range has an LSB of 3.22 mV; a 12-bit one, 0.81 mV. The full-scale range must hold everything at the input, not only the signal. Skin electrodes develop a steady offset of up to a few hundred millivolts, and ECG standards require amplifiers to tolerate ±300 mV. Digitizing an ECG directly, offset and all, in a range of 0.6 V with a resolution of 1 µV takes $\log_2(0.6/10^{-6}) = 19.2$, so 20 bits. Remove the offset with a high-pass filter and amplify the ECG 500 times first, and 5 µV at the electrodes becomes 2.5 mV at a 3.3 V converter: $\log_2(3.3/0.0025) = 10.4$, so 11 bits, and a 12-bit converter is the usual choice. Modern biopotential front ends take the first route with 24-bit converters, and then the amplifier's noise, not the LSB, limits the resolution (Chapters 4 and 8).

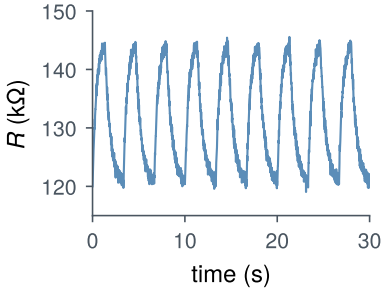


Figure 1.24: Resistance of a PEDOT:PSS humidity sensor under the nose (illustrative). Each exhalation wets the film and raises its resistance; nine breaths in 30 s.

WORKED EXAMPLE 1.4 *A breathing sensor in a divider*

A humidity-sensitive film of PEDOT:PSS, a conducting polymer that takes up water, sits under the nose. Its resistance rises from about 120 kΩ to about 144 kΩ with each humid exhalation (Fig. 1.24). It is the top resistor of a divider with 100 kΩ below it, powered from 3.3 V, and a 10-bit converter with a 3.3 V range reads the output. (a) Find the breathing rate. (b) Find the output swing and the number of codes it spans. (c) What sampling rate is sensible?

SOLUTION. (a) Nine breaths in 30 s is $9 \times 60/30 = 18$ breaths per minute.

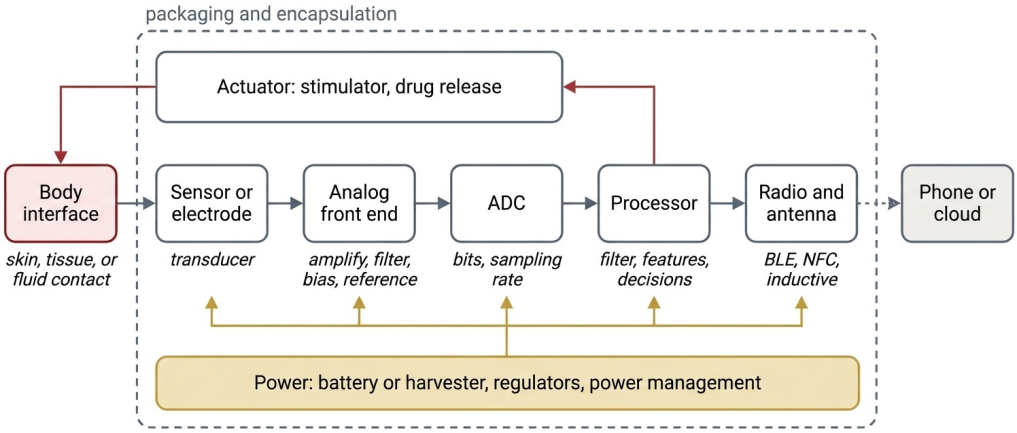


Figure 1.25: The anatomy of a bioelectronic system. The signal path runs left to right from the body to the phone; a closed-loop device returns through an actuator; power feeds every block; the package encloses everything except the body interface, which it must expose.

(b) Equation (1.11) with the sensor as R_1: $v_{\text{out}} = 3.3\,\text{V} \times 100/(100 + R_1)$, which is 1.500 V at 120 kΩ and 1.352 V at 144 kΩ, a swing of 148 mV. With an LSB of 3.22 mV that is $148/3.22 = 46$ codes, plenty for counting breaths. The divider current is at most $3.3\,\text{V}/220\,\text{k}\Omega = 15\,\mu\text{A}$, so the reading costs little power.

(c) Breathing stays below about 1 Hz (60 breaths per minute), so by Eq. (1.20) anything above 2 Hz captures the rate; 10 to 20 Hz keeps the shape of each breath. The film itself takes a large fraction of a second to dry, which is the real bandwidth limit, and an RC low-pass filter at a few hertz in front of the converter keeps electrical noise from aliasing into the band. ■

1.11 *The Anatomy of a Bioelectronic System*

Take any device in this chapter apart and you find the same blocks (Fig. 1.25). Each has its own specifications, its own technology, and its own chapters in this book.

The **body interface** is where device meets tissue: an electrode, an optical window, a microneedle, a fluid inlet. Its impedance, its contact, and its biocompatibility often limit the whole system (Sections 1.5 and 1.8). The **sensor** is the transducer, often printed or flexible because it must be soft and large (Chapters 5 to 7, 11, and 12). The **analog front end** (AFE) amplifies, filters, and references a microvolt or nanoampere signal for the converter (Chapters 4 and 8); biopotential front-end chips, such as the ones in Labs 3 and 4, put the amplifiers and the **ADC** on one chip, with the resolution and rate of Section 1.10. The **processor**, a microcontroller, extracts features (heart rate, the power in an EEG band) and decides. The **radio**, usually Bluetooth Low Energy (BLE) for wearables and an inductive link or a medical-band radio for

implants, often uses the most energy. **Power** comes from a battery or an energy harvester, through regulators (Chapter 13), and the **package** keeps fluids out and the device in place. A closed-loop device adds an **actuator**, a stimulator (Chapter 9) or a drug pump.

Printed or silicon? Flexible hybrid electronics

Not every block should be made the same way. Printed and organic electronics are soft, light, cheap over large areas, and can be made in shapes that fit the body; silicon integrated circuits are fast, dense, low-voltage, and stable for years. In round numbers, organic transistors have carrier mobilities (charge speed per unit field) near $1\,\mathrm{cm^2/(V\,s)}$ against about 1000 for silicon, switch at about a megahertz instead of gigahertz, need about 10 V instead of 1 V, and last months rather than a decade, but they can be printed over square meters of plastic. *Flexible hybrid electronics* use each where it fits: printed or flexible materials for sensors, interconnects, antennas, batteries, and packaging, and small, thinned silicon chips for amplification, conversion, processing, and radio.[20] Silicon wins wherever speed, low noise, and matched devices matter, because organic transistors are slow, vary from device to device, and drift. Printed materials win where the device touches or covers the body. The result is soft where it touches the body and silicon where it computes.

[20] Y. Khan, A. Thielens, S. Muin, J. Ting, C. Baumbauer, and A. C. Arias, *Adv. Mater.* (2020).

Data and power budgets

A wearable measures all day, every day, so its battery life, the volume and privacy of its data, and its false alarms dominate the design. Care at a distance, *telemedicine*, depends on such devices; during the COVID-19 pandemic it filled part of the gap when hospitals were full.

A budget starts from the data. The ADC produces

$$\text{bit rate} = N_{\text{ch}}\, f_s\, b \tag{1.22}$$

for N_{ch} channels sampled at f_s with b bits each, and the processor either sends it all or reduces it to features. Power follows block by block.[21] Each block draws a current I_i for a fraction D_i of the time (its *duty cycle*), so the average battery current and the battery life are

$$I_{\text{avg}} = \sum_i D_i I_i, \qquad t_{\text{life}} = \frac{C_{\text{bat}}}{I_{\text{avg}}}, \tag{1.23}$$

with any converters between the battery and the circuits adding their losses. In most wearable medical systems the optical sensors and the radio draw the most power. Processing on the device and sending only results saves both radio energy and data, as the next example shows; Chapter 13 builds full budgets.

[21] Y. Khan, A. E. Ostfeld, C. M. Lochner, A. Pierre, and A. C. Arias, *Adv. Mater.* (2016), review vital-sign wearables and the power of each block.

KEY EQUATION

$t_{\text{life}} = C_{\text{bat}} / \sum_i D_i I_i$

Battery life from the capacity and the duty-cycled currents of the blocks.

WORKED EXAMPLE 1.5 *A day of data and power for an ECG patch*

An ECG patch streams 2 channels at 250 samples per second and 24 bits. Its front-end chip draws 0.3 mW per channel for 3 channels at 3.3 V; its processor, radio, and regulator draw on average 0.8, 0.4, and 0.2 mA. Everything runs at 3.3 V through a linear regulator, which passes the same current it delivers, from a 3.7 V, 150 mAh battery. (a) Find the data rate and the data per day. (b) Find the average battery current and the battery life. (c) The patch instead sends its heart rate once per second as one byte, which cuts the radio's average current to 0.1 mA. Find the new data per day and battery life.

SOLUTION. (a) Equation (1.22): $2 \times 250 \times 24 = 12\,\text{kbit/s}$, or 1.5 kB/s; over a day of 86 400 s that is 130 MB.

(b) The front end draws $3 \times 0.3\,\text{mW} = 0.9\,\text{mW}$, which at 3.3 V is 0.27 mA. With the other blocks, $I_{\text{avg}} = 0.27 + 0.8 + 0.4 + 0.2 = 1.67\,\text{mA}$, and $t_{\text{life}} = 150\,\text{mAh}/1.67\,\text{mA} = 90\,\text{h}$, nearly four days. The battery delivers $1.67\,\text{mA} \times 3.7\,\text{V} = 6.2\,\text{mW}$; the 0.4 V that the regulator drops turns into heat.

(c) One byte per second is 86.4 kB per day, 1500 times less than the raw ECG. The current falls to 1.37 mA, and the life rises to 109 h, about 4.6 days; the processor now draws the most. Raw data are still needed for diagnosis and for training algorithms, so a practical patch also stores or sends short stretches of raw ECG when it detects an abnormal rhythm. ■

1.12 *Three Devices in Numbers*

SECTION 1.2 SET THREE DEVICES SIDE BY SIDE. With the circuit ideas of Section 1.10 and the blocks of Fig. 1.25 in hand, we can now put numbers on them (Fig. 1.2). For each device we name the part that fills each block and the number that sets its design. The three look nothing alike, yet they are built from the same blocks (Table 1.5) and designed by the same steps (Fig. 1.26).

A wearable: the wireless ECG patch

An ECG patch sticks to the chest and records the heart's electrical activity for days, so that a rhythm problem that comes and goes is caught when it happens (Fig. 1.2a). Its signal is the ECG of Table 1.2: 0.5 to 4 mV between electrodes a few centimeters apart, in a band from 0.05 to 150 Hz.

Body interface and sensor. Three electrodes touch the skin: two pick up the heart's voltage and the third holds the body near the circuit's reference (the right-leg drive of Chapter 8). Each is silver coated with silver chloride (Ag/AgCl), the usual material of skin electrodes, under a pad of hydrogel, a water-swollen polymer that wets the skin and lowers the contact impedance. The electrodes can also be printed, as in Lab 2.

Front end, converter, processor, radio. One chip amplifies, filters, and digitizes. An integrated biopotential front end such as the ADS1293 of Labs 3 and 4 holds three channels, a right-leg drive, and 24-bit converters, and draws about 0.3 mW per channel (Chapter 8). A second chip, a microcontroller with a built-in BLE radio, reads the samples over SPI, a serial link between chips, finds the heartbeats, and sends data to a phone. The budget of Example 1.5 is this patch: two channels at 250 samples per second make 12 kbit/s, and the blocks draw 1.67 mA on average.

Power and package. A 150 mAh lithium-ion cell then lasts nearly four days, or longer if the patch sends heart rate instead of raw ECG. A flexible circuit carries the chips, an adhesive holds the patch to the skin, and a soft cover seals the electronics against sweat and showers while the patch bends with every breath.

Printed sensors and silicon chips meet on one flexible board in such a patch. Khan and coworkers built an early example: ECG electrodes inkjet-printed in gold and a printed temperature sensor on 50 µm polyimide, joined by printed traces directly to the copper pads of the same board, which carried an ECG front end, a Bluetooth chip with its own microcontroller and converter, and a coin cell. Worn on the chest, it recorded the ECG and followed the heart rate from 62 to 108 beats per minute during mild exercise.[22] Commercial patches are cleared medical devices. One single-lead adhesive patch records continuously for up to 14 days; worn together with a 24-hour Holter monitor by 146 patients, it detected 96 arrhythmia events against 61 for the Holter.[23] The ECG option of the design project (Section 1.13) is a patch of the same kind: printed Ag/AgCl electrodes, an ADS1293 board, and BLE firmware.

[22] Y. Khan, M. Garg, Q. Gui, *et al.*, *Adv. Funct. Mater.* (2016). Inkjet-printed gold electrodes: Y. Khan, F. J. Pavinatto, *et al.*, *Adv. Funct. Mater.* (2016).

[23] The Zio patch: P. M. Barrett *et al.*, *Am. J. Med.* (2014).

An implant: a brain-computer interface

A brain-computer interface records the spikes of neurons in the motor cortex, the region of the brain that controls movement, and turns them into the motion of a computer cursor for a person who is paralyzed (Fig. 1.2b). Neuralink's N1 implant is one such device; the first person

received it in January 2024, in the company's PRIME clinical study.[24] Its signals are spikes of 50 to 500 µV with energy up to about 7 kHz (Table 1.2), recorded close to the neurons. The numbers below come from three sources, and the source matters: a 2019 paper on an earlier, wired research system, Neuralink's 2020 public description of the wireless implant, and its 2024 report on the study.[25]

[24] Neuralink, "PRIME study progress update" (April 2024).

[25] 2019: E. Musk and Neuralink, *J. Med. Internet Res.* (2019). 2020: Neuralink public update (August 2020).

Body interface and sensor. The N1 records through 1024 electrodes on 64 flexible polymer threads, 16 per thread on average (2024). In the 2019 system each polyimide thread was 4 to 6 µm thick and 5 to 50 µm wide and carried 32 electrodes. A thread that thin buckles instead of entering the brain, so a surgical robot pushes each one in on a fine needle, steering between blood vessels. In 2019 it inserted up to six threads a minute; at that rate 64 threads take about 11 minutes. Chapter 9 explains why soft threads damage less tissue than stiff needles.

Front end and converter. A custom chip amplifies and digitizes every electrode. The 2019 chip held 256 amplifiers with 5.9 µV of noise (root mean square), each drawing 5.2 µW, and sampled each channel at 19.3 kHz with 10 bits; the whole chip, converters included, drew about 6 mW. At those rates 1024 amplifiers draw about 5.3 mW, and the four chips that 1024 channels need about 24 mW.

Data and radio. Here the implant differs most from the patch. At the 2019 rates, 1024 channels produce $1024 \times 19.3\,\text{kHz} \times 10 = 198\,\text{Mbit/s}$, which the wired 2019 system sent through a cable. The wireless implant talks to a phone or computer over Bluetooth Low Energy at about a megabit per second after compression (2020), so the implant must cut its data about 200-fold before the radio, for example by detecting spikes and sending only when they occur (Chapter 9).

Power and package. A battery inside the implant is recharged inductively, through the skin, by an external charger (2024); in 2020 the company described a full day of use per charge. The sealed enclosure, 23 mm across and 8 mm thick, replaces a disk of skull and sits flush with it, so nothing crosses the skin (2020). It must keep body fluid out for years, and its electronics must not warm the brain by more than a degree or two (Section 1.8 and Chapter 9).

An ingestible: a sensing capsule

The third device is the capsule of Section 1.1 (Fig. 1.2c). Its numbers, from that section: an optical module and a magnetometer for sensors; a BLE microcontroller that sends a few readings a second through

tissue that costs about 8 dB/cm at 2.4 GHz; two silver-oxide cells of 28 mAh that last 16 h at 1.75 mA; and a resin shell 9.8 mm by 26 mm. The battery life is the trade-off of Section 1.9: long enough for the stomach and small intestine, not for the whole colon.

One architecture

Table 1.5 sets the three devices side by side, block by block. Every row is filled in every column: each device has a body interface, a sensor, a front end and converter, a processor, a radio, a power source, and a package. The patch and the capsule even use the same kind of part, a microcontroller with a BLE radio on the same chip.

The blocks are the same; the body site sets the constraints, and the constraints drive different choices for the same block.

Size. A patch can spread over several square centimeters of skin, so its battery and board are easy to fit. The implant must fit in a hole in the skull, and the capsule must be swallowed, so both pack everything into a volume of a few cubic centimeters or less.

Power. The patch is taken off and recharged. The implant cannot be, so its battery is recharged through the skin by an inductive link (Chapter 13). The capsule is used once and runs on cells that are not rechargeable, chosen for safety in the gut rather than for capacity. The patch spends about 6 mW and the capsule about 5 mW; at the 2019 chip's rates, the implant's front end alone would need about 24 mW.

How data leave. The patch's ECG is slow, and BLE carries all of it. The implant's raw data are about 16 000 times larger than the patch's, so it must reduce them in the body before the radio. The capsule sends little, but its signal must cross centimeters of tissue that absorb radio waves.

Lifetime and environment. The patch lasts days on dry, accessible skin, and its adhesive and cover must survive sweat. The implant must last years in warm salt water under an immune attack (Section 1.8). The capsule needs to last only a day or two, but in acid, enzymes, and churning muscle, and it must pass without getting stuck.

One design process

The same steps design all three (Fig. 1.26). Define the need and the signal: what must be measured, where, how large it is, and how fast it changes. Set specifications and budgets: noise, data rate, power, size, and lifetime. Choose the architecture block by block. Build the sensor,

Block	ECG patch (wearable)	Brain implant, N1 (implantable)	Sensing capsule (ingestible)
Body site	skin of the chest	motor cortex, under the skull	stomach and intestines
Body interface and sensor	3 hydrogel Ag/AgCl electrodes; ECG 0.5 to 4 mV	1024 electrodes on 64 polymer threads; spikes 50 to 500 µV	optical window; 4 LEDs and a photodiode; magnetometer
Front end and ADC	one chip (ADS1293): 3 channels, 24 bits, 0.3 mW per channel	custom chip (2019): 256 channels, 10 bits, 19.3 kHz; 5.2 µW per amplifier, 6 mW per chip	inside the optical sensor chip
Processor	BLE microcontroller	on the implant (not published in detail)	BLE microcontroller (nRF5340)
Radio or link	BLE to a phone	BLE to a phone or computer	BLE to a receiver outside the body
Data rate	12 kbit/s raw ECG	198 Mbit/s raw; about 1 Mbit/s sent	a few readings per second
Power source	3.7 V, 150 mAh lithium-ion cell	rechargeable battery, charged inductively	two SR626 silver-oxide cells, 3.1 V, 28 mAh
Average power	6.2 mW (Example 1.5)	not published; front end alone about 24 mW at 2019 rates	5.4 mW (1.75 mA)
Lifetime	nearly 4 days per charge; worn for days to 2 weeks	a day per charge; stays implanted	about 16 h, single use
Package	flexible circuit, adhesive, soft sealed cover	sealed enclosure flush with the skull	3D-printed resin shell
Size	a few centimeters across	23 mm across, 8 mm thick	9.8 mm by 26 mm

Table 1.5: One architecture, three body sites. The blocks of Fig. 1.25 filled in for the devices of Fig. 1.2. Patch numbers follow Example 1.5; implant numbers are from Neuralink's 2019 paper (front end), 2020 description (radio, size, charge), and 2024 report (electrodes, threads); capsule numbers are from Abdigazy *et al.* (2024).

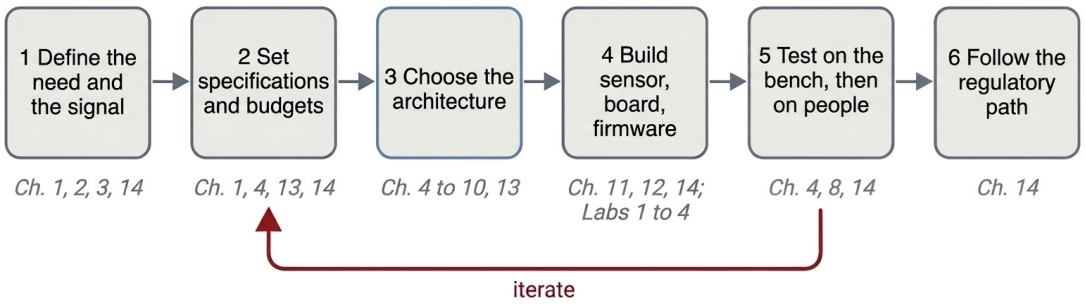
the board, and the firmware. Test on the bench, and then, safely, on people. Follow the regulatory path to patients. Each test sends the team back to an earlier step, so the process is a loop. The table in Fig. 1.26 shows how each of the three devices answers each step.

The process is the same because the questions are the same. A team designing a sweat sensor, a retinal implant, or a capsule that releases a drug walks through the same six steps; only the answers change.

1.13 *How This Book Is Organized*

EVERY CHAPTER OF THIS BOOK TEACHES one or more blocks of the system that the three devices share (Fig. 1.27). The body comes first: Chapters 2 and 3 teach the cells and neurons whose signals a device reads. Sensors and the interfaces to skin, fluid, and tissue fill Chapters 4 to 9; the electronics that amplify, convert, process, and send are built up in this chapter and in Chapters 4 and 8, with the data links of implants in Chapter 9; stimulators, the actuators of the nervous system, come in Chapters 9 and 10; materials and packages in Chapters 11 and 12; power in Chapter 13; and regulation and the design project, which frame the whole system, in Chapter 14.

The four labs are the basic design steps, and each builds one layer of the project hardware: Lab 1 teaches the layout of the circuit boards everything sits on (with a small sensor breakout board), Lab 2 prints the electrodes and a sensor array, Lab 3 designs the analog front end



Step	ECG patch	Brain implant	Sensing capsule
1 Need and signal	catch rhythm problems over days; ECG 0.5 to 4 mV, 0.05 to 150 Hz	move a cursor by thought; spikes 50 to 500 µV, up to 7 kHz	sense along the gut and know where; light and magnetic field
2 Specifications and budgets	12 kbit/s, 1.67 mA, several days	1024 channels, 198 Mbit/s raw, little heat	smaller than a size-000 pill, 1.75 mA, 16 h
3 Architecture	front-end chip, BLE microcontroller, rechargeable cell	custom chip, BLE, inductively charged battery	optical chip, BLE microcontroller, magnetometer, silver-oxide cells
4 Build	printed electrodes, flexible board, firmware	thin-film threads, insertion robot, sealed enclosure	small board in a 3D-printed shell
5 Test	ECG simulator on the bench, then volunteers	bench and animals, then a clinical study	gut model and cow intestine (related study)
6 Regulatory path	class II, 510(k)	clinical study first; class III implants need a PMA	research prototype

Figure 1.26: One design process for every bioelectronic device. Top: six steps, with the chapters and labs that teach each below it; testing sends the design back for another pass (red arrow), and step 3 (blue) is where the blocks of Fig. 1.25 are chosen. Bottom: how the three devices of Fig. 1.2 answer each step. Class II and class III are the FDA's moderate- and high-risk device classes; most class II devices are cleared through a 510(k) and class III devices approved through premarket approval (PMA) (Chapter 14).

and converter, and Lab 4 writes the processor's firmware and its SPI and Bluetooth link. Together they make an ECG patch like that of Fig. 1.2a, or the EEG or sensor glove of the design project.

The chapters are grouped in four parts, with the labs as a fifth. Part I lays the foundations in circuits and biology, Part II turns physical, chemical, optical, and electrical signals into measurements, Part III connects electronics to the nervous system, and Part IV covers how devices are made, powered, and brought to patients.

Part I continues with the biology that sensors detect: the cell, its membrane, proteins, DNA, body fluids, and diffusion in Chapter 2, and in Chapter 3 the resting potential, the membrane as an RC circuit, and the action potential. **Part II** builds sensors and the circuits that read them: figures of merit and the op amp (Chapter 4); temperature, strain, capacitive, and transistor sensors (Chapter 5); electrochemical sensors and the potentiostat (Chapter 6); optical sensors and pulse oximetry (Chapter 7); and biopotential recording, from electrodes and instrumentation amplifiers to wireless links and electrical safety (Chapter 8). **Part III** records from and stimulates the nervous system (Chapter 9) and follows two sensory prostheses, the cochlear implant and the retinal prosthesis (Chapter 10). **Part IV** covers soft materials and their mechanics (Chapter 11), printing (Chapter 12), power

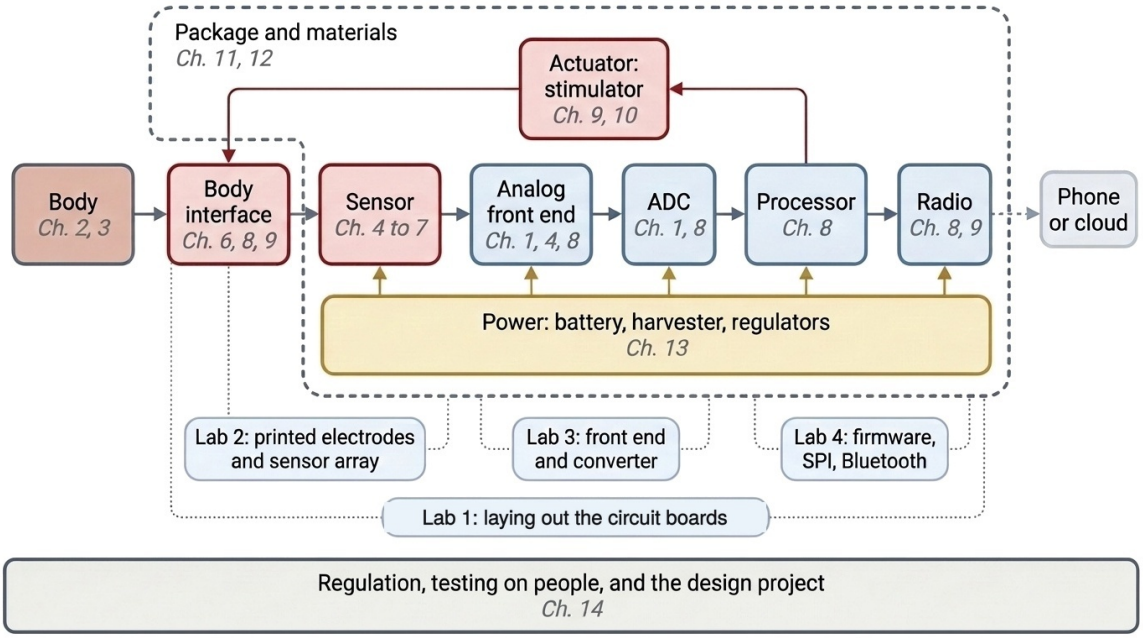


Figure 1.27: The book on the system. The blocks of Fig. 1.25, colored as in Fig. 1.2, carry the chapters that teach them (italic); the labs below build the blocks they point to. The dashed line is the package, which, as in Fig. 1.25, leaves the body interface exposed. Regulation, testing on people, and the design project (Chapter 14) frame the whole system.

(Chapter 13), and the path from prototype to regulated medical device, together with a semester design project (Chapter 14). In **Part V**, the four labs build the project hardware block by block, as described above.

The design project. Teams build one of three complete systems: a glove covered with printed capacitive pressure sensors, whose readings a small neural network on a microcontroller turns into the identity of the object being held; a wearable EEG with printed electrodes around the ear that tracks stress through the power in its frequency bands; or a wearable ECG patch with printed silver/silver chloride electrodes that reports heart rate. Each has every block of Fig. 1.25, and each joins a printed, flexible sensor through a flat flexible cable to a rigid circuit board: the front-end board of Lab 3 for the ECG and EEG, and a capacitance readout board for the glove (Lab 2).

Readers from biology should work through Section 1.10 and Chapter 4 with care, because every later circuit builds on them; readers from electrical engineering will find most of their new material in Chapters 2, 3, and 6.

Summary

- Bioelectronic devices record from, stimulate, or treat tissue, on the skin, inside the body, or in the gut. Their transducers are electrical, optical, electrochemical, or physical. Galvani's animal electricity

(biopotentials) and Volta's metallic electricity (the metal-electrolyte cell) are the two halves of the field.

- Tissue is soft (brain 1 to 2 kPa, skin 0.1 to 1 MPa); silicon is 130 to 190 GPa. Bending stiffness $D = Et^3/[12(1-\nu^2)]$ falls as t^3, surface strain is $t/2R$, and a film wraps radius R when $D \leq 2\gamma R^2$: about 13 µm of polyimide for a fingertip.
- Biosignals: ECG 0.5 to 4 mV (0.05 to 150 Hz), EEG 5 to 300 µV (0.5 to 100 Hz), EMG 10 µV to 5 mV (20 to 500 Hz), spikes to 7 kHz; PPG AC is 0.1 to 2 % of DC; heart rate is $60/\overline{RR}$.
- Body fluids trade access for information (blood, ISF, sweat, saliva, breath); a chemical sensor pairs a recognition element with a transducer. Sweat chloride of 60 mM or more is consistent with cystic fibrosis; iontophoresis delivers $n = t_d It/(zF)$ at up to about $0.5\,\mathrm{mA/cm^2}$.
- Implants face the foreign body response, from protein adsorption in seconds to a fibrous capsule in weeks; soft, thin devices provoke less. Transient electronics dissolve (Mg, Zn at 0.5 to 3 µm/h; W, Mo, Fe at 10^{-4} to 0.02 µm/h), and the encapsulation sets their life. Capsules meet stomach pH 1 to 3.5 and transit the gut in 10 to 73 h.
- Circuits: $I = \mathrm{d}q/\mathrm{d}t$; $V = IR$; $P = VI$; divider $V_s R_2/(R_1 + R_2)$; $Q = CV$, $\tau = RC$; $|Z_C| = 1/(2\pi fC)$; $G_{\mathrm{dB}} = 20\log_{10}(V_2/V_1)$; an RC filter is -3 dB at $f_c = 1/(2\pi RC)$ and rolls off at 20 dB per decade.
- Sample at $f_s > 2f_{\max}$ to avoid aliasing; resolve ΔV with $N \geq \log_2(V_{\mathrm{FS}}/\Delta V)$ bits.
- A bioelectronic system has a body interface, sensor, front end, ADC, processor, radio, power, and package. Bit rate is $N_{\mathrm{ch}} f_s b$; battery life is $C_{\mathrm{bat}}/\sum D_i I_i$; flexible hybrid electronics put printed materials at the body and silicon where it computes.
- An ECG patch, a brain implant, and a sensing capsule share these blocks; the body site sets size, power source, data path, and lifetime. The patch streams 12 kbit/s on about 6 mW; 1024 electrodes in the brain produce about 200 Mbit/s raw, which the implant must reduce before its radio; the capsule draws 1.75 mA from 28 mAh of silver-oxide cells and lasts about 16 h. One design process serves all three: need and signal, specifications and budgets, architecture, build, test, regulatory path, and iteration.

1.14 *Problems*

1.1 • A polyimide film ($E = 2.5$ GPa, $\nu = 0.34$) is 5 µm thick.

(a) With $\gamma = 10\,\mathrm{mJ/m^2}$, what is the smallest radius of curvature it wraps without help?

(b) If the work of adhesion were only $1\,\mathrm{mJ/m^2}$, what would that radius be?

(c) Explain why cutting the film into an open mesh helps it conform.

1.2 • A bioresorbable nerve stimulator must work for 14 days.

(a) How thick must a bare Mo trace be to last 14 days at the fast end of the quoted range, $0.02\,\mu\mathrm{m/h}$?

(b) A 400 nm Mg coil dissolves at $1\,\mu\mathrm{m/h}$ once exposed. How long must the encapsulation last?

(c) A 50 nm silicon membrane is exposed at the same time as the Mg. How long after the Mg is it gone, at 5 nm/day?

1.3 • A chest strap senses breathing at 0.2 to 0.5 Hz, but walking adds motion at about 2 Hz. A first-order RC low-pass filter with $C = 10\,\mu\mathrm{F}$ follows the sensor.

(a) Choose R for a corner frequency of 0.7 Hz, and find the time constant.

(b) Find the gain in decibels at 0.3 Hz and at 2 Hz.

(c) By what factor, and by how many decibels, does the filter improve the ratio of breathing signal to walking motion? Why would a second-order filter do better?

1.4 ◦ A 16-bit ADC has a full-scale range of 2.4 V.

(a) Find its LSB.

(b) Can it digitize a direct-coupled EEG with a $\pm 300\,\mathrm{mV}$ electrode offset to $1\,\mu\mathrm{V}$ resolution? If not, how many bits are needed?

(c) What analog gain, after AC coupling, lets this ADC resolve $1\,\mu\mathrm{V}$ at the electrodes, and what is then the largest input it can take?

1.5 •• You must design a patch that records breathing rate, heart rate, and skin temperature on the chest for 14 days. Using Table 1.2:

(a) Choose a transducer for each quantity (one may serve two).

(b) Choose a sampling rate and resolution for each, with a reason.

(c) Estimate the raw data rate, and propose what to compute on the patch so that it sends fewer than 1 MB per day.

2
Cells and Biomolecules

In this chapter you will learn to

- Describe a cell and its parts, and place molecules, cells, and devices on one length scale.
- Model the cell membrane as a capacitor of about $1\,\mu F/cm^2$ in parallel with a resistor.
- Estimate diffusion times with $t \approx x^2/2D$, find a flux with Fick's first law, and compute an osmotic pressure.
- Convert glucose between mg/dL and mM, and explain why enzyme sensors saturate (Michaelis-Menten kinetics).
- Explain how DNA hybridization and aptamer sensors recognize their targets.
- Read a blood count, compute the oxygen blood carries, and compare plasma, interstitial fluid, and sweat.
- Explain how PCR detects a few copies of a gene, and model a layer of cells on an electrode as an impedance.

Pipette a few thousand skin cells into a dish whose floor is a small gold electrode, set the dish in an incubator at $37\,^{\circ}C$, and record the electrode's impedance every few minutes. Over the next few hours, as the cells settle, spread, and close the gaps between them, the impedance climbs severalfold, because the current must squeeze under and between the cells instead of passing straight into the medium. The cells are not electronic parts, yet they set every number in that measurement: their membranes act as capacitors and the narrow gaps between them as resistors. The same holds for every device in this book that touches the body: how fast glucose diffuses to a sensor, how an enzyme responds when glucose is high, and how cells and proteins coat an implant. An engineer who knows only the electronics

can build the amplifier. Designing the sensor takes the biology.

Chapter 1 named the body fluids that wearable sensors sample and the molecules they look for. This chapter is the biology behind them, written for readers who have not taken it. We start with the cell and its membrane, which is also a capacitor, and with how molecules move: diffusion and osmosis. Then come the molecules that sensors measure or borrow (glucose, proteins and enzymes, DNA and RNA) and the body fluids that carry them. Short sections on bacteria and viruses and on cells grown on electrodes follow, and the chapter closes with the ways cells signal to each other, which lead to the neurons of Chapter 3. Nothing here needs more electronics than the circuit essentials of Chapter 1; where a sensor appears, we point ahead to the chapter that builds it.

2.1 *The Cell*

THE CELL IS THE SMALLEST UNIT that can live on its own. Every cell is a bag of watery *cytoplasm* wrapped in a *plasma membrane* about 5 nm thick, and it carries its own instructions in DNA. About 70 % of a cell's mass is water. Most of the rest is large molecules (proteins, nucleic acids, and polysaccharides), with a few percent lipids and the remainder ions and small molecules such as sugars and amino acids. A typical human cell is 10 to 30 µm across, the size of a pad on a microelectrode array.

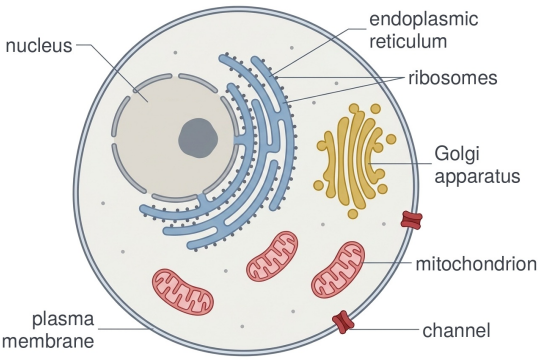


Figure 2.1: An animal cell, cut open. The nucleus holds the DNA; mitochondria make ATP; ribosomes, many of them on the endoplasmic reticulum, build proteins; the Golgi apparatus packages and ships them; the plasma membrane separates the inside from the outside and carries the channels and receptors through which the cell exchanges ions, molecules, and signals.

A cell of a plant, an animal, or a fungus is a *eukaryotic* cell: its DNA sits inside a membrane-bound *nucleus*, and the cytoplasm holds other membrane-bound compartments called *organelles* (Fig. 2.1). It helps to think of the organelles as the departments of a factory (Table 2.1).

Bacteria are *prokaryotic* cells: they have no nucleus and no organelles, and their DNA floats in the cytoplasm. A bacterium is 0.5 to 5 µm long, a tenth the size of a human cell, so it fits into crevices on an implant.

Part	What it does	Engineering analogy
Nucleus	stores the DNA; copies genes into messenger RNA	read-only memory
Ribosomes	read messenger RNA and build proteins	assembly line
Endoplasmic reticulum	folds new proteins (rough ER) and makes lipids (smooth ER)	production floor
Golgi apparatus	modifies, sorts, and ships proteins in vesicles	packaging and shipping
Mitochondria	oxidize sugars and fats to make ATP	power plant
Lysosomes	digest worn-out parts and debris	recycling
Cytoskeleton	sets the shape, moves cargo, lets the cell crawl	frame and rails
Plasma membrane	barrier, transport, signaling, and a charged capacitor	package with I/O ports

Table 2.1: The main parts of a eukaryotic cell. ATP (adenosine triphosphate) is the cell's energy currency: splitting off one phosphate releases about 50 kJ/mol under cell conditions.

Most bacteria wrap their membrane in a rigid wall, while animal cells have none and are fragile. A bacterium can divide in two every 20 min; a human cell in culture divides about once a day, and many, including most neurons, never divide again. Both kinds of cell use the same genetic code, so a human gene put into a bacterium is read correctly; this is how insulin is manufactured.

One scale from molecules to devices

Bioelectronics spans seven orders of magnitude of length (Fig. 2.2). A transistor's gate insulator and a cell membrane are both a few nanometers thick, a virus is about as thick as a sputtered metal film, and a red blood cell is a little smaller than a microelectrode. Keep the figure in mind whenever you choose an electrode size, a film thickness, or a filter pore.

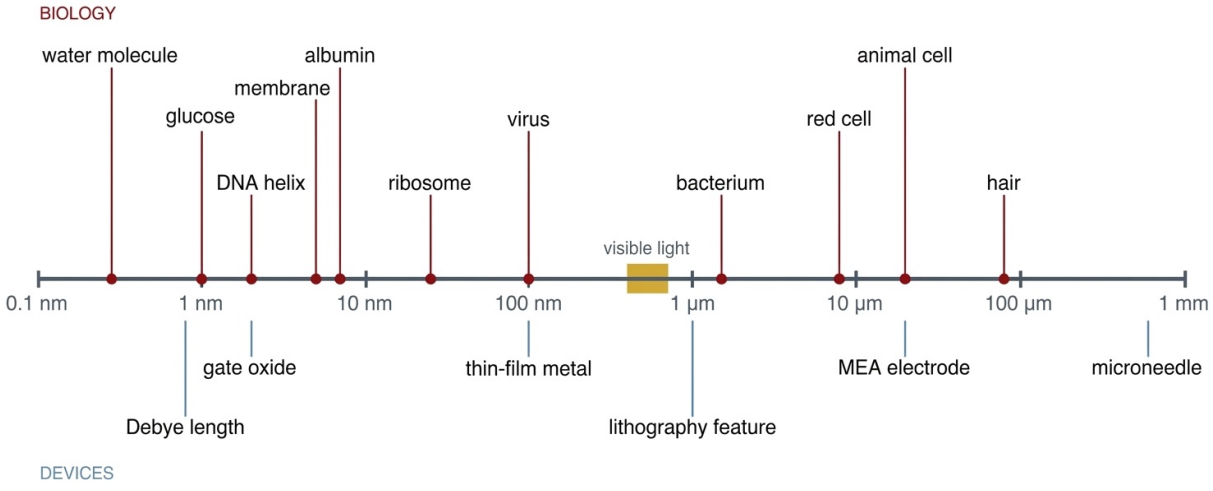
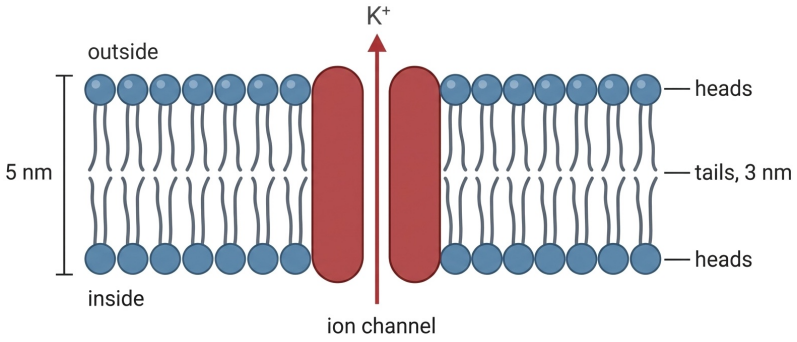


Figure 2.2: Sizes in bioelectronics on a logarithmic scale, from a water molecule (0.28 nm) to a hair. Below the axis: the Debye length in body fluid (Section 2.6), a transistor gate oxide, a sputtered metal film, the smallest feature of routine photolithography, a microelectrode-array pad, and a microneedle. Objects smaller than about half the wavelength of visible light (gold band) cannot be resolved with a light microscope.

2.2 *The Cell Membrane*

THE MEMBRANE IS BUILT FROM LIPIDS. A *phospholipid* has two oily tails, chains of 14 to 24 carbon atoms that avoid water (*hydrophobic*), joined to a small charged head that attracts water (*hydrophilic*). In water these molecules pack into a sheet two molecules thick, the *lipid bilayer*, with the tails inside and the heads facing the water on both sides (Fig. 2.3). The sheet assembles itself: oily tails in water force the water around them into ordered cages, and burying the tails in the middle of the sheet frees that water. This *hydrophobic effect* also folds proteins (Section 2.5). A bilayer closed on itself is a *liposome*, used to deliver drugs; the lipid nanoparticles of the mRNA vaccines are relatives.



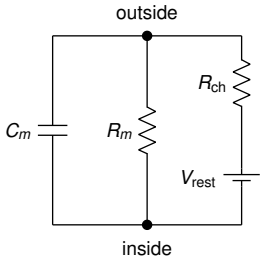


Figure 2.3: Left: a patch of the lipid bilayer, 5 nm thick with an oily core about 3 nm thick, crossed by an ion channel protein. Right: its electrical model. The oily core is the dielectric of a capacitor C_m; leakage through the lipid is R_m; open channels add conductance in series with the battery that the ion gradients create (Chapter 3).

KEY EQUATION

$C_m = \varepsilon_0 \varepsilon_r / d \approx 1\,\mu\text{F/cm}^2$

The specific capacitance of nearly every cell membrane.

A real membrane is a bilayer studded with proteins (channels, carriers, pumps, and receptors) that drift sideways in it, a two-dimensional liquid called the *fluid mosaic*. Sugar chains on its outer face form a fuzzy coat, the first thing an implanted material meets.

The membrane as a capacitor

A capacitor is two conductors separated by a thin insulator (Chapter 1). For flat plates of area A a distance d apart, $C = \varepsilon_0 \varepsilon_r A/d$, where $\varepsilon_0 = 8.85 \times 10^{-12}\,\text{F/m}$ and ε_r is the relative permittivity of the insulator. The bilayer is exactly this. Salt water on both sides is the pair of plates, and the oily core is the insulator. A hydrocarbon has ε_r of about 2 to 3, and the core is about 3 nm thick, so the capacitance per unit area is

$$C_m = \frac{\varepsilon_0 \varepsilon_r}{d} \approx \frac{(8.85 \times 10^{-12})(2.1)}{3 \times 10^{-9}}\ \text{F/m}^2 = 0.62\,\mu\text{F/cm}^2. \qquad (2.1)$$

Measurements on real cells give about $1\,\mu\text{F/cm}^2$ (10 fF per square micrometer); the head regions and embedded proteins add a little to the simple estimate. Every membrane is built from the same bilayer, so the number is almost the same for a neuron, a red cell, and a bacterium,

and close to the $1.7\,\mu F/cm^2$ of the 2 nm silicon dioxide insulator in a transistor.

A thin insulator means a high field: 70 mV across 5 nm is 1.4×10^7 V/m, five times the field at which air breaks down. If a pulse drives the membrane voltage above roughly 0.5 to 1 V, pores open and the cell becomes leaky. This *electroporation* pushes DNA and drugs into cells, and it is one way an overly strong stimulus harms tissue (Chapter 9 sets the safe limits for stimulation). A uniform external field E induces a membrane voltage on a spherical cell of radius r of about

$$\Delta V_m = 1.5\, E\, r \cos\theta, \tag{2.2}$$

largest at the two poles that face the electrodes ($\theta = 0$ and 180°). Putting 1 V across the poles of a cell 20 µm across takes $E = 1/(1.5 \times 10\,\mu m) = 0.67$ kV/cm; a 1 µm bacterium needs 13 kV/cm, twenty times more.

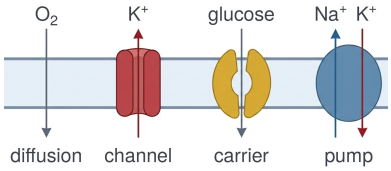


Figure 2.4: Four ways across the membrane (outside at the top). Simple diffusion through the lipid, an ion channel, a carrier that changes shape, and a pump that spends ATP to move 3 Na^+ out and 2 K^+ in against their gradients.

Membrane resistance and transport proteins

A pure lipid bilayer is an excellent insulator, around $10^6\,\Omega\,cm^2$, because dragging a charged ion into an oily core costs too much energy. A living membrane is leakier, typically 1 to $100\,k\Omega\,cm^2$, because its proteins provide routes across (Fig. 2.4); small molecules such as O_2 and CO_2 diffuse through the lipid itself. *Channels* are pores for one kind of ion; an open channel passes millions of ions per second, a current of a few picoamperes (Chapter 3). *Carriers* bind a molecule, change shape, and release it on the other side; the GLUT transporters carry glucose this way. These routes are passive: they move a substance only down its gradient. *Pumps* spend ATP to move ions uphill. The Na^+/K^+ pump moves 3 Na^+ out and 2 K^+ in per ATP, about a hundred cycles per second, and keeps the cell's ion gradients charged like a battery charger.

A cell of membrane area A is therefore a capacitor $C_m A$ in parallel with a resistor R_m/A. Their product, the *membrane time constant* $\tau = R_m C_m$ (the RC time constant of Chapter 1), does not depend on size: with $R_m = 10\,k\Omega\,cm^2$ it is 10 ms for every cell. The 20 µm cell of the next example is a 12.6 pF capacitor in parallel with 0.8 GΩ, which is why recording from one cell takes the gigaohm seals of Chapter 3 and amplifiers that resolve picoamperes (Chapter 4).

> **Worked Example 2.1** *How many ions make the resting potential?*
>
> A spherical cell 20 µm in diameter rests at −70 mV and holds 140 mM of K^+. Its membrane has $C_m = 1\,\mu F/cm^2$. (a) Find the membrane capacitance of the cell. (b) How much charge, and how

many monovalent ions, must sit on the membrane to hold 70 mV? (c) What fraction of the cell's K^+ is that?

SOLUTION. (a) The area is $4\pi r^2 = 4\pi(10\,\mu\text{m})^2 = 1257\,\mu\text{m}^2 = 1.257 \times 10^{-5}\,\text{cm}^2$, so $C = 1\,\mu\text{F/cm}^2 \times 1.257 \times 10^{-5}\,\text{cm}^2 = 12.6\,\text{pF}$.

(b) $Q = CV = 12.6\,\text{pF} \times 70\,\text{mV} = 0.88\,\text{pC}$. Dividing by the elementary charge, $0.88 \times 10^{-12}\,\text{C}/1.602 \times 10^{-19}\,\text{C} = 5.5 \times 10^6$ ions.

(c) The volume is $\frac{4}{3}\pi r^3 = 4.19 \times 10^{-15}\,\text{m}^3 = 4.19\,\text{pL}$, so the cell holds $0.140\,\text{mol/L} \times 4.19 \times 10^{-12}\,\text{L} \times N_A = 3.5 \times 10^{11}$ potassium ions. The membrane charge is $5.5 \times 10^6/3.5 \times 10^{11} = 1.6 \times 10^{-5}$ of them, about 1 in 64 000. A resting potential needs only a thin sheet of unbalanced charge at the membrane surface; the bulk of the cytoplasm stays electrically neutral, and the concentrations hardly change when the voltage does. Chapter 3 uses this fact for the action potential. ■

2.3 *Diffusion and Osmosis*

INSIDE AND BETWEEN CELLS most molecules get around by *diffusion*, the random thermal motion that spreads them from high to low concentration. Blood carries molecules close to every cell, but the last step is always diffusion. Each molecule takes a random walk, and after a time t its mean square displacement along any one direction is

$$\langle x^2 \rangle = 2Dt, \tag{2.3}$$

where D is the *diffusion coefficient* in m^2/s, about $10^{-9}\,\text{m}^2/\text{s}$ for a small molecule in water. (In three dimensions the mean square distance is $6Dt$.) The time to diffuse a distance x therefore grows as its square,

$$t \approx \frac{x^2}{2D}. \tag{2.4}$$

KEY EQUATION

$t \approx x^2/2D$

Diffusion time over a distance x: ten times farther takes a hundred times longer.

This one equation explains why cells are small, why tissues need capillaries every few tens of micrometers, and why a sensor's response time is set by the thickness of its coatings.

Fick's laws

The net flux of molecules J (moles per square meter per second) is proportional to the concentration gradient. This is *Fick's first law*,

$$J = -D\,\frac{\partial c}{\partial x}. \tag{2.5}$$

It has the form of Ohm's law: concentration plays the role of voltage, flux the role of current density, and D the role of conductivity. Through

a layer of thickness L with concentrations c_1 and c_2 on its faces, the steady flux is $J = D(c_1 - c_2)/L$, and the layer has a *diffusion resistance* $L/(DA)$ just as a resistor has $L/(\sigma A)$. Counting the molecules that flow into and out of a thin slice turns the first law into *Fick's second law*,

$$\frac{\partial c}{\partial t} = D\,\frac{\partial^2 c}{\partial x^2}, \tag{2.6}$$

which describes how a concentration profile changes in time; with a leakage term added, the same equation describes how a voltage spreads along an axon, the cable of Chapter 3. If the surface of a layer is suddenly held at c_0, the solution is a front that advances into the layer as $\sqrt{Dt}$.

Table 2.2 lists diffusion times for three molecules that matter to devices: oxygen, the sugar glucose, and albumin, the most common protein in blood. Small molecules cross a cell in tens of milliseconds and a 100 µm coating in seconds; a millimeter of tissue takes minutes for glucose and hours for a protein. In a hydrogel or in tissue, cells and polymer chains act as obstacles. The path becomes longer by a *tortuosity* factor λ, and the effective diffusion coefficient falls to D/λ^2; in the spaces between brain cells $\lambda \approx 1.6$, so $D_{\text{eff}} \approx 0.4\,D$.

Molecule (D at 37 °C)	membrane 5 nm	cell 10 µm	hydrogel 100 µm	tissue 1 mm
O_2 (3.0×10^{-9} m^2/s)	4 ns	17 ms	1.7 s	2.8 min
glucose (9.0×10^{-10} m^2/s)	14 ns*	56 ms	5.6 s	9.3 min
albumin (9.0×10^{-11} m^2/s)	0.14 µs*	0.56 s	56 s	1.5 h

Table 2.2: Diffusion times $t = x^2/2D$ in water at 37 °C. *These times would apply to a 5 nm layer of water. Glucose actually crosses the lipid bilayer only through carriers, and albumin not at all; oxygen dissolves in the lipid and crosses freely. At 25 °C each D is about a quarter smaller (glucose 6.7×10^{-10} m^2/s).

Why is albumin ten times slower than glucose? A sphere of radius r_h in a liquid of viscosity η feels a drag $6\pi\eta r_h$ per unit speed, and dividing the thermal energy $k_B T$ by that drag ($k_B = 1.381 \times 10^{-23}$ J/K is Boltzmann's constant) gives the *Stokes-Einstein relation*, $D = k_B T/6\pi\eta r_h$. Albumin's radius (3.5 nm) is about ten times that of glucose (0.38 nm), so it is ten times slower. Mass grows roughly as r_h^3, so an antibody, more than 800 times heavier than glucose, has a radius (5.3 nm, enlarged by its Y shape) only 14 times larger and diffuses only 14 times more slowly.

Worked Example 2.2 *Diffusion times in a sensor and in tissue*

(a) Using Table 2.2, how long does glucose take to cross a 100 µm hydrogel coating in which $D_{\text{gel}} = 0.5\,D$? (b) An implanted glucose sensor sits 300 µm from the nearest capillary in tissue with tortuosity $\lambda = 1.6$. Estimate the diffusion delay between a change in blood

glucose and the sensor.

SOLUTION. (a) $D_{\text{gel}} = 4.5 \times 10^{-10}\,\text{m}^2/\text{s}$, so $t = (1 \times 10^{-4}\,\text{m})^2/(2 \times 4.5 \times 10^{-10}) = 11\,\text{s}$.

(b) $D_{\text{eff}} = 9.0 \times 10^{-10}/1.6^2 = 3.5 \times 10^{-10}\,\text{m}^2/\text{s}$ and $t = (3 \times 10^{-4}\,\text{m})^2/(2 \times 3.5 \times 10^{-10}) = 128\,\text{s}$, about two minutes. Together with the time glucose takes to cross the capillary wall and the sensor's own coatings, this is why a glucose reading in the fluid between cells lags the blood by several minutes, and why the lag grows when a capsule of scar tissue forms around an implant. ■

Living tissue also consumes what diffuses into it. A slab of cells fed with oxygen from one face at concentration c_0 and consuming it at a rate Q (moles per cubic meter per second) starves at its far side if it is thicker than $L_{\max} = \sqrt{2Dc_0/Q}$, the steady solution of Fick's second law with a uniform sink. For tissue in medium equilibrated with air ($c_0 = 0.20\,\text{mM}$, $D = 2 \times 10^{-9}\,\text{m}^2/\text{s}$ in tissue, $Q = 0.01\,\text{mol}\,\text{m}^{-3}\,\text{s}^{-1}$), $L_{\max} \approx 280\,\mu\text{m}$, so thicker engineered tissue needs its own blood vessels. In the body, capillary blood has less dissolved oxygen and working tissue uses more, so capillaries sit tens of micrometers apart.

Osmosis

Water crosses most membranes far more easily than solutes do. When a membrane that passes water but not solutes separates two solutions, water flows toward the side with more dissolved particles, as if to dilute it. The pressure that would stop the flow is the *osmotic pressure*, given for dilute solutions by van 't Hoff's law,

$$\Pi = c_{\text{osm}}\, RT, \tag{2.7}$$

where c_{osm} is the concentration of all dissolved particles, the *osmolarity* (a salt such as NaCl counts twice because it splits into two ions), $R = 8.314\,\text{J}/(\text{mol}\,\text{K})$, and T is the absolute temperature. Body fluids are at about 290 mOsm/L, so against pure water at 37 °C they pull with $\Pi = (290\,\text{mol}/\text{m}^3)(8.314)(310\,\text{K}) = 7.5 \times 10^5\,\text{Pa}$, about 7.5 bar, the pressure under 75 m of water. A solution at that level is *isotonic*: 0.9 % NaCl (154 mM, which behaves like about 286 mOsm/L because the ions interact) and 5 % glucose (278 mM) are the two isotonic fluids of hospital drips. In a *hypotonic* solution cells take in water, swell, and burst (red cells *lyse*); in a *hypertonic* one they shrink. A cell behaves as an ideal osmometer: its water volume varies inversely with the outside osmolarity, and so with the osmotic pressure Π it sets, while a part V_b of its volume (proteins and other solids) does not change,

$$V = V_b + (V_0 - V_b)\,\frac{\Pi_0}{\Pi}, \tag{2.8}$$

KEY EQUATION

$\Pi = c_{\text{osm}} RT$

Osmotic pressure: about 7.5 bar for body fluids against pure water.

where V_0 is the volume at the normal osmolarity Π_0. A red blood cell of 90 fL, 40 % of it solids, has $V_b = 36$ fL; in a solution of half the normal osmolarity it swells to $36 + 54 \times 2 = 144$ fL, close to the 148 fL sphere that its membrane can just enclose, and red cells indeed start to burst near this point.

Engineers meet osmosis in hydrogels, which swell until the elastic force of the polymer network balances the osmotic pressure of their ions, and in implantable osmotic drug pumps. An artificial kidney, or *dialyzer*, uses diffusion and osmotic balance together: blood flows through thousands of hollow fibers whose porous walls pass urea and potassium but not cells or proteins, and a fluid of normal salt content flowing outside carries the waste away.

2.4 *Sugars and Glucose*

GLUCOSE IS THE MOLECULE that bioelectronic sensors measure most often. It is a simple sugar, $C_6H_{12}O_6$; the five-carbon sugars ribose and deoxyribose form the backbones of RNA and DNA, and long chains of sugars (*polysaccharides*) include glycogen, the glucose store in liver and muscle, and hydrogels such as alginate and hyaluronic acid (Chapter 11). Glucose is the body's main fuel. Burned completely,

$$C_6H_{12}O_6 + 6\,O_2 \longrightarrow 6\,CO_2 + 6\,H_2O, \qquad \Delta G \approx -2870\,\text{kJ/mol}, \quad (2.9)$$

and the cell stores a third to a half of that energy in about 30 ATP per glucose; the rest leaves as heat, which keeps you at 37 °C. The adult brain burns about 120 g of glucose a day, 1.9 MJ, or 22 W averaged over the day, about a fifth of the body's resting power.

Glucose as an analyte

Clinicians in the United States report glucose in milligrams per deciliter; most of the world and the sensor literature use millimolar. With a molar mass of 180.16 g/mol,

$$1\,\text{mM} = 180.16\,\text{mg/L} = 18.0\,\text{mg/dL}. \quad (2.10)$$

KEY IDEA Divide mg/dL by 18 to get mM. Fasting blood glucose: 70 to 100 mg/dL, about 4 to 5.5 mM.

A healthy fasting level is 70 to 99 mg/dL; 126 mg/dL or more on repeated tests indicates diabetes, and below 70 mg/dL is hypoglycemia. Continuous glucose monitors report from about 40 to 400 mg/dL (2.2 to 22 mM), the range their enzyme chemistry must cover (Section 2.5). They measure glucose in the fluid between cells, which follows blood with a delay of several minutes (Worked Example 2.2). Sweat carries only tens to hundreds of micromolar, roughly a hundredth of the blood level (0.1 mM is 1.8 mg/dL), so a sweat sensor needs about

a hundred times better resolution. Glucose also binds slowly to the hemoglobin in red cells, which live about 120 days, so the bound fraction, HbA1c, records the average glucose over two to three months.

The glucose sensors of Chapter 6 borrow an enzyme, glucose oxidase (GOx), that catalyzes

$$\text{glucose} + O_2 \xrightarrow{\text{GOx}} \text{gluconolactone} + H_2O_2, \tag{2.11}$$

and read the hydrogen peroxide at an electrode. GOx is a protein, and its kinetics, the subject of the next section, decide the sensor's range.

2.5 *Proteins and Enzymes*

PROTEINS DO ALMOST EVERY JOB in a cell. They are chains of *amino acids*, small molecules that share a common backbone and differ in one of twenty side chains: some oily, some polar, some charged at body pH. Each amino acid adds about 110 Da (one dalton corresponds to one gram per mole), so a protein of 450 amino acids has a mass of about 50 kDa.

Folding and denaturing

The sequence decides how the chain folds. Hydrogen bonds along the backbone make local spirals (α helices) and pleated strips (β sheets), which pack into a compact shape with the oily side chains buried inside by the hydrophobic effect. A folded 50 kDa protein is a ball about 5 nm across; albumin (66.5 kDa) has a radius of 3.5 nm in water, and an antibody (150 kDa) is Y shaped and 10 to 15 nm across its arms. The fold is held by weak bonds. Heat, extreme pH, solvents, drying, and some surfaces unfold, or *denature*, a protein, and an unfolded enzyme no longer works. That is the engineering problem of every enzyme and antibody sensor: it has a shelf life, it drifts at body temperature, and it cannot be sterilized in an autoclave at 121 °C like a metal electrode. Proteins also stick to almost any surface; within seconds of implantation a protein film coats a device and decides what the body does next.

Antibodies are the proteins that sensors borrow most often after enzymes. White blood cells called B cells (Section 2.7) make them, and the tips of the two arms bind one particular molecule, the *antigen*, tightly and selectively. Affinity sensors use antibodies, or the DNA aptamers of Section 2.6, to capture a target without changing it (Chapter 6).

Enzymes

An *enzyme* is a protein that speeds up one reaction. Its substrate binds in a pocket, the *active site*, shaped and charged to fit that one molecule, which is why an enzyme sensor responds to glucose and not to fructose. Some enzymes need a small helper molecule, a *cofactor*; glucose oxidase carries one called FAD, which takes two electrons from glucose and passes them to oxygen. An enzyme is not used up: one molecule converts substrate over and over, at up to its *turnover number* k_{cat}, which ranges from a few to millions of molecules per second; glucose oxidase manages several hundred.

Michaelis-Menten kinetics

The simplest model has the enzyme E bind the substrate S into a complex ES, which either falls apart or converts S to product P:

$$\mathrm{E} + \mathrm{S} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{ES} \xrightarrow{k_{\text{cat}}} \mathrm{E} + \mathrm{P}. \tag{2.12}$$

In a sensor the substrate is replenished, so the amount of complex quickly settles to a steady level. Setting its rate of formation equal to its rate of loss, with a fixed total amount of enzyme $[\mathrm{E}]_0$, gives the reaction rate

$$v = \frac{V_{\max}[\mathrm{S}]}{K_m + [\mathrm{S}]}, \qquad V_{\max} = k_{\text{cat}}[\mathrm{E}]_0, \qquad K_m = \frac{k_{-1} + k_{\text{cat}}}{k_1}. \tag{2.13}$$

K_m, the *Michaelis constant*, is the substrate concentration at which the enzyme runs at half speed. Below K_m most enzyme molecules are idle and the rate is proportional to the substrate, $v \approx (V_{\max}/K_m)[\mathrm{S}]$; far above it every enzyme is busy and the rate *saturates*. An enzyme-limited sensor follows the same law, $I = I_{\max}\, c/(K_m + c)$, with $I_{\max}$ proportional to the active enzyme on the electrode (Fig. 2.5).

KEY EQUATION

$v = \dfrac{V_{\max}[\mathrm{S}]}{K_m + [\mathrm{S}]}$

Michaelis-Menten: linear for $[\mathrm{S}] \ll K_m$, half speed at K_m, saturated above.

Inverting Eq. (2.13) gives the *Lineweaver-Burk* form,

$$\frac{1}{v} = \frac{K_m}{V_{\max}}\,\frac{1}{[\mathrm{S}]} + \frac{1}{V_{\max}}, \tag{2.14}$$

a straight line against $1/[\mathrm{S}]$ with slope $K_m/V_{\max}$ and intercept $1/V_{\max}$ (Fig. 2.5b). It is a quick graphical fit. It also stretches the points at low concentration, where the relative noise is largest, so for real data a direct nonlinear least-squares fit of Eq. (2.13) is better.

WORKED EXAMPLE 2.3 *Fitting an enzyme sensor*

A glucose oxidase electrode gives these steady currents:

c (mM)	1	2	4	8	16
I (µA)	0.257	0.450	0.720	1.029	1.309

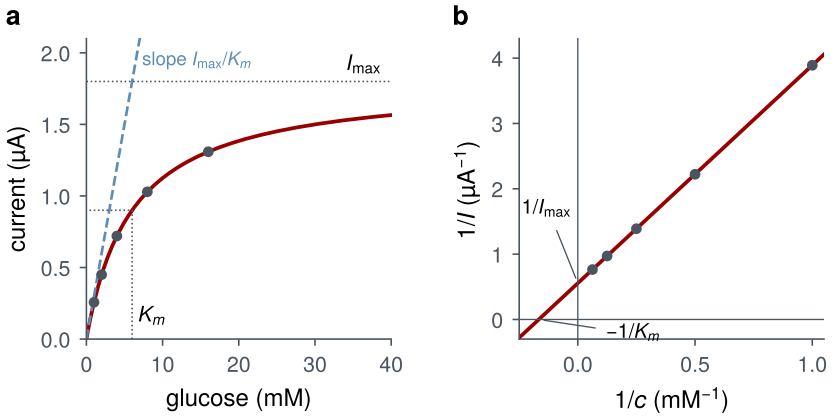


Figure 2.5: Michaelis-Menten response of an enzyme electrode with $I_{max} = 1.8\,\mu A$ and $K_m = 6\,mM$ (the data of Worked Example 2.3). (a) The current rises linearly at first (dashed line), reaches half its maximum at K_m, and saturates. (b) Plotting $1/I$ against $1/c$ turns the curve into a straight line whose intercepts give $1/I_{max}$ and $-1/K_m$.

(a) Find I_{max} and the apparent K_m with a Lineweaver-Burk fit. (b) What is the sensitivity at low concentration? (c) What current does a fasting blood level of 5 mM give, and what would a linear sensor with the low-concentration sensitivity give?

SOLUTION. (a) The reciprocals are $1/c = 1$, 0.5, 0.25, 0.125, and 0.0625, and $1/I = 3.891$, 2.222, 1.389, 0.972, and 0.764. A least-squares line through these points has a slope of 3.336 mM/µA and an intercept of $0.555\,\mu A^{-1}$. So $I_{max} = 1/0.555 = 1.80\,\mu A$ and $K_m =$ slope $\times I_{max} = 3.336 \times 1.80 = 6.0\,mM$.

(b) $I_{max}/K_m = 1.80/6.0 = 0.30\,\mu A/mM$.

(c) $I = 1.80 \times 5/(6.0 + 5) = 0.82\,\mu A$. A linear sensor would give $0.30 \times 5 = 1.50\,\mu A$. At an ordinary blood level this electrode already delivers only 55 % of its linear response: it has saturated. ■

Saturation and the diffusion-limiting membrane

How far does the linear range reach? The true response divided by the low-concentration line is $K_m/(K_m + c)$, so a sensor stays within 10 % of linear only for $c \leq K_m/9$. A continuous glucose monitor that reads up to 22 mM would need an apparent K_m of 200 mM, while glucose oxidase in solution has a K_m of a few tens of millimolar. Oxygen is a second problem: the enzyme needs O_2 (Eq. (2.11)), and the tissue under the skin holds only about 0.06 mM, almost a hundred times less than the glucose (the *oxygen deficit*).

Designers solve both with an outer membrane that passes oxygen easily but restricts glucose so much that the enzyme, which could convert glucose faster than it arrives, keeps the glucose under the membrane near zero. Then the membrane, not the enzyme, sets the signal. By Fick's first law the glucose flux through a membrane of thickness

L is $J = Dc/L$. Each glucose molecule yields n electrons at the electrode (two, through the peroxide of Eq. (2.11)), and a mole of electrons carries the Faraday constant $F = 96\,485\,\mathrm{C}$, so the current from an electrode of area A is

$$I = nFA\,\frac{D}{L}\,c. \tag{2.15}$$

The current is now proportional to c over the whole range, and the response time is about $L^2/2D$. For a $0.5\,\mathrm{mm}^2$ electrode under a $20\,\mu\mathrm{m}$ polyurethane membrane with an effective $D = 1 \times 10^{-11}\,\mathrm{m}^2/\mathrm{s}$ (low, because glucose dissolves poorly in the polymer), the sensitivity is $48\,\mathrm{nA/mM}$ and the response time 20 s. A thinner membrane is faster and gives more current, but it lets in more glucose for the same oxygen. The membrane is the main design variable of a glucose sensor; Chapter 6 builds the electrode underneath it.

2.6 Nucleic Acids and Hybridization

DNA STORES THE INSTRUCTIONS for building every protein. It is a polymer of *nucleotides*, each made of a phosphate group, a five-carbon sugar (deoxyribose), and one of four *bases*: adenine (A), guanine (G), cytosine (C), and thymine (T). Phosphates and sugars alternate along a backbone with a direction, written from its 5′ end to its 3′ end. The bases stick out from the backbone, and the bases of two strands pair by hydrogen bonds:

KEY IDEA A pairs with T (U in RNA), G with C. A strand binds only its complement; this is the recognition rule of every DNA sensor.

$$\mathrm{A{:}T}\ \text{(two hydrogen bonds)}, \qquad \mathrm{G{:}C}\ \text{(three hydrogen bonds)}. \tag{2.16}$$

Two strands with complementary sequences, running in opposite directions, wind into the *double helix*: 2 nm in diameter, with a rise of 0.34 nm per base pair and about 10.5 base pairs per turn. A human cell has about 6.2×10^9 base pairs in its two sets of chromosomes, about 2 m of DNA packed into a nucleus 6 µm across.

Each phosphate carries one negative charge, so DNA is a strongly charged polymer. That charge moves DNA through a gel in electrophoresis, and it changes the surface potential of a transistor gate when DNA binds there (Chapter 5). In salt water, though, positive ions crowd around any negative charge and cancel its field beyond a short distance, the *Debye length*: about 0.8 nm in body fluid (150 mM salt) and 7.9 nm in a sample diluted a hundredfold. A sensor surface feels only the charge within about a Debye length of it.

RNA and the central dogma

RNA is the working copy. It uses ribose and has uracil (U) in place of thymine; U pairs with A. RNA is single stranded and short, and

it folds back on itself wherever two stretches are complementary, the property that aptamers use (below). Information flows from DNA to RNA to protein. *Replication* copies DNA before a cell divides; *transcription* copies one gene into messenger RNA (mRNA); and in *translation* the ribosome reads the mRNA three bases at a time, each three-base *codon* naming one amino acid, and links the amino acids into a protein. Some viruses copy their RNA into DNA with an enzyme called *reverse transcriptase*, which is also the first step of the COVID-19 PCR test (Section 2.8).

Melting and hybridization

Heat a DNA duplex and its strands separate, or *melt*; cool it and complementary strands find each other and re-pair, or *hybridize*. The melting temperature T_m is where half the duplexes have come apart. It rises with length, with G-C content (three hydrogen bonds), and with salt, which screens the repulsion between the two charged backbones. For short probes of 14 to 20 bases in about 1 M salt, the *Wallace rule* is a quick estimate:

$$T_m \approx 2\,(N_A + N_T) + 4\,(N_G + N_C)\ ^\circ\text{C}. \tag{2.17}$$

A single mismatched base pair in a short duplex lowers T_m by several degrees, roughly 1 to 1.5 °C for each percent of mismatched pairs. Running a hybridization or wash step a few degrees below the perfect-match T_m keeps the perfect duplexes and melts the mismatched ones (Fig. 2.6). This *stringency* is how a DNA sensor tells a single-letter mutation from the normal sequence.

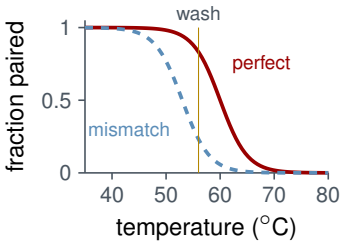


Figure 2.6: Melting curves of a 20-base probe paired with its perfect complement (T_m = 60 °C) and with a target that has one mismatch (T_m about 7 °C lower); illustrative two-state curves. At a wash temperature of 56 °C most perfect duplexes survive and most mismatched ones melt.

DNA and aptamer sensors

A *hybridization sensor* fixes single-stranded *probe* DNA to a surface (on gold, through a sulfur-bearing thiol group at one end) and waits for the complementary *target* to pair with it. The binding can be read by fluorescence (DNA microarrays), by refractive index (Chapter 7), by the charge it brings to a transistor gate (Chapter 5), or electrochemically (Chapter 6). In the last case a reporter molecule such as methylene blue on the probe's free end touches the electrode often while the probe is a flexible single strand; the rigid duplex holds it away, so hybridization lowers the current (Fig. 2.7a)[1].

An *aptamer* is a single strand of DNA or RNA, typically 20 to 80 bases, that folds into a shape that grips one target: a small molecule such as dopamine, serotonin, or cortisol, or a protein. Aptamers are selected from a library of about 10^{14} random sequences by rounds of binding, washing, and copying. When an electrode-bound aptamer

[1] C. Fan, K. W. Plaxco, A. J. Heeger, *Proc. Natl. Acad. Sci. USA* (2003).

folds around its target, its reporter moves closer to the surface and the current changes (Fig. 2.7b). Binding is reversible, so the sensor follows the concentration up and down without added reagents, as a wearable or implant must. Aptamer electrodes have tracked drug levels in the blood of live animals second by second[2], and aptamers on transistor gates detect small molecules at the salt strength of body fluid because the folding moves charge within a Debye length of the gate[3]. Compared with antibodies, aptamers are made chemically, are smaller, survive heating, and refold; enzymes in body fluids digest them, and for some targets they bind more weakly.

[2] B. S. Ferguson, ..., K. W. Plaxco, H. T. Soh, *Sci. Transl. Med.* (2013).

[3] N. Nakatsuka, ..., A. M. Andrews, *Science* (2018).

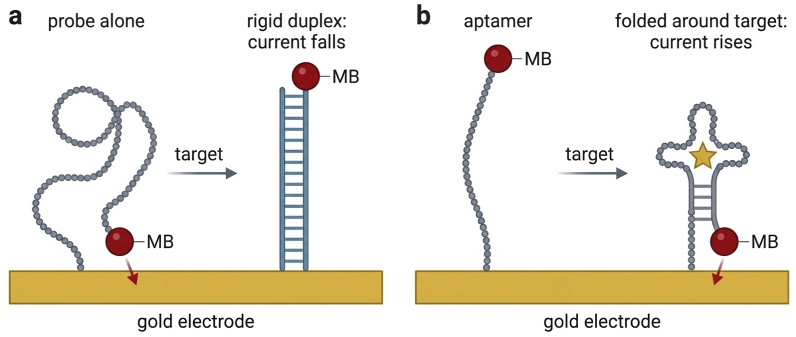


Figure 2.7: Two electrochemical DNA sensors with a methylene blue (MB) reporter. (a) Hybridization: a flexible probe lets MB reach the gold often; pairing with the target makes a rigid duplex that holds MB away, so the current falls. (b) An aptamer that folds around its target (gold star) and brings MB closer to the surface, so the current rises. Both need no added reagents and are reversible.

Worked Example 2.4 *A DNA probe for a sensor*

A sensor uses the 20-base probe

`5'-TGCATCGGAATCCGTAGCTA-3'`.

(a) Write the target sequence it detects. (b) Estimate T_m with the Wallace rule. (c) A mutant target has one mismatched base. Estimate its T_m and choose a wash temperature. (d) How long is the probe-target duplex, and how many negative charges does it carry? How much of that charge can a transistor gate under the probe feel in body fluid?

Solution. (a) The target is the complement read in the opposite direction, `5'-TAGCTACGGATTCCGATGCA-3'`.

(b) The probe has 5 A, 5 T, 5 G, and 5 C, so $T_m \approx 2(10) + 4(10) = 60\,^{\circ}\text{C}$.

(c) One mismatch in 20 pairs is 5 %, so T_m drops by about 5 to 7.5 °C, to about 53 to 55 °C. A wash at about 56 °C sits a few degrees below the perfect-match T_m and above the mismatch T_m (Fig. 2.6). The Wallace rule is good only to a few degrees, so a real design then tests the temperature on the bench.

(d) $20 \times 0.34\,\text{nm} = 6.8\,\text{nm}$. Each strand of 20 nucleotides has 19 phosphates between its bases, so the duplex carries 38 negative

charges. With a Debye length of 0.8 nm in body fluid, the gate feels only the charge in the first nanometer or so of the duplex; the rest is screened by the salt. Diluting the sample a hundredfold makes the Debye length 7.9 nm, comparable to the duplex, which is why transistor DNA sensors often measure in diluted buffer. ■

2.7 *Blood and Body Fluids*

SPIN A TUBE OF BLOOD in a centrifuge and it separates into layers (Fig. 2.8). The pale yellow liquid on top, about 55 % of the volume, is *plasma*: about 91 % water, 7 % proteins (albumin, the antibodies, and the clotting protein fibrinogen), and the rest ions, glucose, lipids, hormones, and waste. Plasma with the clotting proteins removed is *serum*. Its pH is held between 7.35 and 7.45. At the bottom, about 45 % of the volume, are the red cells; this fraction is the *hematocrit*. Between them is a thin whitish *buffy coat* of white cells and platelets, less than 1 % of the volume. A microliter of blood holds 4 to 6 million red cells, 4000 to 11 000 white cells, and 150 000 to 400 000 platelets; red cells outnumber white cells about 700 to 1.

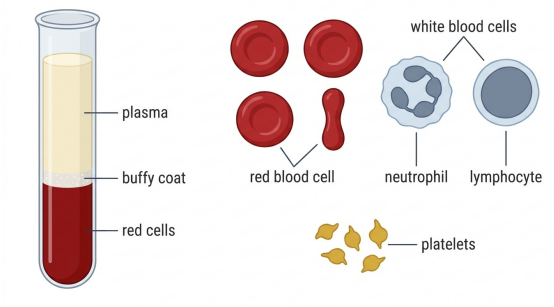


Figure 2.8: Blood. Left: a centrifuged sample separates into plasma, a thin buffy coat, and packed red cells. Right: red cells, white cells (larger, with a nucleus), and platelets.

Red cells and hemoglobin

A red cell (*erythrocyte*) is a biconcave disc about 8 µm across and 2 µm thick, with a volume of about 90 fL. It has no nucleus; it is a flexible bag of hemoglobin that folds to squeeze through capillaries as narrow as 3 µm. *Hemoglobin* is a protein of four chains (64.5 kDa), each holding an iron-containing heme group that binds one O_2, so one hemoglobin carries up to four oxygen molecules. A red cell holds about 280 million of them.

The four sites cooperate: binding one O_2 makes the next bind more easily. The fraction of sites filled, the *saturation* S, therefore follows an S shaped curve against the oxygen partial pressure P_{O_2} (Fig. 2.9), well

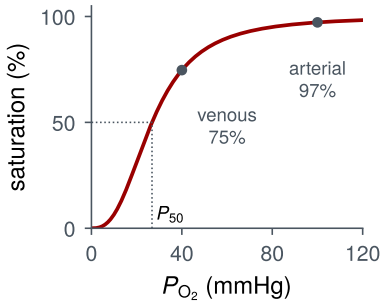


Figure 2.9: Oxygen-hemoglobin dissociation curve, Hill model with $P_{50} = 26.8$ mmHg and $n = 2.7$. Blood leaving the lungs is 97 % saturated; returning venous blood at 40 mmHg is about 75 % saturated.

described by the Hill equation

$$S = \frac{P_{O_2}^n}{P_{50}^n + P_{O_2}^n}, \qquad P_{50} \approx 26.8\,\text{mmHg}, \quad n \approx 2.7. \tag{2.18}$$

The S shape is good engineering: hemoglobin is nearly full at the 100 mmHg of the lungs, and the curve is steep at the 20 to 40 mmHg of working tissue, where a small drop in pressure releases a lot of oxygen; acid and CO_2 shift the curve to release still more. Carbon monoxide binds about 200 times more strongly than oxygen, and a two-wavelength pulse oximeter cannot tell the two apart, so it reads falsely high in carbon monoxide poisoning (Chapter 7).

The oxygen content of blood is the bound oxygen plus a little dissolved in plasma:

$$C_{O_2} = (1.34\,\text{mL/g})\,[\text{Hb}]\,S + (0.003\,\text{mL/dL/mmHg})\,P_{O_2}, \tag{2.19}$$

with [Hb] in g/dL and C_{O_2} in mL of O_2 per dL of blood. The constant 1.34 mL/g is the measured oxygen capacity of hemoglobin.

White cells (*leukocytes*) defend the body. Neutrophils, about half of them, engulf bacteria and are the first cells to reach an implant; lymphocytes (T cells, and B cells that make antibodies) mount immune responses aimed at one invader; monocytes leave the blood and become macrophages, which clean up debris and lead the body's response to an implant. *Platelets* are cell fragments that stick to an injury and start a clot. Every blood-contacting device, from a catheter to a dialyzer, must keep platelets from doing this on its surface.

The complete blood count

The *complete blood count* (CBC) is the most common laboratory test. An analyzer reports the red cell count (RBC), hemoglobin concentration (Hb), and hematocrit (Hct); three red cell indices computed from them,

$$\text{MCV} = \frac{\text{Hct}}{\text{RBC}}, \qquad \text{MCH} = \frac{\text{Hb}}{\text{RBC}}, \qquad \text{MCHC} = \frac{\text{Hb}}{\text{Hct}}, \tag{2.20}$$

the mean cell volume, mean cell hemoglobin, and mean cell hemoglobin concentration (normal ranges about 80 to 100 fL, 27 to 33 pg, and 32 to 36 g/dL); the white cell count with the fraction of each type; and the platelet count. A small MCV with low hemoglobin points to iron deficiency.

Most of the CBC is electrical. The *Coulter principle* (Fig. 2.10) counts and sizes cells by their resistance. A diluted sample flows through an aperture tens of micrometers wide. The saline in it has resistance $R = \rho L/A$ (resistivity ρ, length L, cross-sectional area A), and an insulating cell of volume v raises it by

$$\Delta R \approx \frac{3}{2}\,\frac{\rho\, v}{A^2}, \tag{2.21}$$

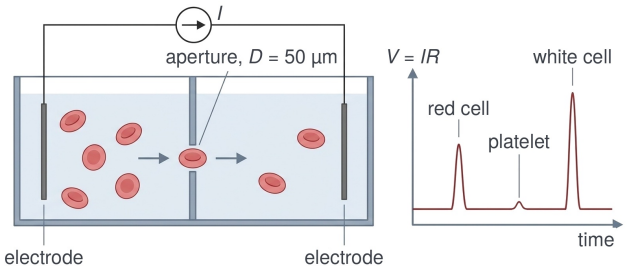


Figure 2.10: The Coulter principle. A constant current flows through a small aperture between two chambers of saline. Each cell that passes displaces conducting fluid and raises the aperture resistance for a moment. The pulse height is proportional to the cell volume, so one instrument counts and sizes red cells, platelets, and white cells.

where 3/2 accounts for the current bending around a small sphere. Pulse counts give RBC and the platelet count, pulse heights give MCV, and the hematocrit is computed as MCV $\times$ RBC; hemoglobin is measured optically (Chapter 7). An aperture 50 µm wide and 60 µm long in saline of $\rho = 0.67\,\Omega\,\text{m}$ has $R = 20.5\,\text{k}\Omega$, or 2.05 V at a constant 100 µA. A 90 fL red cell raises it by 23.5 Ω, a pulse of 2.35 mV, and a 10 fL platelet gives only 0.26 mV, one part in eight thousand of the baseline. Small cells need a small aperture ($\Delta R \propto 1/A^2$) and a quiet amplifier.

Plasma, interstitial fluid, and sweat

Wearable and implanted sensors rarely touch blood. Of the 42 L of water in a 70 kg adult, two thirds is inside cells, about 3 L is plasma, and about 11 L is *interstitial fluid* (ISF), the fluid between cells. Capillary walls pass water, ions, and small molecules but hold back most proteins, so ISF matches plasma in ions and glucose and is poor in protein (Table 2.3). It is the target of continuous glucose monitors and microneedle patches, and its glucose lags blood by several minutes because it must diffuse from the capillaries (Example 2.2). *Sweat* glands draw fluid from the ISF and reabsorb most of its sodium and chloride, so sweat is dilute and varies with the sweat rate, as Chapter 1 described along with the sweat chloride test for cystic fibrosis. Sweat is rich in lactate and carries glucose and hormones such as cortisol far below their blood levels.

2.8 *Bacteria and Viruses*

Bacteria are prokaryotes 0.5 to 5 µm long, most of them spheres, rods, or spirals. One large group, the Gram-negative bacteria, carries an outer membrane with *endotoxin*, a molecule that triggers fever even from dead bacteria, so implants are tested for endotoxin as well as sterility. Few species cause disease. A body carries about 3.8×10^{13} bacteria, mostly in the colon, about as many as its own cells[4]; this gut

[4] R. Sender, S. Fuchs, R. Milo, *PLoS Biol.* (2016).

	Plasma	Interstitial fluid	Sweat
Na^+ (mM)	135 to 145	about 140	10 to 90
K^+ (mM)	3.5 to 5.0	about 4	4 to 8
Cl^- (mM)	98 to 106	about 110	10 to 90
Glucose (mM)	4 to 5.5 (fasting)	close to plasma, lags several minutes	0.01 to 0.2
Lactate (mM)	0.5 to 2	about the same	5 to 40
Protein	60 to 80 g/L	much lower	trace
pH	7.35 to 7.45	about 7.4	4.5 to 7
Volume	3 L	11 L	up to 1 L/h in exercise
Reached by	needle or catheter	microneedle, implanted filament	patch on the skin

Table 2.3: Typical composition of the three fluids that body-worn sensors sample. Sweat values vary with the site, the sweat rate, and the person.

community is a target of the ingestible sensors of Chapter 1.

In a fresh medium a bacterial population grows in three phases (Fig. 2.11). After a lag, every cell divides once per *doubling time* t_d,

$$N(t) = N_0\, 2^{t/t_d}, \tag{2.22}$$

until food runs out. On an implant this arithmetic is alarming: ten bacteria that double every 30 min become $10 \times 2^{16} = 6.6 \times 10^5$ in 8 h and 10^8 in about 12 h, which is why implants are sterilized. Bacteria that reach a surface secrete a slimy matrix and become a *biofilm*, which tolerates a hundred to a thousand times more antibiotic than free bacteria do; the device usually has to be removed.

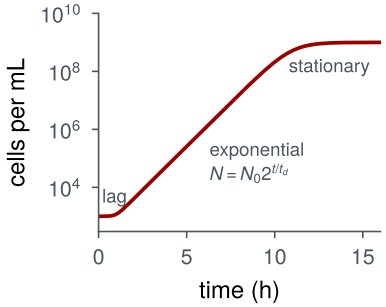


Figure 2.11: A bacterial culture (illustrative): a lag while the cells adapt, exponential growth with a fixed doubling time (here 30 min), and a stationary phase when nutrients run out.

Viruses

Viruses are not cells. A virus particle is a genome of DNA or RNA in a protein shell, sometimes wrapped in a lipid envelope taken from the host. Virus particles are 10 to 300 nm across, below the resolution of a light microscope (Fig. 2.2), and they reproduce only inside a host cell, using the cell's ribosomes and enzymes. SARS-CoV-2, the virus that causes COVID-19, is about 100 nm across and carries one strand of RNA about 30 000 bases long. Its *spike* protein binds a receptor on airway cells and lets the virus in, and the cell's ribosomes then build new virus particles from the viral RNA.

Tests for a virus and PCR

The three kinds of COVID-19 tests detect three different molecules. A *PCR test* detects viral RNA. Reverse transcriptase first copies the RNA into DNA, and the polymerase chain reaction (PCR) then doubles a target region of about a hundred bases in every temperature cycle: heat to about 95 °C to separate the strands, cool to about 60 °C so that short DNA *primers* bind the ends of the target, and let a polymerase enzyme extend them. A fluorescent probe reports the number of copies after each cycle. With an efficiency E per cycle ($E = 1$ is perfect doubling), N_0 starting copies become $N_n = N_0(1+E)^n$ after n cycles. The cycle at which the signal crosses a detection threshold N_{th} is the *cycle threshold*, $C_t = \log(N_{\text{th}}/N_0)/\log(1+E)$, so each tenfold more starting copies crosses $\log 10/\log(1+E)$ cycles earlier: 3.32 cycles at perfect efficiency. With a threshold of 10^{10} copies, a swab that delivers 1000 copies crosses at $C_t = \log_2 10^7 = 23.3$, and even a single copy would cross at 33.2, inside a run of 40 cycles. Amplification is why PCR can detect a few copies.

KEY EQUATION

$N_n = N_0(1+E)^n$

Copies after n PCR cycles with efficiency E ($E = 1$: perfect doubling).

An *antigen test* detects viral protein instead. In a lateral-flow strip, gold nanoparticles coated with antibodies pick up the viral protein, and a second antibody fixed in a line captures the pairs as a red test

line. Without amplification it needs far more virus than PCR, but it takes 15 min. An *antibody test* detects the patient's own antibodies, so it shows a past infection or vaccination.

2.9 Cells in Culture and on Electrodes

MOST BIOELECTRONIC DEVICES MEET CELLS in a dish long before they meet a patient. *Cell culture* is the growth of cells outside the body. *Primary cells*, taken directly from tissue, divide only a limited number of times; *cell lines* divide indefinitely. The first human line, HeLa, was grown in 1951 from a patient's tumor without her consent, an episode that shaped modern rules on consent (Chapter 14). Neurons and heart cells grown from human stem cells now let engineers test electrodes on human cells that fire and beat.

Cells grow in a salt solution, the *medium*, with glucose, amino acids, vitamins, serum, and a bicarbonate buffer (Fig. 2.12). It sits in an incubator at 37 °C and 5 % CO_2, which with the bicarbonate holds the pH near 7.4. *Sterile technique* keeps microbes out: work in a cabinet with filtered air, wipe surfaces with ethanol, and sterilize tools in an autoclave (121 °C steam) or solutions by filtration through 0.22 µm pores, small enough to stop bacteria. Device materials are tested on cells the same way: the ISO 10993 cytotoxicity test exposes cells to a material or its extract and counts the survivors (Chapter 14).

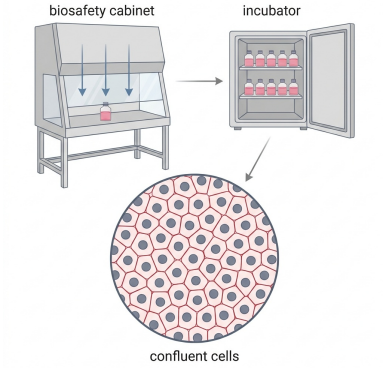


Figure 2.12: Working with cells in culture. Cells grow in a flask of medium in an incubator at 37 °C and 5 % CO_2 and are handled in a biosafety cabinet that blows filtered air over the work surface. Under the microscope, they spread into a single layer that eventually covers the surface (confluence).

Cells on electrodes

Cells are insulators wrapped around a conductor. Their membranes block current at low frequencies, and a cell growing on a surface sits only tens of nanometers above it. When cells cover an electrode, the current must squeeze through the thin gaps under and between them, so the impedance rises (Fig. 2.13). Electric cell-substrate impedance sensing[5] uses this to follow cells as they attach, spread, divide, and migrate, without labels. A small gold electrode and a large counter electrode share a culture well, and an instrument measures the impedance with an AC current of about a microampere, too small to disturb the cells. The metal surface itself behaves like a capacitor: ions gather in a layer about a nanometer thick against it, the *double layer* of Chapter 6, with tens of microfarads per square centimeter.

The same physics rates barrier tissues such as the lining of the gut by their *transepithelial electrical resistance* (TEER), measured across a cell layer grown on a porous membrane and reported as the resistance of the cell layer alone (the reading minus that of a membrane without cells) times the area, in $\Omega\,\mathrm{cm}^2$. On a microelectrode array, the same

[5] I. Giaever, C. R. Keese, *Proc. Natl. Acad. Sci. USA* (1984).

narrow, resistive gap under a neuron lets the cell's own membrane currents build a small voltage at the electrode: the extracellular signal that Chapter 9 records.

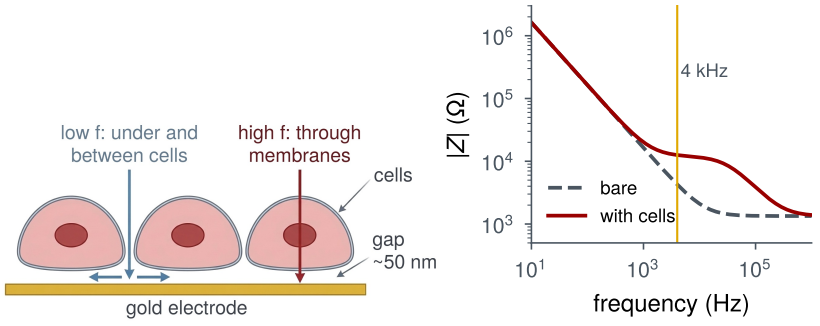


Figure 2.13: Cells on an electrode. (a) At low frequencies the membranes block the current, which must flow through the narrow gaps under and between the cells; at high frequencies it passes through the membranes capacitively. (b) Impedance of a 250 µm gold electrode, bare and covered by a confluent cell layer (illustrative model of Example 2.5); the difference is largest near 10 kHz.

Worked Example 2.5 *Detecting a cell layer by impedance*

A gold disk electrode 250 µm in diameter sits in culture medium. Model it as a double-layer capacitance of 20 µF/cm^2 in series with the resistance of the medium around a small disk, $R_s = \rho/(4a)$, with $\rho = 0.67\,\Omega\,\text{m}$ and radius a. A confluent cell layer adds, in series, the resistance of the gaps under and between the cells, $5\,\Omega\,\text{cm}^2$, in parallel with the capacitance of their membranes, $1\,\mu\text{F/cm}^2$. Find $|Z|$ at 4 kHz with and without cells (Fig. 2.13b).

Solution. The area is $\pi(125\,\mu\text{m})^2 = 4.91 \times 10^{-4}\,\text{cm}^2$, so $C_{\text{dl}} = 9.82\,\text{nF}$, $R_s = 0.67/(4 \times 125 \times 10^{-6}) = 1340\,\Omega$, $R_{\text{gap}} = 5/4.91 \times 10^{-4} = 10.2\,\text{k}\Omega$, and $C_{\text{cell}} = 0.491\,\text{nF}$.

Bare: at 4 kHz the double layer's reactance is $1/(2\pi f C_{\text{dl}}) = 4.05\,\text{k}\Omega$. A resistance and a reactance in series add at right angles (Chapter 1), so

$$|Z| = \sqrt{1.34^2 + 4.05^2}\ \text{k}\Omega = 4.27\,\text{k}\Omega.$$

With cells: the membranes' reactance is $1/(2\pi f C_{\text{cell}}) = 81\,\text{k}\Omega$, eight times R_{gap}, so nearly all the current takes the gaps and the cell layer acts as R_{gap} alone. It adds to the resistive part:

$$|Z| \approx \sqrt{(1.34 + 10.2)^2 + 4.05^2}\ \text{k}\Omega = 12.2\,\text{k}\Omega.$$

(Keeping the small current through the membranes gives 12.5 kΩ.) The impedance almost triples when the cells cover the electrode, and the change is mostly resistive; commercial systems watch the resistance near a few kilohertz. At 100 kHz the membranes' reactance falls to 3.2 kΩ, below R_{gap}; the current then passes through the cells, and the reading measures instead how much of the area they cover. ■

2.10 Cell Signaling

THE CELLS OF A BODY coordinate by sending and receiving signals, most of them molecules. A signal molecule binds a *receptor* protein on or in the target cell, which sets off a chain of reactions inside the cell. Each step can activate many molecules of the next, so the chain is a series of gain stages: one hormone molecule at the surface can produce thousands of messenger molecules inside. Water-soluble signals such as insulin act through receptors in the membrane; small oily signals such as cortisol diffuse through it and switch genes on and off, a slower response that takes hours.

Signals reach their targets in four ways, and the differences are physical: how far the message travels and what carries it. *Contact-dependent* signals stay on the membranes of two touching cells; immune cells use them to inspect the cells they touch. *Synaptic* signals are the job of neurons: an axon carries an electrical signal to its terminal, which releases a *neurotransmitter* that crosses a gap of about 20 nm to one target cell. *Paracrine* signals, such as the growth factors in a healing wound and the inflammatory *cytokines* (small signaling proteins) around an implant, diffuse to nearby cells. *Endocrine* signals are *hormones* (insulin, cortisol, adrenaline) carried by the blood to every tissue, where only cells with the matching receptor respond.

Diffusion sets the time scales (Eq. (2.4)). A transmitter ($D \approx 8 \times 10^{-10}\,\mathrm{m^2/s}$) crosses the synaptic gap in 0.25 µs, so diffusion does not limit synaptic speed. A growth factor ($D \approx 1.5 \times 10^{-10}\,\mathrm{m^2/s}$) needs half a minute to reach cells 100 µm away and an hour for 1 mm, so paracrine signals stay local. Blood circulates in about a minute (5 L pumped at 5 L/min), and a hormone's half-life in plasma sets how quickly its level can fall: about 5 min for insulin and about 1 h for cortisol. An insulin sensor would have to sample every minute or two; sampling cortisol every 10 to 15 min is enough. These messengers are also sensor targets: neurotransmitters such as dopamine and serotonin, cytokines that report inflammation, and hormones such as cortisol, found in sweat and interstitial fluid at nanomolar levels and measured with the affinity sensors of Chapter 6. Neurons and heart muscle cells also signal electrically, with the action potentials of the next chapter.

Summary

- A cell is a membrane-wrapped bag of cytoplasm, about 70 % water, 10 to 30 µm across (bacteria 0.5 to 5 µm). Eukaryotes keep their DNA in a nucleus; prokaryotes have no nucleus.

- The lipid bilayer is a capacitor, $C_m = \varepsilon_0 \varepsilon_r / d \approx 1\,\mu\mathrm{F/cm^2}$, in parallel with R_m of 1 to 100 $\mathrm{k\Omega\,cm^2}$; $\tau = R_m C_m$ does not depend on size. A resting potential needs only 1 ion in 64 000. Channels, carriers, and pumps carry what the lipid cannot.
- Diffusion: $t \approx x^2/2D$; Fick's first law $J = -D\,\partial c/\partial x$ is Ohm's law for molecules; $D = k_B T / 6\pi\eta r_h$. Oxygen consumption limits living tissue to $\sqrt{2Dc_0/Q}$, a few hundred micrometers. Osmotic pressure $\Pi = c_{\mathrm{osm}} RT$ is about 7.5 bar for body fluid.
- Glucose: 1 mM = 18.0 mg/dL; fasting blood 4 to 5.5 mM; sweat about a hundredth of that.
- Enzymes follow $v = V_{\max}[\mathrm{S}]/(K_m + [\mathrm{S}])$ and are within 10 % of linear only for $c \le K_m/9$, so glucose sensors add a diffusion-limiting membrane, $I = nFADc/L$, with response time $L^2/2D$.
- DNA pairs A with T and G with C, 0.34 nm per base pair, one negative charge per phosphate, screened beyond the Debye length (0.8 nm in body fluid). $T_m \approx 2(\mathrm{A}+\mathrm{T}) + 4(\mathrm{G}+\mathrm{C})$ for short probes; a mismatch lowers it by several degrees.
- Blood is 55 % plasma and 45 % red cells; it carries $1.34[\mathrm{Hb}]S + 0.003P_{\mathrm{O_2}}$, about 20 mL/dL of oxygen. A Coulter counter sees each cell as $\Delta R = \frac{3}{2}\rho v / A^2$. ISF matches plasma in small solutes and lags it by minutes; sweat is dilute.
- Bacteria grow as $N_0 2^{t/t_d}$ and form biofilms on implants. PCR copies a target as $N_0(1+E)^n$; each decade of input moves C_t by 3.3 cycles.
- Cells on an electrode raise its impedance, most clearly near 10 kHz. Cells signal by contact, synapses, paracrine diffusion, and hormones, on time scales from microseconds to hours.

2.11 *Problems*

2.1 • Osmosis.

(a) Find the osmolarity of 5 % glucose (50 g/L). Is it isotonic? What happens to its tonicity once the cells take up the glucose?

(b) A red cell has a volume of 90 fL at 290 mOsm/L, and 40 % of it is osmotically inactive. Find its volume in a 600 mOsm/L solution.

(c) Find the osmotic pressure difference, in bar, between plasma (290 mOsm/L) and a 250 mOsm/L solution at 37 °C.

2.2 ◦ An enzyme electrode has $I_{\max} = 2.0\,\mu\mathrm{A}$ and $K_m = 10\,\mathrm{mM}$.

(a) Find the current at 1, 10, and 100 mM.

(b) At what concentration does it reach 90 % of $I_{\max}$?

(c) Up to what concentration is it within 10 % of linear?

2.3 •• Designing a glucose membrane. A glucose electrode with an area of $0.5\,\mathrm{mm}^2$ and $n = 2$ sits under a diffusion-limiting membrane (Eq. (2.15)). It needs a sensitivity of $20\,\mathrm{nA/mM}$.

(a) With $D = 1 \times 10^{-11}\,\mathrm{m}^2/\mathrm{s}$, what membrane thickness does that take, and what is the response time?

(b) Show that at fixed sensitivity the response time is proportional to D. Find the thickness and response time with a less permeable polymer, $D = 2.5 \times 10^{-12}\,\mathrm{m}^2/\mathrm{s}$.

(c) What practical problems limit how thin a membrane can be?

2.4 • A patient's CBC shows RBC $= 3.8 \times 10^6$ per µL, Hb $= 9.5\,\mathrm{g/dL}$, and Hct $= 30\,\%$.

(a) Find MCV, MCH, and MCHC. Which are abnormal?

(b) Find the arterial oxygen content at $S = 98\,\%$ and 100 mmHg, and compare it with the normal $19.8\,\mathrm{mL/dL}$.

(c) What would the heart have to do to deliver the same oxygen per minute?

2.5 • TEER. A cell layer on a porous insert of area $0.33\,\mathrm{cm}^2$ reads $650\,\Omega$; the same insert without cells reads $140\,\Omega$.

(a) Find the TEER in $\Omega\,\mathrm{cm}^2$.

(b) Explain why TEER is multiplied, not divided, by the area.

(c) What would the same tissue read on a $1.12\,\mathrm{cm}^2$ insert whose blank is also $140\,\Omega$?

3
Neurons and the Action Potential

In this chapter you will learn to

- Name the parts of a neuron and state the job of each, from the dendrites to the synapse.
- Explain how a concentration gradient across a membrane makes a voltage, and compute Nernst potentials for K^+, Na^+, Cl^-, and Ca^{2+} at body temperature.
- Find the resting potential with the Goldman equation, and explain why it is negative and what the Na^+/K^+ pump does.
- Draw the membrane as a circuit of batteries, conductances, and a capacitor, and use it to find the resting potential and the ion currents.
- Classify ion channels by what opens them, compute the current through one channel, and explain how the voltage clamp and the patch clamp measure it.
- Describe each phase of the action potential in terms of the sodium and potassium conductances.
- Explain how graded potentials add, and use the length constant to estimate how far a passive signal spreads.
- Explain how myelin speeds conduction, and estimate conduction times along nerves and around a reflex.
- Describe a chemical synapse, and estimate the current and pulse width an electrode needs to bring a neuron to threshold.

Tap the tendon below your kneecap and within about twenty milliseconds your thigh muscle is contracting, before you have decided anything. In that time a stretch sensor in the muscle has fired, an electrical pulse has run more than half a meter up a nerve fiber to

the spinal cord, crossed one junction to a motor neuron, and run back down to the muscle. Every step is ionic current flowing through proteins in a membrane 5 nm thick, and every step can be measured with electrodes and amplifiers.

This chapter explains that signal from the ground up, using only the circuit essentials of Chapter 1 (resistors, capacitors, batteries, and the RC time constant) and the cell membrane of Chapter 2. After the neuron's anatomy come the ion gradients that charge the membrane like a battery, the Nernst and Goldman equations that give its voltage, and the membrane drawn as a circuit. Ion channels are the switches in that circuit, and the voltage clamp and patch clamp measured them. With those pieces the action potential follows. Then we treat the axon as a leaky cable, see why myelin and its nodes act like repeaters on a long line, and end with nerves, reflexes, and synapses, the doorway for the recording and stimulation devices of Chapter 9.

3.1 *The Nervous System and the Neuron*

THE NERVOUS SYSTEM HAS TWO PARTS. The *central nervous system* (CNS) is the brain and spinal cord. The *peripheral nervous system* (PNS) is everything else: the sensory nerves that bring information in, the motor nerves that drive muscles, and the autonomic nerves that regulate the heart, gut, and glands. The human brain has about 86 billion neurons[1], about as many as the transistors on a large processor, but each neuron connects to thousands of others.

[1] F. A. C. Azevedo *et al., J. Comp. Neurol.* (2009). The often quoted 100 billion is an overestimate.

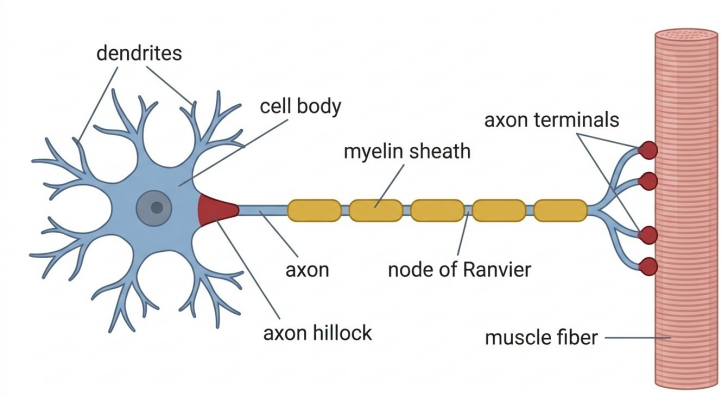


Figure 3.1: A motor neuron. Dendrites collect inputs and carry them to the cell body. The axon begins at the axon hillock, where action potentials start, and is wrapped in myelin broken by nodes of Ranvier. Its terminals form synapses on a muscle fiber.

A neuron is a cell built to receive, combine, and send signals. Its parts (Fig. 3.1) are these:

- *Dendrites*, branched and often hundreds of micrometers long, receive inputs at synapses and carry small voltage changes to the cell body.

- The *cell body* (soma), 10 to 100 µm across, holds the nucleus and makes the proteins for the whole cell.
- The *axon hillock*, where the axon leaves the soma, and the first stretch of axon hold the densest voltage-gated sodium channels. They form the *trigger zone*: the summed inputs fire an action potential here or not at all.
- The *axon*, 0.2 to 20 µm in diameter in humans, carries the action potential away, from less than a millimeter to more than a meter (from the spinal cord to the toes).
- *Myelin* is insulating membrane wrapped around many axons by glial cells (Schwann cells in the PNS, oligodendrocytes in the CNS).
- *Nodes of Ranvier* are gaps in the myelin, about 1 µm long and 0.2 to 2 mm apart, where the action potential is regenerated.
- *Axon terminals* end in *synapses*, where the electrical signal releases a chemical transmitter onto the next neuron, a muscle, or a gland.

An engineer can read the neuron as a signal chain: the dendrites are inputs, the soma an adder, the trigger zone a comparator that fires when the sum crosses a threshold, the axon a transmission line, the nodes repeaters, and the terminals outputs. The rest of this chapter explains each stage with a circuit.

3.2 *Ions and the Nernst Equation*

A RESTING NEURON IS A CHARGED BATTERY. Its inside sits about 70 mV below the outside. The voltage comes from two facts: the concentrations of ions inside and outside the cell are very different (Table 3.1), and the resting membrane lets some ions through much more easily than others.

Ion	Inside (mM)	Outside (mM)	Ratio out/in	E_{ion} at 37 °C
K^+	140	5	1/28	−89 mV
Na^+	12	145	12	+67 mV
Cl^-	7	110	16	−74 mV
Ca^{2+}	0.0001	2	20 000	+132 mV

Table 3.1: Typical ion concentrations in a mammalian neuron, in millimoles per liter (mM), and the Nernst potentials they give. Large anions that cannot cross the membrane (proteins, phosphates) balance the charge inside. Real values vary by cell type, especially for Cl^- (4 to 30 mM inside).

Diffusion against a voltage

Suppose the membrane let only K^+ through. Potassium is 28 times more concentrated inside, so random thermal motion carries more K^+ ions out than in: the ions diffuse down their concentration gradient (Chapter 2). Each K^+ that leaves carries a positive charge out and

leaves an unbalanced negative charge behind, on the large anions that cannot follow. The inside becomes negative, and that voltage pulls K^+ back in. The outflow stops when the electrical pull balances the diffusional push. The voltage at which the two balance is the *equilibrium potential*, or *Nernst potential*, of the ion:

$$E_{\text{ion}} = V_{\text{in}} - V_{\text{out}} = \frac{RT}{zF} \ln \frac{c_{\text{out}}}{c_{\text{in}}}. \tag{3.1}$$

KEY EQUATION

$E = \frac{RT}{zF} \ln \frac{c_{\text{out}}}{c_{\text{in}}}$

Nernst potential: 26.7 mV times the log ratio at 37 °C, or 61.5 mV per decade for a monovalent ion.

Here c_{out} and c_{in} are the concentrations outside and inside, z is the charge number of the ion (+1 for K^+ and Na^+, −1 for Cl^-, +2 for Ca^{2+}), R is the gas constant, F is the Faraday constant (the charge of a mole of electrons, 96 485 C/mol), and T is the absolute temperature. The equation comes from Boltzmann's law: in equilibrium the ratio $c_{\text{in}}/c_{\text{out}}$ across a voltage $V = V_{\text{in}} - V_{\text{out}}$ is $e^{-zFV/RT}$, and taking the logarithm gives Eq. (3.1).

The combination RT/F is the *thermal voltage*, at which the electrical energy of one elementary charge equals the typical thermal energy k_BT of a molecule: 25.7 mV at 25 °C and 26.7 mV at body temperature, 37 °C. (Electrical engineers meet it as k_BT/q in the diode equation.) Each tenfold ratio adds the same voltage: in base-10 form $E = (61.5\,\text{mV}/z)\log_{10}(c_{\text{out}}/c_{\text{in}})$ at body temperature. The sign tells which side ends up positive. A cation more concentrated inside, like K^+, makes the inside negative; a cation more concentrated outside, like Na^+, makes it positive; and the negative charge of Cl^- flips the sign.

Only a tiny charge moves: Chapter 2 found that the membrane charge of a resting cell is about 1 of its potassium ions in 64 000, so the concentrations do not change. The same equation sets the potential of an electrode in Chapter 6.

Worked Example 3.1 *Nernst potentials at body temperature*

Find the Nernst potentials of K^+, Na^+, Cl^-, and Ca^{2+} for the concentrations of Table 3.1 at 37 °C.

Solution. $RT/F = (8.314)(310.15)/96485 = 26.73\,\text{mV}$.

$$\begin{aligned}
E_{\text{K}} &= 26.73 \ln(5/140) = 26.73 \times (-3.332) = -89.1\,\text{mV},\\
E_{\text{Na}} &= 26.73 \ln(145/12) = 26.73 \times 2.492 = +66.6\,\text{mV},\\
E_{\text{Cl}} &= \frac{26.73}{-1} \ln(110/7) = -26.73 \times 2.755 = -73.6\,\text{mV},\\
E_{\text{Ca}} &= \frac{26.73}{2} \ln(2/0.0001) = 13.36 \times 9.903 = +132.3\,\text{mV}.
\end{aligned}$$

K^+ pulls the inside negative; Na^+ and Ca^{2+} pull it positive. Although Cl^- is concentrated outside, its negative charge makes its potential negative, close to rest. A membrane voltage between E_{K} and E_{Na} means that K^+ is pushed out and Na^+ pulled in at the

same time. ■

3.3 *The Resting Potential*

A REAL MEMBRANE PASSES SEVERAL IONS AT ONCE, each in proportion to its *permeability* P, a measure of how easily that ion crosses. The steady voltage at which the currents of all of them add to zero is given by the *Goldman equation* (also called the Goldman-Hodgkin-Katz, or GHK, equation):

$$V_m = \frac{RT}{F} \ln \frac{P_{\mathrm{K}}[\mathrm{K}^+]_o + P_{\mathrm{Na}}[\mathrm{Na}^+]_o + P_{\mathrm{Cl}}[\mathrm{Cl}^-]_i}{P_{\mathrm{K}}[\mathrm{K}^+]_i + P_{\mathrm{Na}}[\mathrm{Na}^+]_i + P_{\mathrm{Cl}}[\mathrm{Cl}^-]_o}. \tag{3.2}$$

KEY IDEA The membrane voltage is a weighted average of the Nernst potentials, weighted by permeability. At rest K^+ dominates; at the peak of an action potential Na^+ does.

It looks like the Nernst equation with each concentration weighted by its permeability. Chloride appears with inside and outside swapped because of its negative charge. If one permeability is much larger than the others, the equation reduces to that ion's Nernst potential. At rest a neuron's membrane is most permeable to K^+, through potassium *leak* channels that are always open, with $P_{\mathrm{K}} : P_{\mathrm{Na}} : P_{\mathrm{Cl}} \approx 1 : 0.04 : 0.45$ (the ratios that Hodgkin and Katz measured on the squid giant axon in 1949, used in this chapter with mammalian concentrations). The resting potential therefore sits near, but above, E_{K}.

WORKED EXAMPLE 3.2 *The resting potential*

(a) Find the resting potential with $P_{\mathrm{K}} : P_{\mathrm{Na}} : P_{\mathrm{Cl}} = 1 : 0.04 : 0.45$ and the concentrations of Table 3.1. (b) At the peak of an action potential the sodium permeability rises about 500-fold, to $P_{\mathrm{Na}}/P_{\mathrm{K}} = 20$. Find the membrane voltage.

SOLUTION. (a) The numerator is $5 + 0.04 \times 145 + 0.45 \times 7 = 5 + 5.8 + 3.15 = 13.95$; the denominator is $140 + 0.04 \times 12 + 0.45 \times 110 = 140 + 0.48 + 49.5 = 189.98$. So

$$V_m = 26.73 \ln(13.95/189.98) = 26.73 \times (-2.611) = -69.8\,\mathrm{mV}.$$

Without the chloride terms the result is -68.6 mV: chloride, being close to equilibrium, barely moves the answer.

(b) Now the numerator is $5 + 20 \times 145 + 3.15 = 2908$ and the denominator $140 + 240 + 49.5 = 429.5$, so $V_m = 26.73 \ln(6.771) = +51.1$ mV. Simply changing which ion the membrane favors swings the voltage by about 120 mV, from near E_{K} to near E_{Na}. That is the whole action potential in one line. ■

The membrane as a circuit

To an engineer the membrane is a circuit (Fig. 3.2). Physiologists describe open channels by their *conductance* $g = 1/R$, measured in siemens (S): 1 nS is the conductance of a 1 GΩ resistor. Conductances in parallel simply add, so opening more channels adds conductance. Each ion species is a battery at its Nernst potential in series with the conductance of its open channels, and all the branches sit in parallel with the membrane capacitance C_m, about 1 µF/cm^2 (Chapter 2). The voltage across one branch's conductance is the membrane voltage minus its battery voltage, so by Ohm's law the branch carries

$$I_{\text{ion}} = g_{\text{ion}}(V_m - E_{\text{ion}}), \tag{3.3}$$

with positive current flowing outward. When the voltage is steady no current flows into the capacitor, so the branch currents must add to zero (Kirchhoff's current law). Solving $\sum g_{\text{ion}}(V_m - E_{\text{ion}}) = 0$ for V_m gives the conductance-weighted average

$$V_m = \frac{g_{\text{K}}E_{\text{K}} + g_{\text{Na}}E_{\text{Na}} + g_{\text{Cl}}E_{\text{Cl}}}{g_{\text{K}} + g_{\text{Na}} + g_{\text{Cl}}}, \tag{3.4}$$

the circuit version of the Goldman equation.

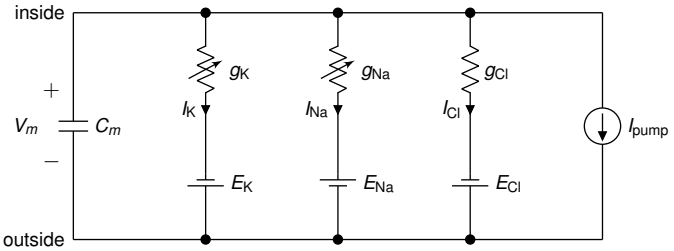


Figure 3.2: Parallel-conductance model of the membrane, inside at the top. Each ion is a battery at its Nernst potential in series with the conductance of its open channels, and its branch carries $I = g(V_m - E)$, positive outward; the pump is a small current source that keeps the batteries charged. The arrows through g_{K} and g_{Na} mark the voltage-gated conductances of the action potential (Section 3.6).

The capacitor matters whenever the voltage changes. A step of current into the cell charges C_m through the parallel conductances with the RC time constant of Chapter 1, $\tau = R_m C_m$, where R_m is the resistance of one square centimeter of membrane: 20 ms for $R_m = 20\,\text{k}\Omega\,\text{cm}^2$ and $C_m = 1\,\mu\text{F/cm}^2$, whatever the cell's size. It sets how fast every small, slow signal of the neuron rises and decays.

The pump

At rest the batteries are not in equilibrium: K^+ leaks out and Na^+ leaks in, continuously, and left alone the gradients would run down over minutes to hours. The Na^+/K^+ pump of Chapter 2 recharges them by moving 3 Na^+ out and 2 K^+ in for each ATP molecule it spends. Because it moves one net charge out per cycle it also adds a

few millivolts of negative voltage directly, but its main job is to keep the batteries charged.

This settles a common misconception. The resting potential is negative because the membrane is mostly permeable to K^+: potassium leaks out through open channels and leaves negative charge behind until the voltage stops it near E_K. The pump only maintains the gradient. The membrane is not charged by sodium ions stuck in its wall; the charge sits on the two faces of the membrane capacitor, positive outside and negative inside.

WORKED EXAMPLE 3.3 *Leak currents and the pump*

A neuron at rest has $g_K = 5\,\text{nS}$, $g_{Na} = 0.75\,\text{nS}$, and $g_{Cl} = 1\,\text{nS}$, with the Nernst potentials of Worked Example 3.1. (a) Find V_m. (b) Find each ion current and check that they cancel. (c) How many Na^+ ions leak in per second, and how many pump cycles per second keep up with them?

SOLUTION. (a) Equation (3.4):

$$V_m = \frac{5(-89.1) + 0.75(66.6) + 1(-73.6)}{6.75} = \frac{-469.0}{6.75} = -69.5\,\text{mV}.$$

(b) From Eq. (3.3), $I_K = 5\,\text{nS} \times (-69.5 + 89.1)\ \text{mV} = +98\,\text{pA}$ (outward), $I_{Na} = 0.75\,\text{nS} \times (-69.5 - 66.6)\ \text{mV} = -102\,\text{pA}$ (inward), and $I_{Cl} = 1\,\text{nS} \times 4.1\,\text{mV} = +4\,\text{pA}$. They sum to zero, as they must at steady state.

(c) $102\,\text{pA}/e = 6.4 \times 10^8$ sodium ions per second leak in. The pump moves 3 per cycle, so it must run 2.1×10^8 cycles per second, spending as many ATP molecules. At about a hundred cycles per second per pump (Chapter 2), that is two million pumps for one neuron at rest; firing adds more. ■

3.4 *Ion Channels and Stimuli*

AN ION CHANNEL IS A PROTEIN that forms a water-filled pore through the membrane. A narrow *selectivity filter* lets one kind of ion through and rejects others; a potassium channel passes K^+ about a thousand times more readily than the smaller Na^+, because its filter is lined with oxygen atoms spaced to replace exactly the water shell of a K^+ ion. A *gate* opens and closes the pore. Channels are named for what moves the gate (Fig. 3.3):

- *Leak* channels are open most of the time. Potassium leak channels set the resting potential.

- *Voltage-gated* channels have charged segments that move in the membrane's electric field, so depolarization opens them. Voltage-gated Na^+, K^+, and Ca^{2+} channels make and shape the action potential.
- *Ligand-gated* channels open when a molecule binds, usually a neurotransmitter on the outside (acetylcholine at muscle, glutamate and GABA in the brain). They turn chemical signals into currents at synapses.
- *Mechanically gated* channels open when the membrane is stretched, in touch receptors, muscle spindles, and the hair cells of the ear (Chapter 10).

Other channels open with temperature (the heat and cold sensors of the skin) or, in engineered cells, with light (the optogenetic tools of Chapter 9). The stimuli that excite neurons are therefore chemical, electrical, or mechanical, and every neurotechnology device reaches the nervous system through one of these routes.

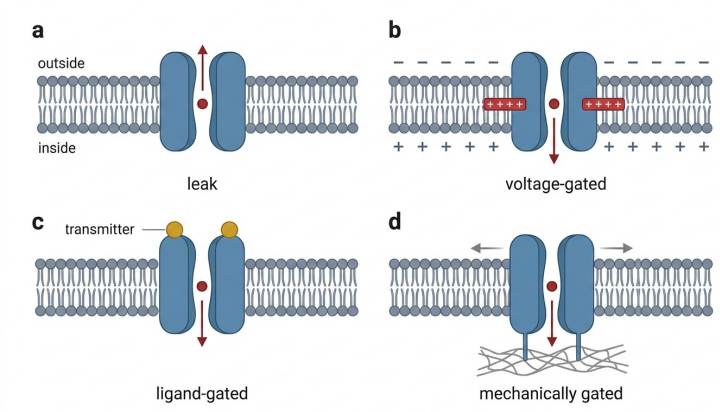


Figure 3.3: Four kinds of ion channel gating. A leak channel is usually open. A voltage-gated channel opens when charged parts of the protein move in the field of a depolarized membrane. A ligand-gated channel opens when a transmitter binds. A mechanically gated channel is pulled open by stretch.

One open channel is a tiny branch of Fig. 3.2: a small conductance γ in series with the battery of its ion, so its current is

$$i = \gamma\,(V_m - E_{\text{ion}}). \tag{3.5}$$

The quantity $V_m - E_{\text{ion}}$ is the *driving force*. Single-channel conductances range from about 1 to 250 pS (picosiemens); many voltage-gated K^+ and Na^+ channels have 10 to 20 pS. With driving forces of tens of millivolts, one channel passes a few picoamperes, about 10^7 ions per second, and it flickers between open and closed at random. The smooth currents of a whole cell are the sum of thousands of such flickering switches. At exactly E_{ion} an open channel passes no net current, because the electrical and diffusional forces on its ion balance. Opening a channel therefore always pulls the membrane toward that ion's Nernst potential.

3.5 *Voltage Clamp and Patch Clamp*

Everything in this chapter was measured with two feedback circuits. Around 1950 Kenneth Cole, and then Alan Hodgkin and Andrew Huxley, threaded a wire along the inside of a squid giant axon, half a millimeter thick, and used an amplifier to hold the membrane voltage at a chosen value, the *voltage clamp*. The clamp is a negative feedback loop (Fig. 3.4), like a thermostat: an amplifier compares the membrane voltage with a command voltage and pushes current in or out until they match. With the voltage held fixed, no current goes into the membrane capacitance and the channels cannot run away with positive feedback, so the current the amplifier supplies equals the current through the open channels. By stepping the voltage and changing the ions in the bath, Hodgkin and Huxley separated the sodium and potassium currents and built the model of Section 3.6.

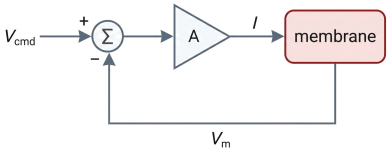


Figure 3.4: The voltage clamp as a feedback loop. A high-gain amplifier injects whatever current I holds V_m equal to the command voltage; at steady state I equals the ionic current through the membrane.

In 1976 Erwin Neher and Bert Sakmann shrank the idea to a patch of membrane small enough to hold one channel[2]. A glass pipette with a tip about 1 µm across, filled with salt solution, is pressed against a cell. Their first seals were only tens of megohms. Around 1980 they found that gentle suction fuses glass and lipid into a seal of more than 1 GΩ, a *gigaseal*, so almost all the current through the patch under the tip flows into the pipette instead of leaking around the rim. The patch can be recorded on the intact cell (*cell-attached*, with the pipette touching the outside face of the patch), torn off the cell, or broken open so that the pipette reaches the interior and records the whole cell (*whole-cell*).

The patch-clamp amplifier holds the pipette at the command voltage and turns the few picoamperes of one channel into a voltage: the current flows through a feedback resistor R_f, typically 10 GΩ, and the output is IR_f, 10 mV per picoampere. Chapter 4 builds this current-to-voltage converter, the transimpedance amplifier, from an op amp, and shows why the resistor must be so large: a smaller one adds more thermal noise current than the signal.

In whole-cell mode the narrow pipette tip adds a series resistance R_s of a few megohms between the amplifier and the cell. A current I through it drops a voltage IR_s, so the membrane sits IR_s away from the command, and the membrane follows a step command with the time constant R_sC of the RC circuit formed by R_s and the cell's capacitance C.

[2] E. Neher, B. Sakmann, *Nature* (1976); the gigaseal: O. P. Hamill *et al.*, *Pflügers Arch.* (1981). Hodgkin and Huxley shared the 1963 Nobel Prize; Neher and Sakmann the 1991 prize.

3.6 *The Action Potential*

INPUTS THAT PUSH THE MEMBRANE above a *threshold*, typically about $-55\,\mathrm{mV}$, trigger an *action potential*: a pulse about 100 mV tall and 1 to 2 ms long in mammalian neurons. It is all or none. A stimulus slightly below threshold does nothing lasting; one slightly above produces the full pulse, and a larger stimulus produces the same pulse. The sequence (Fig. 3.5):

1. *Threshold.* A depolarization opens some voltage-gated Na^+ channels. If the inward sodium current exceeds the outward leak current, the membrane depolarizes further and opens more Na^+ channels: positive feedback.
2. *Rising phase.* Sodium rushes in and the voltage climbs within a fraction of a millisecond toward E_{Na}.
3. *Overshoot.* The inside becomes positive, typically +30 to +40 mV, short of E_{Na} because some K^+ channels are open too.
4. *Repolarization.* Two slower processes end the pulse: the Na^+ channels *inactivate* (a second gate plugs the pore), and delayed voltage-gated K^+ channels open. Potassium flows out and the voltage falls.
5. *After-hyperpolarization.* The K^+ channels close slowly, so for a few milliseconds the voltage dips below rest toward E_K.
6. *Refractory periods.* While the Na^+ channels are inactivated no stimulus can fire a second pulse (the *absolute* refractory period); as they recover, a larger than normal stimulus can (the *relative* refractory period). Refractoriness limits the firing rate to a few hundred hertz and makes a pulse travel only forward along the axon.

The Na^+/K^+ pump is not part of this sequence, because each pulse moves very few ions. The membrane is a capacitor, so the least charge that swings it by 100 mV is $\Delta Q = C_m \Delta V = 100\,\mathrm{nC/cm^2}$, or 6.2×10^{11} sodium ions per square centimeter. In a squid giant axon 500 µm across that raises the internal sodium (50 mM) by two parts per million; in a thin axon 0.5 µm across (12 mM) it raises it by 0.7 %. The real influx is a few times this minimum, but a large axon can fire for a long time without its pump, while a thin one depends on the pump within tens of spikes.

The Hodgkin-Huxley picture

Hodgkin and Huxley described the squid membrane with the circuit of Fig. 3.2, but with sodium and potassium conductances that change with voltage and time. Kirchhoff's current law at the inside of the

membrane says that any injected current not carried away by the ion branches charges the capacitor:

$$C_m \frac{dV}{dt} = I_{ext} - g_{Na}(V - E_{Na}) - g_K(V - E_K) - g_L(V - E_L), \qquad (3.6)$$

where g_L is a small constant leak. All the physiology is in how g_{Na} and g_K behave, which Hodgkin and Huxley measured with the voltage clamp. Each behaves like a switch with a delay. A sodium channel has a fast *activation* gate that opens within a fraction of a millisecond of a depolarization, and a slower *inactivation* gate that closes about a millisecond later and stays closed until the membrane repolarizes. A potassium channel has one slow activation gate that opens a millisecond or so after the depolarization and closes slowly. Hodgkin and Huxley described each gate by an opening probability that relaxes toward a voltage-dependent value at a voltage-dependent rate, fitted those rates to their clamp data, and integrated Eq. (3.6) on a hand-cranked calculator.

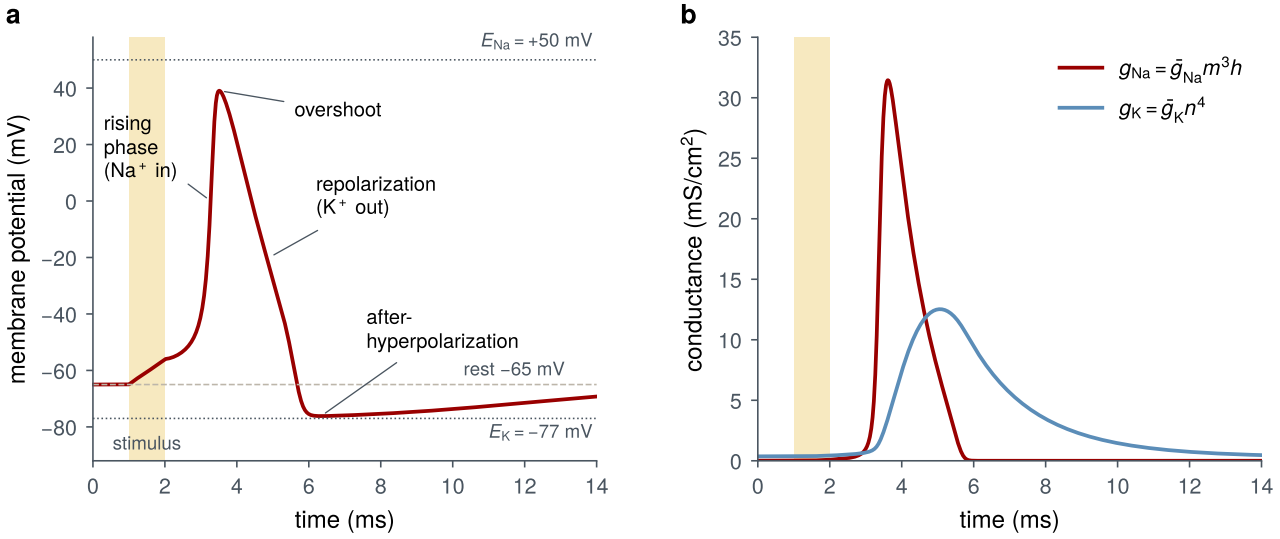


Figure 3.5: An action potential computed from the Hodgkin-Huxley model of the squid axon at 6.3 °C for a 10 µA/cm², 1 ms current pulse (gold band). (a) The membrane potential rises from −65 mV to a peak of +39 mV and undershoots to −76 mV, close to E_K. (b) The sodium conductance rises first and inactivates within about a millisecond; the potassium conductance rises later and decays slowly, which produces the after-hyperpolarization. In the legend, m, h, and n are the opening probabilities of the gates.

Figure 3.5 shows the result, computed today in a fraction of a second. (The squid's ion concentrations give $E_{Na} = +50$ mV and $E_K = -77$ mV, and its rest is −65 mV.) The model reproduces every feature of the list above, and it gives numbers. A 1 ms pulse must exceed about 6.9 µA/cm² to fire: 6.8 only nudges the membrane, and 7.0 fires a full pulse. A second pulse within about 6 ms of the first cannot fire at any amplitude, and over the next ten milliseconds it needs extra current. These times belong to a squid at 6.3 °C. Channel gates speed up about threefold for every 10 °C, so at 37 °C they would be about 29 times faster; mammalian neurons, with their own channels, fire pulses

about 1 ms wide with absolute refractory periods of about 1 ms.

3.7 *Graded Potentials and Summation*

MOST VOLTAGE CHANGES IN A NEURON are not action potentials. A transmitter that opens ligand-gated channels at a synapse produces a *postsynaptic potential* of a fraction of a millivolt to a few millivolts: excitatory (EPSP) if it depolarizes, inhibitory (IPSP) if it hyperpolarizes or holds the membrane near rest. A stretched muscle spindle or a pressed touch receptor produces a *receptor potential*. These are *graded* potentials: their size follows the size of the stimulus, they decay in time with the membrane time constant, and they spread passively, shrinking with distance (Section 3.8).

The trigger zone adds them up. Inputs that arrive close together in time add (*temporal summation*), because each has not yet decayed when the next arrives. Suppose each EPSP jumps the membrane by 5 mV and decays with $\tau = 20$ ms, and threshold is 15 mV above rest. With inputs every 5 ms, the previous depolarization has decayed only to $e^{-5/20} = 0.78$ of its value when the next input adds its 5 mV, so the total grows as 5.0, 8.9, 11.9, 14.3, and 16.1 mV and crosses threshold at the fifth input (Fig. 3.6). With inputs 30 ms apart each EPSP has almost vanished before the next, and the sum never exceeds about 6 mV. Inputs at different synapses add too (*spatial summation*), and IPSPs subtract. If the sum crosses threshold, the trigger zone fires.

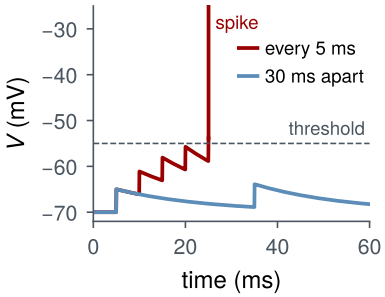


Figure 3.6: Temporal summation. EPSPs of 5 mV that decay with $\tau = 20$ ms, arriving every 5 ms, pile up and cross threshold (15 mV above rest) at the fifth input; the same inputs 30 ms apart never do.

A neuron is therefore an analog adder followed by a comparator, and it reports the strength of a sustained input as a *firing rate*: a weak stretch of a muscle spindle gives a few spikes per second, a strong stretch many. The information is in the timing of identical pulses, a pulse-frequency code, which survives the attenuation and noise of a long line.

3.8 *The Axon as a Cable*

HOW FAR CAN A SIGNAL TRAVEL along an axon before it dies out? An axon is a tube of salty cytoplasm, a decent conductor, wrapped in a thin membrane that is a leaky insulator, immersed in salty extracellular fluid. It is a cable like a submarine telegraph line, and the same analysis applies. Cut it into segments of length Δx (Fig. 3.7). Each segment has a resistance along the inside and a resistance and capacitance across the membrane, described by three quantities per unit length:

- r_a, the axial resistance per unit length of the core, in Ω/cm;

- r_m, the membrane resistance times unit length, in $\Omega\,\mathrm{cm}$ (a longer segment has more membrane in parallel, so its resistance is lower);
- c_m, the membrane capacitance per unit length, in $\mathrm{F/cm}$.

The resistance of the outside fluid is small and we neglect it.

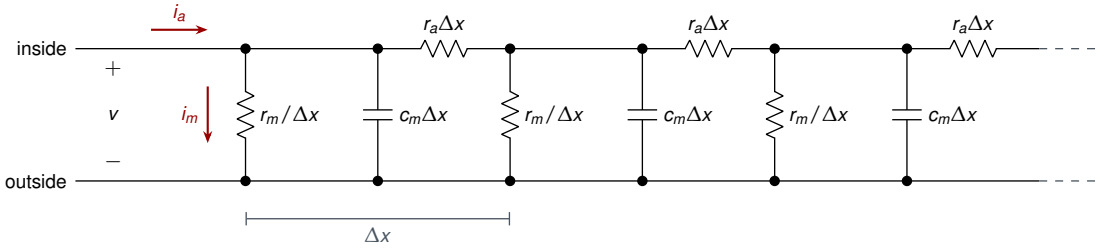


Figure 3.7: The axon as a ladder network. Along the inside, each length Δx has axial resistance $r_a\Delta x$; across the membrane it has resistance $r_m/\Delta x$ in parallel with capacitance $c_m\Delta x$. Current i_a flows along the core and leaks out as membrane current i_m, so the membrane voltage v (measured from rest) shrinks with distance.

The length constant

Inject a steady current into the axon at one point. It flows along the core and, on the way, leaks out through the membrane. How far does it get? Over a length ℓ the core has resistance $r_a\ell$, which grows with ℓ, while the membrane of that length has resistance r_m/ℓ, which falls with ℓ. Close to the injection point the core is the easier path and the current flows along; far away, leaking out is easier. The two are equal at the *length constant* (or space constant) λ, where $r_a\lambda = r_m/\lambda$:

$$\lambda = \sqrt{\frac{r_m}{r_a}}, \qquad \tau = r_m c_m. \tag{3.7}$$

KEY EQUATION

$\lambda = \sqrt{r_m/r_a}$, $\tau = r_m c_m$

Length and time constants of a passive cable.

The units work out: r_m/r_a is $(\Omega\,\mathrm{cm})/(\Omega/\mathrm{cm}) = \mathrm{cm}^2$, so λ is a length. A full analysis of the ladder (the *cable equation*, the same mathematics as a telegraph line) shows that a steady voltage v_0 held at one point of a long axon dies away exponentially with distance,

$$v(x) = v_0\, e^{-x/\lambda}, \tag{3.8}$$

falling to e^{-1}, or 37 %, one length constant away. The time constant $\tau = r_m c_m$ is the membrane's own RC time constant of Section 3.3: how long a patch takes to charge or discharge. A changing signal therefore arrives at a distant point both smaller and later.

The per-length quantities follow from the geometry. For an axon of diameter d with cytoplasm resistivity R_i ($\Omega\,\mathrm{cm}$), specific membrane resistance R_m ($\Omega\,\mathrm{cm}^2$), and specific capacitance C_m ($\mathrm{F/cm}^2$), the core has cross section $\pi d^2/4$ and each unit length has membrane area πd:

$$r_a = \frac{4R_i}{\pi d^2}, \qquad r_m = \frac{R_m}{\pi d}, \qquad c_m = C_m \pi d. \tag{3.9}$$

Substituting into Eq. (3.7),

$$\lambda = \sqrt{\frac{R_m\, d}{4R_i}}, \qquad \tau = R_m C_m. \tag{3.10}$$

KEY IDEA Doubling the diameter halves r_m but quarters r_a, so λ grows as $\sqrt{d}$. The time constant does not depend on size at all.

A thick axon has a core of much lower resistance, while its membrane, though larger in area, leaks only in proportion to d, so the length constant grows with the square root of the diameter. With the membrane of the next example, a squid giant axon 500 µm across has $\sqrt{500} \approx 22$ times the length constant of a 1 µm axon, about 1.6 cm. The squid evolved its giant axon for a fast escape jet; vertebrates solved the same problem with myelin (Section 3.9).

The same attenuation weighs the inputs of a neuron. On a dendrite with $\lambda = 0.5$ mm, a slow 8 mV synaptic potential 0.2 mm from the soma arrives as $8e^{-0.4} = 5.4$ mV, but one 1 mm away arrives as only $8e^{-2} = 1.1$ mV. Distant synapses count for less.

WORKED EXAMPLE 3.4 *Cable constants of a thin axon*

An unmyelinated axon is 1 µm in diameter with $R_m = 20\,\text{k}\Omega\,\text{cm}^2$, $R_i = 100\,\Omega\,\text{cm}$, and $C_m = 1\,\mu\text{F/cm}^2$. (a) Find r_a and r_m. (b) Find λ and τ. (c) What fraction of a steady signal survives 1 mm of passive spread, and 1 m?

SOLUTION. (a) With $d = 1 \times 10^{-4}$ cm: $r_a = 4(100)/(\pi \times 1 \times 10^{-8}) = 1.27 \times 10^{10}\,\Omega/\text{cm}$ and $r_m = 2 \times 10^4/(\pi \times 1 \times 10^{-4}) = 6.4 \times 10^7\,\Omega\,\text{cm}$.

(b)

$$\lambda = \sqrt{\frac{6.4 \times 10^7}{1.27 \times 10^{10}}}\ \text{cm} = \sqrt{5.0 \times 10^{-3}}\ \text{cm} = 0.071\ \text{cm} = 0.71\ \text{mm},$$

the same as $\sqrt{R_m d/4R_i}$, and $\tau = R_m C_m = (2 \times 10^4)(1 \times 10^{-6})$ s $= 20$ ms.

(c) $e^{-1/0.71} = 0.24$ after 1 mm, and $e^{-1000/0.71}$, effectively zero, after 1 m. Passive spread works over the length of a dendrite, not a nerve. To travel a meter the signal must be regenerated, which is what the action potential does. ■

Conduction velocity

An action potential travels because its own current spreads passively ahead of it and brings the next stretch of membrane to threshold. The farther the current spreads (λ) and the faster the membrane charges (τ), the faster the pulse moves, so the speed scales as λ/τ. In an unmyelinated axon $\lambda \propto \sqrt{d}$ and τ is fixed, so

$$v_{\text{unmyelinated}} \propto \sqrt{d}, \qquad v \approx 1\,\text{m/s} \times \sqrt{d/\mu\text{m}}. \tag{3.11}$$

The coefficient is an engineering fit: it gives about 1 m/s for the 1 µm C fibers that carry slow pain, and about 22 m/s for the squid giant axon, which measures 21 m/s at 18 °C. A hundredfold thicker axon is only ten times faster.

3.9 *Myelin and Saltatory Conduction*

MYELIN IS A GLIAL CELL that has wrapped its own flattened membrane around the axon 10 to more than 100 times. The layers are in series, like resistors in series and capacitors in series. With N membrane layers, the membrane resistance of the internode rises about N-fold and its capacitance falls about N-fold:

$$r_m \to N r_m, \qquad c_m \to c_m / N, \qquad \lambda \to \sqrt{N}\,\lambda, \qquad \tau \to \tau. \tag{3.12}$$

The length constant grows severalfold, and the charge needed to depolarize the internode falls N-fold. Every 0.2 to 2 mm, about a hundred fiber diameters, the myelin stops at a *node of Ranvier*, a bare ring about 1 µm long packed with sodium channels at densities near a thousand per square micrometer (Fig. 3.8). An action potential at one node drives current down the low-resistance core to the next node, which reaches threshold within tens of microseconds. The pulse seems to jump from node to node: *saltatory conduction*, from the Latin for leaping.

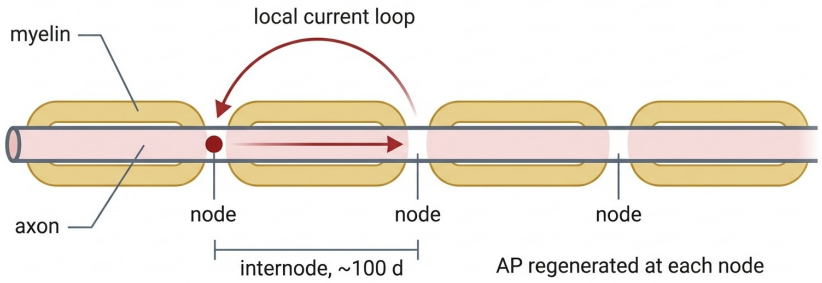


Figure 3.8: Saltatory conduction. Myelin (gold) insulates the internodes. An action potential at one node (dot) drives current along the core and back through the outside fluid, depolarizing the next node above threshold, where sodium channels regenerate the full pulse (AP, action potential).

A long wire attenuates and smears a digital pulse, so a long link places *repeaters* along it that detect each bit and retransmit it at full amplitude. Nodes of Ranvier are the axon's repeaters, and the myelinated internodes between them are nearly lossless links. Take a 10 µm fiber (a 7 µm axon inside its myelin) with 100 layers, $R_m = 20\,\mathrm{k\Omega\,cm^2}$, and $R_i = 100\,\Omega\,\mathrm{cm}$. The internode membrane has $100 \times 20\,\mathrm{k\Omega\,cm^2} = 2\,\mathrm{M\Omega\,cm^2}$, so by Eq. (3.10) $\lambda = 1.9$ cm, and a steady signal loses only about 5 % over a 1 mm internode. The node's 100 mV action potential needs to deliver only some 20 mV to the next node, a *safety factor* of

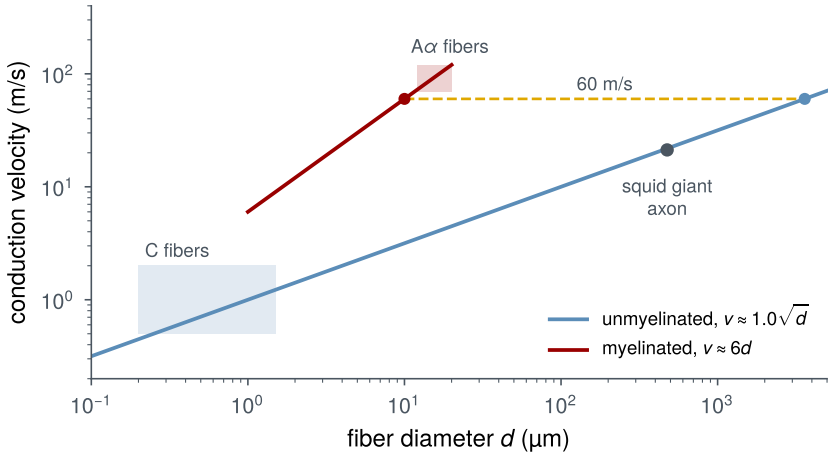


Figure 3.9: Conduction velocity against fiber diameter, from the engineering fits $v \approx 1.0\sqrt{d}$ (unmyelinated) and $v \approx 6d$ (myelinated), with d in micrometers and v in m/s. Shaded boxes show the measured ranges of human C and Aα fibers, and the dot the squid giant axon. A 10 µm myelinated fiber matches an unmyelinated axon 3.6 mm thick.

about five, and because the pulse is rebuilt at every node, its size and shape do not degrade with distance.

Myelin pays three times. It is fast: conduction velocity grows in proportion to diameter instead of its square root, about 6 m/s per micrometer of fiber diameter[3], so a 10 µm fiber conducts at about 60 m/s and crosses each 1 mm internode in 17 µs. It is compact: an unmyelinated axon would have to be about 3.6 mm thick to match that speed (Fig. 3.9). And it is efficient: ions cross the membrane only at the nodes, a small fraction of the axon's surface, so the pump has far less to restore. The cost is vulnerability: in multiple sclerosis the immune system strips CNS myelin, current leaks out of the bare internodes, and conduction slows or fails.

[3] J. B. Hursh, *Am. J. Physiol.* (1939).

You can feel the difference. Stub your toe and the pain comes twice: a sharp pain carried by myelinated Aδ fibers (3 µm, about 18 m/s, 56 ms for the meter to the spinal cord), then a dull pain a second later, carried by unmyelinated C fibers (1 µm, about 1 m/s).

3.10 *Nerves and Reflexes*

All neurons share the plan of Fig. 3.1. *Sensory* (afferent) neurons carry information toward the CNS, *motor* (efferent) neurons carry commands to muscles, and *interneurons*, most of the neurons of the brain, connect neurons to each other. A *nerve* is a cable of many axons, from a few hundred to more than a hundred thousand, bundled by connective tissue into groups called *fascicles*. Like a multicore cable it carries many independent channels in parallel. Nerves are *afferent* (sensory), *efferent* (motor), or, most often, *mixed*. Their fibers are classified by diameter and speed (Table 3.2), and a cuff electrode around a

nerve records or stimulates all of them at once (Chapter 9).

Fiber	Diameter	Velocity	Carries
Aα (myelinated)	12 to 20 µm	70 to 120 m/s	muscle spindle and tendon sensors; motor neurons
Aβ (myelinated)	6 to 12 µm	35 to 75 m/s	touch, vibration
Aδ (myelinated)	1 to 5 µm	5 to 30 m/s	sharp pain, cold
C (unmyelinated)	0.2 to 1.5 µm	0.5 to 2 m/s	dull pain, warmth, autonomic

Table 3.2: Classes of peripheral nerve fibers. Speed follows diameter and myelin as in Fig. 3.9.

The knee-jerk reflex

The simplest circuit in the body is the *stretch reflex* (Fig. 3.10). A tap on the patellar tendon stretches the quadriceps muscle and its *muscle spindles*, sensors whose mechanically gated channels produce a receptor potential and a burst of action potentials. The spindle's large Aα afferent fiber runs to the spinal cord, where it does two things. It makes excitatory synapses directly on the motor neurons of the quadriceps, one synapse in the whole path, so the quadriceps contracts. It also excites an inhibitory interneuron that silences the motor neurons of the hamstring, the opposing muscle, so the hamstring relaxes. The leg kicks. All of this happens in the spinal cord; the brain learns about it afterwards.

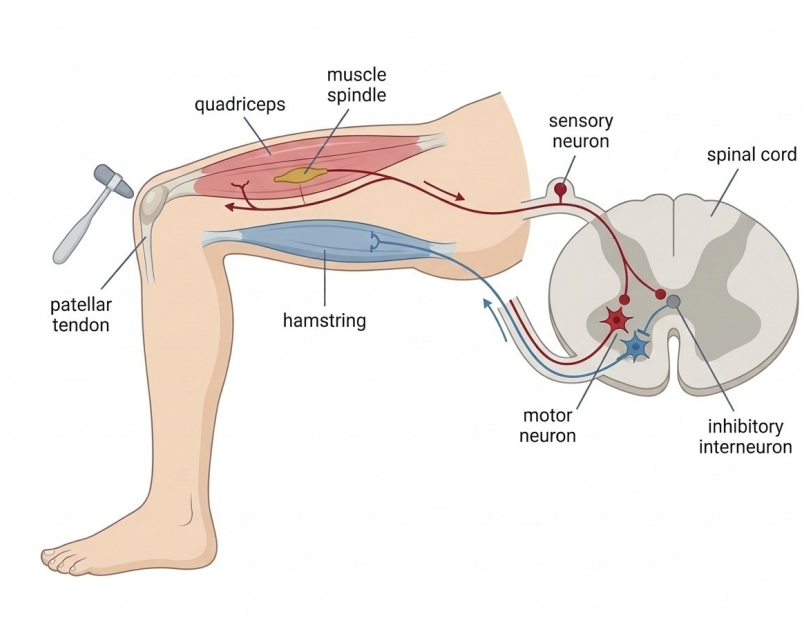


Figure 3.10: The knee-jerk reflex. A tap stretches the quadriceps and its muscle spindles; the sensory neuron excites the quadriceps motor neuron in the spinal cord and, through an inhibitory interneuron, silences the hamstring motor neuron.

An engineer reads the reflex as a negative feedback controller of muscle length (Fig. 3.11): the spindle is the sensor, the spinal cord the controller, the motor neuron and muscle the actuator, and the leg the

plant (the system being controlled). A stretch is sensed and opposed by a contraction. Clinicians tap tendons because the reflex tests the whole loop: an absent reflex points to a broken sensor, wire, or actuator, and an exaggerated one to a loss of the brain's inhibitory control, which turns up the gain of the loop.

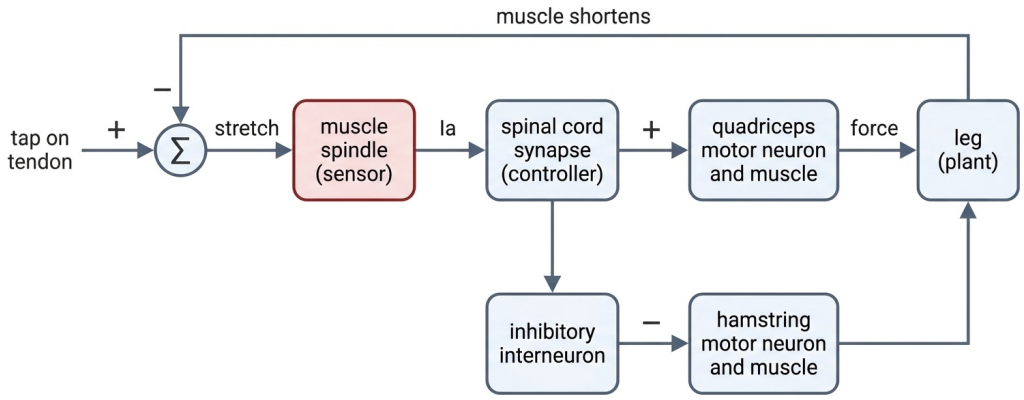


Figure 3.11: The stretch reflex as a block diagram. The muscle length is the controlled variable: a tap adds a stretch, the spindle senses it, and the spinal cord drives the quadriceps to shorten the muscle, closing a negative feedback loop. A feedforward branch through an inhibitory interneuron relaxes the antagonist. Ia names the spindle's large Aα afferent fiber.

Worked Example 3.5 *Timing the knee-jerk reflex*

The afferent path from the quadriceps spindles to the spinal cord and the efferent path back are each about 0.6 m. The afferent fibers conduct at 90 m/s and the motor fibers at 80 m/s; the spinal synapse takes about 0.7 ms and the neuromuscular junction 0.5 ms. Estimate the delay from tap to muscle activation, and compare it with a voluntary reaction.

Solution. Afferent conduction $0.6/90 = 6.7$ ms, synapse 0.7 ms, efferent conduction $0.6/80 = 7.5$ ms, neuromuscular junction 0.5 ms: about 15 ms. The spindle's own response and the spread of excitation in the muscle add a few milliseconds, so the electrical response of the quadriceps starts some 20 ms after the tap. A voluntary response to the same tap, which must pass through the brain, takes 150 ms or more. If disease slows conduction (for example by stripping peripheral myelin), the reflex delay grows in proportion, which is what a nerve conduction study measures. ■

3.11 *Synapses and Tweaking the Body*

At the axon terminal the signal becomes chemical (Fig. 3.12). The action potential opens voltage-gated Ca^{2+} channels; with E_{Ca} at +132 mV, calcium rushes in and makes vesicles holding a few thousand transmitter molecules fuse with the membrane. The transmitter

diffuses across the 20 to 40 nm cleft in 0.25 to 1 µs ($t \approx x^2/2D$, Chapter 2) and binds receptors on the target. Fast receptors are ligand-gated channels that produce a postsynaptic potential within a millisecond; slow receptors start chemical cascades that tune the target for seconds to hours. Enzymes or reuptake pumps then clear the cleft. The whole synaptic delay is half a millisecond to a millisecond. Some cells, especially in the heart, are also joined by *electrical synapses*, gap junctions that pass current directly from cell to cell.

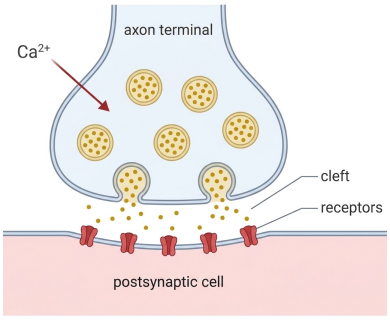


Figure 3.12: A chemical synapse. Calcium entering the terminal makes vesicles of transmitter (gold) fuse with the membrane; the transmitter crosses the 20 nm cleft and opens ligand-gated channels (red) on the target.

The main transmitters are few. *Glutamate* excites and *GABA* inhibits most neurons of the brain; *glycine* inhibits in the spinal cord, including the interneuron of the knee jerk. *Acetylcholine* drives skeletal muscle at the neuromuscular junction and acts in the autonomic system. The *monoamines* dopamine, serotonin, and norepinephrine modulate whole circuits: dopamine is lost in Parkinson's disease, and most of the body's serotonin is made in the gut. The monoamines oxidize at a carbon electrode at a few tenths of a volt, which is why they can be measured electrochemically in the brain (Chapters 6 and 9).

Tweaking the body

The nervous system is an electrical network whose wires run to every organ, and an electrode that passes current through the tissue can drive those wires directly. Near a negative electrode (the cathode), current flows out through the membrane and depolarizes it; if a node of Ranvier or an axon hillock reaches threshold, the neuron fires an action potential that is indistinguishable from a natural one. Devices already use this. Deep brain stimulators quiet the tremor of Parkinson's disease, cochlear implants give hearing to the deaf (Chapter 10), spinal cord stimulators block chronic pain, and vagus nerve stimulators treat epilepsy and depression. The idea of *electroceuticals* is to replace some drugs with devices that modulate the action potentials in a chosen nerve or circuit[4].

[4] K. Famm, B. Litt, K. J. Tracey, E. S. Boyden, M. Slaoui, *Nature* (2013).

How strong and how long must a stimulus be? Treat the membrane that the electrode charges as the RC circuit of Fig. 3.2: a resistance R in parallel with a capacitance C, with time constant $\tau = RC$. A current step I raises its voltage as $IR(1 - e^{-t/\tau})$, the RC charging curve of Chapter 1. To fire, the depolarization must reach threshold, about $\Delta V_{\text{th}} = 15\,\text{mV}$, before the pulse ends at time T. A very long pulse needs only the *rheobase* current $I_{\text{rh}} = \Delta V_{\text{th}}/R$. A shorter pulse must charge the membrane faster and needs more:

$$I_{\text{th}} = \frac{I_{\text{rh}}}{1 - e^{-T/\tau}}. \tag{3.13}$$

This *strength-duration curve* (Fig. 3.13) has two regimes. For long pulses the threshold is the rheobase. For short pulses $e^{-T/\tau} \approx 1 - T/\tau$, so

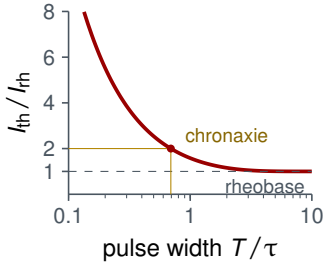


Figure 3.13: The strength-duration curve of an RC membrane, Eq. (3.13). Long pulses need only the rheobase current; short pulses need more. At the chronaxie, 0.69τ, the threshold is twice the rheobase.

$I_{\text{th}}T \approx I_{\text{rh}}\tau$: what matters is the charge, about $C\,\Delta V_{\text{th}}$, delivered before the membrane can leak it away. The pulse width at which the threshold is twice the rheobase is the *chronaxie*, $T_{\text{ch}} = \tau \ln 2 \approx 0.69\tau$. Measured chronaxies range from about 0.1 ms for large myelinated axons to a few milliseconds for cell bodies, so short pulses favor axons. A stimulator must also reach the right fibers among thousands (large myelinated fibers, with their widely spaced nodes, fire first) without damaging the tissue or the electrode; Chapter 9 designs the electrodes and charge-balanced waveforms that do this.

Recording these signals, and every other one in this book, starts with a sensor and the circuit that reads it. Part II begins there: Chapter 4 says what makes a sensor good and builds the op-amp circuits, among them the current amplifier of the patch clamp, that turn a sensor's output into a clean voltage.

Summary

- A neuron sums its inputs at the trigger zone, fires there, conducts along the axon, and releases transmitter at its terminals.
- Mammalian neuron, inside and outside (mM): K^+ 140 and 5, Na^+ 12 and 145, Cl^- 7 and 110, Ca^{2+} 0.0001 and 2. Diffusion balanced by voltage gives the Nernst potential $E = (RT/zF)\ln(c_o/c_i)$, with $RT/F = 26.7\,\text{mV}$ at 37 °C (61.5 mV per decade): $E_{\text{K}} = -89$, $E_{\text{Na}} = +67$, $E_{\text{Cl}} = -74$, $E_{\text{Ca}} = +132\,\text{mV}$.
- The Goldman equation with $P_{\text{K}} : P_{\text{Na}} : P_{\text{Cl}} = 1 : 0.04 : 0.45$ gives $-70\,\text{mV}$; with $P_{\text{Na}}/P_{\text{K}} = 20$, about $+51\,\text{mV}$. Rest is negative because K^+ leaks out through open channels; the Na^+/K^+ pump (3 out, 2 in) keeps the gradients charged.
- The membrane is a capacitor ($\approx 1\,\mu\text{F/cm}^2$) in parallel with one battery and conductance per ion: $I = g(V_m - E)$, $V_m = \sum gE / \sum g$ at rest, time constant $R_m C_m \approx 20\,\text{ms}$.
- Channels are leak, voltage-, ligand-, or mechanically gated; $i = \gamma(V - E)$, a few picoamperes. The voltage clamp holds V_m by feedback and measures the channel current; the patch clamp (gigaseal, 10 GΩ current-to-voltage converter) resolves one channel.
- The action potential: threshold near $-55\,\text{mV}$, Na^+ influx with positive feedback, overshoot, Na^+ inactivation and delayed K^+ efflux, after-hyperpolarization, refractory periods. Graded potentials decay with τ, sum in time and space, and code strength as firing rate.
- Cable: $\lambda = \sqrt{r_m/r_a} = \sqrt{R_m d/4R_i}$, $v = v_0 e^{-x/\lambda}$; 0.71 mm for a 1 µm axon. Unmyelinated $v \approx \sqrt{d}$ m/s; myelinated $v \approx 6d$ m/s (d in µm). Nodes regenerate the pulse like repeaters.

- The knee jerk is a one-synapse negative feedback loop of about 20 ms. Synapses release transmitter when Ca^{2+} enters. A stimulus must charge the membrane to threshold: $I_{th} = I_{rh}/(1 - e^{-T/\tau})$, chronaxie $\tau \ln 2$.

3.12 *Problems*

3.1 ° Nernst slope.

(a) Find the Nernst slope in mV per decade for a monovalent ion at 25 °C and 37 °C, and for Ca^{2+} at 37 °C.

(b) By how much does E_K change if $[K^+]_o$ rises tenfold?

3.2 • Hyperkalemia. Use the Goldman equation with $P_K : P_{Na} : P_{Cl} = 1 : 0.04 : 0.45$ and the concentrations of Table 3.1.

(a) Find the resting potential when $[K^+]_o$ rises from 5 to 8 mM.

(b) Find the resting potential if instead P_{Na}/P_K rises to 0.1.

(c) Why is a high blood potassium dangerous for the heart?

3.3 • A dendrite is 2 µm in diameter with $R_m = 30\,k\Omega\,cm^2$, $R_i = 150\,\Omega\,cm$, and $C_m = 1\,\mu F/cm^2$.

(a) Find r_a, r_m, λ, and τ.

(b) What fraction of a slow synaptic potential at 500 µm reaches the soma (steady-state approximation)?

3.4 • A patient with a demyelinating neuropathy has conduction velocities of 30 m/s in both the afferent and efferent fibers of the knee jerk. Using the path lengths and delays of Worked Example 3.5, find the new conduction-and-synapse delay and compare it with the normal one.

3.5 •• Choosing a stimulus pulse. A nerve fiber stimulated through a nearby electrode follows the strength-duration curve of Eq. (3.13). A long pulse needs at least 0.40 mA to fire it, and a 0.10 ms pulse needs 1.00 mA.

(a) Find the time constant τ and the chronaxie.

(b) Find the threshold current and the charge per pulse for pulse widths of 0.05, 0.2, and 1 ms.

(c) What is the smallest charge that can ever fire the fiber, and with what kind of pulse?

(d) If the electrode and tissue act as a fixed resistance, the energy per pulse is proportional to I^2T. Which of the three pulses

uses the least energy? Suggest a pulse width for a battery-powered stimulator.

Part II

Sensing and Measurement

4
Sensors and Signal Conditioning

In this chapter you will learn to

- Describe a biosensor as a bioreceptor, a transducer, and readout electronics.
- Find sensitivity, linear range, and selectivity from calibration data, and read the response time from a step.
- Estimate thermal, shot, and quantization noise and compute a signal-to-noise ratio.
- Tell accuracy from precision, calibrate with standards, and compute the decision and detection limits.
- Remove shared drift with a reference sensor.
- Analyze the basic op-amp circuits with the two golden rules.
- Design readouts for high-impedance electrodes and for small currents, and fit their output into an ADC.

A small disk on the back of the upper arm reports a glucose number every five minutes for two weeks. Under the disk, a flexible filament a few millimeters long sits in the fluid between the cells of the skin. On the filament an enzyme oxidizes glucose, an electrode turns the products of that reaction into a current of a few nanoamps, and the electronics in the disk turn the current into a number on a phone (Fig. 4.1). People decide how much insulin to inject from that number, so it is fair to ask how good the number is. How much current does each millimolar of glucose give? Does a painkiller change the reading? How long after a meal does the sensor catch up? What is the smallest amount it can see? And how do a few nanoamps become a number at all?

This chapter answers those questions in two parts. The first part (Sections 4.1 to 4.7) defines a biosensor's *figures of merit*, the numbers

on its data sheet. The second part builds the electronics that read it: the operational amplifier from scratch (Section 4.8), readouts for electrodes with hundreds of megohms of internal resistance (Section 4.9), and amplifiers for currents of nanoamps and below (Section 4.10). You need only the circuit essentials of Chapter 1 (Ohm's law, the voltage divider, the RC filter, decibels, and the ADC) and the Michaelis-Menten kinetics of Chapter 2. Chapters 5 to 8 then apply these tools to physical, electrochemical, optical, and biopotential sensors.

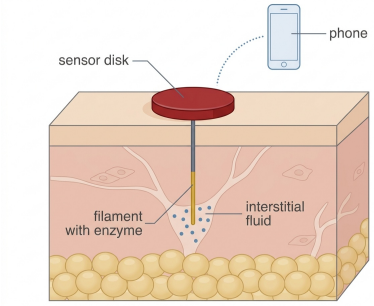


Figure 4.1: A continuous glucose monitor. A filament under the skin carries the enzyme and electrodes; the disk on the skin holds the amplifier, the ADC, and the radio.

4.1 *What Is a Biosensor?*

A *biosensor* is a measuring device that couples a biological recognition element to a transducer, so that a biochemical event becomes an electrical signal. It has three parts. The *bioreceptor* recognizes the target molecule, the *analyte*. The *transducer* converts the recognition event into a change of current, voltage, resistance, capacitance, frequency, or light. The *electronics* amplify, filter, and digitize that change and turn it into a concentration.

Bioreceptors come in two families (Fig. 4.2). *Catalytic* receptors, mostly enzymes, convert the analyte into a product again and again; the signal reports a reaction rate and keeps flowing as long as analyte arrives. Glucose oxidase (Chapter 2) is the classic example. *Affinity* receptors bind the analyte and hold it: antibodies, membrane receptors and whole cells, and synthetic receptors such as the aptamers of Chapter 2 and *molecularly imprinted polymers* (polymer networks cast around the target and then washed, leaving shaped cavities). An affinity sensor reports how many binding sites are occupied at equilibrium, not a rate.

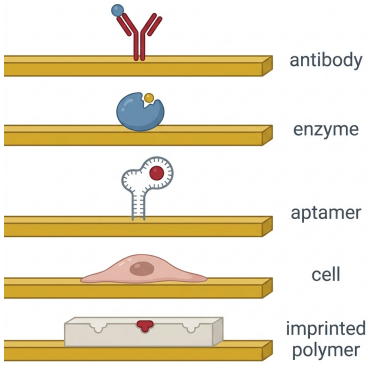


Figure 4.2: Bioreceptors on an electrode: an antibody, an enzyme, a folded aptamer, a living cell, and an imprinted polymer with cavities shaped like the analyte.

Table 4.1: Biosensor classes by transduction principle, with the chapter that treats each one.

Class	What changes	Examples (chapter)
Potentiometric	electrode potential at zero current	pH glass electrode, ion-selective electrode (6)
Amperometric	current at a fixed potential	glucose enzyme electrode (6)
Impedimetric	impedance versus frequency	antibody-coated electrode (5, 6)
Conductometric	conductance between electrodes	enzyme on interdigitated electrodes
Field effect	threshold voltage, drain current	ISFET, OECT (5, 11)
Optical	absorbance, fluorescence, refractive index	oximeter, fluorescent dye (7)
Mass	resonant frequency	quartz crystal microbalance (5)
Thermal	temperature	enzyme thermistor

Transducers are sorted by what they measure (Table 4.1). An amperometric enzyme electrode is held at a potential that oxidizes the

product of the enzyme reaction, so its current measures the reaction rate. A potentiometric electrode develops a voltage at zero current that follows the logarithm of an ion's concentration. A field-effect sensor lets the charge of bound molecules change the current of a transistor.

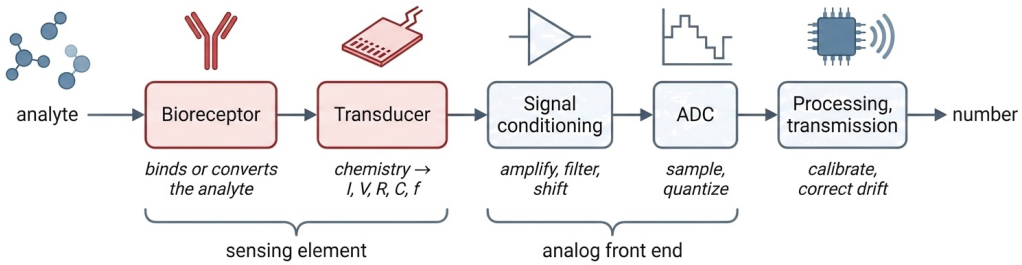


Figure 4.3: The parts of a biosensor system. In a *smart* sensor the signal conditioning and the ADC sit next to the sensing element, so only numbers travel over the cable or the radio link. The sensing element is treated in Sections 4.2 to 4.7, the analog front end in Sections 4.8 to 4.10.

Figure 4.3 puts the parts in order. A smart sensor digitizes beside the sensing element, so the long link carries numbers, which are immune to the pickup and leakage that spoil tiny analog currents; wearable and implanted devices are almost always built this way. The next six sections describe the sensing element with numbers. (Problem 4.1 designs one analyte's sensor five different ways.)

4.2 *Sensitivity and Selectivity*

Every sensor has a transfer function, $S_{\text{out}} = f(S_{\text{in}})$, found by *calibration*: apply known inputs, record the outputs, and fit a curve. The *sensitivity* is its slope,

$$S = \frac{\mathrm{d}S_{\text{out}}}{\mathrm{d}S_{\text{in}}} \approx \frac{\Delta S_{\text{out}}}{\Delta S_{\text{in}}}, \tag{4.1}$$

in units such as nA/mM for an amperometric sensor or mV/pH for a potentiometric one. Many sensors saturate, so the sensitivity depends on the operating point (Section 4.3).

The current of a planar electrode grows with its area, so electrodes of different sizes are compared by the *sensitivity per area*, S/A, in $\mu\text{A}\,\text{mM}^{-1}\,\text{cm}^{-2}$. For a potentiometric sensor, whose output follows the logarithm of concentration, the sensitivity is quoted per decade: an ideal pH electrode gives 59.2 mV per decade at 25 °C, the Nernst slope of Chapter 3. A large slope helps only if the noise does not grow with it; their ratio sets the resolution and the detection limit (Sections 4.4 and 4.6).

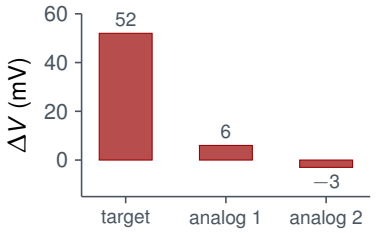


Figure 4.4: An interference test (illustrative data). The sensor sees the target and two structurally similar molecules at the same concentration. The selectivity against analog 1 is $52/6 \approx 9$.

Selectivity

Many sensors respond to things other than the analyte: temperature, pH, ionic strength, flow, and other molecules. The *selectivity* of a sensor

is its sensitivity to the analyte divided by its sensitivity to an interfering parameter. Its inverse is the *selectivity coefficient*,

$$K_{A,B} = \frac{S_B}{S_A}, \tag{4.2}$$

where A is the analyte and B the interferent. A small $K_{A,B}$ is good. For a sensor that is linear in both species, the reading interpreted as analyte is

$$c_{\text{apparent}} = c_A + K_{A,B}\, c_B, \tag{4.3}$$

so the error in concentration units is $K_{A,B}c_B$. Ion-selective electrodes use the same coefficient inside a logarithm (Chapter 6).

Selectivity is tested by exposing the sensor to the analyte and to each likely interferent at the same concentration (Fig. 4.4). A steroid sensor, for example, must be checked against the other steroids in the same fluid, which differ from the target by a few atoms.[1]

[1] One such sensor is the wearable aptamer field-effect transistor for sweat cortisol of B. Wang, *et al.*, *Sci. Adv.* (2022).

Poor selectivity has three common causes. Other molecules may bind the receptor because they look like the analyte. Other species may react at the transducer itself: at the $+0.6\,\text{V}$ that oxidizes hydrogen peroxide on a glucose electrode, ascorbate (vitamin C), urate, and acetaminophen are oxidized too. And physical parameters such as temperature change the response. The remedies follow from the causes: better receptors, *permselective membranes* that exclude interferents by charge or size, a *blank* electrode whose signal is subtracted (Section 4.7), and an extra sensor that measures the interfering parameter so the *firmware*, the program in the device's microcontroller, can correct for it.

Worked Example 4.1 *Sensitivity and selectivity of a glucose electrode*

A printed glucose electrode with an area of $3.0\,\text{mm}^2$ gives 1.0, 181, and 361 nA in 0, 5, and 10 mM glucose. Adding 0.20 mM acetaminophen (a common painkiller) to the 5 mM solution raises the current by 10.8 nA. (a) Find the sensitivity, the sensitivity per area, and the selectivity coefficient. (b) A therapeutic dose of acetaminophen gives about 0.10 mM in blood. What glucose error results, and how large is it at the hypoglycemia threshold of 3.9 mM? (c) A membrane halves the glucose sensitivity and cuts the acetaminophen sensitivity twentyfold. Repeat (b).

Solution. (a) The response is linear: $S_A = (361 - 1.0)/10 = 36\,\text{nA/mM}$. With $A = 0.030\,\text{cm}^2$, the sensitivity per area is

$$S_A/A = 36\,\text{nA/mM}/0.030\,\text{cm}^2 = 1.2\,\mu\text{A}\,\text{mM}^{-1}\,\text{cm}^{-2}.$$

For acetaminophen $S_B = 10.8\,\text{nA}/0.20\,\text{mM} = 54\,\text{nA/mM}$, so $K_{A,B} = 54/36 = 1.5$ (a selectivity of 0.67). The bare electrode is *more* sensitive to the drug than to glucose.

(b) By Eq. (4.3) the error is $K_{A,B}c_B = 1.5 \times 0.10\,\text{mM} = 0.15\,\text{mM}$, or 3.8 % of a 3.9 mM reading, in the direction that hides a low.

(c) Now $S_A = 18\,\text{nA/mM}$ and $S_B = 2.7\,\text{nA/mM}$, so $K_{A,B} = 0.15$ and the error falls to 0.015 mM. The membrane costs half the signal and buys a tenfold better selectivity. Sensitivity sets how much signal a change produces; selectivity sets how wrong the answer is in a real sample that contains more than the analyte. ■

Temperature is the most common interfering parameter. Enzymes speed up with temperature until the protein unfolds (*denatures*) and loses activity, often for good, so the response of an enzyme electrode rises, peaks, and falls (Fig. 4.5).

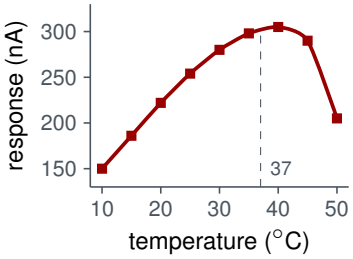


Figure 4.5: Response of a glucose electrode at 4.0 mM versus temperature (illustrative data). Activity rises until about 40 °C; above that the enzyme denatures.

The body holds its core near 37 °C, close to the flat top of this curve, where the slope is only about 0.5 % per kelvin. But a glucose monitor sits in tissue under the skin, which runs a few degrees cooler and follows the room. Between 30 and 35 °C the slope is 3.6 nA/K, or 1.2 % per kelvin, so a 3 K change (a cold walk, a hot shower) shifts the reading by about 4 %. That is why commercial monitors carry a thermistor and correct for temperature in firmware.

4.3 *Range, Linearity, and Response Time*

A CALIBRATION CURVE IS STRAIGHT only over part of its range. At the bottom, the signal disappears into the noise at the detection limit (Section 4.6). At the top, the sensor saturates. The span between them is the *dynamic range*, often quoted as a ratio or in decades. Inside it, the *linear range* is the part where the calibration stays within a stated tolerance of a straight line, often 5 %. Outside the linear range the sensor still works, but it needs a curved calibration, and the falling slope costs resolution.

Saturation

Enzyme sensors saturate because each enzyme molecule can convert only so many analyte molecules per second. Their response follows the Michaelis-Menten law of Chapter 2,

$$I = I_{\max}\,\frac{c}{K_m + c}, \tag{4.4}$$

where $I_{\max}$ is the current with every enzyme busy and K_m is the concentration at half of it. Affinity sensors follow the same shape with the fraction of occupied binding sites, $\theta = c/(K_D + c)$ (the *Langmuir isotherm*), where the *dissociation constant* K_D is the concentration that fills half the sites. At low concentration $I \approx (I_{\max}/K_m)c$, so the initial

sensitivity is $I_{\max}/K_m$, and the curve falls below that line by the fraction $c/(K_m + c)$. The response therefore stays within a tolerance δ of linear for

$$c \le \frac{\delta}{1-\delta}\, K_m, \tag{4.5}$$

only $0.053\,K_m$ for $\delta = 5\,\%$ ($K_m/9$ for $10\,\%$). Affinity sensors are therefore usually read against $\log c$: their response rises from $10\,\%$ to $90\,\%$ of full scale over an 81-fold range of concentration, almost two decades centered on K_D.

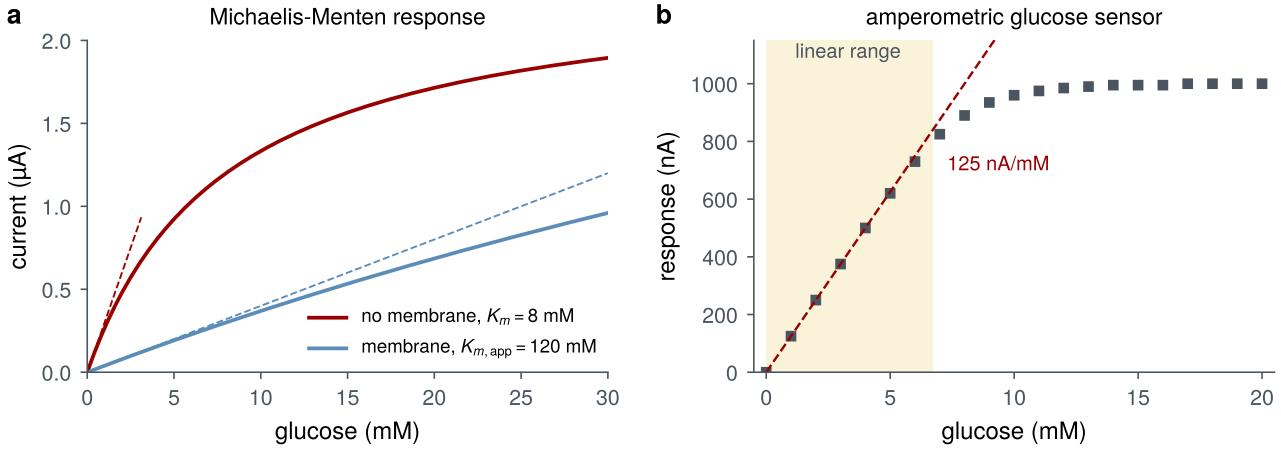


Figure 4.6: Saturating calibration curves (illustrative models). (a) A Michaelis-Menten enzyme electrode with and without a diffusion-limiting membrane (the membrane version carries twice the enzyme); dashed lines are the initial slopes. The membrane lowers the sensitivity from 0.30 to 0.04 µA/mM and stretches the linear range from 0.4 mM to about 6 mM. (b) An amperometric glucose sensor with 125 nA/mM, linear to about 6.7 mM (shaded) and saturating near 1 µA; Example 4.4 designs its readout.

To use an enzyme sensor over a wide range, designers put a membrane in front of the enzyme that limits how fast the analyte arrives (Chapter 2). The enzyme then never runs near saturation, the apparent K_m rises, and the response becomes nearly linear over the physiological range, at the cost of sensitivity (Fig. 4.6a).

Figure 4.6b shows how a linear range is read from data. From 0 to 4 mM the points lie on a line through the origin with a slope of 125 nA/mM. Compared with that line, the reading is 0.8 % low at 5 mM, 2.7 % at 6 mM, 5.7 % at 7 mM, and 11 % at 8 mM. With a 5 % tolerance the linear range ends near 6.7 mM. Above 10 mM the slope is only tens of nanoamps per millimolar, so the same current noise means several times worse glucose resolution.

Response time

A sensor does not follow a step change of concentration instantly: diffusion through membranes and binding take time. The simplest model is first order, like the charging of the RC circuit of Chapter 1. After a step from y_0 to y_∞,

$$y(t) = y_\infty + (y_0 - y_\infty)\, e^{-t/\tau}. \tag{4.6}$$

The *response time* is quoted as the time to reach a stated fraction of the change: $t_{63} = \tau$, $t_{90} = \tau \ln 10 = 2.30\tau$, and $t_{95} = \tau \ln 20 = 3.00\tau$.

For example, when the lactate around a sensor steps from 0 to 10 mM, its current rises from 0.40 toward 2.40 µA and reaches 1.66 µA, 63 % of the change, after 15 s (Fig. 4.7). So $\tau = 15$ s, $t_{90} = 34.5$ s, and $t_{95} = 45$ s. A reading taken 30 s after the step is still $e^{-2} = 13.5$ % short of its final value, 1.35 mM on the 10 mM step. Coming back down is often slower, because washing analyte out of a membrane takes longer than loading it: this sensor needs 120 s to return within 5 % of its baseline, a fall time constant of $120 / \ln 20 = 40$ s.

A first-order sensor behaves like an RC low-pass filter with its corner at $f_c = 1/(2\pi\tau)$. For the lactate sensor that is 10.6 mHz, enough for sweat lactate, which changes over minutes during exercise; a sensor that had to follow the shape of each heartbeat would need about a thousand times more. Finally, a sensor with *hysteresis* gives an output that depends on whether the input is rising or falling, as the viscoelastic polymer of a soft pressure sensor does. It is quoted as the largest gap between the rising and falling branches, as a fraction of full scale.

KEY EQUATION

$t_{63} = \tau \quad t_{90} = 2.30\,\tau \quad t_{95} = 3.00\,\tau$

First-order response times.

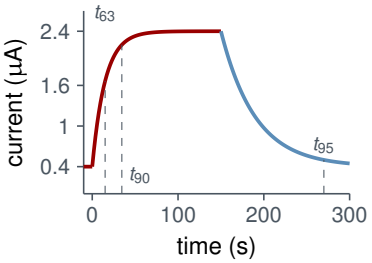


Figure 4.7: First-order response of a lactate sensor: a step from 0 to 10 mM at $t = 0$ ($\tau = 15$ s, red) and back down at 150 s ($\tau = 40$ s, blue).

4.4 Noise and Signal-to-Noise Ratio

Noise is the random part of a sensor's output, the part unrelated to the input. It comes from the random motion of charges in the sensor and the electronics, from electrical pickup, from vibration, and from the chemistry itself (bubbles, slow surface reactions). Record n samples with the input held fixed, take the deviation x_i of each from their mean, and the noise is their root mean square (rms),

$$x_{\text{rms}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}, \tag{4.7}$$

which for many samples equals the standard deviation. The *signal-to-noise ratio* compares a signal amplitude with it, often in decibels:

$$\text{SNR} = \frac{\text{signal}}{x_{\text{rms}}}, \qquad \text{SNR}_{\text{dB}} = 20 \log_{10} \frac{\text{signal}}{x_{\text{rms}}}. \tag{4.8}$$

Four kinds of electronic noise

Some noise has nothing to do with the sensor design; it is physics, and it sets a floor.

Thermal noise (Johnson noise) comes from the thermal jiggling of the electrons in any resistance R at absolute temperature T: a resistor with nothing connected to it shows a small random voltage. In a bandwidth

B it is

$$v_n = \sqrt{4kTRB}, \qquad i_n = \sqrt{\frac{4kTB}{R}}, \tag{4.9}$$

as an open-circuit voltage or a short-circuit current, with Boltzmann's constant $k = 1.381 \times 10^{-23}\,\mathrm{J/K}$ (the k_B of Chapters 2 and 3; noise formulas drop the subscript). At 300 K, $4kT = 1.66 \times 10^{-20}\,\mathrm{J}$. A 1 MΩ electrode at body temperature (310 K) produces 1.31 µV rms in a 100 Hz band. Thermal noise is spread evenly over frequency (it is *white*), so its rms value grows as $\sqrt{B}$, and narrowing the band is the first cure. Here B is the *noise bandwidth* of the filter that sets the band. For a filter with a sharp edge it equals the cutoff frequency. A first-order RC low-pass filter falls off gradually and passes as much white noise as a sharp filter $\pi/2 \approx 1.57$ times wider than its corner frequency, so its rms noise is $\sqrt{\pi/2} \approx 1.25$ times what B equal to the corner would predict.

Shot noise comes from the graininess of charge. A current I that crosses a barrier, as in a photodiode or an electrode reaction, is a stream of electrons arriving at random times, like raindrops on a roof:

$$i_n = \sqrt{2qIB}, \tag{4.10}$$

with $q = 1.602 \times 10^{-19}\,\mathrm{C}$, the elementary charge. It is also white. A 1 nA current carries 56.6 fA rms of shot noise in 10 Hz.

KEY EQUATION

$v_n = \sqrt{4kTRB} \quad i_n = \sqrt{2qIB}$

Thermal noise of a resistor; shot noise of a current I.

Flicker noise, or $1/f$ noise, grows as the frequency falls. Its origin varies (traps in a transistor channel, fluctuating contacts, slow chemistry at an electrode), but its effect is always the same: it dominates below a *corner frequency* $f_{1/f}$ where it equals the white floor (Fig. 4.8a). Biosensor signals change over minutes, so they live in the $1/f$ region. For an amplifier with a white floor of $10\,\mathrm{nV}/\sqrt{\mathrm{Hz}}$ and a corner at 100 Hz, the band from 0.1 to 10 Hz holds about seven times the noise the white floor alone predicts. That is why sensitive instruments move the signal to a higher frequency (with an AC excitation, for example) before amplifying it.

Quantization noise comes from the ADC, which rounds every sample to the nearest step, its LSB (Chapter 1). The error is spread evenly between ± half an LSB, and its rms value is

$$q_n = \frac{\mathrm{LSB}}{\sqrt{12}}. \tag{4.11}$$

A 12-bit converter with a 3.3 V reference has an LSB of 0.806 mV and $q_n = 0.233\,\mathrm{mV}$.

Independent noise sources add in power, not in amplitude: $n_{\mathrm{tot}} = \sqrt{n_1^2 + n_2^2 + \cdots}$. The largest one usually dominates, so a noise budget is a hunt for the biggest term.

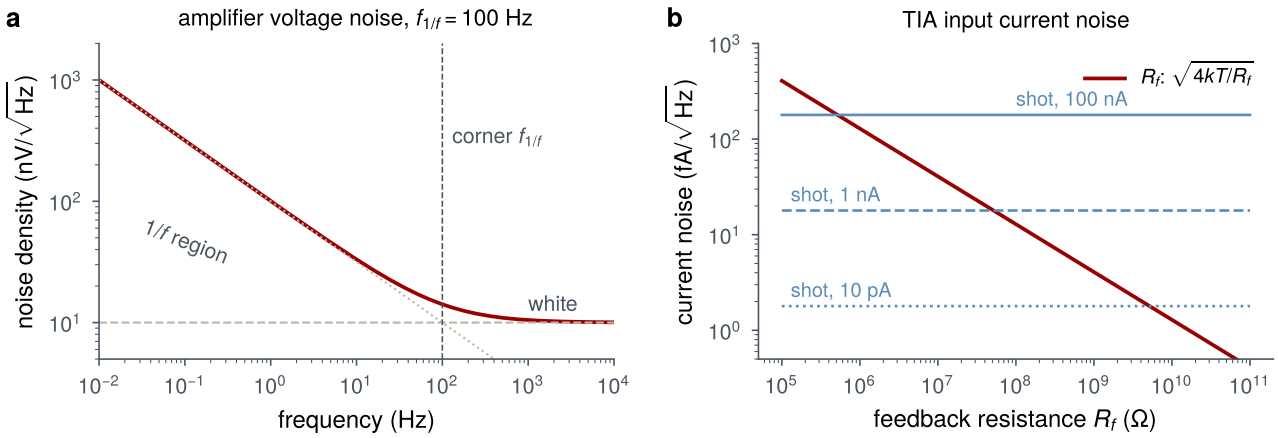


Figure 4.8: Noise densities. (a) The voltage noise of a typical amplifier: white above the corner frequency, rising as $1/\sqrt{f}$ below it. (b) The input current noise of a transimpedance amplifier (Section 4.10): the thermal current noise of its feedback resistor at 300 K falls as $1/\sqrt{R_f}$, while the shot noise of a signal current (horizontal lines) does not depend on R_f. Where the red line lies below a blue one, the signal's own shot noise is the floor.

Input-referred noise and resolution

Noise measured at the output of a chain includes the gains of every stage. To compare it with the signal, divide it by the gain from the input to the output. This *input-referred noise* is expressed in the units of the input: amps for a current sensor, volts for an electrode, millimolar for the whole system. A later stage's noise counts for less when the stages before it have gain, so good designs put the gain early (Example 4.5).

Dividing the input-referred noise by the sensitivity gives the *resolution*, the smallest change of concentration one reading can resolve (or the ADC step divided by the sensitivity, if that is larger). With 0.36 nA rms of noise, the electrode of Example 4.1 resolves $0.36/36 = 0.01$ mM. Averaging N independent readings lowers white noise by $\sqrt{N}$, so resolution improves at the cost of response time; averaging does nothing for drift and little for $1/f$ noise.

4.5 *Accuracy, Precision, and Calibration*

Accuracy and precision answer two different questions. The *accuracy* of a measurement is how close it is to the true value. The *precision* is how well repeated measurements of the same quantity agree with each other. With many readings, accuracy is the offset between their mean and the true value, the *bias*, and precision is their spread, the standard deviation. Bias is a *systematic* error and spread a *random* one. Figure 4.9 shows the four combinations.

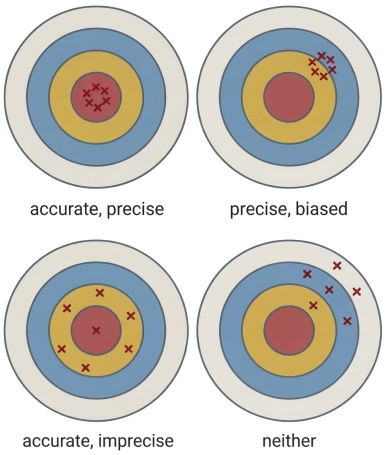


Figure 4.9: Accuracy and precision as hits on a target whose center is the true value. A tight cluster off center is precise but biased; a wide cluster around the center is accurate on average but imprecise.

For n readings x_i with mean $\bar{x}$, the sample standard deviation is

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}, \tag{4.12}$$

the *coefficient of variation* is CV $= s/\bar{x}$, and the *standard error of the mean* is $s/\sqrt{n}$, the spread you would expect of the mean of n readings. Precision across sensors, days, or operators (*reproducibility*) is always worse than with one sensor over a short time (*repeatability*).

Both figures are measured against *standards*, samples whose value is known: buffers of known pH, solutions made by weight, or a laboratory analyzer in a clinical study. A high correlation with the reference shows only that the two track each other: a sensor that reads 20 % high everywhere correlates perfectly. Glucose monitors are judged instead by the *mean absolute relative difference* (MARD), the average of $|x_{\text{sensor}} - x_{\text{ref}}|/x_{\text{ref}}$ over paired readings.

Calibration

Poor precision is fought by averaging; poor accuracy is fought by *calibration*. A one-point calibration measures one standard and removes an offset. A two-point calibration fixes both the offset and the gain. A multipoint calibration measures several standards across the working range and fits a straight line, or a curve, through them by least squares (the line that minimizes the sum of the squared misses; a spreadsheet or `numpy.polyfit` finds it). For example, five glucose standards at 0, 2, 4, 6, and 8 mM that give 1.1, 9.0, 17.2, 24.8, and 33.1 nA fit a slope $m = 3.99$ nA/mM and an intercept $b = 1.08$ nA. A sample that then reads 21.0 nA contains $(21.0 - 1.08)/3.99 = 4.99$ mM. Measure the standards quickly, so that drift does not pass for imprecision, and repeat the calibration as the sensor ages (Section 4.7). A calibration also holds only for samples like its standards. The calibration curve of a capacitive sweat-rate sensor shifts with the ion concentration of the sweat, so one design adds a second capacitor of fixed size at the inlet, measures the incoming sweat with it, and uses that reading to choose the matching curve.[2]

[2] M. S. Islam, S. Cha, W. Cai, M. Ferdoushi, Y. Khan, *IEEE Sens. Lett.* (2025).

4.6 The Detection Limit

How little analyte can a sensor see? Measure many *blanks*, samples with no analyte, and the readings scatter around a mean $\bar{y}_B$ with a standard deviation s_B. A sample is reported as containing analyte when its reading rises far enough above the blank scatter. That

simple idea hides two different questions, and the International Union of Pure and Applied Chemistry (IUPAC) gives each its own limit.

The first question: *is there analyte in this sample?* Choose the probability α of a false positive you will accept, a blank that reads above the threshold. Only 5 % of normally distributed blanks lie more than 1.645 standard deviations above their mean, so the *decision limit* (or critical level) for $\alpha = 5\,\%$ is

$$L_C = \bar{y}_B + 1.645\, s_B. \tag{4.13}$$

For $\alpha = 1\,\%$ the factor is 2.326.

The second question: *what is the smallest true signal that will be detected reliably?* A sample whose mean signal sits exactly at L_C falls below it half the time. Ask instead that a sample be missed (a false negative) only 5 % of the time. Its readings must then sit another 1.645 standard deviations higher (Fig. 4.10a), and the *detection limit* is

$$L_D = L_C + 1.645\, s_B = \bar{y}_B + 3.29\, s_B. \tag{4.14}$$

Both limits are converted to concentration with the calibration slope, $c = (L - \bar{y}_B)/S$, so $c_D = 3.29\, s_B / S$.

The popular rule "blank plus three standard deviations", $\bar{y}_B + 3s_B$, sits close to L_D. Trouble starts when one number serves as both limits: with $\bar{y}_B + 3s_B$ as the decision threshold, false positives fall to 0.13 %, but a sample at that level is missed half the time. When you report a limit, say which convention you use; this book uses Eqs. (4.13) and (4.14). (With only a few blanks the factors grow a little; statistics texts tabulate them.) A third limit, the *quantification limit* $\bar{y}_B + 10\, s_B$, marks where a reading has a relative uncertainty of about 10 %.

KEY EQUATION

$L_C = \bar{y}_B + 1.645\, s_B$
$L_D = \bar{y}_B + 3.29\, s_B$

Decision and detection limits (5% false positives and negatives). Divide $L - \bar{y}_B$ by the slope for concentration.

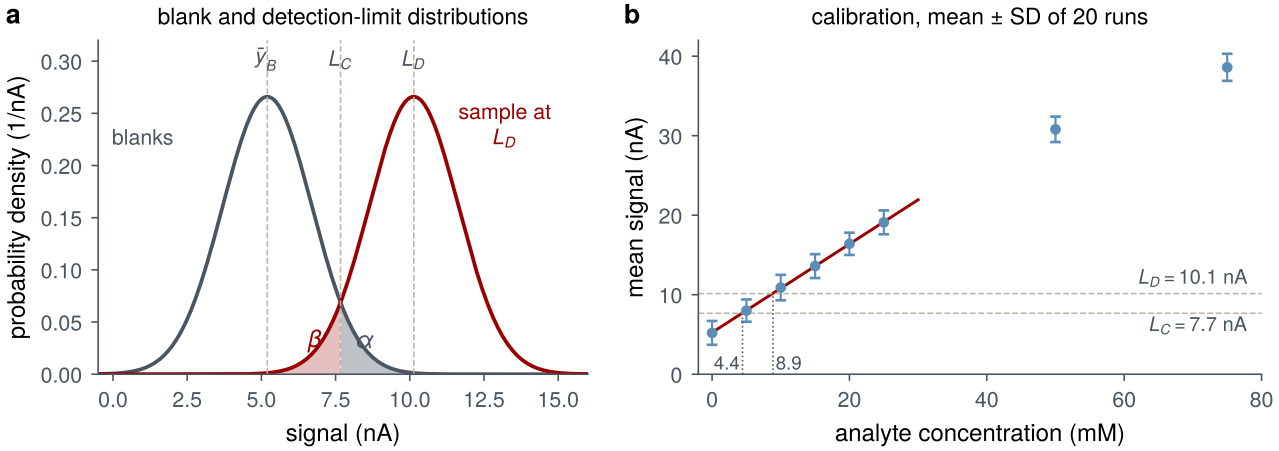


Figure 4.10: Decision and detection limits for the sensor of Example 4.2 (illustrative data). (a) Blank readings (gray) and readings of a sample whose mean is at L_D (red), both with $s = 1.5$ nA. Blanks above L_C are false positives (α); sample readings below L_C are false negatives (β). (b) The calibration, with the line fitted over its linear part, maps L_C and L_D to 4.4 and 8.9 mM.

WORKED EXAMPLE 4.2 *Decision and detection limits from 20 runs*

A sweat lactate sensor is calibrated with 20 runs at each concentration. The means and standard deviations are (illustrative data):

Lactate (mM)	0	5	10	15	20	25	50	75
Mean (nA)	5.2	8.0	10.9	13.6	16.4	19.1	30.8	38.6
SD (nA)	1.5	1.4	1.6	1.5	1.4	1.5	1.6	1.7

(a) Find the sensitivity over the linear part. (b) Find the decision limit and the detection limit, in nanoamps and in millimolar. (c) Is a 5 mM sample reliably detected? A 10 mM sample?

SOLUTION. (a) From 0 to 25 mM the means rise by about 2.8 nA per 5 mM; above that the curve bends over. A least-squares line through the first six points gives $S = 0.557\,\text{nA/mM}$ (the end points alone give $13.9/25 = 0.556\,\text{nA/mM}$).

(b) With $\bar{y}_B = 5.2\,\text{nA}$ and $s_B = 1.5\,\text{nA}$:

$$L_C = 5.2 + 1.645 \times 1.5 = 7.67\,\text{nA} \quad \rightarrow \; 2.47/0.557 = 4.4\,\text{mM},$$
$$L_D = 5.2 + 3.29 \times 1.5 = 10.1\,\text{nA} \quad \rightarrow \; 4.94/0.557 = 8.9\,\text{mM}.$$

The $3s_B$ rule gives 9.7 nA, or 8.1 mM. Figure 4.10b shows the construction.

(c) A 5 mM sample (mean 8.0 nA) lies above the decision limit but below the detection limit: it is often reported, but not reliably. A 10 mM sample lies above c_D and is detected at least 95 % of the time. A common error is to compute $5.2 + 2.33 \times 1.5 = 8.7\,\text{nA}$ and call it the detection limit "blank + 3 SD": $k = 2.33$ belongs to the 1 % decision limit. Another is to name the next calibration level, 10 mM, as the limit. The limits come from the blank statistics and the slope, not from the list of standards. ■

4.7 *Drift and Differential Measurement*

DRIFT IS A SLOW CHANGE OF THE OUTPUT while the input stays constant (Fig. 4.11). A drifting *baseline* adds an offset that grows with time; a drifting *sensitivity* changes the slope. In the body the causes are many: proteins and cells foul the surface, the enzyme slowly loses activity, membranes take up water, the reference electrode potential wanders, and the electronics warm up. Drift is the main reason worn and implanted sensors need recalibration.

Much of what disturbs a sensor disturbs its neighbor in the same way. Two identical sensing elements a millimeter apart see the same temperature, the same pH, the same nonspecific adsorption, and the same electrical pickup. If only one of them carries the receptor, the

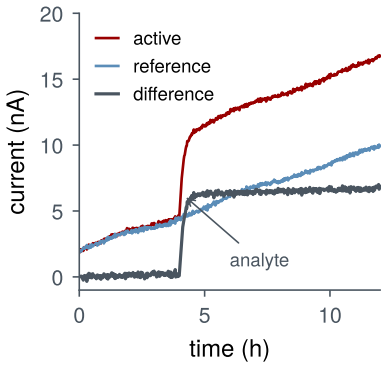


Figure 4.11: Drift correction with a reference electrode (illustrative data). Only the active electrode responds to the analyte at 4 h; the difference keeps the step and loses almost all of the drift.

difference of their outputs keeps the response to the analyte and cancels the rest. The second element is the *reference* (or *blank*) sensor: an electrode coated with an inactive protein instead of the enzyme, an aptamer with a scrambled sequence, or a polymer cast without the template molecule.

Suppose an active electrode with a sensitivity of 2.0 nA/mM drifts by $+0.80$ nA/h and its reference by $+0.74$ nA/h. After 10 h the active baseline has risen by 8.0 nA, which reads as 4.0 mM of analyte that is not there. The difference drifts by only $(0.80 - 0.74) \times 10 = 0.6$ nA, or 0.3 mM: the subtraction removes 92.5 % of the error, and it works only as well as the two drift rates match. It cannot fix a loss of sensitivity in the active element, because the reference has no receptor: a 10 % loss of sensitivity makes every reading 10 % low until the sensor is recalibrated against a known sample, such as a fingerstick reading for a glucose monitor.

Call the reference output V_1 and the active output V_2. Their *differential* part is $V_d = V_2 - V_1$ and their *common-mode* part is $V_{\text{cm}} = (V_1 + V_2)/2$. A real subtracting circuit responds a little to the common mode as well,

$$V_{\text{out}} = A_d V_d + A_{\text{cm}} V_{\text{cm}}, \qquad \text{CMRR} = \frac{A_d}{A_{\text{cm}}}, \tag{4.15}$$

and the *common-mode rejection ratio* measures how well it ignores what the inputs share. It is usually quoted in decibels, $20 \log_{10} \text{CMRR}$: with $A_d = 10$ and a CMRR of 1000 (60 dB), a shared 1 V puts 10 mV at the output, as much as a 1 mV difference. The subtracting circuit is the difference amplifier of the next section.

4.8 The Operational Amplifier

A sensor's signal is almost never ready for an ADC. A glass electrode delivers a fraction of a volt through hundreds of megohms; an enzyme electrode delivers a few nanoamps; a pair of electrodes delivers a small difference on top of a large shared voltage. One part handles all of these: the *operational amplifier*, or op amp. This section introduces it from scratch and builds the six circuits that the rest of the book uses.

An op amp is an integrated circuit of a few dozen transistors with two inputs and one output (Fig. 4.12). It is powered from two supply rails, for example $+5$ V and -5 V, or 3.3 V and ground on a battery device. The input marked $+$ is the *non-inverting input*, the one marked $-$ the *inverting input*. The op amp amplifies the difference between them,

$$V_{\text{out}} = A\,(V_+ - V_-), \tag{4.16}$$

where the *open-loop gain* A is enormous, typically 10^5 to 10^6 (100 to 120 dB). Used this way the op amp is nearly useless as an amplifier. With $A = 10^5$ and ± 5 V rails, an input difference of only 50 µV drives the output to a rail, where it stays (it *saturates*). And A varies several-fold from part to part and with temperature.

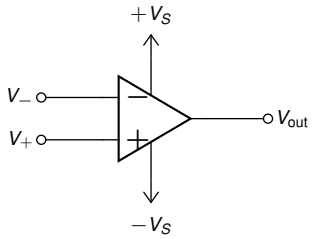


Figure 4.12: The op amp. It amplifies the difference between its non-inverting (+) and inverting (−) inputs by its open-loop gain A; the supply rails $\pm V_S$ bound the output. Schematics usually leave the supply pins out.

Negative feedback and the two golden rules

The trick is *negative feedback*: connect the output back to the inverting input, directly or through resistors. Suppose the output rises too far. Part of that rise reaches V_-, which shrinks $V_+ - V_-$ and pulls the output back down. The loop settles where the input difference is just V_{out}/A, 10 µV for a 1 V output with $A = 10^5$: to the rest of the circuit, zero. The gain of the whole circuit is then set by the resistors, which are stable and accurate, and not by A.

An *ideal op amp* has infinite gain, infinite input resistance, and zero output resistance. With negative feedback, two rules, often called the *golden rules*, then describe every circuit in this book:

1. No current flows into either input.
2. The output does whatever it must to make $V_- = V_+$.

Rule 1 holds because the inputs are the control terminals of transistors (Chapter 5), which draw only femtoamps to nanoamps (Section 4.9 says when that matters). Rule 2 is the feedback at work, as the 10 µV estimate shows. To analyze a circuit, use rule 2 to find the voltage at the inverting input, then rule 1 to say that any current arriving there must flow on through the feedback path. Figure 4.13 collects the six circuits that follow.

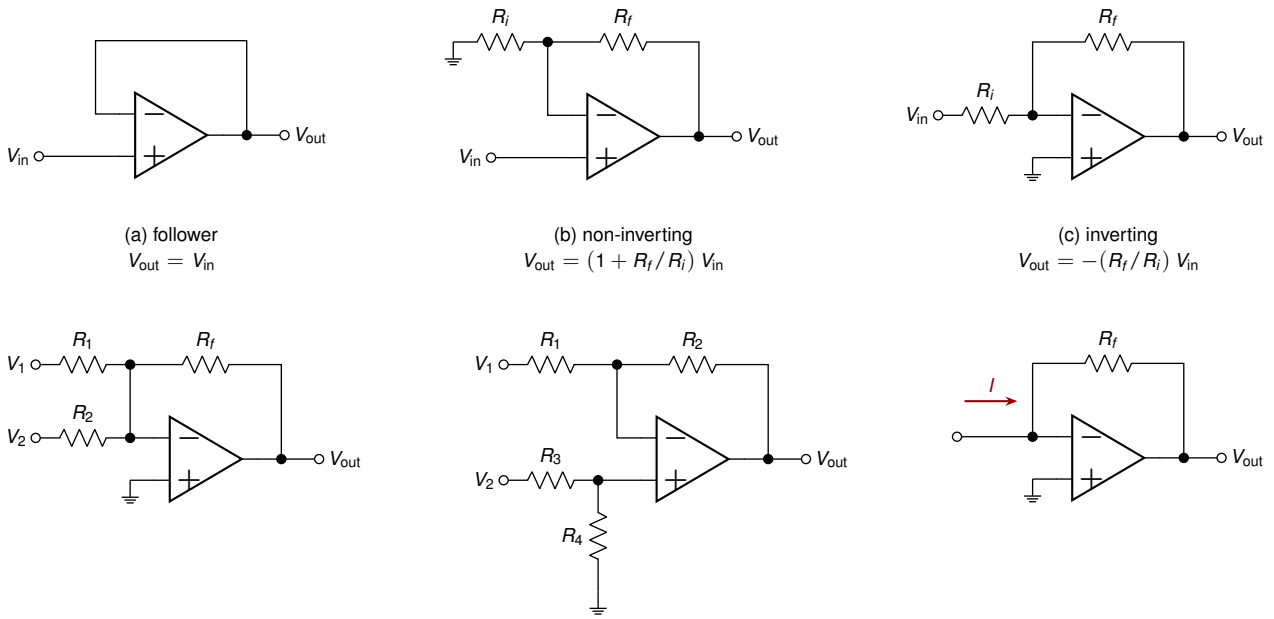
KEY IDEA With negative feedback: no current into the inputs, and $V_- = V_+$.

The voltage follower

Connect the output straight to the inverting input and the signal to the non-inverting input (Fig. 4.13a). Rule 2 gives $V_{out} = V_- = V_+ = V_{in}$: the gain is 1. That looks pointless until rule 1 is added. The input draws no current, while the output can deliver milliamps to whatever follows, so the follower is a *buffer*. A sensor with an internal resistance of 1 MΩ connected directly to a 10 kΩ load forms a voltage divider and delivers only $10/1010 = 1\,\%$ of its voltage; with a follower between them, the sensor sees no load and its full voltage reaches the 10 kΩ.

The non-inverting amplifier

Now feed back only a fraction of the output, through the divider R_f, R_i (Fig. 4.13b). By rule 1 no current enters the inverting input, so the



(a) follower
$V_{out} = V_{in}$

(b) non-inverting
$V_{out} = (1 + R_f/R_i)\, V_{in}$

(c) inverting
$V_{out} = -(R_f/R_i)\, V_{in}$

(d) summing
$V_{out} = -R_f(V_1/R_1 + V_2/R_2)$

(e) difference
$V_{out} = (R_2/R_1)(V_2 - V_1)$

(f) transimpedance
$V_{out} = -IR_f$

Figure 4.13: The six basic op-amp circuits. Each uses negative feedback, so the two golden rules give its output. (a) The follower copies its input and draws no input current. (b) The non-inverting amplifier. (c) The inverting amplifier; its inverting input is a virtual ground. (d) The summing amplifier adds the currents of its inputs at the virtual ground. (e) The difference amplifier, with $R_4/R_3 = R_2/R_1$. (f) The transimpedance amplifier turns a current into a voltage.

divider is unloaded and $V_- = V_{out} R_i/(R_i + R_f)$. Rule 2 sets this equal to V_{in}, so the gain is

$$G = \frac{V_{out}}{V_{in}} = 1 + \frac{R_f}{R_i}. \tag{4.17}$$

With $R_i = 1.0\,\text{k}\Omega$ and $R_f = 99\,\text{k}\Omega$, $G = 100$: a 10 mV sensor signal becomes 1.00 V. The input still draws no current, so the circuit keeps the follower's virtue.

The inverting amplifier and the virtual ground

Ground the non-inverting input and bring the signal to the inverting input through R_i (Fig. 4.13c). Rule 2 holds the inverting input at 0 V, although it is not connected to ground: it is a *virtual ground*. The current through R_i is therefore V_{in}/R_i. By rule 1 none of it enters the op amp, so all of it flows on through R_f, and the output must sit $R_f V_{in}/R_i$ below zero:

$$V_{out} = -\frac{R_f}{R_i}\, V_{in}. \tag{4.18}$$

With $R_i = 10\,\text{k}\Omega$ and $R_f = 100\,\text{k}\Omega$ the gain is -10: the output is ten times larger and of opposite sign. The price is the input resistance. The source sees R_i to the virtual ground, only $10\,\text{k}\Omega$ here, so a high-impedance sensor must not drive it directly (Example 4.3 shows how badly that fails).

The summing amplifier

Because the virtual ground stays at 0 V, several inputs can feed it through their own resistors without disturbing each other (Fig. 4.13d). Their currents add in R_f:

$$V_{\text{out}} = -R_f \left(\frac{V_1}{R_1} + \frac{V_2}{R_2} \right). \tag{4.19}$$

When one input carries the sensor signal and the other a fixed reference voltage, the output is shifted by a constant. That is how a sensor's offset is removed before the ADC (Section 4.10).

The difference amplifier

To subtract a reference sensor from an active one we need $V_2 - V_1$. The *difference amplifier* does it with one op amp and four resistors (Fig. 4.13e). By rule 1, R_3 and R_4 form an unloaded divider, $V_+ = V_2 R_4/(R_3 + R_4)$, and rule 2 puts V_- at the same voltage. The current $(V_1 - V_-)/R_1$ through R_1 flows on through R_2, so $V_{\text{out}} = V_- - (R_2/R_1)(V_1 - V_-)$. Together,

$$V_{\text{out}} = \frac{R_4}{R_3 + R_4} \left(1 + \frac{R_2}{R_1} \right) V_2 - \frac{R_2}{R_1} V_1. \tag{4.20}$$

When $R_4/R_3 = R_2/R_1 = G$ this collapses to

$$V_{\text{out}} = G\,(V_2 - V_1). \tag{4.21}$$

The inputs draw current through R_1 and through $R_3 + R_4$, so the circuit's input resistance is only that of its resistors, tens of kilohms.

The rejection of the common mode depends on how well the resistors match. Put $V_1 = V_2 = V_{\text{cm}}$ in Eq. (4.20) to find the common-mode gain, $A_{\text{cm}} = (R_1 R_4 - R_2 R_3)/[R_1(R_3 + R_4)]$, which is zero only when $R_4/R_3 = R_2/R_1$. If each resistor is off by up to a fraction ϵ (its tolerance) in the worst direction, then $A_{\text{cm}} \approx 4\epsilon G/(1 + G)$ and

$$\text{CMRR} \approx \frac{1 + G}{4\epsilon}. \tag{4.22}$$

Take an active and a reference electrode that both sit near 1.0 V and differ by 5 mV, amplified by $G = 10$ ($R_1 = R_3 = 10\,\text{k}\Omega$, $R_2 = R_4 = 100\,\text{k}\Omega$) to a wanted 50 mV. With 1 % resistors the worst-case CMRR is $11/0.04 = 275$ (48.8 dB), and the shared volt adds up to $(G/\text{CMRR}) \times 1.0\,\text{V} = 36\,\text{mV}$ of error (the exact worst case from Eq. (4.20) is 35.7 mV). With 0.1 % resistors the CMRR is 2750 (68.8 dB) and the error 3.6 mV. The *instrumentation amplifier* of Chapter 8 fixes both weaknesses: it puts two buffers in front of a difference amplifier and reaches a CMRR of 100 dB or more. Chapter 5 uses the same subtraction to read the Wheatstone bridge of a resistive sensor.

The transimpedance amplifier

Leave out the input resistor of the inverting amplifier and bring a current straight to the virtual ground (Fig. 4.13f). By rule 1 the current I flows through R_f, and by rule 2 the inverting input stays at 0 V, so $V_{\text{out}} = -IR_f$. The gain has units of resistance (volts per amp), which is why the circuit is called a *transimpedance amplifier* (TIA), or current-to-voltage converter. It reads amperometric sensors, photodiodes, and ion channels, and Section 4.10 designs it.

KEY IDEA Follower 1; non-inverting $1 + R_f/R_i$; inverting $-R_f/R_i$; difference $(R_2/R_1)(V_2 - V_1)$; transimpedance $-IR_f$.

Real op amps

The output of a real op amp cannot swing beyond the supply rails (a "rail-to-rail" part comes within tens of millivolts of them), and on a single supply, such as 0 to 3.3 V, signals are referred to a mid-supply reference instead of ground. The output can supply only a few to tens of milliamps. And the gain falls with frequency: above a few hertz the open-loop gain of most op amps falls as $1/f$, so the product of gain and bandwidth, the *gain-bandwidth product* GBW, is constant. A stage whose *noise gain* is G_n (the gain the op amp applies to a voltage at its own input, $1 + R_f/R_i$ for both the inverting and the non-inverting amplifier) has a bandwidth of about

$$f_{-3\,\text{dB}} \approx \frac{\text{GBW}}{G_n}. \tag{4.23}$$

A gain of 1000 from an op amp with GBW = 1 MHz leaves 1 kHz of bandwidth in one stage; two stages of $\sqrt{1000} = 31.6$ keep about 20 kHz. Finally, the inputs draw a small current and carry a small offset voltage, the subject of the next section.

4.9 *Bias Current, Offset, and the Electrometer Buffer*

RULE 1 IS NOT QUITE TRUE. The input transistors of a real op amp draw a small current, the *input bias current* I_B. The two inputs draw slightly different currents, and their difference is the *input offset current* I_{OS}. A real op amp also behaves as if a small voltage V_{OS}, the *input offset voltage*, sat in series with one input, because its input transistors are not perfectly matched. Op amps are grouped by the kind of input transistor: *bipolar* transistors need a small steady input current to work, while *JFET* and *CMOS* inputs (the latter built from the MOSFETs of Chapter 5) are controlled by voltage and draw almost none. Table 4.2 lists typical values. Each error appears at the output multiplied by a gain: V_{OS} by the noise gain $1 + R_f/R_i$; the drop $I_B R_S$ that the bias current makes in the source resistance R_S by the signal gain;

and the bias current of the inverting input adds $I_B R_f$ across the feedback resistor. For sensors the term $I_B R_S$ is the one to watch. A bipolar op amp with 10 nA of bias current reading a 1 kΩ source makes 10 µV of error; the same op amp reading a 100 MΩ electrode makes 1 V. The bias currents of bipolar parts are large but well matched, so a resistor equal to R_S in the feedback path cancels most of the error, leaving $I_{OS} R_S$. For JFET and CMOS parts the cure is to choose a part with a smaller I_B. Bipolar inputs suit low-impedance sources; JFET, CMOS, and electrometer inputs suit photodiodes and glass electrodes.

Input	I_B	V_{OS}
Bipolar	0.1 to 100 nA	10 to 500 µV
JFET	1 to 50 pA	0.1 to 3 mV
CMOS	0.1 to 10 pA	0.01 to 5 mV
Electrometer	below 0.1 pA	0.05 to 1 mV

Table 4.2: Typical input bias current and offset voltage of op amps at 25 °C, by input stage. The bias current of JFET and CMOS inputs roughly doubles for every 10 °C of warming.

The electrometer buffer

Some sensors are voltage sources with an enormous internal resistance. A glass pH electrode has 50 MΩ to 500 MΩ, ion-selective microelectrodes reach gigaohms, and the floating gate of an extended-gate FET sensor (Chapter 5) is almost a pure capacitance. Model the sensor as its voltage E in series with its resistance R_S. An amplifier with input resistance R_{in} forms a divider with it, the loading of Chapter 1, and reads

$$V = E \frac{R_{in}}{R_{in} + R_S}, \tag{4.24}$$

so R_{in} must exceed R_S by a factor of 10^3 to 10^4 for an error below 0.1 % to 0.01 %, and the bias current must be small enough that $I_B R_S$ is negligible. An *electrometer buffer* is a follower built for this: an op amp with an input resistance above 10^{12} Ω and a bias current in the femtoamp range. Its circuit board needs care too. A fingerprint or humid air can leak 10^{10} Ω to the input, so the input pin sits inside a *guard ring*, a copper trace driven by the buffer's output at the input potential. With no voltage across the leakage path, no leakage current flows; a cable shield driven the same way keeps the cable from loading the electrode.

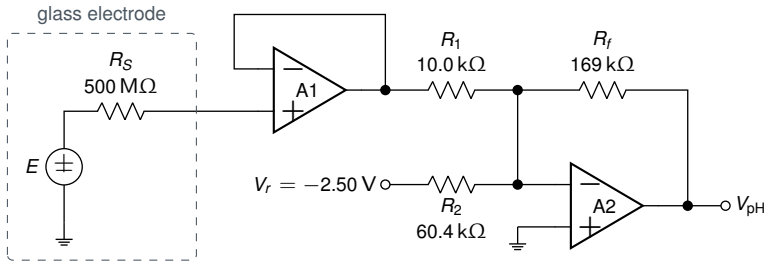


Figure 4.14: A pH meter that reads 1 V per pH. The glass electrode is modeled as its DC potential E behind 500 MΩ. A1 is an electrometer follower; A2 is an inverting summing amplifier that applies the gain −16.9 and adds the 7.0 V offset that makes V_{pH} = pH.

Worked Example 4.3 *A pH meter that reads volts as pH*

A glass pH electrode follows

$$E = E^0 - \frac{2.303RT}{F}\,\mathrm{pH},$$

with $E = 0$ at pH 7.00 and an internal resistance of 500 MΩ. (a) Find the slope at 25 °C. (b) Design an amplifier whose output reads the pH directly, 1 V per pH unit. (c) Why not connect the electrode straight to an inverting amplifier? (d) The op amps have input bias currents of 1 pA. Find the error at the output. Repeat for 1 nA.

SOLUTION. (a) This is the Nernst equation of Chapter 3 for hydrogen ions in base-10 form ($2.303 = \ln 10$). At 298.15 K the slope is $2.303RT/F = 59.16$ mV per pH, and E falls as the pH rises; $E^0 = 7 \times 59.16\,\mathrm{mV} = 0.414$ V puts the zero at pH 7.

(b) The output must be $V_{\mathrm{pH}} = 7 - E/(0.059\,16\,\mathrm{V})$, a gain of −16.90 plus an offset of +7.00 V. An electrometer follower (A1 in Fig. 4.14) feeds an inverting summing amplifier (A2), which by Eq. (4.19) gives

$$V_{\mathrm{pH}} = -\frac{R_f}{R_1}E - \frac{R_f}{R_2}V_r.$$

With $R_1 = 10.0$ kΩ and $R_f = 169$ kΩ (both standard 1 % values) the gain is −16.9. A reference $V_r = -2.50$ V and $R_2 = R_f \times 2.50/7.00 = 60.36$ kΩ give the offset; the standard 60.4 kΩ gives 6.995 V. Check: at pH 4, $E = 3 \times 0.05916 = 0.1775$ V and $V_{\mathrm{pH}} = -16.9 \times 0.1775 + 6.995 = 4.00$ V. In practice R_2 and R_f include small trimmers (adjustable resistors), set with pH 7 and pH 4 buffers: a two-point calibration (Section 4.5).

(c) An inverting amplifier's input resistance is $R_1 = 10$ kΩ. By Eq. (4.24) it would see $E \times 10^4/(5 \times 10^8) = 2 \times 10^{-5}\,E$: the electrode's whole signal would drop across its own glass. The follower's input resistance of 10^{12} Ω loses only $500 \times 10^6/10^{12} = 0.05$ % of E, a small slope error that the calibration absorbs. A leakage of 10^{11} Ω on the board would already cost 0.5 %, which is why the input needs a guard.

(d) A1's bias current flows through the electrode resistance and shifts the voltage A1 sees by $I_B R_S = 1\,\mathrm{pA} \times 500\,\mathrm{M\Omega} = 0.5$ mV. A2 multiplies that by 16.9: 8.45 mV at the output, 0.0085 pH, acceptable for a meter that displays 0.01 pH. A2's own bias current through R_f adds only $1\,\mathrm{pA} \times 169\,\mathrm{k\Omega} = 0.17\,\mu\mathrm{V}$. With 1 nA, $I_B R_S = 0.5$ V at the input and 8.45 V at the output, an error of 8.5 pH units with the amplifier close to its rail. A bipolar op amp cannot read a glass electrode; a JFET or CMOS electrometer part is required. ■

Temperature matters too: the Nernst slope changes by 0.198 mV per pH per kelvin, so a reading at pH 4 taken at 37 °C with a 25 °C

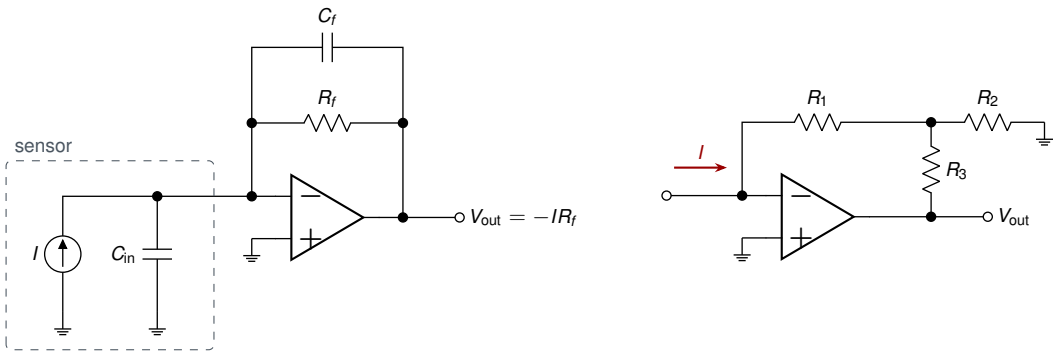
calibration is off by 0.12 pH. Real meters measure the temperature and scale the slope.

4.10 Measuring Small Currents

AMPEROMETRIC SENSORS, PHOTODIODES, AND ION CHANNELS produce currents from microamps down to picoamps. All of them are read with the transimpedance amplifier of Section 4.8: the sensor connects straight to the virtual ground, and its current, which cannot enter the op amp, flows through the feedback resistor,

$$V_{\text{out}} = -IR_f \tag{4.25}$$

(Fig. 4.15a). The virtual ground also holds the sensor at a fixed potential whatever its current, which is what an electrochemical working electrode needs; Chapter 6 builds the potentiostat around this. Small currents need a large R_f and an op amp whose bias current is far below the signal, and a large R_f brings two problems of its own: it is slow, and it can make the loop unstable.



(a) transimpedance amplifier

(b) T-network, $R_{\text{eff}} = R_1 + R_3 + R_1R_3/R_2$

Figure 4.15: Low-current amplifiers. (a) The TIA with its feedback capacitor and the sensor's capacitance C_{in} (electrode, cable, op-amp input). (b) A T-network makes three modest resistors act like one very large one.

Choosing the feedback resistor

For a full-scale current I_{FS} and an output swing V_{FS}, the feedback resistor is $R_f = V_{\text{FS}}/I_{\text{FS}}$. Table 4.3 lists the decades for a 1 V output. A sensor whose current spans several decades needs several feedback resistors, switched by the firmware.

Resistors above about 100 MΩ are large, expensive, and sensitive to humidity and surface contamination. A *T-network* (Fig. 4.15b) replaces one with three. The junction of R_1 and R_3 sits at $-IR_1$, so R_2 adds a

current IR_1/R_2 that flows on through R_3 with I, and

$$V_{\text{out}} = -I\left(R_1 + R_3 + \frac{R_1 R_3}{R_2}\right) = -IR_{\text{eff}}. \tag{4.26}$$

With $R_1 = R_3 = 1\,\text{M}\Omega$ and $R_2 = 1\,\text{k}\Omega$, $R_{\text{eff}} = 1.002\,\text{G}\Omega$. The price is in the errors: the op amp's offset and voltage noise see a gain of $1 + R_3/R_2 = 1001$, so a 1 mV offset becomes 1 V at the output, and the current noise is that of R_1 (Table 4.3), $129\,\text{fA}/\sqrt{\text{Hz}}$ instead of the $4\,\text{fA}/\sqrt{\text{Hz}}$ of a real $1\,\text{G}\Omega$ resistor. A T-network buys gain, not low noise.

I_{FS}	R_f (1 V)	C_f (10 Hz)	$\sqrt{4kT/R_f}$ (fA/$\sqrt{\text{Hz}}$)
1 µA	1 MΩ	15.9 nF	129
100 nA	10 MΩ	1.59 nF	40.7
10 nA	100 MΩ	159 pF	12.9
1 nA	1 GΩ	15.9 pF	4.07
100 pA	10 GΩ	1.59 pF	1.29

Table 4.3: Feedback resistor for a 1 V output, capacitor for a 10 Hz bandwidth, and the resistor's current noise at 300 K.

The feedback capacitor, bandwidth, and stability

A capacitor C_f across R_f makes the stage a first-order low-pass filter, the RC filter of Chapter 1 with its corner at

$$f_c = \frac{1}{2\pi R_f C_f}. \tag{4.27}$$

Every resistor already has a stray capacitance of about 0.1 pF to 0.3 pF across it, so a large R_f is slow even without C_f: 1 GΩ with 0.2 pF has a corner at 800 Hz, 10 GΩ at 80 Hz.

C_f also keeps the circuit stable. Together with R_f, the capacitance C_{in} of the sensor, its cable, and the op-amp input delays the feedback signal, and without C_f the loop can ring or oscillate. A common rule gives a well-behaved response when

$$C_f \geq \sqrt{\frac{C_{\text{in}}}{2\pi R_f\,\text{GBW}}}. \tag{4.28}$$

A slow sensor easily satisfies it, because the C_f chosen for bandwidth is larger. For an electrode of 10 nF on a 500 MΩ resistor and an op amp with GBW = 1 MHz, the rule asks for at least 1.8 pF, while a 5 Hz bandwidth already needs 63.7 pF.

Worked Example 4.4 *A glucose readout with 0.1 mM resolution*

The glucose sensor of Fig. 4.6b (125 nA/mM, linear to about 6.7 mM) will be read by a digital voltmeter with a resolution of 0.1 V. A glucose resolution of 0.1 mM is required. Design an op-amp circuit.

Solution. A change of 0.1 mM changes the current by 12.5 nA, and that must move the output by at least 0.1 V. A TIA needs $R_f \geq 0.1\,\text{V}/12.5\,\text{nA} = 8.0\,\text{M}\Omega$. Choose exactly 8.0 MΩ (for example 6.8 MΩ and 1.2 MΩ in series): the output is then 1.0 V per mM, and the display reads millimolar directly. The working electrode connects to the virtual ground, which holds its potential; an oxida-

tion current flows out of the summing node into the electrode, so the output is positive, $+IR_f$.

The other choices follow. The linear range ends near 6.7 mM, or 6.7 V, so the op amp needs supplies of at least ± 9 V or a rail-to-rail part on a single 12 V supply. Glucose changes over minutes, so a 2.2 nF capacitor across R_f ($f_c = 9.0$ Hz) removes interference without slowing the reading. The bias current must be far below the 12.5 nA resolution step: a JFET or CMOS op amp with $I_B < 10$ pA contributes less than 0.1 % of it. (A 100 kΩ TIA followed by a gain of 80.6 would also give 1.01 V/mM, but its second stage multiplies its own offset by 81.6: 1 mV becomes 82 mV, nearly one resolution step.) ■

Noise in a transimpedance amplifier

The feedback resistor sets the noise floor of a TIA. Its thermal current noise, $\sqrt{4kT/R_f}$, falls as R_f grows (Table 4.3 and Fig. 4.8b), so in a TIA the gain belongs in the first stage: a larger R_f gives both more signal and less input-referred noise. (In a voltage amplifier it is the other way around: a larger resistor is a noisier one.) The resistor's noise equals the signal's own shot noise when $4kT/R_f = 2qI$, that is at $I = 2kT/(qR_f)$: 5.2 nA for $R_f = 10$ MΩ at 300 K. Above that current the shot noise is the floor and a larger R_f would not help; below it, R_f sets the floor, which is why picoamp amplifiers, such as the patch clamp of Chapter 3, use gigaohm resistors.

WORKED EXAMPLE 4.5 *A noise budget that the ADC wins*

An amperometric sensor delivers 2.0 nA to a TIA with $R_f = 100$ MΩ, followed by a sharp low-pass filter with a noise bandwidth of 10 Hz. The op amp adds an input current noise of 1 fA/$\sqrt{\text{Hz}}$; its voltage noise is negligible here. A 12-bit ADC with a 3.3 V reference digitizes the output. Work at 300 K. (a) Find each noise current referred to the input and the SNR of the analog chain. (b) Include the ADC. (c) What changes with a 16-bit ADC?

SOLUTION. (a) With $B = 10$ Hz, R_f gives $\sqrt{4kTB/R_f} = 40.7$ fA, the signal's shot noise is $\sqrt{2qIB} = 80.1$ fA, and the op amp adds $1 \times \sqrt{10} = 3.2$ fA. In power, $\sqrt{40.7^2 + 80.1^2 + 3.2^2} = 89.9$ fA. The SNR is 2.0 nA/89.9 fA $= 2.2 \times 10^4$, or 86.9 dB. At the output this noise is only 89.9 fA $\times$ 100 MΩ $= 9.0$ µV on a 0.20 V signal. (A first-order filter with its corner at 10 Hz would give $B = 15.7$ Hz and 1.25 times this noise.)

(b) The ADC's 0.233 mV of quantization noise is 26 times larger than the analog noise at the output. Referred to the input it is $0.233\,\text{mV}/100\,\text{M}\Omega = 2.33\,\text{pA}$, so the total is 2.33 pA and the SNR drops to 58.7 dB. The converter, not the physics, limits this design.

(c) A 16-bit converter has $q_n = 14.5\,\mu\text{V}$, or 145 fA at the input; with the analog noise the total is 171 fA and the SNR 81.4 dB. The rule: give the first stage enough gain that the ADC's noise, referred to the input, falls below the sensor's own, or use more bits. ■

Fitting the signal into the ADC

An ADC accepts 0 to V_{ref}. Most sensor signals sit on a large offset, or have the wrong sign, or cover only part of that range. The conditioning chain must subtract the offset and scale the rest to fill the converter:

$$V_{\text{ADC}} = G\,(V_s - V_{\text{offset}}) + V_{\text{low}}. \tag{4.29}$$

A summing amplifier with one input from the sensor and one from a reference does both, as in the pH meter of Example 4.3; taking the reference from a DAC (a digital-to-analog converter) lets the firmware retrim the offset as the sensor ages. On a single supply, the non-inverting input of a TIA can sit at a reference voltage V_b instead of ground, so that $V_{\text{out}} = V_b - IR_f$ swings around V_b. By rule 2 the sensor electrode then sits at V_b as well.

The figures of merit and the six op-amp circuits of this chapter are the tools of the rest of Part II. Chapter 5 puts them to work first on transducers whose output is a resistance, a capacitance, or the current of a transistor.

Summary

- A biosensor couples a bioreceptor to a transducer and electronics. Catalytic receptors report a rate; affinity receptors report occupancy.
- Sensitivity is the calibration slope, $S = \mathrm{d}S_{\text{out}}/\mathrm{d}S_{\text{in}}$; per area, S/A. The selectivity coefficient $K_{A,B} = S_B/S_A$ gives the error $K_{A,B}c_B$.
- Enzyme and binding curves are linear within a tolerance δ only up to $\delta K_m/(1-\delta)$, or $0.053\,K_m$ at 5 %; membranes trade sensitivity for range. First order: $t_{63} = \tau$, $t_{90} = 2.30\tau$, $t_{95} = 3.00\tau$, $f_c = 1/(2\pi\tau)$.
- Noise: thermal $\sqrt{4kTRB}$ or $\sqrt{4kTB/R}$, shot $\sqrt{2qIB}$ (noise bandwidth B, $\pi/2$ times the corner of a first-order filter), flicker below a corner $f_{1/f}$, quantization $\text{LSB}/\sqrt{12}$. Add in power and refer to the input; resolution is input-referred noise divided by the sensitivity.

- Accuracy is the bias of the mean; precision is the spread (s, CV). Calibrate to fight bias, average to fight noise. Decision limit $L_C = \bar{y}_B + 1.645\,s_B$; detection limit $L_D = \bar{y}_B + 3.29\,s_B$; in concentration, $(L - \bar{y}_B)/S$.
- A reference sensor and a differential measurement cancel shared drift; sensitivity drift needs recalibration. CMRR $= A_d/A_{\text{cm}}$.
- Op amp: $V_{\text{out}} = A(V_+ - V_-)$, $A \approx 10^5$; with negative feedback, no current enters the inputs and $V_- = V_+$. Follower 1; non-inverting $1 + R_f/R_i$; inverting $-R_f/R_i$; difference $(R_2/R_1)(V_2 - V_1)$, CMRR $\approx (1+G)/(4\epsilon)$; transimpedance $-IR_f$; bandwidth $\approx \text{GBW}/G_n$.
- A high-impedance electrode needs an electrometer follower with $R_{\text{in}} \gg R_S$, a small $I_B R_S$ (1 pA through 500 MΩ gives 0.5 mV), and a guarded input.
- TIA: $R_f = V_{\text{FS}}/I_{\text{FS}}$, $f_c = 1/(2\pi R_f C_f)$; its current noise $\sqrt{4kT/R_f}$ falls with R_f, so the gain belongs in the first stage. A summing stage shifts and scales the signal into the ADC.

4.11 *Problems*

4.1 • The enzyme urease hydrolyzes urea:

$$\text{CO(NH}_2\text{)}_2 + 2\,\text{H}_2\text{O} + \text{H}^+ \longrightarrow 2\,\text{NH}_4{}^+ + \text{HCO}_3{}^-.$$

(a) Check that the equation is balanced in atoms and charge.

(b) Propose five sensors for urea with different transduction principles (Table 4.1) that use this reaction, and say what each one measures.

(c) Which principle fits worst, and why?

4.2 ◦ An electrode with a resistance of 5 MΩ at body temperature (310 K) picks up a signal of 200 µV, read with a noise bandwidth of 50 Hz.

(a) Find the electrode's thermal noise voltage and the SNR in decibels.

(b) A gain of 100 feeds a 12-bit, 3.3 V ADC. Find its quantization noise referred to the input, the total noise, and the SNR.

(c) Which source dominates, and what would you change?

4.3 • A pH electrode reads a pH 7.00 buffer eight times: 7.04, 7.06, 7.03, 7.07, 7.05, 7.04, 7.06, and 7.05.

(a) Find the mean, the standard deviation, the CV, and the bias.

(b) Is the electrode more in need of averaging or of calibration?

(c) In a pH 4.00 buffer the mean reading is 4.09. Find a two-point correction $\mathrm{pH}_{\mathrm{true}} = a\,\mathrm{pH}_{\mathrm{read}} + b$.

4.4 •• A pH electrode with 200 MΩ output resistance is used at 15 °C, where its slope is Nernstian.

(a) Find the slope in mV/pH.

(b) Design the gain and offset of Fig. 4.14 for 1 V per pH with $E = 0$ at pH 7 and $V_r = -2.50$ V, keeping $R_1 = 10.0\,\mathrm{k\Omega}$.

(c) Find the output error from A1's bias current for an op amp with 100 fA and one with 50 pA.

(d) The board leaks $5 \times 10^{10}\,\Omega$ from the input to ground. Find the resulting slope error in percent and the pH error at pH 10.

4.5 • A sensor current flows out of a TIA's summing node, from 0 to 200 nA. The circuit runs on a single 3.3 V supply and feeds a 0 to 3.3 V ADC.

(a) Where should the non-inverting input sit, and what R_f maps the range onto 0.1 V to 3.1 V?

(b) Choose C_f for a 2 Hz bandwidth.

(c) What potential does the sensor electrode sit at?

5
Physical and Field-Effect Transducers

In this chapter you will learn to

- Explain how a Wheatstone bridge reads small changes of resistance, and find the output of quarter, half, and full bridges.
- Compare RTDs, thermistors, and thermocouples, and remove the resistance of long leads with a four-wire connection.
- Relate the gauge factor of a strain gauge to its geometry and its piezoresistance, and choose a gauge bridge.
- Describe how a silicon diaphragm senses pressure and how a quartz crystal weighs a layer of molecules.
- Compute the change of a differential capacitor and read it with a bridge or a charge amplifier, with attention to parasitic capacitance.
- Measure the impedance of tissue with four electrodes, and explain why it changes with frequency.
- Describe the MOSFET as a voltage-controlled switch and resistor, and use the square law to find its current and transconductance.
- Explain how an ion-sensitive FET senses pH, why its gate oxide sets the sensitivity, and why the Debye length limits FET biosensors.
- Read an ISFET at constant current and voltage, and cancel drift with a reference FET.

Slide a hand into a glove lined with thin stretchable capacitors, one over each knuckle, and bend a finger. The capacitor over that knuckle stretches, its plates grow and come closer, and its capacitance rises by a few percent. A small chip on the back of the hand turns that change into a number fifty times a second (Fig. 5.1). One design project

of Chapter 14 builds a cousin of it, a glove with printed pressure-sensing capacitors on the palm, and Lab 2 prints and reads a small array of such capacitors. The same chain, a physical change turned into a change of resistance, capacitance, or threshold voltage and then into a voltage, runs a fever thermometer, the pressure sensor at the tip of a catheter, a body-composition scale, and a transistor whose gate is a drop of blood.

Chapter 4 set out what a good sensor must do and built the op amps that read one. This chapter builds the transducers themselves and reads them with those op amps. It starts with the Wheatstone bridge and uses it for temperature, strain, and pressure sensors. It then turns to capacitors and to the impedance of tissue itself. The last three sections introduce the transistor from scratch, the ion-sensitive FET that turns pH into a shift of its threshold voltage, and the circuits that read it.

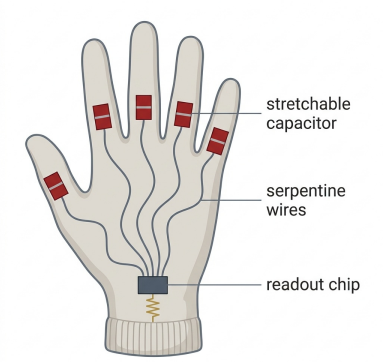


Figure 5.1: A capacitive sensor glove. Stretchable capacitors over the knuckles change capacitance as the fingers bend; a readout chip on the back of the hand digitizes them.

5.1 *The Wheatstone Bridge*

A RESISTIVE SENSOR CHANGES by a fraction of a percent per degree or per unit of strain. Read in a voltage divider (Chapter 1), that change is a small step on top of a large fixed voltage, and an amplifier that magnifies the step magnifies the fixed part too. The *Wheatstone bridge* puts two dividers side by side and takes the difference of their midpoints, so the fixed part cancels (Fig. 5.2). With the arms named as in the figure,

$$V_o = V_D - V_B = V_s \left(\frac{R_2}{R_1 + R_2} - \frac{R_4}{R_3 + R_4} \right) = V_s \frac{R_2 R_3 - R_1 R_4}{(R_1 + R_2)(R_3 + R_4)}. \tag{5.1}$$

The bridge is *balanced*, $V_o = 0$, when $R_1 R_4 = R_2 R_3$, whatever the supply voltage. A drifting supply therefore does not move the zero; it only scales the output.

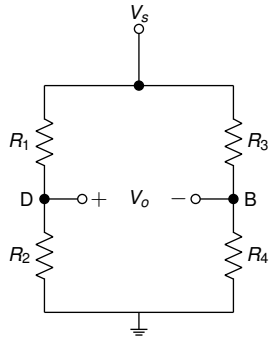


Figure 5.2: The Wheatstone bridge. The left divider sets node D, the right one node B, and $V_o = V_D - V_B$.

Quarter, half, and full bridges

Let all four arms equal R at the reference condition (zero strain, or a chosen temperature), and let each *active* arm, a sensor, change by a small fraction δ of R. With one active arm, $R_2 = R(1+\delta)$, the *quarter bridge* gives

$$V_o = V_s \frac{\delta}{4 + 2\delta} \approx V_s \frac{\delta}{4}. \tag{5.2}$$

The simple form $V_s\delta/4$ overestimates the output by a fraction $\delta/2$, the *nonlinearity*: 1 % at $\delta = 0.02$. Two sensors that change in opposite

directions in the same divider, $R_2 = R(1+\delta)$ and $R_1 = R(1-\delta)$, make a *half bridge* with $V_o = V_s\delta/2$ exactly, because the divider's total resistance stays $2R$. Four active arms, with $R_3 = R(1+\delta)$ and $R_4 = R(1-\delta)$ as well, make a *full bridge* with $V_o = V_s\delta$, again exactly linear (Fig. 5.3).

Both output nodes of the bridge sit near $V_s/2$, so the amplifier that follows must amplify their small difference and ignore the large voltage they share, the *common-mode* voltage. The difference amplifier of Chapter 4 does this; an *instrumentation amplifier*, a difference amplifier with a follower on each input, does it better and draws no current from the bridge (Chapter 8 builds one). Seen from the amplifier, an equal-arm bridge has an output resistance of $(R \parallel R) + (R \parallel R) = R$.

KEY EQUATION

$V_s\delta/4, \quad V_s\delta/2, \quad V_s\delta$

Quarter, half, and full bridge outputs for a small fractional change δ.

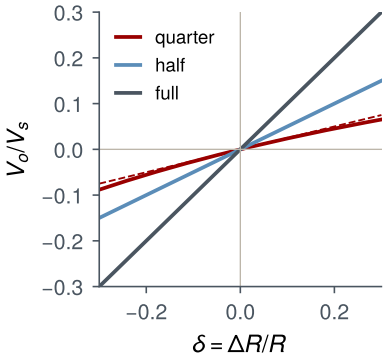


Figure 5.3: Bridge output against the fractional change of the active arms. The quarter bridge (red) bends away from its linear approximation (dashed); the half and full bridges are straight.

5.2 *Temperature Sensors*

Three sensor types cover nearly all medical temperature measurement. A *resistance temperature detector* (RTD) is a thin film or wire of platinum whose resistance rises almost linearly with temperature,

$$R_T = R_0\,[1 + \alpha(T - T_0)], \qquad \alpha = 0.003\,85\,\mathrm{K}^{-1}, \tag{5.3}$$

where R_0 is the resistance at $T_0 = 0\,^\circ\mathrm{C}$: $100\,\Omega$ for a Pt100, $1000\,\Omega$ for a Pt1000. (Over wide ranges a small quadratic term is added.) RTDs are stable and interchangeable, but their sensitivity is small: $0.34\,\%/\mathrm{K}$ at body temperature.

Every resistive thermometer must carry a current to be read, and that current heats it. The temperature rise is the power I^2R divided by the sensor's *dissipation constant*, the power that warms it by one kelvin, typically a few milliwatts per kelvin. A Pt100 at $37\,^\circ\mathrm{C}$ carrying $1\,\mathrm{mA}$ dissipates $0.11\,\mathrm{mW}$ and, at $2\,\mathrm{mW/K}$, reads $0.06\,\mathrm{K}$ high. Small currents or short pulses keep this *self-heating* negligible.

A *thermistor* is a sintered metal-oxide semiconductor whose resistance falls steeply with temperature (a negative temperature coefficient, hence NTC), $R_T = R_0 \exp[B(1/T - 1/T_0)]$, with T in kelvin, R_0 the resistance at T_0, usually $25\,^\circ\mathrm{C}$ (a "10K" thermistor has $10\,\mathrm{k}\Omega$ there), and B a material constant. Its relative sensitivity is

$$\frac{1}{R}\frac{\mathrm{d}R}{\mathrm{d}T} = -\frac{B}{T^2}, \tag{5.4}$$

$-4.1\,\%/\mathrm{K}$ at $37\,^\circ\mathrm{C}$ for $B = 3950\,\mathrm{K}$, twelve times the RTD's (Fig. 5.4). Thermistors are the usual choice in catheters, skin patches, and wearable thermometers; their curved response is straightened by a fitted equation or a lookup table in software.

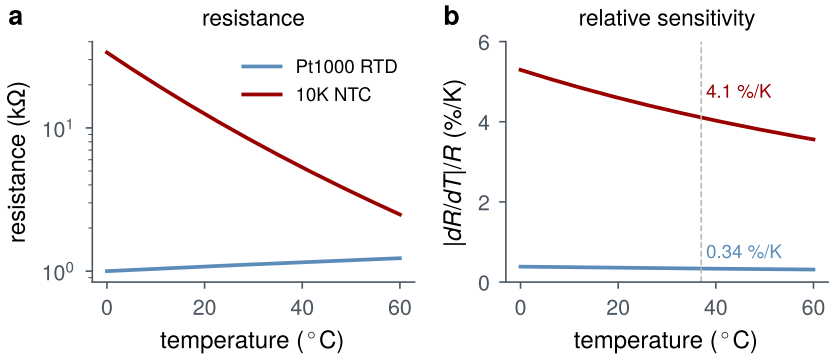


Figure 5.4: A Pt1000 RTD and a 10K NTC thermistor ($B = 3950\,\mathrm{K}$). (a) The thermistor changes by a factor of 13.5 from 0 to 60 °C, the RTD by 23 %. (b) At body temperature the thermistor's relative sensitivity is twelve times larger.

Thermocouples

A *thermocouple* is a pair of wires of two different metals joined at the measuring point. A temperature difference along a wire drives a small voltage (the Seebeck effect), and because the two metals respond differently, the free ends see a net voltage $V_{\mathrm{TC}} \approx S\,(T_h - T_c)$. Here T_h is the temperature of the measuring junction, T_c that of the terminals where the wires meet the copper of the instrument (the *cold junction*), and S is about $41\,\mu\mathrm{V/K}$ for type K wire. A thermocouple therefore measures a *difference* of temperature: a junction in tissue at 43 °C with its terminals at 25 °C gives only 0.74 mV, which an instrument that assumes a 0 °C cold junction would read as 18 °C. *Cold-junction compensation* measures T_c with a thermistor at the terminals and adds $S\,T_c$ in firmware. Thermocouples need no excitation current and can be made very fine, which suits hyperthermia and ablation probes.

Reading a sensor over long leads

A sensor at the end of a cable adds the resistance of its leads to its own. In a *two-wire* connection both leads sit in series with the sensor. A Pt100 changes by only $0.385\,\Omega/\mathrm{K}$, so a few ohms of lead mean degrees of error: a Pt100 at the tip of a 1 m catheter, on copper wire 0.10 mm in diameter, has $2.1\,\Omega$ in each lead, and the two together read as 11 K too warm. A calibration can subtract that, but not its drift as the leads warm in the body.

A *four-wire* or Kelvin connection separates the current from the measurement. A known current flows through one pair of leads, and a second pair carries the sensor's voltage to an amplifier with a high input resistance. No current flows in the sense leads, so their resistance drops no voltage, and $R_T = V/I$ exactly. Section 5.6 uses the same idea to remove the electrodes from a measurement of tissue, and Fig. 5.11 there draws it.

WORKED EXAMPLE 5.1 *Nonlinearity of an RTD bridge*

A Pt1000 RTD sits in the R_2 position of a bridge whose other arms are 1000 Ω, excited by $V_s = 2.5$ V. (a) The bridge is balanced at 0 °C. Find the exact output at 37 °C and the error of the linear formula $V_s\delta/4$. (b) For a fever thermometer the fixed arms are changed to $R(37\,°\text{C})$. Find the sensitivity near balance, and the output and its nonlinearity at 41 °C.

SOLUTION. (a) $\delta = \alpha T = 0.00385 \times 37 = 0.1425$. Equation (5.2) gives $V_o = 2.5 \times 0.1425/(4 + 0.285) = 83.1$ mV; the linear formula gives 89.0 mV, 7.1 % too high, as $\delta/2$ predicts.

(b) Now $R_1 = R_3 = R_4 = R(37) = 1142.45\,\Omega$. Each kelvin changes the RTD by $\alpha R_0 = 3.85\,\Omega$, so near balance the output changes by $V_s\alpha R_0/(4R_{37}) = 2.11$ mV/K. At 41 °C the arm has changed by $4 \times 3.85\,\Omega$, so $\delta = 15.4/1142.45 = 0.0135$. The output is $2.5 \times 0.0135/4.027 = 8.37$ mV, and the nonlinearity $\delta/2$ is only 0.67 %. Balancing the bridge at the center of the range of interest keeps δ small and the output linear. ■

5.3 Strain Gauges

STRETCH A WIRE AND ITS RESISTANCE RISES, for three reasons: it gets longer, it gets thinner, and the resistivity of its material changes. Start from $R = \rho L/A$ and take the fractional change:

$$\frac{\mathrm{d}R}{R} = \frac{\mathrm{d}\rho}{\rho} + \frac{\mathrm{d}L}{L} - \frac{\mathrm{d}A}{A}. \tag{5.5}$$

The *strain* is the fractional stretch, $\varepsilon = \mathrm{d}L/L$. Poisson's ratio ν says that each transverse dimension shrinks by $\nu\varepsilon$, so the area changes by $\mathrm{d}A/A = -2\nu\varepsilon$. The change of resistivity, the *piezoresistive* effect, is proportional to the stress $\sigma = E\varepsilon$, where E is Young's modulus: $\mathrm{d}\rho/\rho = \pi E\varepsilon$, with π the piezoresistive coefficient. Together they give the *gauge factor*,

$$\mathrm{GF} = \frac{\mathrm{d}R/R}{\varepsilon} = 1 + 2\nu + \pi E. \tag{5.6}$$

The first two terms are geometry; with $\nu \approx 0.3$ they give about 1.6 for metals. Metal alloys add a small resistivity term (constantan, with $\nu = 0.33$ and GF ≈ 2.1, has $\pi E \approx 0.44$), so metal foil gauges have GF near 2. In a semiconductor, strain changes how easily the carriers move, and πE dominates: for lightly doped p-type silicon (silicon with a trace of boron added; Section 5.7 explains doping) along a favorable crystal direction, $\pi = 72 \times 10^{-11}$ Pa^{-1} and $E = 169$ GPa give $\pi E \approx 120$, so silicon piezoresistors reach gauge factors of about 100. Stretchable

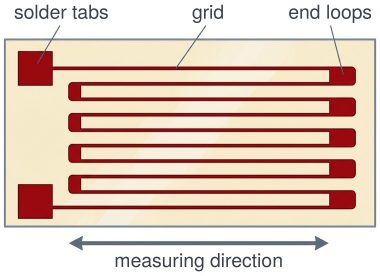


Figure 5.5: A metal foil strain gauge. The long grid lines lie along the measuring direction, so stretching along them raises the resistance most; the wide end loops add little resistance.

KEY EQUATION

$\mathrm{GF} = 1 + 2\nu + \pi E$

Geometry gives $1 + 2\nu$; piezoresistance gives πE.

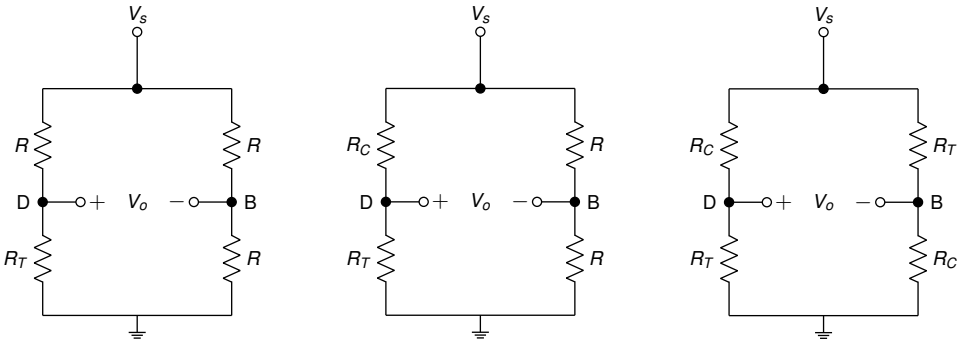
strain sensors use the cracking films and the percolating networks of Chapter 11 to reach gauge factors of tens to thousands, at the cost of hysteresis and drift.[1]

A foil gauge (Fig. 5.5) folds a long thin conductor into a grid, 120 to 1000 Ω, that is sensitive along one direction. Glued to a structure, it follows the surface strain. A strain of 1000 microstrain (10^{-3}) changes a GF 2 gauge by only 0.2 %, so gauges are always read in a bridge.

[1] One such sensor, a cracked gold film on an elastomer wrapped around tumors in mice, told treated from untreated tumors within 5 hours of the start of therapy: A. Abramson, C. T. Chan, Y. Khan, *et al.*, Z. Bao, *Sci. Adv.* (2022).

Bridges with one, two, and four active gauges

Figure 5.6 shows the three arrangements. On a bending beam the top surface stretches and the bottom compresses by the same amount, so gauges on the two faces give $+\delta$ and $-\delta$, with $\delta = \mathrm{GF}\,\varepsilon$. One active gauge makes a quarter bridge. One gauge in tension and one in compression in the same divider make a half bridge with twice the output and no nonlinearity. Four active gauges, two in tension and two in compression, make a full bridge with four times the quarter-bridge output. The extra gauges do a second job: all of them see the same temperature, so changes of resistance with temperature cancel, just as a balanced bridge ignores its supply. A quarter bridge gets the same benefit from a *dummy* gauge, mounted on an unstrained piece of the same material, in the adjacent arm.



(a) quarter: $V_s\,\delta/(4+2\delta)$ (b) half: $V_s\,\delta/2$ (c) full: $V_s\,\delta$

Figure 5.6: Strain-gauge bridges with one, two, and four active gauges. $R_T = R(1+\delta)$ is a gauge in tension, $R_C = R(1-\delta)$ a gauge in compression, and R a fixed resistor or a dummy gauge.

Real gauges differ from each other by a few tenths of a percent, so an unstrained bridge is not quite balanced, and its offset can be as large as the signal. A trimming potentiometer, or a large resistor in parallel with the higher arm, zeroes it before the measurement.

Worked Example 5.2 *One, two, and four active gauges*

A beam carries 350 Ω constantan gauges (GF = 2.1). Under load the top surface is strained by +1500 and the bottom by −1500 mi-

crostrain. The bridge is excited by 5 V. (a) Find δ and the output of a quarter, a half, and a full bridge. (b) What amplifier gain makes the full-bridge output 1.0 V?

SOLUTION. (a) $\delta = 2.1 \times 1.5 \times 10^{-3} = 3.15 \times 10^{-3}$, a change of only 1.10 Ω. Quarter bridge: $5 \times 0.00315/4.0063 = 3.93$ mV (the linear formula gives 3.94 mV). Half bridge: $5 \times 0.00315/2 = 7.88$ mV. Full bridge: $5 \times 0.00315 = 15.75$ mV.

(b) 1.0 V / 15.75 mV = 63.5, a job for an instrumentation amplifier, whose gain one resistor sets (Chapter 8). ■

5.4 *Pressure and Mass Sensors*

A BLOOD-PRESSURE SENSOR AT THE TIP OF A CATHETER is a strain gauge on a drum. Silicon micromachining etches a thin square diaphragm out of a silicon chip and implants four piezoresistors at the middle of its edges, where pressure stresses it most (Fig. 5.7). Two resistors carry their current across the edge and two along it; in p-type silicon the edge stress moves the two pairs by about the same amount in opposite directions, so the four make the full bridge of Section 5.1, $V_{\text{out}} = V_s\,\Delta R/R$, with

$$\frac{\Delta R}{R} \approx \frac{\pi_{44}}{2}\,(1-\nu)\,\sigma, \qquad \sigma \approx 0.308\,P\left(\frac{a}{h}\right)^2, \tag{5.7}$$

where $\pi_{44} \approx 138 \times 10^{-11}\,\text{Pa}^{-1}$ is the piezoresistive coefficient of p-type silicon, $\nu = 0.28$, and σ is the stress at the middle of the edge of a diaphragm of side a and thickness h under pressure P. A diaphragm 0.40 mm square and 10 µm thick at 300 mmHg (40.0 kPa) sees 19.7 MPa at its edges, so $\Delta R/R \approx 1\,\%$ and a full bridge excited by 3 V gives 29 mV: 33 µV per volt of excitation per mmHg, more than the standard 5 µV/V/mmHg of disposable blood-pressure transducers.

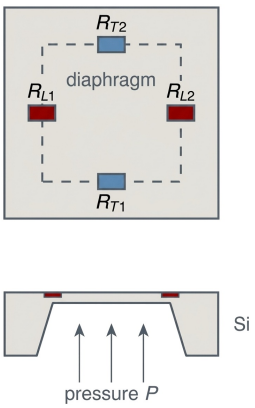


Figure 5.7: A piezoresistive pressure sensor. Top: four resistors at the edge centers of a square diaphragm, two with their current across the edge (R_{L1}, R_{L2}, red) and two along it (R_{T1}, R_{T2}, blue). Bottom: cross section; pressure from below bows the thin membrane up and compresses its top surface at the edges, so the red pair's resistance falls and the blue pair's rises.

The quartz crystal microbalance

Quartz is *piezoelectric*: squeezing it puts a voltage across it, and a voltage deforms it. A thin quartz disk with electrodes on both faces vibrates at a frequency set by its thickness (a 5 MHz crystal is 334 µm thick). A film added to a face lowers that frequency in proportion to its mass per area, $\Delta f = -C_S\,\Delta m/A$ (the Sauerbrey equation), with the Sauerbrey constant $C_S = 56.6\,\text{Hz}\,\text{cm}^2/\mu\text{g}$ at 5 MHz. Coated with antibodies, this *quartz crystal microbalance* (QCM) weighs the antigen that binds: $0.10\,\mu\text{g}/\text{cm}^2$ of protein lowers the frequency by 5.7 Hz. In water, though, the liquid's drag shifts the frequency by about 710 Hz,

and that shift changes by about 9 Hz per kelvin, more than the binding signal. The cure is differential, as for the strain gauges: a reference crystal with a non-binding layer in the same liquid, whose frequency is subtracted. Surface acoustic wave (SAW) sensors weigh with a wave that runs along the surface of a piezoelectric chip instead.

5.5 *Capacitive Sensing*

A PARALLEL-PLATE CAPACITOR of area A and gap d has $C = \varepsilon_0 \varepsilon_r A/d$, with $\varepsilon_0 = 8.854 \times 10^{-12}\,\mathrm{F/m}$ and ε_r the relative permittivity of the material between the plates (1 for air, about 3 for many polymers). A sensor changes C by moving a plate (displacement, pressure, acceleration), by changing the overlap area (a stretched elastomer), or by changing ε_r (humidity, a touching finger). Capacitive sensors draw no DC current and are cheap to make on silicon or plastic, which is why the glove of the opening and the pressure array of Lab 2 use them. Their weakness is size: a sensor of a few picofarads changes by femtofarads, and its wires add picofarads of their own.

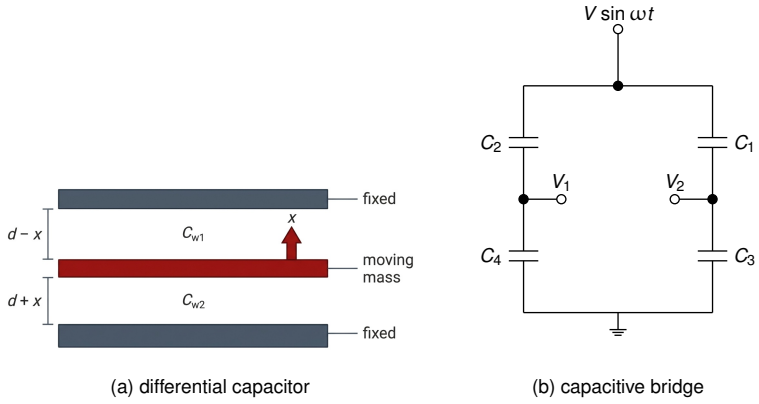


Figure 5.8: (a) A moving plate between two fixed plates forms two capacitors whose gaps change in opposite directions. (b) A capacitive bridge driven by an AC source. In the half bridge $C_2 = C_4 = C_0$, $C_1 = C_{w1}$, $C_3 = C_{w2}$; in the full bridge $C_1 = C_4 = C_{w1}$ and $C_2 = C_3 = C_{w2}$.

Differential capacitors and the capacitive bridge

A MEMS (microelectromechanical system) accelerometer, a tiny machine etched into silicon, suspends a proof mass on springs between two fixed electrodes (Fig. 5.8a). An acceleration a moves the mass by $x = a/\omega_0^2$, where $\omega_0 = \sqrt{k/m}$ is the resonant frequency of the mass m on springs of stiffness k; for a mass that resonates at 5 kHz, 1 g ($9.81\,\mathrm{m/s^2}$) moves it by 9.9 nm. The motion narrows one gap and widens the other:

$$C_{w1} = \frac{\varepsilon A}{d-x} = \frac{C_0}{1 - x/d}, \qquad C_{w2} = \frac{\varepsilon A}{d+x} = \frac{C_0}{1 + x/d}, \qquad C_0 = \frac{\varepsilon A}{d}. \tag{5.8}$$

For small motion both change by about $\Delta C = C_0 x/d$, in opposite directions.

A capacitor divider driven by an AC voltage V divides inversely with capacitance, because the impedance of a capacitor is $1/(j\omega C)$ (Chapter 1): the node between a top capacitor C_t and a bottom capacitor C_b sits at $VC_t/(C_t + C_b)$. Each capacitance here is inversely proportional to its gap, so that ratio becomes a ratio of gaps. In the half bridge of Fig. 5.8b the left divider holds $V_1 = V/2$, and the right one holds

$$V_2 = V\,\frac{d+x}{(d+x)+(d-x)} = \frac{V}{2}\left(1+\frac{x}{d}\right), \tag{5.9}$$

so $V_{\text{out}} = V_2 - V_1 = Vx/(2d)$. In the full bridge the left divider moves the other way, $V_1 = (V/2)(1 - x/d)$, and $V_{\text{out}} = Vx/d$. Both outputs are exactly linear in x, even though each capacitance is not.

KEY EQUATION

$V\,x/(2d)$ (half), $V\,x/d$ (full)

Capacitive bridge outputs for a differential capacitor.

Parasitic capacitance C_p from each output node to ground spoils this. It comes from bond pads over the substrate, from the wiring, and from the input of the amplifier. Each node then divides against an extra capacitor to ground, and with $C_1 + C_3 \approx C_2 + C_4 \approx 2C_0$ the full bridge gives

$$V_{\text{out}} = V\,\frac{2\,\Delta C}{2C_0 + C_p}. \tag{5.10}$$

Reading capacitance with a charge amplifier

Instead of a bridge, drive the two capacitors with opposite AC voltages, $+V$ on one and $-V$ on the other, and connect their common node, the moving mass, to the inverting input of an op amp whose feedback path is a capacitor C_f (Fig. 5.9). This is the inverting amplifier of Chapter 4 with capacitors in place of resistors, and the same rule applies: the op amp holds its inverting input at virtual ground. The two drive voltages push a net charge $V(C_{w1} - C_{w2})$ into that node, and with nowhere else to go it collects on C_f:

$$V_{\text{out}} = -V\,\frac{C_{w1} - C_{w2}}{C_f}. \tag{5.11}$$

Because the node stays at zero volts, a parasitic capacitance from it to ground has no voltage across it and carries no charge: it drops out. A large resistor R_b across C_f sets the DC level. This *charge amplifier* is the front end of most capacitive sensor chips, and its gain, a ratio of capacitors, does not depend on the drive frequency.

Integrated *capacitance-to-digital converters* do the same with switches in place of a sine wave. In one clock phase they charge the sensor C_x to a reference voltage V_{ref}; in the next they dump that charge, $C_x V_{\text{ref}}$,

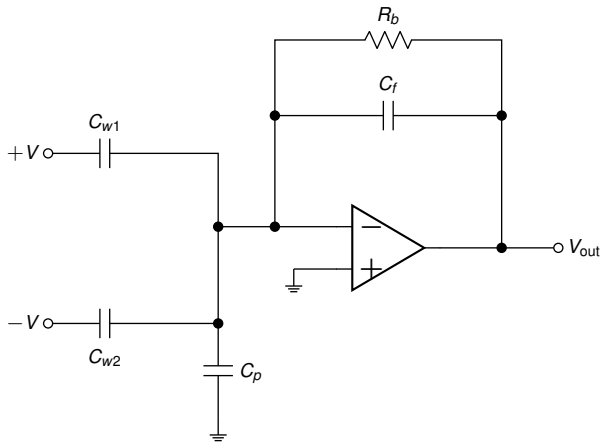


Figure 5.9: A charge amplifier for a differential capacitor. The two sensor capacitors are driven by AC voltages of amplitude V in antiphase. Their common node is a virtual ground, so the parasitic capacitance C_p never changes its voltage, and the output is $-V(C_{w1} - C_{w2})/C_f$.

onto C_f at the virtual ground, so

$$V_{\text{out}} = V_{\text{ref}} \frac{C_x}{C_f}, \tag{5.12}$$

and an ADC digitizes the result. The sensor's terminals only ever switch between fixed voltages and the virtual ground, so the parasitics drop out here too: with $V_{\text{ref}} = 1.8\,\text{V}$ and $C_f = 2.2\,\text{pF}$, each femtofarad of the sensor gives 0.82 mV. A third method connects the sensor to an inductor L (a small coil); the pair oscillates at $f = 1/(2\pi\sqrt{LC})$, and a counter reads that frequency. The converter of Lab 2 works this way.

5.6 *Bioimpedance*

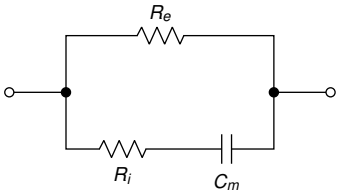


Figure 5.10: A simple model of tissue. Current at low frequency flows only through the extracellular fluid (R_e); at high frequency it also crosses the cell membranes (C_m) and the fluid inside the cells (R_i).

Tissue is a resistor and a capacitor at once. The fluids inside and outside its cells are salt water and conduct like resistors; the cell membranes, lipid bilayers of about $1\,\mu\text{F/cm}^2$ (Chapter 2), are capacitors that block low-frequency current. At low frequency, current flows only through the fluid around the cells; at high frequency the membranes' impedance falls and current also flows through the cells. Figure 5.10 captures this with three elements: the resistance R_e of the extracellular fluid, in parallel with the resistance R_i of the intracellular fluid in series with the membrane capacitance C_m. The impedance falls from R_e at low frequency to $R_e \parallel R_i$ at high frequency, and the change happens around $f = 1/[2\pi(R_e + R_i)C_m]$, tens of kilohertz for most tissues.

Bioimpedance measurements track whatever changes those fluids. Each heartbeat adds blood to a limb and lowers its impedance slightly; each breath raises the impedance of the chest, which bedside monitors use to count breaths through the ECG electrodes; and body-composition scales estimate body water from the impedance at about

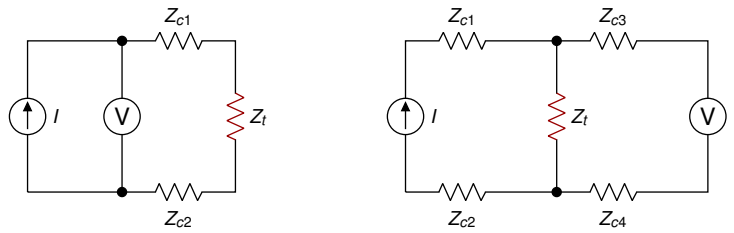
50 kHz. A flexible array of printed electrodes detected pressure damage in skin through its impedance before the damage was visible.[2] The impedance of an electrode in a solution, which affinity biosensors use, follows the same ideas and is the subject of Chapter 6.

[2] S. L. Swisher *et al.*, *Nat. Commun.* (2015).

Four electrodes

The difficulty is the electrodes. Each electrode-skin contact has an impedance of its own (Chapter 8 models it): at the tens of kilohertz used here, hundreds of ohms for a gel electrode and a kilohm or more for a dry one, and far more at low frequency. It also changes with sweat, pressure, and motion. A *two-electrode* measurement drives a current through two electrodes and reads the voltage across the same two, so it measures both contacts in series with the tissue (Fig. 5.11a).

The *four-electrode* (tetrapolar) measurement is the four-wire connection of Section 5.2 applied to the body. An AC current source drives a known current I through two outer electrodes, and two inner electrodes carry the voltage across the tissue between them to an amplifier with a high input impedance (Fig. 5.11b). Almost no current flows through the sensing contacts, so their impedance drops no voltage, and $Z = V/I$ is the tissue alone. Typical drives are tens to hundreds of microamperes at tens of kilohertz, far below what a person can feel at that frequency (Chapter 8 gives the safety limits; above 1 kHz they rise in proportion to frequency).



(a) two electrodes (b) four electrodes

Figure 5.11: Measuring the tissue impedance Z_t through electrode contacts Z_c. (a) The voltmeter sees the tissue and both contacts in series. (b) A second pair of electrodes senses the voltage; almost no current flows through Z_{c3} and Z_{c4}, so the voltmeter sees only the tissue.

Worked Example 5.3 *Two electrodes or four*

A wristband drives 200 µA rms at 50 kHz through a forearm segment. The segment's impedance is 300 Ω, and each dry electrode adds a contact impedance of 1.5 kΩ; treat all of them as resistances. (a) What does a two-electrode measurement report? (b) What does a four-electrode measurement report if the sense amplifier's input impedance is 1 MΩ? What voltage must the current source deliver? (c) A frequency sweep of the same segment gives 330 Ω at 5 kHz and 220 Ω at 1 MHz. Taking these as the low- and high-frequency

limits of the model in Fig. 5.10, find R_e and R_i.

SOLUTION. (a) $300 + 2 \times 1500 = 3300\,\Omega$, eleven times the tissue value; and the contacts change with sweat and pressure, so the error is not even constant.

(b) The tissue voltage is $200\,\mu A \times 300\,\Omega = 60\,mV$. The sense path is the amplifier's $1\,M\Omega$ in series with two $1.5\,k\Omega$ contacts, so the amplifier receives $10^6/(10^6 + 3000) = 0.997$ of it and reports $299.1\,\Omega$, 0.3 % low. The drive current still passes through the two outer contacts, so the source must deliver $200\,\mu A \times 3.3\,k\Omega = 0.66\,V$.

(c) At low frequency only the extracellular path conducts: $R_e = 330\,\Omega$. At high frequency $R_e \parallel R_i = 220\,\Omega$, so $1/R_i = 1/220 - 1/330$ and $R_i = 660\,\Omega$. Tracking the two separately tells a change of fluid outside the cells, such as swelling, which lowers R_e, from a change inside them. ■

5.7 The MOSFET

THE TRANSISTOR IS THE SWITCH AND THE AMPLIFIER inside every chip in this book, and the field-effect sensors of the next two sections are transistors too. The most common kind is the MOSFET, short for metal-oxide-semiconductor field-effect transistor. Two facts about doped silicon are enough to see how it works. Silicon with a small amount of boron added is *p-type*: its mobile charges are positive *holes*, places where an electron is missing. Silicon with phosphorus added is *n-type* (n^+ when heavily doped): its mobile charges are electrons.

An n-channel MOSFET (Fig. 5.12) is a p-type substrate with two n^+ islands, the *source* and the *drain*. Between them the silicon surface is covered by a thin insulator, the *gate oxide* (silicon dioxide a few to tens of nanometers thick), and on top of that sits the *gate* electrode. Gate and silicon form a parallel-plate capacitor with a capacitance per area $C_{ox} = \varepsilon_0 \varepsilon_r / t_{ox}$. The gate draws no DC current: it acts only through its electric field, hence *field effect*.

With the gate at the source voltage, the p-type region between source and drain blocks current in both directions: the transistor is an open switch. Raising the gate-source voltage V_{GS} pushes holes away from the surface and pulls electrons toward it. Above a *threshold voltage* V_T, typically 0.3 to 1 V, the surface holds more electrons than holes: a thin n-type layer, the *channel*, connects source and drain, and the switch closes. The electron charge in the channel per area is about $C_{ox}(V_{GS} - V_T)$, the charge on the gate capacitor beyond what reaching threshold took.

For a small drain-source voltage V_{DS} the channel is simply a resistor,

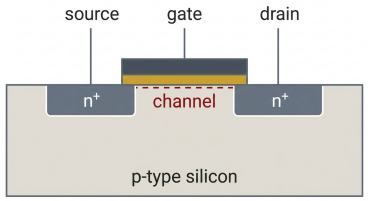


Figure 5.12: Cross section of an n-channel MOSFET. A thin oxide (gold) separates the gate from the p-type silicon. Above the threshold voltage a thin layer of electrons, the channel (red), connects the n^+ source and drain.

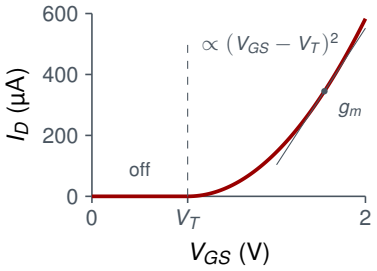


Figure 5.13: Saturation current of an n-channel MOSFET with $\beta = 0.69\,mA/V^2$ and $V_T = 0.7\,V$. Below threshold it is off; above, the current grows as the square of the overdrive. The tangent's slope is the transconductance g_m.

and its conductance is proportional to that charge: $G = \beta(V_{GS} - V_T)$, with $\beta = \mu C_{ox} W/L$, where μ is the mobility of the electrons (their drift speed per unit electric field) and W/L the width-to-length ratio of the channel. A MOSFET is therefore a *voltage-controlled resistor*. As V_{DS} grows, the channel thins near the drain, where the gate-to-channel voltage is smallest, and once V_{DS} exceeds $V_{GS} - V_T$ the current stops growing. The transistor is then in *saturation*, and its current is set by the gate alone:

$$I_D = \frac{\beta}{2}(V_{GS} - V_T)^2, \qquad \beta = \mu C_{ox}\frac{W}{L} \qquad (V_{DS} > V_{GS} - V_T). \quad (5.13)$$

KEY EQUATION
$I_D = \frac{\beta}{2}(V_{GS} - V_T)^2$
The square law, in saturation; $g_m = \beta(V_{GS} - V_T)$.

This is the *square law*. In words: raising the gate overdrive $V_{GS} - V_T$ both adds charge to the channel and allows a larger voltage along it before saturation, and the product of the two grows as the square (Fig. 5.13). A sensor cares about how far the current moves for a small change of gate voltage, the slope of that curve, called the *transconductance*: $g_m = \mathrm{d}I_D/\mathrm{d}V_{GS} = \beta(V_{GS} - V_T) = 2I_D/(V_{GS} - V_T)$.

The threshold voltage is where sensing enters. It depends on the gate material and on any charge in or on the oxide: positive charge near the channel helps the gate attract electrons and lowers V_T, and negative charge raises it. If a chemical reaction or a binding molecule can change that charge, V_T moves, and at a fixed V_{GS} the current moves with it.

Worked Example 5.4 *Current and transconductance of a MOSFET*

An n-channel MOSFET has a 50 nm SiO_2 gate oxide ($\varepsilon_r = 3.9$), an electron mobility of $500\,\mathrm{cm^2/(V\,s)}$ in the channel, and $W/L = 20$. (a) Find C_{ox} and β. (b) At $V_{GS} - V_T = 1.0\,\mathrm{V}$, find the resistance of the channel for a small V_{DS}, and the current and transconductance in saturation. (c) A sensing layer raises V_T by 55 mV at fixed V_{GS}. By how much does the saturation current fall?

Solution. (a) $C_{ox} = 3.9 \times 8.854 \times 10^{-12}/50 \times 10^{-9} = 6.91 \times 10^{-4}\,\mathrm{F/m^2}$, or $69\,\mathrm{nF/cm^2}$. With $\mu = 0.050\,\mathrm{m^2/(V\,s)}$, $\beta = 0.050 \times 6.91 \times 10^{-4} \times 20 = 0.69\,\mathrm{mA/V^2}$.

(b) For small V_{DS} the channel conductance is $\beta(V_{GS} - V_T) = 0.69\,\mathrm{mS}$, a resistance of $1.45\,\mathrm{k\Omega}$ that the gate sets. In saturation $I_D = (0.69/2) \times 1.0^2 = 345\,\mu\mathrm{A}$, and $g_m = \beta(V_{GS} - V_T) = 0.69\,\mathrm{mS}$: each millivolt on the gate moves $0.69\,\mu\mathrm{A}$.

(c) $g_m \Delta V_T = 0.69\,\mathrm{mS} \times 55\,\mathrm{mV} = 38\,\mu\mathrm{A}$. Exactly, $I_D = 0.345 \times 0.945^2 = 308\,\mu\mathrm{A}$, a fall of $37\,\mu\mathrm{A}$; the estimate from g_m is good to a few percent. Section 5.8 shows that 55 mV is what one pH unit does to a good ion-sensitive FET. ■

5.8 The Ion-Sensitive FET

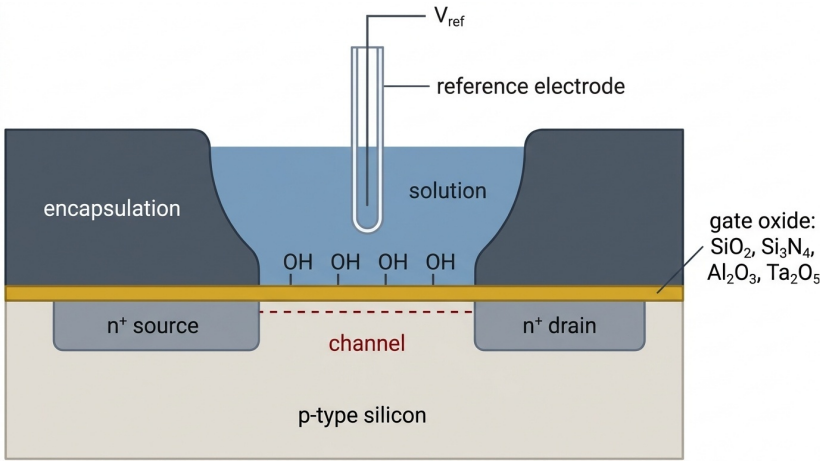


Figure 5.14: Cross section of an ISFET. The gate oxide touches the solution; a reference electrode sets the solution potential. Hydroxyl groups on the oxide surface exchange protons with the solution and set its surface potential ψ_0.

Remove the metal gate of a MOSFET, let the gate oxide touch a solution, and the transistor becomes a pH sensor: the *ion-sensitive field-effect transistor* (ISFET). A reference electrode, whose potential relative to the solution is fixed (Chapter 6 explains how), takes the place of the gate terminal, and encapsulation covers all but the gate area (Fig. 5.14). The threshold voltage keeps the fixed terms of a MOSFET and gains one that is not fixed, the potential ψ_0 of the oxide surface relative to the bulk of the solution:

$$V_T = V_{T0} - \psi_0, \tag{5.14}$$

where V_{T0} collects the potential of the reference electrode, the materials, and the charges in the oxide. Only ψ_0 depends on the solution, and it depends on the pH.

How the oxide senses pH

An oxide surface in water is covered with hydroxyl groups, which are *amphoteric*: each can give up a proton or take one,

$$\mathrm{A{-}OH} \rightleftharpoons \mathrm{A{-}O^-} + \mathrm{H_s{}^+}, \qquad \mathrm{A{-}OH} + \mathrm{H_s{}^+} \rightleftharpoons \mathrm{A{-}OH_2{}^+}, \tag{5.15}$$

where A is the metal or silicon atom and $\mathrm{H_s{}^+}$ a proton at the surface. As the pH of the solution rises (fewer protons), the surface gives up protons and becomes more negative; as the pH falls it takes protons and becomes more positive. The surface acts as a buffer, and its *intrinsic buffer capacity* β_{int} counts how many charged groups appear per unit area when the surface pH changes by one unit. The surface charge

pulls ions of the opposite sign out of the solution into a thin layer facing it, and the two layers of charge form a capacitor, the *electrical double layer*, with a capacitance C_{dl} of about 10 to 40 µF/cm^2 (Chapter 6 develops it). Balancing the charge that the surface groups release against the charge that this capacitor holds, and using Boltzmann's law (Chapter 3) for the protons at the charged surface, gives the change of surface potential per pH unit:

$$\frac{d\psi_0}{dpH} = -2.3\,\frac{kT}{q}\,\alpha, \qquad \alpha = \frac{1}{\dfrac{2.3\,kT\,C_{dl}}{q^2\beta_{int}} + 1}. \tag{5.16}$$

KEY EQUATION

$$\frac{d\psi_0}{dpH} = -2.3\,\alpha\,\frac{kT}{q}$$

ISFET sensitivity; $0 < \alpha < 1$, and $\alpha \to 1$ gives 59.2 mV/pH at 25 °C.

Here kT/q is the thermal voltage RT/F of Chapter 3 written per charge (k is Boltzmann's constant and q the elementary charge), 25.7 mV at 25 °C. The sensitivity factor α lies between 0 and 1. A surface with many reactive sites (large β_{int}) compared with the double-layer capacitance gives $\alpha \to 1$ and the *Nernstian* limit, $2.3kT/q = 59.16$ mV/pH at 25 °C (61.5 mV/pH at 37 °C). That is the slope of the Nernst equation for a singly charged ion (Chapter 3) and of the glass pH electrode of Example 4.3. Since V_T contains $-\psi_0$, the threshold of an n-channel ISFET rises by $2.3\alpha kT/q$ per pH unit.

The gate material sets β_{int} (Fig. 5.15a). Silicon dioxide has few active sites near neutral pH and gives only 25 to 40 mV/pH, not quite linearly. Silicon nitride gives about 45 to 56 mV/pH, aluminum oxide 53 to 57 mV/pH, and tantalum pentoxide 55 to 59 mV/pH, close to Nernstian. A large β_{int} also makes the sensitivity depend little on C_{dl}, which changes with the salt content of the sample, so Ta_2O_5 and Al_2O_3 are the preferred gates.

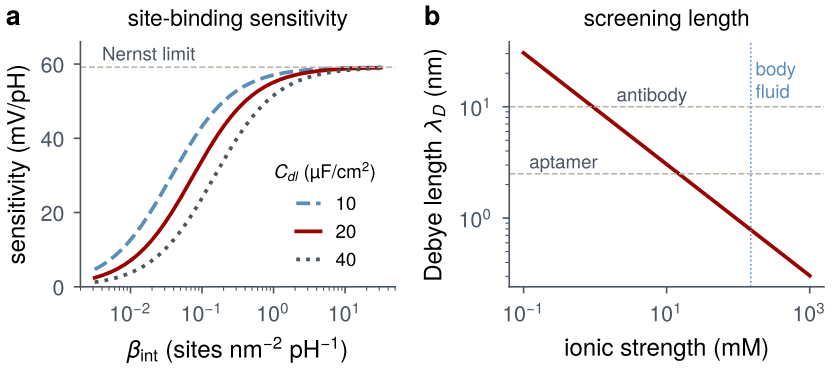


Figure 5.15: (a) ISFET sensitivity from Eq. (5.16) at 25 °C against the intrinsic buffer capacity, for three double-layer capacitances. (b) The Debye length falls as one over the square root of the ionic strength; in body fluids (150 mM) it is under a nanometer, smaller than an antibody or even an aptamer.

The Debye length

The ions of the double layer also limit what a FET can sense. Chapter 2 noted that ions in a salt solution crowd around any charge and cancel

its field beyond a short distance, the *Debye length*. For singly charged ions in water at 25 °C,

$$\lambda_D \approx \frac{0.304\,\text{nm}}{\sqrt{I/\text{M}}}, \tag{5.17}$$

where the ionic strength I is, for a simple salt such as NaCl, its molar concentration. More salt screens over a shorter distance. In blood or interstitial fluid, with $I \approx 150\,\text{mM}$, $\lambda_D = 0.79\,\text{nm}$; at 15 mM it is 2.5 nm, and at 1.5 mM, 7.9 nm (Fig. 5.15b). This is why FET biosensors for proteins are hard. An antibody is about 10 nm tall, so the charge of a molecule it captures sits many Debye lengths from the gate and is screened before it reaches it. Practical FET biosensors use small receptors such as aptamers (Chapter 2), whose change of shape happens within 1 to 2 nm of the surface;[3] or they dilute the sample, or add a polymer layer that lengthens the screening distance near the surface.

[3] N. Nakatsuka, …, A. M. Andrews, *Science* (2018), detected serotonin and dopamine this way at the salt concentration of body fluids.

The extended-gate FET

The sensing surface does not have to be on the transistor. In an *extended-gate FET* (EGFET), a sensing film (a pH-sensitive oxide, an ion-selective membrane, or an aptamer layer on gold) sits on a conductor that runs to the gate of an ordinary MOSFET (Fig. 5.16). The transistor stays dry, and the sensing strip can be printed on plastic and thrown away. The extended gate is a floating node of very high impedance, so it must be guarded like the input of the electrometer buffer of Chapter 4.

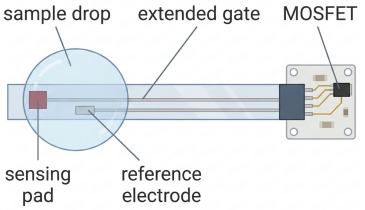


Figure 5.16: An extended-gate FET. A sensing pad printed on a flexible strip touches the sample next to a reference electrode; a trace carries its potential to the gate of a MOSFET on a separate readout board.

5.9 *Reading an ISFET*

An ISFET can be read as a current: the transistor of Example 5.4 moves 38 µA per pH unit. But that current scales with g_m, which changes with temperature. The cleaner way is to hold the drain current and the drain-source voltage constant. Then $V_{GS} = V_T + \sqrt{2I_D/\beta}$ by Eq. (5.13), and any change of V_T must appear as an equal change of V_{GS}. With the reference electrode (the gate) at ground, the source moves by $-\Delta V_T$, and it becomes the output: millivolts per pH, one for one.

Figure 5.17 shows this *constant-current, constant-voltage* (CCCV) readout. A current sink fixes I_D. A follower (an op amp wired as in Chapter 4) copies the source voltage to the output and, through a floating voltage source V_{DS}, holds the drain a fixed amount above the source.

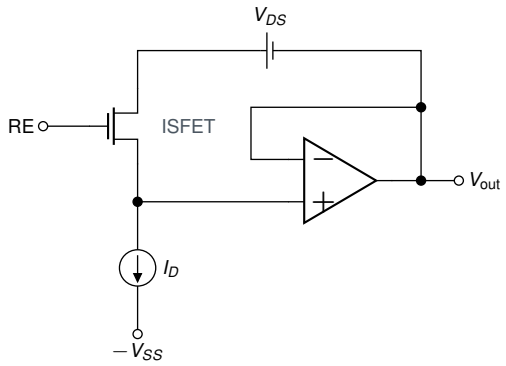


Figure 5.17: Constant-current, constant-voltage ISFET readout. The reference electrode (RE) is at 0 V; I_D is fixed by the current sink and V_{DS} by the follower and the floating source, so the output, equal to the source voltage, follows $-\Delta V_T$.

Reference FETs and drift

A miniature implant or patch has no room for a glass reference electrode. A bare platinum or gold wire can serve as a *pseudo-reference*, but its potential wanders by tens of millivolts with the sample. The fix is the differential measurement of Chapter 4: pair the ISFET with a *reference FET* (REFET), an identical transistor whose gate is covered by a pH-insensitive layer, and subtract their outputs (Fig. 5.18). The pseudo-reference potential, temperature, light, and much of the drift are common to both and cancel; the pH sensitivity is the difference of the two.

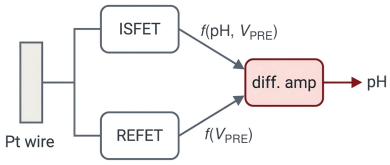


Figure 5.18: ISFET and REFET share a pseudo-reference electrode, a bare metal wire at the wandering potential V_{PRE}; the difference keeps the pH and drops what the two share.

Three other problems need an answer in every design. *Temperature* enters through kT/q (a Ta_2O_5 gate with α near 0.97 goes from 57.4 to 59.6 mV/pH between 25 and 37 °C) and through the transistor itself, so an ISFET needs a temperature sensor beside it. *Drift* is slow: water hydrates the gate insulator and moves V_T by tenths of a millivolt to about a millivolt per hour for days, so sensors are soaked and recalibrated. And the *encapsulation* must seal every contact from the liquid, and shade the silicon from light, for the life of the device.

Worked Example 5.5 *An ISFET pH meter with a REFET*

An ISFET with a sensitivity of 52 mV/pH sits in the CCCV circuit of Fig. 5.17, set so that $V_{out} = -1.20$ V at pH 7. (a) Find V_{out} at pH 4 and pH 9. (b) The ISFET and a REFET (3 mV/pH) share a platinum pseudo-reference. Over 10 h the platinum potential drifts by +12 mV, which both transistors see; the ISFET's gate also drifts by 1.2 mV/h and the REFET's by 1.0 mV/h. Find the pH error after 10 h for the ISFET alone and for the difference.

Solution. (a) V_T rises 52 mV per pH unit, so V_{out} falls by as much. At pH 4, $-1.20 + 3 \times 0.052 = -1.044$ V; at pH 9, $-1.20 - 2 \times 0.052 = -1.304$ V.

(b) The ISFET alone sees $12 + 1.2 \times 10 = 24$ mV of drift, which

reads as $24/52 = 0.46\,\text{pH}$. The difference has a sensitivity of $52 - 3 = 49\,\text{mV/pH}$. The 12 mV of pseudo-reference drift cancels, and the gate drifts leave $(1.2 - 1.0) \times 10 = 2\,\text{mV}$, or $2/49 = 0.04\,\text{pH}$, an elevenfold improvement. What remains is set by how well the two gates match. ■

Organic electrochemical transistors

An *organic electrochemical transistor* (OECT) drops the silicon altogether: it replaces the channel and its oxide with a film of a conducting polymer such as PEDOT:PSS in direct contact with the electrolyte. A gate electrode in the electrolyte pushes ions into or out of the whole volume of the film, so the channel acts as a capacitor that fills its volume rather than a thin sheet, and its transconductance is large: a few millisiemens below 1 V from a channel 100 µm wide, several times the MOSFET of Example 5.4, from a device that can be printed on plastic. The price is speed, since ions must move through the film, and a stability that depends on the polymer. Chapter 11 describes the material and works out the OECT's transconductance.

Summary

- Bridge: $V_o = V_s(R_2R_3 - R_1R_4)/[(R_1 + R_2)(R_3 + R_4)]$, balanced at $R_1R_4 = R_2R_3$. Quarter $V_s\delta/(4 + 2\delta)$ (nonlinearity $\delta/2$), half $V_s\delta/2$, full $V_s\delta$. Balance near the operating point; the output rides on a common-mode voltage of $V_s/2$.
- RTD: $R_0[1 + \alpha(T - T_0)]$, $\alpha = 0.003\,85\,\text{K}^{-1}$, 0.34 %/K at 37 °C. Thermistor: $-B/T^2$, about -4 %/K. Thermocouple: $S(T_h - T_c)$, plus cold-junction compensation. A four-wire connection removes lead resistance.
- Gauge factor $1 + 2\nu + \pi E$: metals about 2, silicon about 100. Gauges in tension and compression make half and full bridges that are linear and cancel temperature.
- A silicon diaphragm with four piezoresistors makes a full-bridge pressure sensor. A QCM weighs a film by its frequency shift, $\Delta f = -C_S\Delta m/A$, measured against a reference crystal.
- Differential capacitor $C_0/(1 \mp x/d)$: a half bridge gives $Vx/(2d)$ and a full bridge Vx/d, exactly linear, but parasitics reduce it to $2V\Delta C/(2C_0 + C_p)$. A charge amplifier, $-V(C_{w1} - C_{w2})/C_f$, holds the sensor node at virtual ground and ignores them.
- Tissue: R_e at low frequency, $R_e \parallel R_i$ at high frequency, because membranes are capacitors. Four electrodes, two to drive and two to

sense, remove the contact impedances.

- MOSFET: off below V_T; above it, a voltage-controlled resistor of conductance $\beta(V_{GS} - V_T)$; in saturation $I_D = (\beta/2)(V_{GS} - V_T)^2$ and $g_m = \beta(V_{GS} - V_T)$, with $\beta = \mu C_{ox} W/L$.
- ISFET: $V_T = V_{T0} - \psi_0$, and the threshold rises by $2.3\alpha kT/q$ per pH unit, where α, between 0 and 1, grows with the buffer capacity of the oxide (Eq. (5.16)). The Nernst limit, 59.16 mV/pH at 25 °C, is nearly reached by Ta_2O_5 and Al_2O_3. The Debye length, $0.304\,\text{nm}/\sqrt{I/\text{M}}$, is 0.79 nm in body fluids.
- CCCV readout holds I_D and V_{DS}, so the source follows $-\Delta V_T$; a REFET on a pseudo-reference cancels common drift.

The ISFET borrowed two ideas from electrochemistry, the reference electrode and the double layer. Chapter 6 builds both from the ground up and uses them to turn a concentration into a current or a voltage at the surface of an electrode.

5.10 Problems

5.1 ° A 10K thermistor ($B = 3950$ K, 10.0 kΩ at 25 °C) and a Pt100 are compared.

(a) Find the thermistor's resistance and relative sensitivity at 37 °C and 42 °C.

(b) Find the Pt100's at the same temperatures.

5.2 • A cantilever carries four 120 Ω gauges (GF = 2.0) in a full bridge, two on top at +800 and two below at −800 microstrain, excited by 3.3 V.

(a) Find the output.

(b) What amplifier gain gives 2.0 V?

(c) The beam warms by 10 K, and each gauge's resistance rises by 0.2 %. Find the output change for the full bridge and for a quarter bridge with fixed resistors.

5.3 •• An accelerometer's proof mass resonates at 3.0 kHz; its gaps are 2.0 µm.

(a) Find the displacement at 2 g.

(b) Find the fractional change of one capacitor, and its deviation from the linear approximation $C_0(1 + x/d)$.

(c) Find the full-bridge output at 2 g with 2.5 V, and explain why the bridge is linear even though each capacitor is not.

5.4 • A four-electrode sensor on the calf measures $|Z| = 52\,\Omega$ at 5 kHz and $38\,\Omega$ at 1 MHz.

(a) Treating these as the low- and high-frequency limits of the model in Fig. 5.10, find R_e and R_i.

(b) The drive is 300 µA rms, each contact 2 kΩ, and the amplifier input 500 kΩ. At 5 kHz, find the tissue voltage, the voltage the amplifier receives, and the source voltage.

(c) Why could two electrodes not follow a $1\,\Omega$ change of R_e?

5.5 •• An ISFET has $C_{dl} = 15\,\mu\text{F/cm}^2$.

(a) Find α and the sensitivity at 37 °C for $\beta_{\text{int}} = 0.2$ and $5\,\text{nm}^{-2}$.

(b) Find the Debye length at 15 mM and at 150 mM (25 °C).

(c) The better ISFET drifts by 0.4 mV/h. After 48 h, what pH error results without correction?

6
Electrochemical Biosensors

In this chapter you will learn to

- Identify oxidation and reduction in a half reaction, and use standard reduction potentials and $\Delta G = -nFE$ to predict which way a redox reaction runs.
- Apply the Nernst equation to an electrode, and convert potentials between the hydrogen, Ag/AgCl and calomel scales.
- Describe the electrode-electrolyte interface (double layer, charge transfer, diffusion) and model it with the Randles circuit.
- Read a cyclic voltammogram, an amperometric current transient and an impedance spectrum.
- Explain the three-electrode cell, and why reference electrodes, printed pseudo-references above all, drift.
- Explain how enzymatic glucose sensors work, from test strips to continuous monitors, and calibrate an amperometric sensor.
- Find a concentration from an ion-selective electrode and from an impedimetric affinity sensor.
- Analyze and design a potentiostat, including a single-supply version, from the op-amp circuits of Chapter 4.
- Estimate the flow, Reynolds number and refresh time of a microfluidic channel for sweat.

Prick a fingertip and touch the drop to the end of a test strip. Capillary action pulls less than a microliter of blood over two printed electrodes coated with an enzyme and a mediator, and five seconds later the meter shows a glucose number. In those five seconds the enzyme oxidizes glucose, the mediator carries the electrons to the

electrode, and the meter holds the electrode at a fixed potential while it measures the current that results. A continuous glucose monitor runs the same chemistry for two weeks on a filament under the skin of the upper arm, and a sweat patch on a runner's forearm runs a cousin of it for sodium, potassium and lactate.

All of these devices turn a chemical concentration into a current or a voltage at the surface of an electrode. This chapter builds that conversion from the ground up and assumes no chemistry beyond high school. We start with redox reactions and the potential of an electrode, where the Nernst equation of Chapter 3 reappears. Then we look at the interface between a metal and salt water, the three-electrode cell that controls it, and the three measurements that read it: voltammetry, amperometry, and impedance spectroscopy. With these tools we build enzymatic glucose sensors, ion-selective electrodes, and affinity sensors. The chapter closes with the circuit that runs them all, the potentiostat, built from the op amps of Chapter 4, and with the microfluidics that bring fresh sweat to a wearable sensor. Enzymes and diffusion come from Chapter 2, the figures of merit from Chapter 4; optical readout is the subject of Chapter 7.

6.1 *Redox Reactions and Electrode Potentials*

In a metal wire the current is carried by electrons; in salt water, and in the body, it is carried by ions. Wherever the two meet, at the surface of an electrode, the charge must change carriers. An electron cannot simply swim off into the water: it can leave the metal only by landing on a molecule, and it can enter the metal only by being taken from one. Each of these steps is a chemical reaction, and electrochemistry is the study of them.

A species that gains electrons is *reduced*; a species that loses electrons is *oxidized*:[1]

[1] A memory aid: OIL RIG, oxidation is loss, reduction is gain (of electrons).

$$\text{reduction:}\quad \mathrm{A} + \mathrm{e}^- \longrightarrow \mathrm{A}^- \qquad \text{(A receives an electron)}, \tag{6.1}$$

$$\text{oxidation:}\quad \mathrm{B} \longrightarrow \mathrm{B}^+ + \mathrm{e}^- \qquad \text{(B loses an electron)}. \tag{6.2}$$

Added together, the two *half reactions* give the overall reaction $\mathrm{A} + \mathrm{B} \longrightarrow \mathrm{A}^- + \mathrm{B}^+$, in which no free electrons remain. This pair is a *redox* reaction, short for reduction-oxidation. A rusting bolt is one: iron is oxidized, $\mathrm{Fe} \longrightarrow \mathrm{Fe}^{2+} + 2\,\mathrm{e}^-$, and oxygen is reduced, $\mathrm{O_2} + 2\,\mathrm{H_2O} + 4\,\mathrm{e}^- \longrightarrow 4\,\mathrm{OH}^-$. The oxidized and reduced forms of one species make a *redox couple*, written

$$\mathrm{Ox} + n\mathrm{e}^- \rightleftharpoons \mathrm{Red}, \tag{6.3}$$

where n is the number of electrons transferred. The couple $\mathrm{Fe}^{3+}/\mathrm{Fe}^{2+}$ has $n = 1$; the couple $\mathrm{O_2}/\mathrm{H_2O}$ has $n = 4$.

The electrochemical cell

An *electrochemical cell* is two electrodes in an electrolyte, joined by an external circuit. Electrons move through the wire; ions move through the solution. By definition the electrode where oxidation happens is the *anode* and the electrode where reduction happens is the *cathode*. In a *galvanic* cell (a battery) the reaction runs by itself and pushes current through a load; the anode is the negative terminal. In an *electrolytic* cell an external source drives a reaction that would not run by itself, such as splitting water (Fig. 6.1); the anode is then the positive terminal. In both, electrons in the wire flow from anode to cathode, opposite to the conventional current.[2]

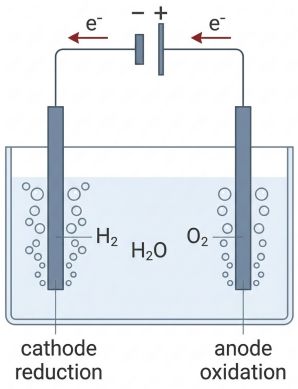


Figure 6.1: An electrolytic cell splitting water. The source pulls electrons out of the anode, where water is oxidized to O_2, and pushes them into the cathode, where water is reduced to H_2. Ions carry the current through the electrolyte.

[2] The names follow the chemistry, not the sign. A sensor electrode that oxidizes glucose is an anode whether the circuit drives it or not.

The electrode as an electron reservoir

To a molecule in solution, a metal electrode is a reservoir of electrons whose energy can be set from outside. The electrons fill the energy levels of the metal up to the *Fermi level* E_F. A molecule has its own levels: its highest occupied orbital (HOMO) and its lowest empty one (LUMO). A potential E on the electrode shifts its electron energies by $-eE$. Make the electrode more negative and the Fermi level rises; once it passes the LUMO of a molecule at the surface, electrons tunnel into the molecule and reduce it (Fig. 6.2a). Make it more positive and the Fermi level drops below the HOMO, so electrons leave the molecule: oxidation (Fig. 6.2b). The electrode voltage is a knob that selects which reactions can happen. *Inert* electrodes (platinum, gold, carbon) only supply or accept the electrons, which is why they are the working electrodes of most sensors.

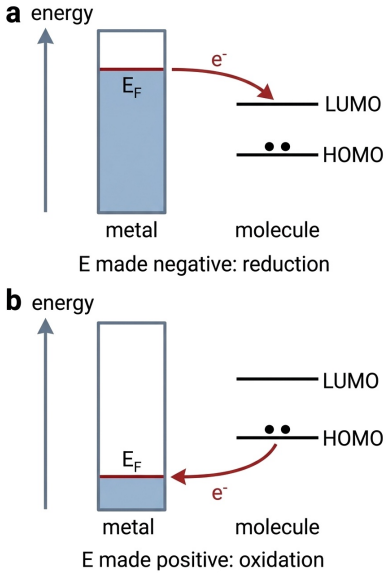


Figure 6.2: Electron transfer seen through energy levels. (a) A negative electrode raises its Fermi level above the LUMO of a nearby molecule, and an electron tunnels in: reduction. (b) A positive electrode lowers it below the HOMO, and an electron leaves the molecule: oxidation.

Standard reduction potentials

The tendency of a couple to accept electrons is its *reduction potential*, measured between an inert electrode in the solution and a reference electrode joined to it by a salt bridge. A voltage needs a zero, and the zero of this scale is the *standard hydrogen electrode* (SHE): platinum in acid of unit hydrogen-ion activity under one bar of hydrogen, for $2\,H^+ + 2\,e^- \rightleftharpoons H_2$. The *standard* potential E° has every species at unit activity (about 1 M) and 25 °C (Table 6.1). The more positive E°, the more strongly the oxidized form pulls electrons: peroxide and oxygen sit near the top, sodium near the bottom.

Two half reactions make a cell, with $E_{\text{cell}} = E_{\text{cathode}} - E_{\text{anode}}$. A volt is a joule per coulomb, so each electron that passes through the cell releases eE_{cell} of energy, and a mole of electrons releases FE_{cell}, about 96.5 kJ per volt. Chemists write this as the Gibbs free energy of the

Half reaction	vs SHE	vs Ag/AgCl
$H_2O_2 + 2\,H^+ + 2\,e^- \rightleftharpoons 2\,H_2O$	+1.776	+1.566
$O_2 + 4\,H^+ + 4\,e^- \rightleftharpoons 2\,H_2O$	+1.229	+1.019
$Fe^{3+} + e^- \rightleftharpoons Fe^{2+}$	+0.771	+0.561
$O_2 + 2\,H^+ + 2\,e^- \rightleftharpoons H_2O_2$	+0.695	+0.485
$[Fe(CN)_6]^{3-} + e^- \rightleftharpoons [Fe(CN)_6]^{4-}$	+0.361	+0.151
$Cu^{2+} + 2\,e^- \rightleftharpoons Cu$	+0.342	+0.132
$AgCl + e^- \rightleftharpoons Ag + Cl^-$	+0.222	+0.012
$2\,H^+ + 2\,e^- \rightleftharpoons H_2$ (SHE)	0.000	−0.210
$Zn^{2+} + 2\,e^- \rightleftharpoons Zn$	−0.762	−0.972
$Na^+ + e^- \rightleftharpoons Na$	−2.71	−2.92
Reference electrodes in use		
SCE, saturated KCl	+0.241	+0.031
Ag/AgCl, saturated KCl	+0.197	−0.013
Ag/AgCl, 3 M KCl	+0.210	0.000

Table 6.1: Standard reduction potentials at 25 °C, in volts, from standard tables. The last column is the same potential measured against an Ag/AgCl reference in 3 M KCl, the reference used in this chapter (subtract 0.210 V). The H_2O_2 and O_2 couples matter for glucose sensors; ferricyanide is the usual redox probe and test-strip mediator. SCE is the saturated calomel electrode.

overall reaction,

$$\Delta G = -nFE_{\text{cell}}, \tag{6.4}$$

where $F = 96\,485\,\text{C/mol}$ is the charge of one mole of electrons (the Faraday constant). A negative E_{cell} means the reaction must be driven by a source of at least $|E_{\text{cell}}|$.

KEY EQUATION

$\Delta G = -nFE$

A positive cell potential means a negative ΔG: the reaction runs by itself.

Two results follow. First, splitting water at pH 0 pairs the hydrogen couple ($E^\circ = 0$, the cathode) with the reverse of the oxygen couple ($E^\circ = 1.229\,\text{V}$, the anode): $E_{\text{cell}} = -1.229\,\text{V}$, so a source must supply at least 1.23 V, and the hydrogen made stores $2F \times 1.229\,\text{V} = 237\,\text{kJ}$ per mole. Real electrolyzers need 1.8 to 2 V; the extra voltage, the *overpotential*, drives the reaction at a useful rate (Section 6.3). Second, when an inert electrode is swept negative in a mixture of oxidants, the one with the most positive E° is reduced first: in acid with Fe^{3+}, Cu^{2+} and H^+, iron(III) below +0.771 V, then copper below +0.342 V, then hydrogen near 0 V. An amperometric sensor picks its potential the same way: past the analyte, short of the next species.

6.2 *The Nernst Equation at an Electrode*

STANDARD POTENTIALS ASSUME UNIT ACTIVITIES, about one mole per liter of everything, but a sensor sees millimolar or micromolar concentrations. Chapter 3 met the same problem for ions on the two sides of a cell membrane and solved it with the Nernst equation: the potential at which the electrical energy balances the push of a concentration difference. At an electrode the two sides are the oxidized and the reduced forms of a couple. The free energy of the reduction depends on their ratio,

$$\Delta G = \Delta G^\circ + RT \ln \frac{a_{\text{Red}}}{a_{\text{Ox}}}, \tag{6.5}$$

and substituting $\Delta G = -nFE$ and $\Delta G^\circ = -nFE^\circ$ gives the potential the electrode settles at when no current flows:

$$E = E^\circ + \frac{RT}{nF}\ln\frac{a_{\mathrm{Ox}}}{a_{\mathrm{Red}}} = E^\circ + \frac{2.303\,RT}{nF}\log_{10}\frac{a_{\mathrm{Ox}}}{a_{\mathrm{Red}}}. \qquad (6.6)$$

KEY EQUATION

$E = E^\circ + \frac{2.303RT}{nF}\log_{10}\frac{a_{\mathrm{Ox}}}{a_{\mathrm{Red}}}$

The Nernst equation: 59.16 mV per decade per electron at 25 °C, 61.54 mV at 37 °C.

Here $R = 8.314\,\mathrm{J\,mol^{-1}\,K^{-1}}$ is the gas constant, T the absolute temperature, a_{Ox} and a_{Red} the *activities* (effective concentrations) of the two forms, and $2.303 = \ln 10$. It is the membrane equation of Chapter 3 with the oxidized form in the place of the outside concentration. The sign is easy to check: adding oxidant makes the electrode more positive, more eager to take electrons. The factor RT/F is the thermal voltage, 25.69 mV at 25 °C, so

$$\frac{2.303\,RT}{F} = 59.16\,\mathrm{mV}\ \text{at}\ 25\,^\circ\mathrm{C}, \qquad 61.54\,\mathrm{mV}\ \text{at}\ 37\,^\circ\mathrm{C}: \qquad (6.7)$$

a tenfold change of $a_{\mathrm{Ox}}/a_{\mathrm{Red}}$ moves the potential by this amount divided by n.

The activity is the concentration times an *activity coefficient* that falls below 1 as the salt content rises (about 0.75 for a monovalent ion in saline). In practice one uses the *formal potential* $E^{\circ\prime}$, measured in a stated medium, which absorbs the activity coefficients, and writes Eq. (6.6) with concentrations. Pure solids (a metal, a silver chloride coating) have unit activity and drop out. When protons take part, the pH enters: for $O_2 + 4\,H^+ + 4\,e^- \rightleftharpoons 2\,H_2O$, $E = 1.229 - 0.0592\,\mathrm{pH}$, or 0.791 V at the pH of blood, 7.4. The slope also grows by about 0.2 mV per decade per kelvin, so a careful sensor measures its own temperature.

Worked Example 6.1 *The Nernst equation for the iron couple*

A platinum electrode sits in an acid solution with $[Fe^{3+}] = 5.0\,\mathrm{mM}$ and $[Fe^{2+}] = 0.50\,\mathrm{mM}$; take $E^\circ = 0.771\,\mathrm{V}$ vs SHE at both temperatures below. (a) Find the electrode potential at 25 °C and at 37 °C. (b) Express both against an Ag/AgCl reference in 3 M KCl. (c) What is the potential at 25 °C when the two concentrations are swapped?

Solution. (a) One electron is transferred, $n = 1$. The ratio $[Fe^{3+}]/[Fe^{2+}] = 5.0/0.50 = 10$, so the log term is $+1$:

$$E_{25} = 0.771 + 0.05916 = 0.830\,\mathrm{V\ vs\ SHE},$$
$$E_{37} = 0.771 + 0.06154 = 0.833\,\mathrm{V\ vs\ SHE}.$$

The excess of the oxidized form makes the electrode more positive than E°.

(b) The Ag/AgCl reference sits 0.210 V above SHE (Table 6.1), so subtract it: 0.620 V and 0.623 V vs Ag/AgCl.

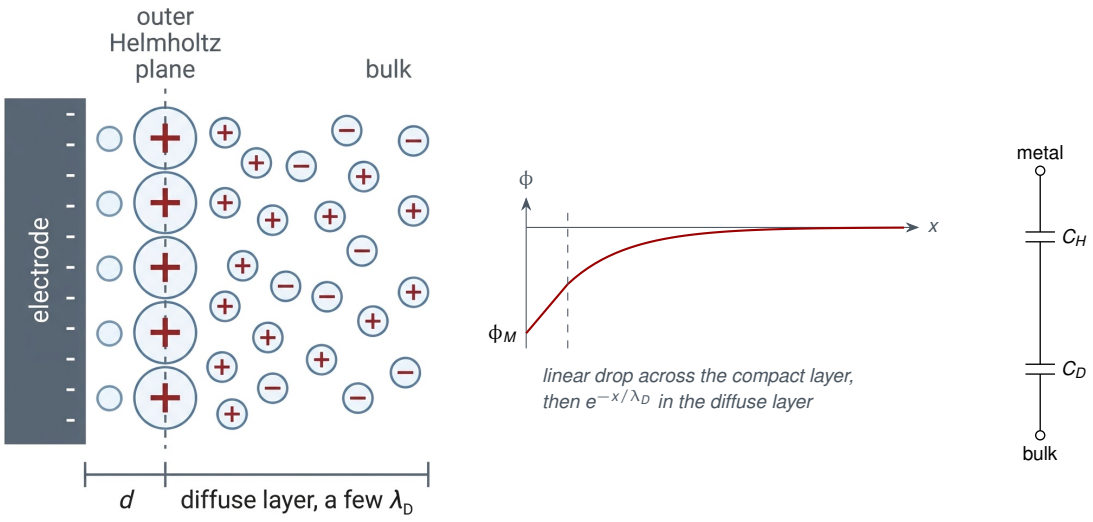


Figure 6.3: The double layer at a negatively charged electrode. Cations with their shells of water approach no closer than the outer Helmholtz plane at distance d (a few tenths of a nanometer; small circles are water); beyond it, a diffuse cloud with excess cations fades into the neutral bulk over a few Debye lengths. The potential falls linearly across the compact layer and exponentially in the diffuse layer (center). The two regions act as capacitors in series (right).

(c) The ratio is now 0.1 and the potential $0.771 - 0.0592 = 0.712\,\text{V}$ vs SHE, or 0.502 V vs Ag/AgCl. A hundredfold change of the ratio moved the electrode by only 118 mV: the logarithm compresses concentration into a small voltage range. ■

6.3 *The Electrode-Electrolyte Interface*

Dip a metal into salt water and any excess charge on its surface is balanced at once by a layer of ions of the opposite sign in the solution. This *electrical double layer* is a capacitor about a nanometer thick, present at every electrode in the book: the biopotential electrodes of Chapter 8, the stimulation electrodes of Chapter 9, and every sensor in this chapter. Next to it run two processes that a sensor relies on: electrons crossing the interface in a reaction, and molecules diffusing to it.

The electrical double layer

The simplest model, due to Helmholtz, treats the layer as a parallel-plate capacitor (Fig. 6.3). The countercharge sits in one plane of ions at the distance d of closest approach, set by the size of an ion with its shell of water, so

$$C_H = \frac{\varepsilon_r \varepsilon_0}{d}. \tag{6.8}$$

The water molecules squeezed into this gap are lined up by the field and have a relative permittivity near 6, far below the 78 of bulk water.

With $\varepsilon_r = 6$ and $d = 0.3\,\text{nm}$, $C_H = 18\,\mu\text{F/cm}^2$, a huge capacitance per area because the plates are so close.

Thermal motion spreads the countercharge into a diffuse cloud (the Gouy-Chapman picture). Balancing the spreading of the ions, which follow Boltzmann statistics, against their pull toward the charge (Poisson's equation) gives a potential that decays as e^{-x/λ_D} over the *Debye length*, the screening distance that limited the field-effect sensors of Chapter 5:

$$\lambda_D = \sqrt{\frac{\varepsilon_r \varepsilon_0 RT}{2F^2 I}}, \qquad I = \tfrac{1}{2}\sum_i c_i z_i^2, \tag{6.9}$$

where I is the ionic strength (in mol/m^3 here) and z_i the charge numbers. For a 1:1 salt such as NaCl at 25 °C, I is the salt concentration and $\lambda_D = 0.304\,\text{nm}/\sqrt{I/\text{M}}$: 0.79 nm in saline (150 mM), 1.4 nm in sweat with 50 mM NaCl, 9.6 nm in a 1 mM buffer. A charge more than a few Debye lengths away is invisible. A protein captured by an antibody, about 10 nm out, is screened by about $e^{-10/0.79} \approx 3 \times 10^{-6}$ in saline, which is why the field-effect sensors of Chapter 5 see so little of it; the impedance sensors of Section 6.9 measure how a bound layer blocks electron transfer instead.

KEY EQUATION

$\lambda_D \approx \dfrac{0.304\,\text{nm}}{\sqrt{I/\text{M}}}$

Debye length of a 1:1 electrolyte in water at 25 °C: 0.79 nm in 150 mM saline.

Stern joined the two pictures: the compact and the diffuse layer are capacitors in series, $1/C_{dl} = 1/C_H + 1/C_D$. In saline the diffuse layer's capacitance, $C_D = \varepsilon_r \varepsilon_0/\lambda_D \approx 88\,\mu\text{F/cm}^2$ (with the bulk $\varepsilon_r = 78$), is much the larger, so the compact layer dominates and $C_{dl} \approx 15\,\mu\text{F/cm}^2$. Measured values on smooth metals fall between about 10 and 40 $\mu\text{F/cm}^2$ of real area; rough or porous electrodes (platinum black, carbon, PEDOT) show far more per square centimeter of footprint.

Faradaic and non-faradaic current

Current can cross the interface in two ways. A *non-faradaic* current charges the double layer, $i = C_{dl} A\, \text{d}E/\text{d}t$, like the current into any capacitor; no electron crosses the interface. A *faradaic* current transfers electrons in a redox reaction and obeys Faraday's law, $Q = nFN$ for N moles converted, so it counts molecules. Sensors measure the faradaic current; the non-faradaic part is background. An electrode that passes no faradaic current over a range of potentials is *ideally polarizable*, a pure capacitor (gold or platinum in clean saline, over about a volt). One with a fast, reversible reaction available is *nonpolarizable*: a large current hardly moves its potential. Silver coated with silver chloride in a chloride solution is the standard example, a good reference and a good biopotential electrode.

In circuit form the interface is C_{dl} in parallel with a faradaic path, the *charge-transfer resistance* R_{ct}, in series with the solution resistance

R_s (Section 6.6). For a disk electrode of radius a far from the other electrodes, $R_s = \rho/(4a)$, where ρ is the resistivity of the solution, about 70 Ω cm for saline. The double layer charges through R_s quickly: a gold disk 1 mm across in saline has $R_s \approx 350\,\Omega$ and, at 20 µF/cm², $C = 0.16\,\mu F$, so $\tau = 55\,\mu s$.

Charge transfer

At equilibrium a redox electrode is not idle: electrons cross in both directions at equal rates, and the current each way is the *exchange current* i_0. Shift the electrode from its equilibrium potential by the *overpotential* $\eta = E - E_{eq}$, and one direction wins.

Each rate grows exponentially with the part of η that lowers its energy barrier, so the net current is a difference of two exponentials, the *Butler-Volmer equation*:

$$i = i_0 \left[\exp\left(\frac{\alpha n F \eta}{RT}\right) - \exp\left(-\frac{(1-\alpha) n F \eta}{RT}\right) \right], \tag{6.10}$$

with anodic (oxidation) current positive in this book and a *transfer coefficient* α near 0.5 (Fig. 6.4). Two consequences are all we need. Near equilibrium, for $|\eta|$ well below RT/F (26 mV), the curve is a straight line and the interface acts as a resistor,

$$R_{ct} = \frac{RT}{nFi_0}. \tag{6.11}$$

Far from equilibrium one exponential wins, and the current grows tenfold for every 118 mV of overpotential ($\alpha = 0.5$, $n = 1$, 25 °C): the straight *Tafel line* of $\log i$ against η. A couple with $i_0 = 0.5\,\mu A$, for example, has $R_{ct} = 25.69\,mV/0.5\,\mu A = 51.4\,k\Omega$ and needs $\eta = 118\,mV \times \log_{10}(100/0.5) = 272\,mV$ to pass 100 µA; a catalyst that raises i_0 tenfold saves 118 mV. The exponential cannot grow forever: once the reaction consumes the reactant faster than it arrives, mass transport sets the current.

Mass transport

A reaction can only consume what reaches the electrode, and molecules arrive in three ways. *Diffusion* moves them down a concentration gradient with the flux of Fick's first law, $J = -D\,\partial C/\partial x$ (Chapter 2); small molecules in water have $D \approx 10^{-5}\,cm^2/s$ (at 25 °C: glucose $6.7 \times 10^{-6}\,cm^2/s$, oxygen $2 \times 10^{-5}\,cm^2/s$). *Migration* moves ions in the electric field of the solution; an excess of an inert salt, the *supporting electrolyte* (0.1 M KCl in the laboratory, the salts of the body fluid in the body), carries almost all of the ionic current and makes migration of the analyte negligible. *Convection* moves the whole fluid:

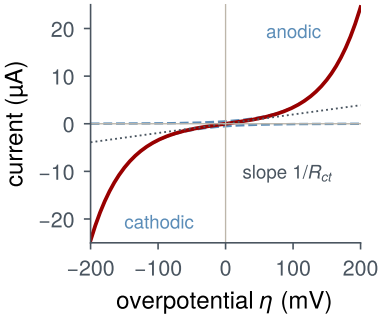


Figure 6.4: Current against overpotential for $i_0 = 0.5\,\mu A$, $\alpha = 0.5$, $n = 1$ (the numbers in the text). The net current (cardinal) is the sum of an anodic and a cathodic branch (dashed). Near $\eta = 0$ it follows the dotted line of slope $1/R_{ct}$; far from it, one branch grows exponentially.

KEY EQUATION

$R_{ct} = RT/(nFi_0)$

Charge-transfer resistance near equilibrium: a fast reaction has a small R_{ct}.

stirring, flow in a channel, or the slow natural convection that disturbs a quiet solution after ten seconds or so.

When an electrode consumes a species, a depleted *diffusion layer* grows in front of it, about $\delta \approx \sqrt{\pi D t}$ thick: 56 µm after one second and 177 µm after ten for $D = 10^{-5}\,\mathrm{cm^2/s}$. If the reaction is fast enough to hold the surface concentration at zero, the current is set by the gradient across this layer,

$$i_L = \frac{nFADC^*}{\delta}, \tag{6.12}$$

where C^* is the bulk concentration. This *limiting current* is proportional to concentration, which is the basis of amperometric sensing. Behind a membrane, in a flowing channel, or in a stirred solution, δ stays fixed and the current is steady; in a quiet solution δ grows and the current decays (Section 6.5).

6.4 The Three-Electrode Cell and Reference Electrodes

With only two electrodes, the current that the reaction needs must flow through the electrode you use as a reference. That current polarizes the reference, shifting its potential, and it drops a voltage iR_s across the solution; the potential of the working electrode is then not what you think it is. The *three-electrode cell* (Fig. 6.5) separates the two jobs. The *working electrode* (WE) is the interface under study, inert and polarizable. The *reference electrode* (RE) is nonpolarizable and holds a constant potential because essentially no current flows through it. The *counter electrode* (CE, or auxiliary electrode) supplies the current the WE demands; it is much larger than the WE so that its own polarization never limits the reaction. A feedback amplifier, the *potentiostat*, drives the CE until the WE-RE potential equals the set value, and measures the current (Section 6.10).

The Ag/AgCl reference electrode

The silver-silver chloride electrode, a silver wire coated with solid silver chloride in a chloride solution, is the reference of nearly every biosensor. Its half reaction,

$$\mathrm{AgCl(s) + e^- \rightleftharpoons Ag(s) + Cl^-}, \tag{6.13}$$

involves two solids of unit activity and one ion, so the Nernst equation gives a potential that depends on chloride alone:

$$E = E^\circ_{\mathrm{AgCl/Ag}} - \frac{RT}{F}\ln a_{\mathrm{Cl^-}} = 0.222\,\mathrm{V} - (0.0592\,\mathrm{V})\log_{10} a_{\mathrm{Cl^-}} \quad \text{at } 25\,^\circ\mathrm{C}. \tag{6.14}$$

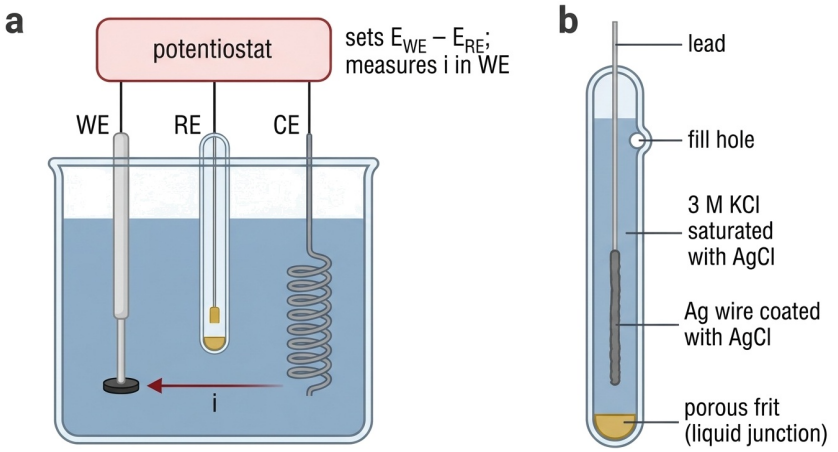


Figure 6.5: (a) A three-electrode cell. The potentiostat holds the working electrode at a set potential against the reference and supplies the current through the counter electrode, a platinum coil or mesh larger than the WE. (b) An Ag/AgCl reference electrode: a silver wire coated with silver chloride in a filling solution of fixed chloride concentration, in contact with the sample through a porous frit.

The potential is constant only while the chloride activity is, so the electrode carries its own filling solution behind a frit: +0.210 V vs SHE with 3 M KCl, +0.197 V with saturated KCl. The filling is pre-saturated with AgCl so that the coating does not dissolve.

Papers quote potentials against whatever reference they used, so converting between scales is a daily task (Fig. 6.6). Peroxide oxidation at +0.60 V vs Ag/AgCl (3 M KCl) is $0.60 + 0.210 = +0.81$ V vs SHE, and $0.81 - 0.241 = +0.57$ V vs the saturated calomel electrode. Always state the reference: the "same" potential differs by tens of millivolts from one reference to another.

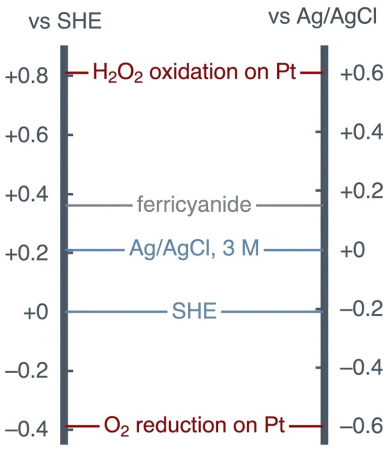


Figure 6.6: Two potential scales side by side. Potentials vs Ag/AgCl in 3 M KCl are 0.210 V lower than vs SHE. Sensor potentials quoted in this chapter: H_2O_2 oxidation near +0.6 V and O_2 reduction near −0.6 V vs Ag/AgCl.

Pseudo-references for wearables

A wearable has no room for a glass tube. Most patches use a bare film of printed Ag/AgCl ink as a *pseudo-reference*, which by Eq. (6.14) moves 59 mV per tenfold change of sample chloride, and sweat chloride is not constant. If it rises from 20 to 60 mM during a run, at a skin temperature of 33 °C (60.7 mV per decade) the reference moves by $-60.7\,\text{mV} \times \log_{10} 3 = -29\,\text{mV}$, and the WE, held at a set voltage above it, moves with it. An amperometric sensor running on its diffusion-limited current hardly notices; a potentiometric sensor read against the same reference takes the 29 mV as error, about half a decade of concentration (Problem 6.3). Solid-state references cover the Ag/AgCl with a polymer membrane loaded with KCl, which holds the chloride at the silver nearly constant for hours to days. A bare platinum wire has no defined couple at all and is acceptable only when potentials are calibrated in the same run.

Where the filling solution meets the sample, at the frit, the ions diffuse across the boundary at different speeds, a little charge separates,

and a small *liquid junction potential* appears: about 39 mV for a tenfold step of HCl, whose protons outrun the chloride, but only about a millivolt for KCl, whose two ions move at nearly the same speed. That is why references and salt bridges use concentrated KCl: its ions carry nearly all the current at the junction, which then stays at a few millivolts.

6.5 Voltammetry and Amperometry

Sweep the potential of an electrode and watch the current: that is *voltammetry*, the way electrochemists study a new sensor. Hold the potential fixed and record the current against time: that is *amperometry*, the way most sensors are run.

Cyclic voltammetry

In *linear sweep voltammetry* the working-electrode potential ramps at a scan rate v (volts per second). Take 5 mM ferricyanide in KCl. As the potential approaches $E^{\circ\prime}$ from the positive side, the Nernst equation forces the surface ratio $C_{\mathrm{Ox}}/C_{\mathrm{Red}}$ down, ferricyanide is reduced, and a cathodic current flows. It grows, peaks, and decays, because once the surface concentration is near zero the current is limited by diffusion across a depletion layer that keeps growing. In *cyclic voltammetry* (CV) the sweep then reverses, and the ferrocyanide made on the way out is oxidized back on the way in, giving an anodic peak (Fig. 6.7).

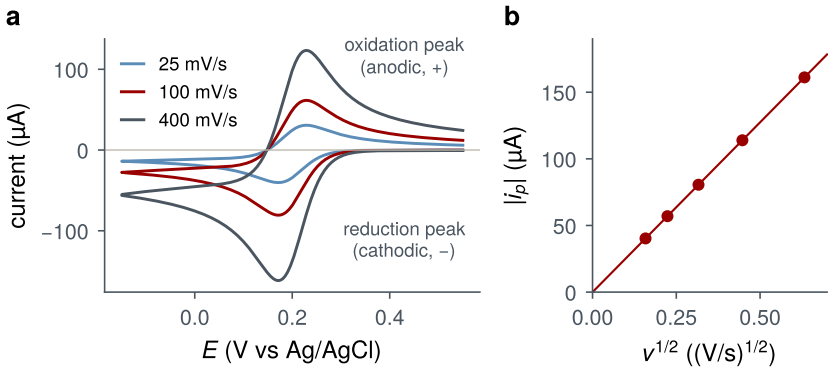


Figure 6.7: Simulated cyclic voltammograms of a reversible one-electron couple (5 mM, $D = 7.2 \times 10^{-6}\,\mathrm{cm^2/s}$, a 3 mm disk, $E^{\circ\prime} = +0.20$ V); illustrative, from a finite-difference model. (a) Three scan rates, starting at +0.55 V and sweeping negative first. Anodic current is positive. The peaks stay 58 mV apart, the exact value for a sweep reversed this far past the peak (Eq. (6.15) rounds it to 59 mV). (b) The peaks grow as $\sqrt{v}$ (Randles-Sevcik line).

A reversible (Nernstian) couple, one whose electron transfer is fast compared with the sweep, leaves a fingerprint in the CV. The two peaks are separated by

$$\Delta E_p = E_{pa} - E_{pc} \approx \frac{2.3\,RT}{nF} = \frac{59\,\mathrm{mV}}{n} \text{ at } 25\,^{\circ}\mathrm{C}, \tag{6.15}$$

their midpoint is close to the formal potential, and the two peak currents are equal. The peak current is given by the *Randles-Sevcik equation*:

$$i_p = 0.4463\, nFAC^* \sqrt{\frac{nFvD}{RT}} = (2.69\times 10^5)\, n^{3/2} A C^* D^{1/2} v^{1/2} \quad \text{at } 25\,^\circ\text{C}, \tag{6.16}$$

proportional to concentration, and to $\sqrt{v}$ because a faster sweep reaches the peak before the diffusion layer has grown as thick. Every CV also carries a non-faradaic background, the charging current $i_c = C_{dl} A v$, which grows faster with v than the peak does.

KEY EQUATION

$i_p = (2.69\times 10^5)\, n^{3/2} A C^* \sqrt{Dv}$

Randles-Sevcik, with A in cm^2, C^* in mol/cm^3, D in cm^2/s, v in V/s, i_p in amperes.

Changing the scan rate is the standard diagnostic. If $i_p \propto \sqrt{v}$, the species diffuses to the electrode from solution. If $i_p \propto v$ and the two peaks nearly coincide, the redox molecules are attached to the surface, as in the aptamer sensors of Section 6.9. If ΔE_p grows with v, either the electron transfer is not fast enough to stay in equilibrium, or the uncompensated solution resistance adds an iR shift that grows with the current.

Worked Example 6.2 *Peak currents and a scan-rate test*

A glassy carbon disk 3.0 mm in diameter sits in 5.0 mM ferricyanide in 0.1 M KCl, with $D = 7.2\times 10^{-6}$ cm^2/s and $C_{dl} = 20\,\mu$F/cm^2, at 25 °C. (a) Find the cathodic peak current and the charging current at 100 mV/s and at 10 V/s. (b) A student records peak currents of 40.3, 57.0, 80.6, 113.9 and 161.1 µA at 25, 50, 100, 200 and 400 mV/s, with ΔE_p rising from 59 to 68 mV. What controls the current, how many electrons are transferred, and what is D?

Solution. (a) $A = \pi(0.15\,\text{cm})^2 = 0.0707\,\text{cm}^2$ and $C^* = 5.0\times 10^{-6}$ mol/cm^3. At 0.1 V/s,

$$i_p = (2.69\times 10^5)(1)(0.0707)(5.0\times 10^{-6})\sqrt{(7.2\times 10^{-6})(0.1)}$$
$$= 80.7\,\mu\text{A}.$$

The charging current is $C_{dl} A v = (20\,\mu\text{F/cm}^2)(0.0707)(0.1) = 0.14\,\mu$A, 570 times smaller. At 10 V/s the peak grows by $\sqrt{100} = 10$ to 807 µA, and the charging current by 100 to 14 µA; the ratio has dropped to 57. Slow scans favor the faradaic signal.

(b) Each peak divided by $\sqrt{v}$ gives the same 255 µA/(V/s)$^{1/2}$ (for example $40.3/\sqrt{0.025}$ and $161.1/\sqrt{0.4}$), so the current is controlled by diffusion from solution. A separation near 59 mV at low scan rate means $n = 1$; its slow growth points to a small uncompensated resistance or slightly slow kinetics at the faster sweeps.

Solving Eq. (6.16) for D,

$$D = \left(\frac{2.55 \times 10^{-4}\,\mathrm{A}/(\mathrm{V/s})^{1/2}}{(2.69 \times 10^{5})(0.0707)(5.0 \times 10^{-6})} \right)^{2} = 7.2 \times 10^{-6}\,\mathrm{cm^2/s}.$$

Run backward with a known D, the same equation measures the real electrochemical area of an electrode, which is how the roughness of a printed or porous electrode is checked. ■

Amperometry and the Cottrell equation

An amperometric measurement steps the potential to a value where the analyte reacts as fast as it arrives, so that its surface concentration is zero, and records the current. The potential is chosen as in Section 6.1: past the analyte's wave, short of the next species. For a planar electrode in a quiet solution, the diffusion layer grows as $\sqrt{\pi D t}$ and the current decays according to the *Cottrell equation*:

$$i(t) = nFAC^{*}\sqrt{\frac{D}{\pi t}}. \tag{6.17}$$

For 1.0 mM of a one-electron analyte with $D = 1.0 \times 10^{-5}\,\mathrm{cm^2/s}$, a disk 3.0 mm across carries 12.2 µA one second after the step and 3.85 µA after ten. The current is proportional to concentration at every instant, but it never settles (Fig. 6.8a). Sensors avoid that in three ways: they sample at a fixed time after the step (test strips), they stir or flow the sample, or they put a membrane in front of the electrode whose fixed thickness sets δ in Eq. (6.12). The Clark oxygen electrode,[3] the ancestor of every enzyme electrode, is a platinum cathode at about −0.6 V vs Ag/AgCl behind an oxygen-permeable membrane; it reduces oxygen, $O_2 + 4\,H^+ + 4\,e^- \longrightarrow 2\,H_2O$, at a rate set by diffusion through the membrane, so its current is proportional to the oxygen partial pressure.

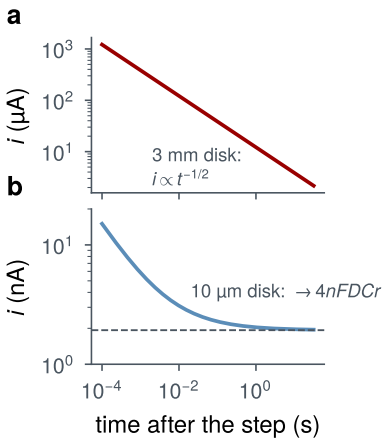


Figure 6.8: Current after a potential step in 1 mM of analyte ($D = 10^{-5}\,\mathrm{cm^2/s}$). (a) A 3 mm disk follows Cottrell, falling as $t^{-1/2}$. (b) A 10 µm disk levels off within a few tenths of a second at $4nFDC^{*}r = 1.9$ nA (dashed).

[3] L. C. Clark, *Trans. Am. Soc. Artif. Intern. Organs* (1956).

Microelectrodes reach a steady state

Shrink the electrode to a disk a few micrometers in radius r and molecules soon arrive from all sides; the diffusion field becomes hemispherical and the current settles at

$$i_{ss} = 4nFDC^{*}r \quad \text{(disk)}, \tag{6.18}$$

approached once t is long compared with r^2/D (Fig. 6.8b). For the same 1.0 mM analyte, a disk 10 µm across ($r^2/D = 25$ ms) comes within 10 % of $4FDC^{*}r = 1.93$ nA after about a third of a second, some 13 times r^2/D: about 6300 times less than the large disk at one second, but steady. Microelectrodes give steady currents without stirring,

small iR drops and fast response; their currents are nanoamperes or less, which sets the gain of the readout (Section 6.10). Carbon-fiber microelectrodes a few micrometers across, swept at hundreds of volts per second, oxidize dopamine and serotonin and follow them in the brain ten times a second (Chapter 9).

6.6 *Electrochemical Impedance Spectroscopy*

VOLTAMMETRY PUSHES THE INTERFACE HARD; impedance spectroscopy listens to it gently. A sine of 5 to 10 mV, small enough to keep the interface linear, rides on a DC bias, and the current is recorded at each frequency from about 0.1 Hz to 100 kHz. The impedance is the ratio of the sine of voltage to the sine of current, amplitude and phase together, $Z(\omega) = V/I = Z' + jZ''$: the same complex impedance as for any circuit (Chapter 1), with real part Z' and imaginary part Z''. A *Nyquist plot* shows $-Z''$ against Z', one point per frequency, high frequencies at the left, with equal axis scales; a *Bode plot* shows $|Z|$ and the phase against frequency.

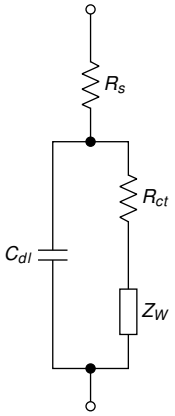


Figure 6.9: The Randles circuit: solution resistance R_s in series with the double-layer capacitance C_{dl}, which is in parallel with the faradaic path, the charge-transfer resistance R_{ct} and the Warburg impedance Z_W of diffusion.

The Randles circuit

The interface of Section 6.3 becomes the *Randles circuit* (Fig. 6.9), the one impedance model this book uses. Start without diffusion, $Z_W = 0$: a resistor R in parallel with a capacitor C, so $Z = R/(1 + j\omega RC)$. As the frequency falls, Z traces a semicircle of diameter R in the Nyquist plane, from $Z \to 0$ at high frequency (the capacitor shorts R) to $Z \to R$ at low frequency. Its top, where $-Z'' = R/2$, is at

$$\omega_{\max} = \frac{1}{R_{ct}C_{dl}}, \qquad f_{\max} = \frac{1}{2\pi R_{ct}C_{dl}}, \tag{6.19}$$

so the diameter gives R_{ct} and the apex frequency then gives C_{dl}. The series solution resistance shifts the whole semicircle to the right by R_s: the high-frequency intercept is R_s, and the low-frequency one is $R_s + R_{ct}$ (Fig. 6.10). For example, a semicircle that starts at the origin, returns to the real axis at 27 kΩ and peaks at 1.1 kHz has $R_{ct} = 27\,\text{k}\Omega$ and $C_{dl} = 1/(2\pi \times 1.1\,\text{kHz} \times 27\,\text{k}\Omega) = 5.4\,\text{nF}$, which at $20\,\mu\text{F/cm}^2$ is a disk about 185 µm across.

KEY EQUATION

$\omega_{\max} = 1/(R_{ct}C_{dl})$

The apex of the semicircle gives C_{dl} once R_{ct} is read from its diameter.

At low frequencies each half cycle lasts long enough to deplete the surface, and diffusion adds the *Warburg impedance* Z_W to the faradaic path. Its real and imaginary parts are equal and both grow as $1/\sqrt{\omega}$, so on the Nyquist plot it is a straight line at 45° that rises without bound as the frequency falls. The plot reads like a map of what limits the current: kinetics at high frequency (the semicircle) and mass transport at low frequency (the line). A fast reaction has a small semicircle

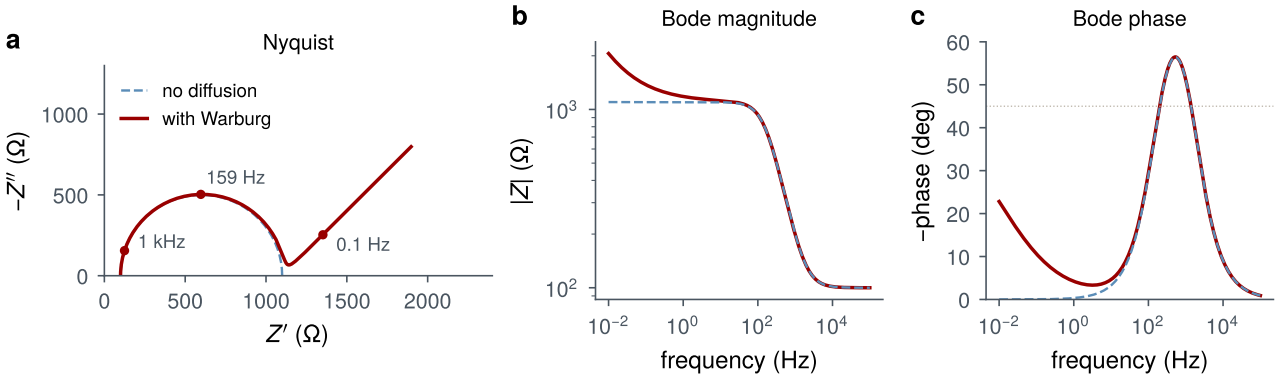


Figure 6.10: Impedance of a Randles circuit with $R_s = 100\,\Omega$, $R_{ct} = 1\,\text{k}\Omega$, $C_{dl} = 1\,\mu\text{F}$ and a Warburg element (computed). (a) Nyquist plot: a semicircle from R_s to $R_s + R_{ct}$ with its apex at 159 Hz, then the 45° Warburg line at low frequency. (b, c) Bode magnitude and phase: the same data against frequency. Dashed curves: the same circuit without diffusion.

and a long Warburg line; a slow reaction, or one behind a blocking film, has a large semicircle (Problem 6.2).

Real electrodes are rough, and their semicircles look flattened; fitting programs then replace C_{dl} by a *constant phase element*, an imperfect capacitor whose phase is a little less than 90°.

6.7 *Enzymatic Glucose Sensors*

More than half a billion adults live with diabetes, and glucose is by far the most measured molecule in medicine. A healthy fasting value is about 4 to 5.5 mM (70 to 100 mg/dL, with 1 mM = 18.0 mg/dL, Chapter 2); below 3.9 mM is hypoglycemia, and people who manage diabetes aim to stay between 3.9 and 10 mM (70 to 180 mg/dL).

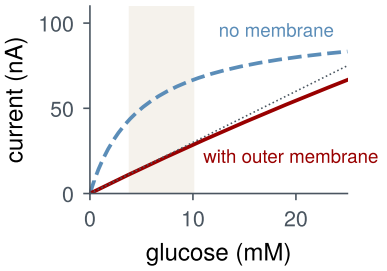


Figure 6.11: Response of an enzyme electrode without and with an outer diffusion-limiting membrane (model curves, illustrative). The membrane cuts the glucose flux and raises the apparent K_m, trading sensitivity for a linear range that covers the clinical band (shaded) and beyond.

The enzyme

Chapter 2 introduced glucose oxidase (GOx), the enzyme of most glucose sensors, and the Michaelis-Menten law that sets its rate. Its cofactor, flavin adenine dinucleotide (FAD), takes two electrons and two protons from glucose and hands them to oxygen:

$$\text{glucose} + (\text{FAD})_{\text{enz}} \longrightarrow \text{gluconolactone} + (\text{FADH}_2)_{\text{enz}}, \tag{6.20}$$

$$(\text{FADH}_2)_{\text{enz}} + \text{O}_2 \longrightarrow (\text{FAD})_{\text{enz}} + \text{H}_2\text{O}_2. \tag{6.21}$$

The net reaction is glucose + O_2 ⟶ gluconolactone + H_2O_2, and the lactone hydrolyzes to gluconic acid. The sensor current follows the Michaelis-Menten form, linear only well below the Michaelis constant K_m. A membrane that limits how fast glucose reaches the enzyme raises the *apparent* K_m and stretches the linear range (Fig. 6.11; Chapter 4 treats the linear range of a saturating sensor).

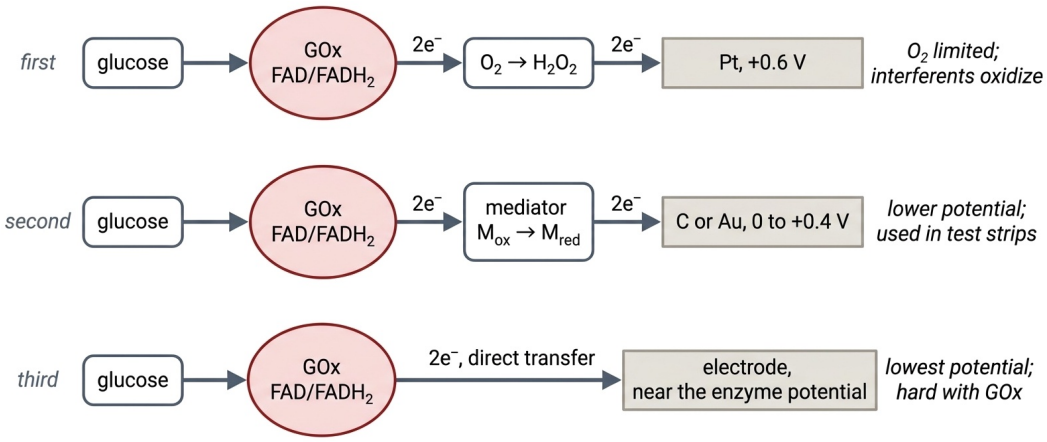


Figure 6.12: Three generations of enzymatic glucose electrodes. First generation: oxygen carries the electrons and the electrode oxidizes the peroxide. Second: an artificial mediator shuttles them to an electrode at a lower potential. Third: electrons pass directly from the enzyme to the electrode. The first enzyme electrodes: L. C. Clark and C. Lyons, *Ann. N.Y. Acad. Sci.* (1962); S. J. Updike and G. P. Hicks, *Nature* (1967).

Three generations of glucose electrodes

The electrons that GOx takes from glucose must reach an electrode, and the three ways of getting them there are called the three *generations* (Fig. 6.12).

First generation sensors let oxygen collect the electrons, as in Eq. (6.21), and measure the peroxide made. A platinum electrode at about +0.6 V vs Ag/AgCl oxidizes it,

$$H_2O_2 \longrightarrow O_2 + 2\,H^+ + 2\,e^-, \tag{6.22}$$

two electrons per glucose; alternatively a cathode near −0.6 V measures the oxygen used, as in the Clark electrode. (The peroxide couple itself sits near +0.07 V at pH 7; the rest of the 0.6 V is overpotential.) Two problems follow. Tissue holds far less oxygen (about 0.06 mM) than glucose (about 5 mM), the *oxygen deficit*, so the response saturates. And at +0.6 V ascorbate (vitamin C), urate and the drug acetaminophen are oxidized too, and look like glucose.

Second generation sensors replace oxygen with an artificial *mediator*, a small redox molecule that carries electrons from the reduced enzyme to the electrode: ferricyanide, ferrocene derivatives, or osmium complexes. The electrode then sits at 0 to +0.4 V, where fewer interferents react, and the oxygen supply matters much less. Nearly all test strips work this way. *Third generation* sensors pass electrons directly from the cofactor to the electrode, which needs the cofactor within about a nanometer of the surface; the FAD of glucose oxidase is buried deep in the protein, so this works best with engineered enzymes.

Two films improve a practical sensor. An inner *permselective* film blocks interferents: Nafion, a negatively charged polymer, repels the anions ascorbate and urate, and size-exclusion films pass the small peroxide molecule but not larger ones such as acetaminophen. An outer *diffusion-limiting* membrane (Chapter 2), often a polyurethane, passes

oxygen far more easily than glucose: it cures the oxygen deficit, keeps the response linear past 20 mM (Fig. 6.11), and protects the enzyme.

Test strips

A finger-stick strip is a second-generation sensor. Capillary action draws the blood over printed carbon electrodes covered with dried enzyme and potassium ferricyanide, and each glucose oxidized leaves two ferrocyanide ions.

After a few seconds the meter steps the electrode to a few hundred millivolts and measures their oxidation current at a fixed time, or integrates it (a *coulometric* strip; Fig. 6.13). The signals are comfortable: if all the glucose in a 0.30 µL drop at 5.5 mM is converted, the charge is $Q = 2FcV = 318\,\mu\text{C}$, and a Cottrell estimate for 11 mM of ferrocyanide ($D = 6.5 \times 10^{-6}\,\text{cm}^2/\text{s}$) at a 1 mm^2 electrode, 5 s after the step, gives 6.8 µA. Errors come from hematocrit (red cells slow diffusion), temperature, and oxygen competing with the mediator; the standard for strips asks that 95 % of readings fall within 15 % of a laboratory analyzer (15 mg/dL at low glucose).

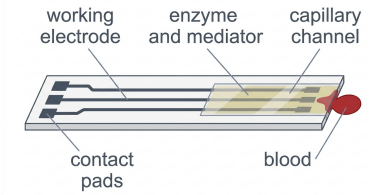


Figure 6.13: A glucose test strip: a capillary channel that holds less than a microliter, over printed carbon electrodes coated with a dried layer of enzyme and mediator.

Continuous glucose monitors

A continuous glucose monitor (CGM) puts a filament carrying the enzyme layer and its membranes a few millimeters into the tissue under the skin (Fig. 6.14) and reads interstitial glucose every one to five minutes for one to two weeks. Glucose must diffuse out of the capillaries, so with the sensor's own response the reading lags blood by about 5 to 15 min (Fig. 6.15). The currents are nanoamperes, the sensor drifts as the enzyme ages and protein coats the membrane, and the potentiostat draws microwatts. Modern sensors are calibrated at the factory and corrected by algorithms that model the lag.

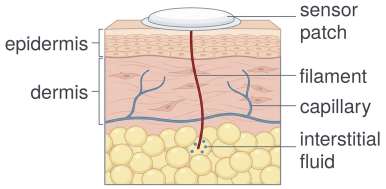


Figure 6.14: A continuous glucose monitor. The filament under the patch sits in the interstitial fluid a few millimeters below the skin.

Calibrating an amperometric sensor

An amperometric sensor is described by the figures of merit of Chapter 4: sensitivity (often per area, in $\mu\text{A}\,\text{mM}^{-1}\,\text{cm}^{-2}$), background current (charging, residual reactions, interferents), linear range, detection limit, and stability. Diffusion coefficients rise about 2 % per kelvin and enzyme activity rises too, so a sensor at skin temperature reads differently than on the bench. With a linear response, two calibration points give both slope and background.

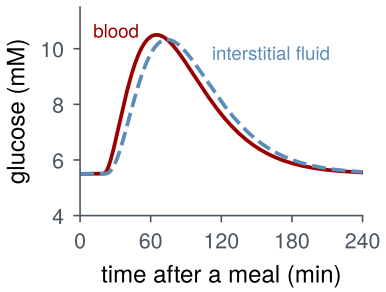


Figure 6.15: Blood and interstitial glucose after a meal with a first-order lag of 8 min between them (model, illustrative). While glucose rises, a sensor in interstitial fluid reads low.

Worked Example 6.3 *Calibrating a glucose sensor*

An implantable glucose sensor with an electrode area of $0.20\,\mathrm{mm}^2$ reads 6.1 nA in a 4.0 mM calibration solution and 15.7 nA at 12.0 mM. (a) Find the sensitivity, the area-normalized sensitivity and the background current. (b) What glucose does a reading of 9.0 nA represent, in mM and mg/dL? (c) After a meal blood glucose is rising at 2 mg/dL per minute and the sensor lags by 8 min (Fig. 6.15). How far behind is the reading?

Solution. (a) $S = (15.7 - 6.1)/(12.0 - 4.0) = 1.20\,\mathrm{nA/mM}$. Per area, $1.20\,\mathrm{nA/mM}/0.0020\,\mathrm{cm}^2 = 0.60\,\mu\mathrm{A\,mM^{-1}\,cm^{-2}}$. The background is $6.1 - 1.20 \times 4.0 = 1.3\,\mathrm{nA}$.

(b) $c = (9.0 - 1.3)/1.20 = 6.42\,\mathrm{mM}$, or $6.42 \times 18.0 = 116\,\mathrm{mg/dL}$.

(c) A first-order lag τ under a steady ramp leaves the reading behind by the ramp rate times τ: $2 \times 8 = 16\,\mathrm{mg/dL}$, about 0.9 mM. The same lag makes the sensor read high while glucose falls, which matters most when it falls toward hypoglycemia. The calibration itself assumes a linear response through both points; a third point near 20 mM would test it. ■

6.8 Potentiometric Sensors

A potentiometric sensor draws no current. It measures the equilibrium voltage between an indicator electrode, whose potential depends on the analyte, and a reference electrode. For an ideal sensor selective to an ion i of charge z_i, the Nernst equation gives

$$E = K + \frac{2.303\,RT}{z_i F} \log_{10} a_i, \tag{6.23}$$

where K collects the standard, reference and junction potentials. The slope is $+59.2\,\mathrm{mV}$ per decade for a monovalent cation at 25 °C and $+29.6\,\mathrm{mV}$ for a divalent one. For an anion z_i is negative, so the slope is *negative*: more chloride or fluoride makes the potential fall. Because the response is logarithmic, 1 mV is about 4 % of concentration for a monovalent ion at any level.

key equation

$$E = K + \frac{2.303RT}{z_i F} \log_{10} a_i$$

An ion-selective electrode: +59.2 mV per decade for a monovalent cation at 25 °C, negative for an anion.

The glass pH electrode

The oldest potentiometric sensor is the glass pH electrode: a thin bulb of special glass separates the sample from an internal solution of fixed pH that holds an internal Ag/AgCl reference. Hydrogen ions exchange with sodium ions in a thin hydrated layer on each glass surface, so the potential across the bulb follows the pH difference between

the two sides:

$$E = E^0 - \frac{2.303\,RT}{F}\,\mathrm{pH}, \tag{6.24}$$

59.16 mV per pH unit at 25 °C. Electrically the electrode is a voltage source behind the resistance of the glass, 50 to 500 MΩ. A 10 MΩ voltmeter on a 500 MΩ electrode would read only 2 % of its voltage, so the electrode needs the electrometer buffer of Chapter 4, with an input resistance of 10^{12} Ω or more (as in the pH meter of Example 4.3).

Ion-selective membranes

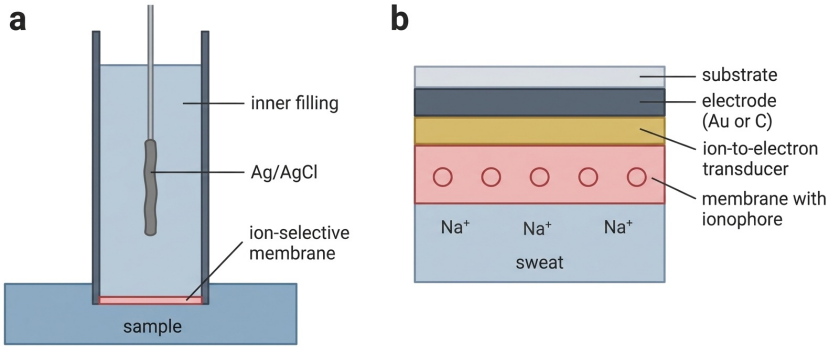


Figure 6.16: Ion-selective electrodes. (a) The classic design: a membrane closes a tube of inner filling solution with an Ag/AgCl wire. (b) A solid-contact ISE for a wearable: a conducting polymer or carbon layer replaces the inner solution under a polymer membrane that carries the ionophore (circles), which binds only the target ion.

Ions other than H^+ are sensed with membranes that carry an *ionophore*, a molecule that binds one ion selectively and shuttles it into the membrane. The classic membrane is plasticized poly(vinyl chloride) with about one percent ionophore: valinomycin for K^+, synthetic crown ethers for Na^+. Fluoride uses a crystal of lanthanum fluoride instead. Ion exchange at the membrane surface sets up a Nernstian potential (Fig. 6.16a). No membrane is perfect; the *Nikolsky-Eisenman equation* adds the interfering ions j:

$$E = K + \frac{2.303\,RT}{z_i F}\log_{10}\left(a_i + \sum_j K_{ij}^{\mathrm{pot}}\, a_j^{z_i/z_j}\right), \tag{6.25}$$

where the selectivity coefficient K_{ij}^{pot} is small for a good membrane (about 10^{-4} for Na^+ on a valinomycin K^+ membrane). The same formula explains the detection limit: at low activity the interferents flatten the response, near 10^{-6} to 10^{-5} M (Fig. 6.17).

For a wearable, the inner solution is replaced by a *solid contact*: a conducting polymer (such as PEDOT) or carbon layer between the membrane and the metal that converts ionic to electronic charge (Fig. 6.16b). Solid-contact ISEs for Na^+ and K^+, together with enzymatic glucose and lactate sensors and a temperature sensor, make up multiplexed sweat patches that read out over Bluetooth.[4] Sweat

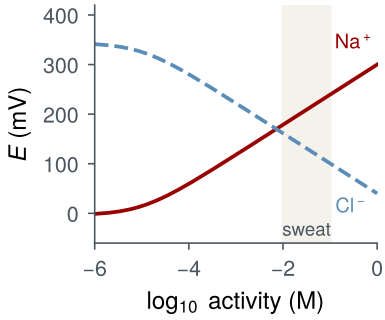


Figure 6.17: Ideal calibration lines at 33 °C: a Na^+ ISE (+60.7 mV per decade) and a Cl^- ISE (−60.7 mV per decade), with a detection limit near 10^{-5} M (model, illustrative); sweat lies in the shaded band.

[4] W. Gao, S. Emaminejad, …, A. Javey, *Nature* (2016).

sodium runs from about 10 to 90 mM and potassium from about 4 to 8 mM, far above the detection limits; the hard problems are a stable reference (Section 6.4) and a water layer that forms under the membrane and makes the potential drift.

WORKED EXAMPLE 6.4 *A fluoride electrode calibrated at one point*

A fluoride-selective electrode reads +0.152 V against its reference in a 1.00 mM fluoride standard at 25 °C. In a sample it reads +0.105 V. (a) Assuming a Nernstian response, find the fluoride concentration. (b) A second standard at 10.0 mM reads +0.0952 V. Recompute the sample concentration and comment on single-point calibration.

SOLUTION. (a) Fluoride is an anion ($z = -1$), so $E = K - 0.059\,16\,\text{V}\log_{10}[\text{F}^-]$. From the standard, $K = 0.152 + 0.05916\log_{10}(1.00\times10^{-3}) = -0.0255\,\text{V}$. For the sample,

$$\log_{10}[\text{F}^-] = \frac{K-E}{0.05916} = \frac{-0.0255-0.105}{0.05916} = -2.206,$$

so $[\text{F}^-] = 6.2$ mM. The potential fell by 47 mV, which for an anion means the concentration rose, by $10^{0.794} = 6.2$ times. With the sign of the slope wrong, the same fall would be read as a 6.2-fold *decrease*, to 0.16 mM, wrong by a factor of 39: always check the sign of the slope against the charge of the ion.

(b) The measured slope is $(0.0952 - 0.152)/\log_{10}(10) = -56.8$ mV per decade, sub-Nernstian, as real electrodes often are. Then $\log_{10}(c/1\,\text{mM}) = (0.105-0.152)/(-0.0568) = 0.828$, and $c = 6.7$ mM, 8 % more than the one-point result. One standard fixes only K and assumes the theoretical slope, and every millivolt of slope error grows with the distance from the standard. Two standards that bracket the sample measure the slope too, and a third checks linearity. ■

6.9 *Affinity Biosensors*

MANY TARGETS HAVE NO ENZYME that turns them over: proteins, hormones, drugs, viruses. Affinity sensors detect them by binding. The receptor holds the target; the transducer reports that binding has happened.

Antibodies, aptamers, and binding

An *antibody* (immunoglobulin G) is a Y-shaped protein of about 150 kDa, made by the immune system: two heavy and two light

chains, with a binding site for one *antigen* at the tip of each arm (Fig. 6.18). A second, labeled antibody that binds another part of the antigen forms a *sandwich*, the basis of most immunoassays. An *aptamer* (Chapter 2) is a single strand of DNA or RNA, typically 20 to 80 nucleotides, selected from a random library to fold around its target; it is made chemically, is a few nanometers in size, carries labels at chosen positions, and often refolds when it binds. For either receptor the fraction of occupied sites follows the binding curve of Chapter 4, the Langmuir isotherm,

$$\theta = \frac{C}{C + K_D}, \tag{6.26}$$

where K_D is the dissociation constant, the concentration at which half the sites are filled (nanomolar or lower for good antibodies, nanomolar to micromolar for aptamers). A sensor is most useful from about $0.1K_D$ to $10K_D$.

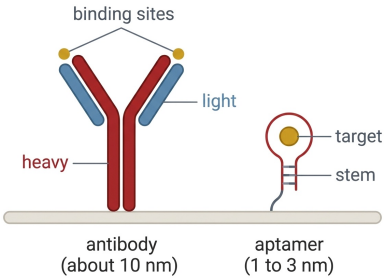


Figure 6.18: Two receptors on a surface. An IgG antibody has two heavy and two light chains and two identical binding sites at the tips. An aptamer is a short folded strand of DNA or RNA that wraps around its target.

Impedimetric sensing

In a *faradaic* impedance sensor the receptor layer sits on an electrode in a solution of a redox probe, usually 5 mM each of ferri- and ferrocyanide. Every layer on the electrode makes it harder for the probe to reach the metal, so coating the electrode raises R_{ct} and binding the target raises it again (Fig. 6.19). The change ΔR_{ct}, often normalized as $\Delta R_{ct}/R_{ct,0}$ to cancel electrode-to-electrode differences, is the signal. Because it measures blocking, not charge, it works in saline, where the Debye length is under a nanometer. A non-faradaic variant, without a probe, tracks the fall of the interfacial capacitance instead.

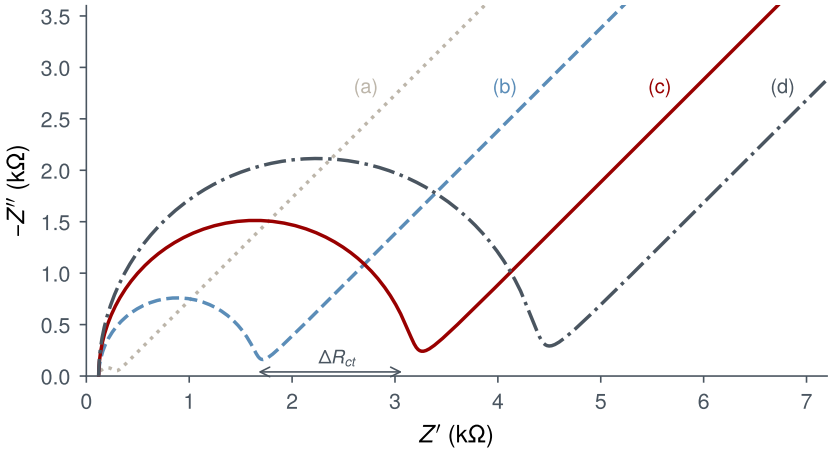


Figure 6.19: Faradaic impedance spectra of an affinity sensor with a ferri/ferrocyanide probe (computed from Randles circuits, illustrative): (a) bare electrode, $R_{ct} = 0.15\,\text{k}\Omega$; (b) coated with the receptor layer, $1.5\,\text{k}\Omega$; (c) after the target binds, $3.0\,\text{k}\Omega$; (d) at a higher target concentration, $4.2\,\text{k}\Omega$. The Warburg line shifts right with each increase of R_{ct}.

Optical labels and real-time sensors

Binding can also switch a light signal. In *Förster resonance energy transfer* (FRET) an excited donor dye hands its energy to an acceptor dye a few nanometers away, with an efficiency that falls with the sixth power of their distance, so a receptor that changes shape when it binds reports binding as a change of color (Chapter 7 returns to FRET).

Most affinity sensors are used once. *Electrochemical aptamer-based* (EAB) sensors are reversible and need no added reagents, so they follow a concentration second by second, even in the blood of a living animal.[5] The aptamer is attached at one end to a gold electrode through a thiol and carries a redox reporter, usually methylene blue, at the other (Fig. 6.20). When the target binds, the aptamer folds, the reporter's distance to the gold changes, and so does its electron-transfer rate. A fast voltammetric scan reads the reporter's peak current, which follows Eq. (6.26) and, because the reporter is fixed to the surface, grows in proportion to scan rate (Section 6.5). The current also drifts as the monolayer degrades and proteins foul it, so real devices measure at two scan frequencies and use their difference, the differential trick of Chapter 4.

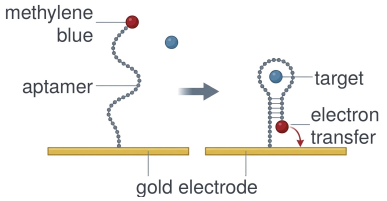


Figure 6.20: An electrochemical aptamer-based sensor. The aptamer, tethered to gold, folds when it binds the target and brings its methylene blue reporter closer to the electrode, which speeds up electron transfer.

[5] B. S. Ferguson, ..., K. W. Plaxco, and H. T. Soh, *Sci. Transl. Med.* (2013); N. Arroyo-Currás, ..., K. W. Plaxco, *Proc. Natl. Acad. Sci. USA* (2017).

6.10 *Potentiostat Circuits*

To the circuit, the cell is a network of resistors and capacitors (Fig. 6.21): the solution resistance R_c from the counter electrode to the point the reference senses, the uncompensated resistance R_u from there to the working electrode, and the interface, C_{dl} in parallel with R_{ct}. The reference adds its own resistance (the frit, kilohms to megohms). No current flows in the RE lead, so its resistance causes no error; the drop iR_u does, which is why the reference tip sits close to the WE. A potentiostat needs three op-amp circuits from Chapter 4: a follower to read the reference without loading it, an inverting summer to compare it with the command, and a transimpedance amplifier (TIA) to hold the working electrode at a virtual ground and turn its current into a voltage.

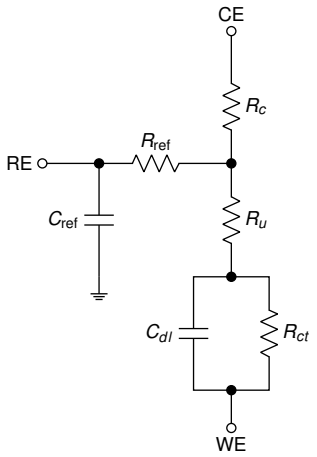


Figure 6.21: Equivalent circuit of a three-electrode cell. The reference senses the solution through its own resistance, and C_{ref} is the stray capacitance of its lead; R_u is the uncompensated resistance between the reference tip and the working electrode.

The adder potentiostat

Figure 6.22 is the classic three-amplifier potentiostat. The working electrode sits at the virtual ground of OA-3, and the cell current flows through R_{out}: $V_{out} = -i_{WE}R_{out}$, with i_{WE} drawn out of the WE as in the figure, the direction of a reduction (cathodic) current; an anodic current gives a positive output. OA-2, a follower, reads the reference without drawing current. OA-1 is an inverting summer whose node S

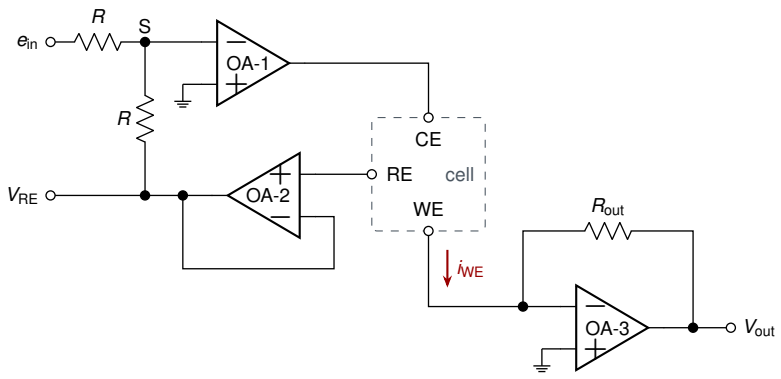


Figure 6.22: The adder potentiostat. The dashed box is the cell of Fig. 6.21, seen at its three terminals. OA-1, the control amplifier, drives the counter electrode. OA-2 buffers the reference electrode, so no current flows in it. OA-1 sums e_{in} and the buffered reference potential at node S, which forces $V_{RE} = -e_{in}$. OA-3, a transimpedance amplifier, holds the working electrode at virtual ground and converts its current to $V_{out} = -i_{WE}R_{out}$.

is a virtual ground, so $e_{in}/R + V_{RE}/R = 0$: OA-1 moves the counter electrode until $V_{RE} = -e_{in}$. Then

$$E_{WE} - E_{RE} = 0 - (-e_{in}) = +e_{in}. \tag{6.27}$$

More resistors into node S add more inputs (a DC bias, a CV ramp, a small sine for EIS), and the cell sees their sum. R_{out} sets the current range: 1 kΩ gives 1 V/mA, 10 MΩ gives 10 mV/nA, and instruments switch it in decades from picoamperes to milliamperes, with a capacitor across each. Because the loop includes the cell's capacitance, a small capacitor around OA-1 keeps it stable with large electrodes.

KEY IDEA A potentiostat is a feedback loop through the cell: the control amplifier moves the counter electrode until the reference reads the command, and the working-electrode current is measured separately.

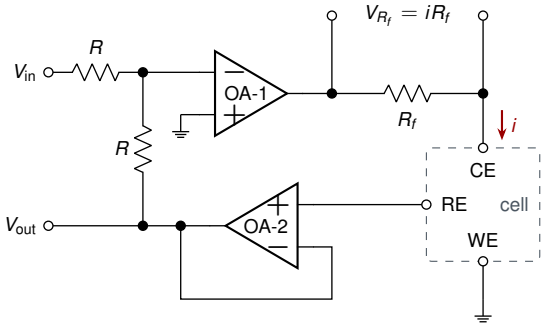


Figure 6.23: A potentiostat with a grounded working electrode, drawn like Fig. 6.22. The cell current i is measured as the voltage V_{R_f} between the two terminals across R_f, in series with the counter electrode.

WORKED EXAMPLE 6.5 *Analyzing a potentiostat*

For the circuit of Fig. 6.23: (a) What is V_{out} in terms of V_{in}, and why? (b) What is the potential of the working electrode against the reference? (c) How is the working-electrode current measured? (d) The cell current is 1 µA. Design a circuit that produces 1 V for this current, using only resistors between 100 Ω and 220 kΩ.

SOLUTION. (a) The inverting input of OA-1 is a virtual ground that draws no current, so the current V_{in}/R through the input resistor returns through the second R to OA-2: $V_{in}/R + V_{out}/R = 0$, or $V_{out} = -V_{in}$. OA-1 drives the counter electrode until this holds.

(b) OA-2 is a follower, so $V_{RE} = V_{out} = -V_{in}$, and with the WE grounded, $E_{WE} - E_{RE} = +V_{in}$: for $V_{in} = 0.40\,V$ the reference sits at $-0.40\,V$.

(c) No current enters the follower, so all of the current that OA-1 supplies flows through R_f, the counter electrode, the solution and the working electrode to ground. It is measured as the voltage across R_f, $V_{R_f} = iR_f$. That voltage floats (neither end is at ground), so it must be read by a differential amplifier whose inputs draw no current from the cell path, such as the instrumentation amplifier of Chapter 8.

(d) The largest allowed resistor, $220\,k\Omega$, gives only $0.22\,V$ for $1\,\mu A$, so the job takes a trick. Move the measurement to a transimpedance amplifier at the working electrode, as in Fig. 6.22, and give it the T-network feedback of Chapter 4 (Fig. 6.24): R_1 and R_3 in series from the summing node to the output, R_2 from their junction to ground. Its effective resistance is

$$R_{eff} = R_1 + R_3 + \frac{R_1 R_3}{R_2} = 100 + 100 + \frac{(100)(100)}{12.5}\ k\Omega = 1.00\,M\Omega$$

with $R_1 = R_3 = 100\,k\Omega$ and $R_2 = 12.5\,k\Omega$ (the standard $12.4\,k\Omega$ gives $1.006\,M\Omega$). The output is $-1.0\,V$ for a reduction current of $1\,\mu A$; an inverting stage flips the sign if needed. The T-network multiplies the op amp's offset and voltage noise by $1 + R_3/R_2 = 9$, a fair price for avoiding a megohm resistor. (A $100\,k\Omega$ TIA and an inverting gain of 10 would also work.) ■

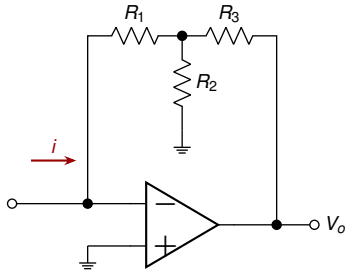


Figure 6.24: T-network feedback (Chapter 4): $V_o = -i\,(R_1 + R_3 + R_1 R_3/R_2)$, a large transimpedance from moderate resistors.

A single-supply potentiostat

A wearable runs from one battery, so there is no negative rail. The fix is a *virtual ground*: a buffered voltage V_z, often half the supply, that plays the role of zero (Fig. 6.25). The TIA holds the working electrode at V_z and outputs $V_{out} = V_z + i_a R_{TIA}$ for an anodic current i_a, so currents of either sign stay inside the rails. The control amplifier compares the reference with a set point $V_z - E_{bias}$ from a DAC and drives the counter electrode until they match. Integrated potentiostat front ends build exactly this, with a programmable bias and a TIA resistor switched in steps.

Take a wearable lactate sensor that oxidizes peroxide at $+0.60\,V$ vs its Ag/AgCl reference and makes up to $2\,\mu A$, on $3.3\,V$ with $V_z = 1.65\,V$ and a 12-bit ADC. The reference set point is $1.65 - 0.60 = 1.05\,V$. Keeping the output $0.15\,V$ inside the rails at $\pm 2\,\mu A$ allows $1.50\,V$ each way, so $R_{TIA} = 1.50\,V/2\,\mu A = 750\,k\Omega$, and one LSB ($3.3\,V/4096 = 0.806\,mV$) is $1.07\,nA$; a nanoampere CGM signal would use $10\,M\Omega$ or

more. The counter electrode, which must reduce something (oxygen or water) to balance the anodic current, is driven below the reference toward ground: about 1 V of room, which a lower V_z would take away.

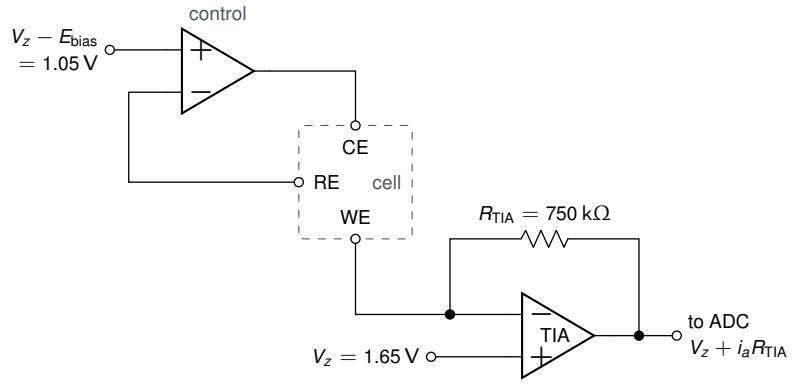


Figure 6.25: A single-supply potentiostat on 3.3 V. The working electrode is held at the virtual ground $V_z = 1.65\,V$ by the TIA; the control amplifier holds the reference at the DAC set point of 1.05 V, 0.60 V below the WE, by driving the counter electrode.

6.11 *Microfluidics for Wearable Sweat Sensing*

A SWEAT SENSOR NEEDS FRESH SWEAT. Sweat that pools on the skin evaporates and mixes with old sweat, so a patch collects it through a small inlet and routes it through microfluidic channels past the electrodes (Fig. 6.26).[6] Local sweat rates during exercise are of order $1\,\mu L\,min^{-1}\,cm^{-2}$, much lower at rest, where sweat can be induced by iontophoresis (Chapter 1). Three facts about small channels set the design; we follow a typical patch, $0.80\,cm^2$ of skin sweating at $0.60\,\mu L\,min^{-1}\,cm^{-2}$ into a channel 200 µm wide and 100 µm deep, a flow of $Q = 0.48\,\mu L/min$.

[6] Z. Sonner, ..., J. Heikenfeld, *Biomicrofluidics* (2015), review how the sweat gland works as a microfluidic source and what that means for sensors.

Flow is laminar. The Reynolds number, the ratio of inertial to viscous forces,

$$\mathrm{Re} = \frac{\rho v D_h}{\mu}, \qquad D_h = \frac{2wh}{w+h}, \tag{6.28}$$

with density ρ, mean velocity v, viscosity μ and the hydraulic diameter D_h of a $w \times h$ channel, is far below the value of about 2000 where flow turns turbulent. In our channel the sweat moves at $v = Q/(wh) = 0.40\,mm/s$, $D_h = 133\,\mu m$, and with the density and viscosity of water $\mathrm{Re} = 0.053$. Streams flow side by side without turbulence, and they mix only by diffusion, which across 200 µm takes $w^2/2D \approx 15\,s$ for sodium.

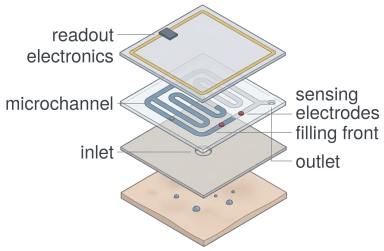


Figure 6.26: A wearable sweat patch. Sweat enters through an inlet over the skin, flows through a serpentine microchannel past the sensing electrodes, and leaves at an outlet. The position of the filling front over time gives the sweat rate.

Surface tension fills the channel. In a channel with hydrophilic walls (a *contact angle* $\theta < 90°$ between the liquid surface and the wall, as on the silicone rubber PDMS after an oxygen plasma; Chapter 12 explains wetting), the curved meniscus pulls the liquid in with a capillary pressure $\Delta P = 2\gamma\cos\theta\,(1/w + 1/h)$: 1.9 kPa for our channel, with the

surface tension of water, $\gamma = 72\,\mathrm{mN/m}$, and $\theta = 30°$. Sweat wicks along as fast as the glands supply it, so the supply, not capillarity, sets the pace; abrupt widenings act as capillary valves that let a patch fill chambers in sequence.

The refresh time is volume over flow. A sensor chamber of volume V is flushed in a time of order $\tau = V/Q$. A chamber $2\,\mathrm{mm} \times 2\,\mathrm{mm} \times 100\,\mu\mathrm{m}$ holds $0.40\,\mu\mathrm{L}$ and is refreshed in $\tau = 50\,\mathrm{s}$; if it mixes well, its contents approach the new sweat as $1 - e^{-t/\tau}$ and reach 95 % after $3\tau = 150\,\mathrm{s}$, so a change in sweat sodium shows up a minute or two late. A long serpentine channel of known cross section turns the position of the filling front into the total sweat volume, and its speed into the sweat rate.[7] At rest, a front that moves 12 mm in 2 min along our channel has swept $0.24\,\mu\mathrm{L}$: $0.12\,\mu\mathrm{L/min}$, or $0.15\,\mu\mathrm{L\,min^{-1}\,cm^{-2}}$, a quarter of the exercise rate. A printed patch can also read the front electrically. Its channel runs between two printed silver plates, and sweat has a far higher permittivity than the air it replaces, so the capacitance rises as the channel fills, and the slope of capacitance against time gives the sweat rate.[8]

[7] A. Koh, …, J. A. Rogers, *Sci. Transl. Med.* (2016).

[8] M. S. Islam, S. Cha, M. F. Hassan, *et al.*, Y. Khan, *Adv. Intell. Syst.* (2024).

Summary

- Reduction gains electrons, oxidation loses them; a more positive $E°$ is a stronger oxidant. $E_{\mathrm{cell}} = E_{\mathrm{cathode}} - E_{\mathrm{anode}}$, $\Delta G = -nFE$.
- Nernst at an electrode: $E = E° + (2.303RT/nF)\log_{10}(a_{\mathrm{Ox}}/a_{\mathrm{Red}})$, 59.16 mV per decade per electron at 25 °C, 61.54 mV at 37 °C.
- Interface: $\lambda_D = 0.304\,\mathrm{nm}/\sqrt{I/\mathrm{M}}$ (0.79 nm in saline), C_{dl} 10 to $40\,\mu\mathrm{F/cm^2}$, $R_{ct} = RT/(nFi_0)$, limiting current $i_L = nFADC^*/\delta$.
- Three electrodes: WE, RE (no current), CE. Ag/AgCl is 0.210 V vs SHE in 3 M KCl; a bare pseudo-reference moves 59 mV per decade of chloride.
- CV: $\Delta E_p \approx 59/n$ mV, $i_p = 2.69 \times 10^5\, n^{3/2} AC^* \sqrt{Dv}$. Cottrell $i = nFAC^*\sqrt{D/\pi t}$; microdisk $i_{ss} = 4nFDC^*r$.
- EIS: Randles semicircle from R_s to $R_s + R_{ct}$, apex at $\omega = 1/(R_{ct}C_{dl})$, then a 45° Warburg line.
- Glucose: GOx makes H_2O_2, oxidized near +0.6 V vs Ag/AgCl; mediators work at 0 to +0.4 V; CGMs lag blood by 5 to 15 min.
- Ion-selective electrodes: $E = K + (2.303RT/z_iF)\log a_i$, negative slope for anions; calibrate with two standards. Affinity sensors: $\theta = C/(C + K_D)$, binding raises R_{ct}.
- Potentiostat: control amplifier drives CE, buffered RE, TIA at WE; T-network $R_1 + R_3 + R_1R_3/R_2$; single supply with a virtual ground.

Sweat channels: Re $\ll 1$, refresh time V/Q.

Every sensor in this chapter reads chemistry through electrons crossing an electrode. Chapter 7 reads it with light instead: how much a sample absorbs, how it fluoresces, and, in the pulse oximeter, the color of blood inside a finger.

6.12 *Problems*

6.1 ° Redox basics.

(a) A cell joins a zinc electrode in 1 M Zn^{2+} to a copper electrode in 1 M Cu^{2+}. Write the two half reactions, name the anode, and find E_{cell} and ΔG per mole of zinc.

(b) Ascorbic acid is oxidized as $C_6H_8O_6 \longrightarrow C_6H_6O_6 + 2\,H^+ + 2\,e^-$. At a sensor electrode, is its current anodic or cathodic, and how many electrons does each molecule give?

6.2 •• Figure 6.27 shows the impedance of two electrode reactions. The top of the semicircle is at 3.2 Hz for reaction A and at 280 Hz for reaction B.

(a) Estimate R_s and R_{ct} for both reactions.

(b) What process controls the current in each case over most of the frequency range?

(c) Find C_{dl} for each.

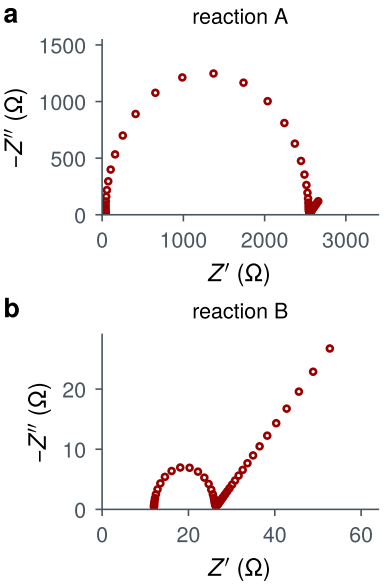


Figure 6.27: Impedance-plane plots of two electrode reactions, (a) A and (b) B, from 10^{-4} Hz to 10^{4} Hz (Problem 6.2; computed, illustrative).

6.3 • A sweat patch reads a Na^+ ISE against a bare Ag/AgCl pseudo-reference at 33 °C.

(a) Show that the cell voltage is $K' + s\log_{10}(a_{Na^+}a_{Cl^-})$, with $s = 2.303RT/F$.

(b) Sweat NaCl rises from 20 to 60 mM (both ions). By how much does the voltage change? What sodium ratio would someone infer who assumes a slope of s?

(c) Suggest two fixes.

6.4 • An impedimetric aptasensor has $R_{ct} = 2.0\,\text{k}\Omega$ before and 2.9 kΩ after a sample. Its calibration follows Eq. (6.26) with $K_D = 50$ nM and a saturating $\Delta R_{ct}/R_{ct,0} = 1.2$.

(a) Find the target concentration.

(b) Over what range of concentration is the sensor most useful?

(c) Why is the signal normalized by $R_{ct,0}$?

6.5 • A sweat patch has a spiral channel 300 µm wide, 150 µm deep and 60 mm long, fed from an inlet over 1.2 cm^2 of skin sweating at 0.80 $\mu L\,min^{-1}\,cm^{-2}$.

(a) How long does the channel take to fill?

(b) Find the mean velocity and the Reynolds number.

7
Optical Biosensing

In this chapter you will learn to

- Convert between wavelength, photon energy, and photon rate, and say which molecular events absorb which part of the spectrum.
- Use the Beer-Lambert law to find a concentration from an absorbance, for one absorber and for a mixture, and say when the law fails.
- Read a Jablonski diagram and explain fluorescence, phosphorescence, the Stokes shift, lifetime, quantum yield, and quenching.
- Explain how a fluorometer rejects its own excitation light and how a lifetime measurement senses oxygen.
- Compute the penetration depth of an evanescent wave, and explain how total internal reflection fluorescence and surface plasmon resonance see only the molecules bound to a surface.
- Compute SpO_2 from the ratio of ratios, explain why every oximeter is calibrated empirically, and describe its bias in people with darker skin.
- Size a photodiode and transimpedance front end from a photocurrent budget, and reject ambient light by pulsing the source.

In a hospital laboratory, a technician drops a blood sample into a cuvette, slides it into a spectrophotometer, and reads the hemoglobin concentration a few seconds later. On the ward, a clip on a patient's finger reports blood oxygen once a second, all day, with no sample at all. In a research lab, a gold-coated glass chip reports, in real time, how fast an antibody grabs a viral protein. In each case the instrument

shines light at molecules and measures what comes back: how much is absorbed, what color is emitted, or how a reflection changes. Light is a favorite transducer in biosensing: photons do not disturb the sample, cross air, glass, and tissue, and are cheap to make and to detect.

Chapter 6 read a concentration through the electrons that cross an electrode; this chapter reads it with light. It starts with light and molecules (Section 7.1) and with absorption, which the Beer-Lambert law turns into a concentration (Section 7.2). Sections 7.3 and 7.4 follow a molecule that absorbs and then glows, and build the instrument that measures the glow. Sections 7.5 and 7.6 use the boundary between glass and a liquid to see only the molecules bound to a surface. Section 7.7 applies the Beer-Lambert law to blood inside a finger: the pulse oximeter, whose signal Chapter 1 introduced. Section 7.8 closes with the electronics every optical sensor shares: the light source, the photodiode, and the transimpedance amplifier of Chapter 4.

7.1 *Light and Molecular Energy Levels*

Light is an electromagnetic wave whose frequency ν and wavelength λ obey $c = \nu\lambda$, with $c = 2.998 \times 10^8\,\mathrm{m/s}$ in vacuum. It also comes in packets, photons, each carrying the energy

$$E = h\nu = \frac{hc}{\lambda}, \tag{7.1}$$

where $h = 6.626 \times 10^{-34}\,\mathrm{J\,s}$ is Planck's constant. A photon is absorbed only when its energy matches the gap between two energy levels of the molecule it meets. That single rule is the basis of every technique in this chapter.

KEY EQUATION

$E = h\nu = hc/\lambda$

Photon energy. In practical units, $E\,[\mathrm{eV}] = 1240/\lambda\,[\mathrm{nm}]$.

One *electron volt* (eV), the energy an electron gains in crossing one volt, is $1.602 \times 10^{-19}\,\mathrm{J}$. Energy in electron volts compares directly with the energy gaps of molecules and semiconductors; with $hc = 1240\,\mathrm{eV\,nm}$, a 620 nm photon carries 2.0 eV, and the visible band runs from about 400 nm (violet, 3.1 eV) to 700 nm (red, 1.8 eV).

A molecule stores energy in three main ways, and each one absorbs a different part of the spectrum (Fig. 7.1). *Electronic* transitions move a valence electron from one molecular orbital to a higher one. Their gaps are a few electron volts, so they absorb visible and ultraviolet light; they cause color and fluorescence and are the subject of most of this chapter. *Vibrational* transitions change how strongly bonded atoms oscillate, with gaps of 0.04 to 0.4 eV, in the infrared. *Rotational* transitions change how fast the molecule spins and absorb microwaves. Because vibrational gaps are small, each electronic level carries a ladder of vibrational levels, which in a liquid smear an electronic absorption line into a band tens of nanometers wide.

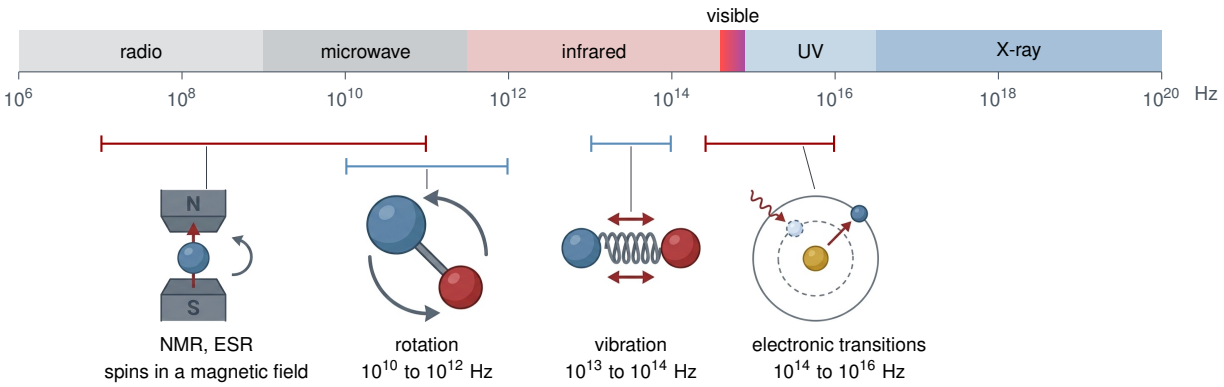


Figure 7.1: The electromagnetic spectrum and the molecular events that absorb each part of it. Electronic transitions of valence electrons sit in the visible and ultraviolet, bond vibrations in the infrared, molecular rotations in the microwave, and spin flips of nuclei in a magnetic field (NMR) at radio frequencies. This chapter works mostly in the visible and near infrared.

Optical instruments look at one of three things. *Absorption* instruments (the spectrophotometer, the colorimetric test strip, the pulse oximeter) measure the fraction of light a sample removes. *Emission* instruments measure light the sample gives off after it has been excited, by light (fluorescence) or by a chemical reaction (chemiluminescence, used in many clinical immunoassays); they work against a dark background and so reach lower concentrations. *Scattering* instruments measure light the sample redirects: mostly at the same color (elastic scattering, which makes the sky blue and tissue look white), and in a tiny fraction at a slightly shifted color (Raman scattering).

Worked Example 7.1 *Photon energies of common sources*

Find the photon energy of a 660 nm red LED, a 940 nm infrared LED, and the 254 nm line of a mercury lamp. How many photons per second does a 1.0 mW red beam carry?

Solution. For 660 nm, Eq. (7.1) gives

$$E = \frac{(6.626 \times 10^{-34}\,\mathrm{J\,s})(2.998 \times 10^{8}\,\mathrm{m/s})}{660 \times 10^{-9}\,\mathrm{m}} = 3.01 \times 10^{-19}\,\mathrm{J} = 1.88\,\mathrm{eV}.$$

The shortcut $1240/\lambda$ gives 1.32 eV at 940 nm and 4.88 eV at 254 nm. The ultraviolet photon carries enough energy to break chemical bonds, which is why 254 nm light sterilizes surfaces and why biological samples are measured with as little UV as possible.

The photon rate is the power divided by the energy per photon:

$$\dot{N} = \frac{P}{E} = \frac{1.0 \times 10^{-3}\,\mathrm{W}}{3.01 \times 10^{-19}\,\mathrm{J}} = 3.3 \times 10^{15}\,\mathrm{photons/s}.$$

Even a weak beam carries an enormous number of photons, so in most biosensors the limit is not the number of photons but how many of them carry information. ■

Infrared, Raman, and magnetic resonance

The rest of the spectrum has its own instruments, which this book only names. Bond vibrations absorb in the infrared, where spectroscopists quote the *wavenumber* $1/\lambda$ in cm^{-1}, but only if the vibration changes the molecule's dipole moment: the N_2 and O_2 of air are transparent, while CO_2 absorbs near 4.3 µm, the band a capnograph uses to monitor exhaled CO_2 during anesthesia. Raman spectroscopy reads the same vibrations as small color shifts of scattered visible or near-infrared laser light, and suits tissue because water scatters weakly. Rotations matter mostly for gases. Nuclei with spin, such as the proton, act as tiny magnets: in a magnetic field they absorb radio waves at 42.58 MHz per tesla (63.9 MHz in a 1.5 T scanner), the nuclear magnetic resonance (NMR) behind magnetic resonance imaging.[1]

[1] An optical pulse sensor can even work during a scan: a Bluetooth photoplethysmography wearable (Section 7.7) recorded the pulse inside a 0.55 T scanner with no observable effect on either its signal or the image. F. Muñoz, K. S. Nayak, Y. Khan, *IEEE Sens. Lett.* (2025).

7.2 *Absorption and the Beer-Lambert Law*

THE SIMPLEST OPTICAL MEASUREMENT asks how much light a sample removes. Clinical analyzers measure hemoglobin, bilirubin, and the colored products of enzyme reactions this way, and every molecular biology lab measures DNA by its absorbance at 260 nm and protein at 280 nm. The Beer-Lambert law turns the measurement into a concentration.

Picture a beam of intensity I crossing a thin slice $\mathrm{d}x$ of a sample that holds N absorbing molecules per unit volume, each blocking an area σ, its *absorption cross section* (Fig. 7.2). The slice blocks a fraction $\sigma N\,\mathrm{d}x$ of the beam. Every slice removes the same fraction of whatever light reaches it, so the intensity falls exponentially along a path ℓ:

$$\frac{\mathrm{d}I}{I} = -\sigma N\,\mathrm{d}x \quad\Longrightarrow\quad I = I_0\,e^{-\sigma N \ell}. \tag{7.2}$$

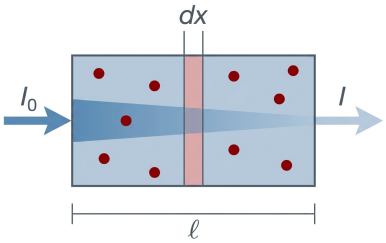


Figure 7.2: Each thin slice $\mathrm{d}x$ removes the same fraction $\sigma N\,\mathrm{d}x$ of the light that reaches it, so the intensity falls exponentially with depth.

Chemists write the same result with base-10 logarithms and molar concentration c:

$$A = -\log_{10} T = -\log_{10}\frac{I}{I_0} = \varepsilon c \ell. \tag{7.3}$$

Here $T = I/I_0$ is the *transmittance*, A the *absorbance*, and ε the *molar absorption coefficient* in $\mathrm{M}^{-1}\,\mathrm{cm}^{-1}$, which says how strongly the molecule absorbs at that wavelength; strongly colored dyes reach 10^4 to 10^5. The fraction absorbed is $1 - T = 1 - 10^{-A}$: an absorbance of 1 absorbs 90 % of the light, an absorbance of 2 absorbs 99 %.

KEY EQUATION

$A = -\log_{10}(I/I_0) = \varepsilon c \ell$

Beer-Lambert law; the fraction absorbed is $1 - 10^{-A}$.

Absorbances add. A solution of several species, each with its own spectrum, has

$$A(\lambda) = \ell \sum_i \varepsilon_i(\lambda)\, c_i, \tag{7.4}$$

so absorbances measured at as many wavelengths as there are species give a set of linear equations for the concentrations. The pulse oximeter (Section 7.7) is the two-species case; a CO-oximeter, which fits more wavelengths, also reports the carboxyhemoglobin of carbon monoxide poisoning from a blood sample.

When the law fails

Equation (7.3) assumes that each molecule absorbs independently and that monochromatic light is absorbed rather than scattered. Above about 10 mM molecules perturb each other; a source band wider than the absorption feature bends the calibration line; stray light that bypasses the sample sets a ceiling (with 0.1 % stray light a true absorbance of 2.5 reads 2.38, and nothing reads above 3.0); and turbid samples such as whole blood scatter light out of the beam, which looks like absorption. The best accuracy comes at absorbances between about 0.1 and 1: below 0.1 the signal is a small difference between two large intensities, and above 1 so little light gets through that noise and stray light dominate. Change the path length or dilute the sample to land in that window.

Color and its complement

A solution looks colored because it removes part of the visible spectrum from white light; the eye sees what is left. The color seen is therefore the *complement* of the color absorbed (Table 7.1). A solution that absorbs red light looks blue-green; one that absorbs blue looks yellow; one that absorbs green looks purple, which is why blood, whose hemoglobin absorbs blue and green strongly, looks red. Absorption in the ultraviolet, below 400 nm, does not affect the color at all.

band (nm)	absorbed	seen
400 to 435	violet	yellow-green
435 to 480	blue	yellow
480 to 490	green-blue	orange
490 to 500	blue-green	red
500 to 560	green	purple
560 to 580	yellow-green	violet
580 to 595	yellow	blue
595 to 650	orange	green-blue
650 to 750	red	blue-green

Table 7.1: Absorbed and observed colors. The color of a solution in white light is the complement of the band it absorbs most strongly. Boundaries between color names are approximate.

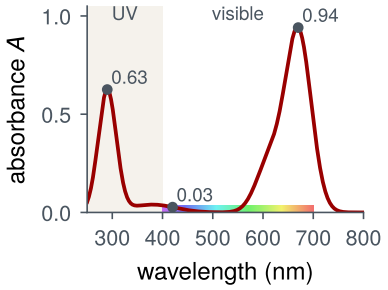


Figure 7.3: Absorption spectrum of the 10 µM dye of Example 7.2 in a 1 cm quartz cuvette (illustrative data). Dots mark the absorbances used in the example.

Worked Example 7.2 *A dye with peaks at 290 and 670 nm*

Figure 7.3 is the absorption spectrum of a 10 µM aqueous dye in a 1 cm cuvette. (a) What color does the eye see? (b) The absorbances are 0.63 at 290 nm, 0.03 at 420 nm, and 0.94 at 670 nm. What percentage of the light is absorbed at each wavelength, and what are the molar absorption coefficients at the two peaks?

Solution. (a) Ignore the 290 nm peak: it is in the ultraviolet and invisible. In the visible the dye removes red light near 670 nm and transmits violet, blue, and green almost completely. By Table 7.1, absorbing at 650 to 750 nm gives the complementary color blue-green: the dye looks blue to blue-green, like methylene blue, which absorbs near the same wavelength.

(b) The fraction absorbed is $1 - 10^{-A}$:

$$\begin{aligned} 290\,\text{nm}: &\quad 1 - 10^{-0.63} = 1 - 0.234 = 76.6\,\%, \\ 420\,\text{nm}: &\quad 1 - 10^{-0.03} = 1 - 0.933 = 6.7\,\%, \\ 670\,\text{nm}: &\quad 1 - 10^{-0.94} = 1 - 0.115 = 88.5\,\%. \end{aligned}$$

With $c\ell = (1 \times 10^{-5}\,\text{M})(1\,\text{cm})$, $\varepsilon_{290} = 0.63/10^{-5} = 6.3 \times 10^4\,\text{M}^{-1}\,\text{cm}^{-1}$ and $\varepsilon_{670} = 9.4 \times 10^4\,\text{M}^{-1}\,\text{cm}^{-1}$. The red peak, at $A = 0.94$, sits just inside the 0.1 to 1 window; a more concentrated sample would call for a shorter cuvette. ■

The instrument that records such a spectrum is the UV-visible spectrophotometer. A deuterium lamp (ultraviolet) and a tungsten-halogen lamp (visible) feed a grating monochromator, which selects a band a few nanometers wide, narrow enough for the Beer-Lambert law to hold. The cuvette is quartz, because glass and most plastics absorb below about 300 nm and would hide the 290 nm peak. A second beam crosses a cuvette of plain solvent, and the ratio of the two photocurrents gives the absorbance while canceling lamp drift and the absorption of the cell and the water. A fixed-wavelength version for one dye needs only an LED, a photodiode, and the transimpedance amplifier of Section 7.8.

7.3 *Fluorescence and the Jablonski Diagram*

When a molecule absorbs a photon, the photon disappears and its energy lifts an *electron* of the molecule into an excited electronic state. What happens next decides whether the molecule glows, heats up, or reacts, and it is summarized in a *Jablonski diagram* (Fig. 7.4). The thick lines are electronic states: the ground state S_0 and the excited states S_1 and S_2, in which the two electrons of a pair have opposite spins (*singlets*), and the *triplet* T_1, in which the excited electron's spin has flipped so that the two spins are parallel. The thin lines above each are vibrational levels.

The sequence runs as follows.

1. **Absorption** takes about 10^{-15} s and can land the electron in any vibrational level of S_1 or S_2.
2. **Relaxation.** Within about 10^{-12} s the molecule hands its extra vibrational energy to the solvent as heat and drops to the lowest level of S_1, whatever color was absorbed.
3. **Fluorescence** is the jump from S_1 back to S_0 with the emission of a photon. The two spins stay paired, so the transition is allowed and

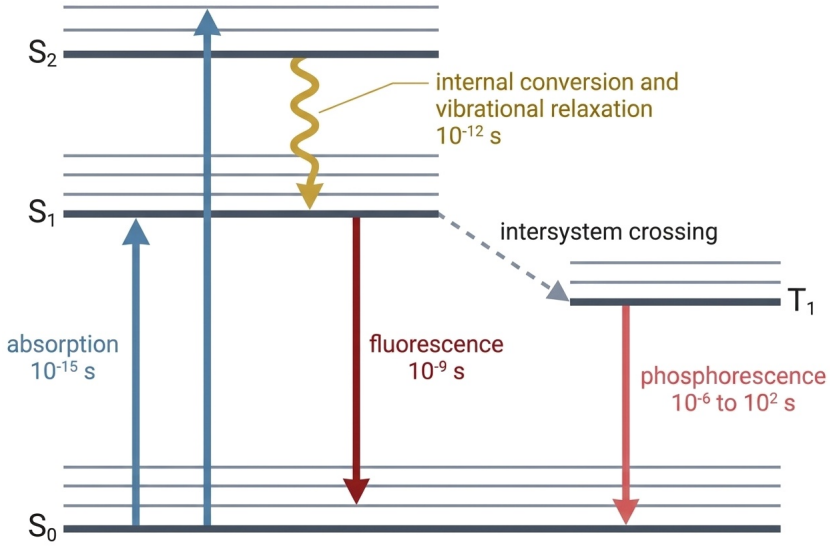


Figure 7.4: Jablonski diagram. Absorption (steel) lifts an electron into a vibrational level of S_1 or S_2. Internal conversion and vibrational relaxation (wavy) drain it to the bottom of S_1 within picoseconds. Fluorescence returns it to S_0 in nanoseconds. Intersystem crossing flips its spin into the triplet T_1, from which phosphorescence is slow because the triplet-to-singlet transition is spin forbidden. Typical time scales are marked.

fast: S_1 lifetimes are 1 to 10 ns.

4. **Intersystem crossing and phosphorescence.** Sometimes the excited electron flips its spin and moves into the triplet T_1, below S_1. To return to the singlet ground state it must flip its spin again, a spin-forbidden and therefore slow transition: the triplet lives for microseconds to seconds, and its emission, *phosphorescence*, persists long after the excitation stops.

Because energy is lost to heat before emission, the emitted photon has less energy, and therefore a longer wavelength, than the absorbed one. The gap between the absorption and emission maxima is the *Stokes shift* (Fig. 7.5). It is what makes fluorescence sensitive: an optical filter can block the excitation wavelength and pass only the emission, so the detector sees a bright signal on a dark background. Phosphorescence is shifted further still, because T_1 lies below S_1.

Lifetime and quantum yield

A molecule at the bottom of S_1 can leave by competing paths: a radiative one with rate constant k_r (fluorescence) and nonradiative ones with total rate k_{nr} (conversion to heat, intersystem crossing, and collisions). A population N of excited molecules decays as $\mathrm{d}N/\mathrm{d}t = -(k_r + k_{nr})N$, an exponential with the *fluorescence lifetime*

$$\tau = \frac{1}{k_r + k_{nr}}, \tag{7.5}$$

KEY EQUATION

$$\tau = \frac{1}{k_r + k_{nr}}, \quad \Phi = k_r \tau$$

Lifetime and quantum yield of an excited state.

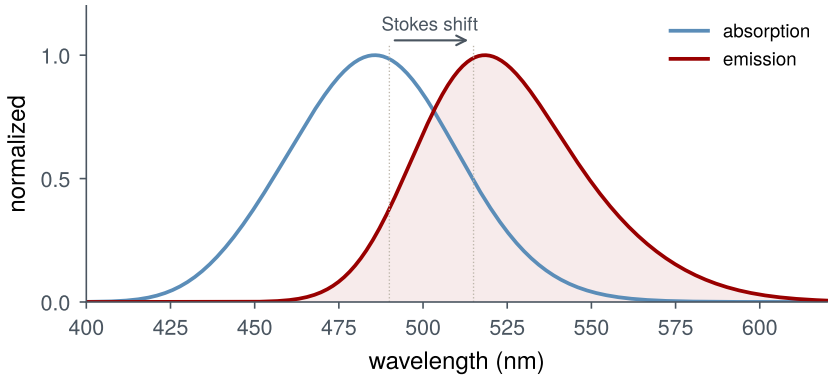


Figure 7.5: Absorption and emission spectra of a fluorescein-like dye (illustrative band shapes, peaks at 490 and 515 nm). The emission is a rough mirror image of the absorption, shifted to longer wavelength by the energy lost to vibrational relaxation.

and the *quantum yield*, the fraction of excited molecules that emit a photon, is

$$\Phi = \frac{k_r}{k_r + k_{nr}} = k_r\,\tau. \tag{7.6}$$

A bright dye such as fluorescein has $\tau = 4.0\,\text{ns}$ and $\Phi = 0.93$, so $k_r = \Phi/\tau = 2.3 \times 10^8\,\text{s}^{-1}$ and $k_{nr} = 1/\tau - k_r = 1.8 \times 10^7\,\text{s}^{-1}$: nearly every excited molecule emits. Its Stokes shift, from 490 to 515 nm, is $1240/490 - 1240/515 = 0.123\,\text{eV}$. Anything that adds a nonradiative path (a quencher, a nearby metal, a change in the dye's surroundings) shortens τ and lowers Φ in the same proportion. Many sensors measure exactly that change.

Quenching: sensing by lifetime

Some molecules switch off an excited dye by colliding with it. Molecular oxygen is an efficient collisional quencher. With a quencher at concentration $[Q]$, the nonradiative rate grows by $k_q[Q]$, and the ratio of unquenched to quenched intensity or lifetime follows the *Stern-Volmer* equation

$$\frac{I_0}{I} = \frac{\tau_0}{\tau} = 1 + K_{SV}[Q], \qquad K_{SV} = k_q\tau_0. \tag{7.7}$$

Measuring the lifetime instead of the intensity makes a sensor immune to LED drift, dye bleaching, and the amount of dye: those change the brightness but not τ. A simple way to measure a lifetime of nanoseconds to microseconds is to modulate the excitation sinusoidally at frequency f and measure the phase lag of the emission, $\tan\phi = 2\pi f\tau$. A second dye can quench too: in Förster resonance energy transfer (FRET), one readout for the affinity sensors of Chapter 6, an acceptor within a few nanometers takes the donor's energy without a photon.

7.4 The Fluorometer

A FLUOROMETER measures the light emitted by a sample while rejecting the much brighter light used to excite it. Two features do the rejection (Fig. 7.6). First, the emission is collected at 90° to the excitation beam, so the transmitted beam never enters the detector. Second, optical filters select the excitation band on one side and the Stokes-shifted emission band on the other. A solid-state fluorometer puts the same idea on a chip: an LED excites a dye layer that responds to the analyte, and a filter over a photodiode passes only the emission.

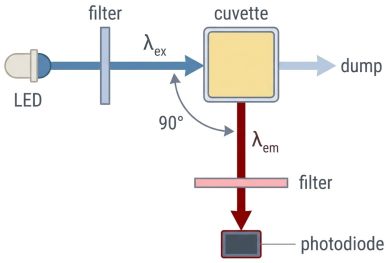


Figure 7.6: Right-angle fluorometer. The excitation filter passes the dye's absorption band and the emission filter its emission band; the detector looks at the cuvette (square) from the side, and the transmitted beam goes on to a beam dump.

For a dilute sample the emitted power is proportional to the absorbed power:

$$P_{\mathrm{em}} = \Phi\, P_0 \left(1 - 10^{-A_{\mathrm{ex}}}\right) \approx 2.303\, \Phi\, P_0\, \varepsilon c \ell, \tag{7.8}$$

where P_0 is the excitation power and A_{ex} the absorbance at the excitation wavelength; the approximation holds for A_{ex} below about 0.05. In this range the signal is linear in concentration and grows with the excitation power, which an absorption measurement cannot do (there, the signal is a small difference between two large numbers). That is why fluorescence assays reach nanomolar and lower concentrations.

At higher concentrations the response bends over, because the solution absorbs the excitation beam before it reaches the middle of the cuvette, where the detector looks, and reabsorbs part of the emission on its way out (the *inner-filter effect*). A sample with absorbance 0.10 at the excitation and 0.04 at the emission wavelength already reads about 15 % low; a tenfold dilution cuts the error to about 1.6 %. Keep the absorbance below about 0.1.

7.5 Total Internal Reflection Fluorescence

A BIOSENSOR WANTS TO SEE only the molecules bound to its surface, not the thousand times more that float in the sample above it. Total internal reflection lights up a layer about 100 nm thick and nothing else.

Snell's law and the critical angle

A ray that crosses from a medium of refractive index n_1 into one of index n_2 bends according to Snell's law,

$$n_1 \sin\theta_1 = n_2 \sin\theta_2, \tag{7.9}$$

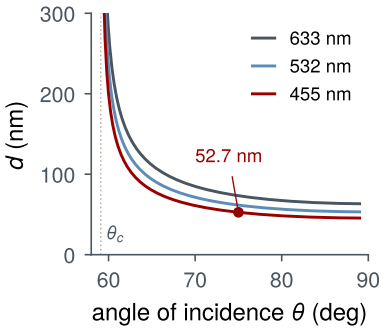


Figure 7.7: Intensity penetration depth, Eq. (7.11), from crystalline quartz ($n_1 = 1.55$) into an aqueous liquid ($n_2 = 1.33$) at three wavelengths. The dot is Example 7.3.

with angles measured from the normal. Going from glass ($n_1 = 1.5$) into water ($n_2 = 1.33$), the ray bends away from the normal. At the *critical angle*

$$\sin\theta_c = \frac{n_2}{n_1} \tag{7.10}$$

the refracted ray runs along the interface, and beyond it no light is transmitted: the beam is totally reflected back into the glass. For glass at $n_1 = 1.5$ against water at $n_2 = 1.33$, $\theta_c = \arcsin(0.887) = 62.5°$.

The evanescent wave

Total reflection does not mean that the light stops dead at the interface. The electromagnetic field must be continuous across the boundary, so a field leaks a short way into the second medium and decays exponentially with the distance z from the surface. This *evanescent wave* carries no power away from the interface, but it can excite a fluorophore that sits inside it. Its intensity falls as

$$I(z) = I_0\, e^{-z/d}, \qquad d = \frac{\lambda}{4\pi\sqrt{n_1^2 \sin^2\theta - n_2^2}}, \tag{7.11}$$

where λ is the vacuum wavelength. The depth d is a fraction of a wavelength, shrinks as the angle of incidence grows, and grows without limit as the angle approaches the critical angle (Fig. 7.7). (Some sources quote the depth of the field, twice as large.)

KEY EQUATION

$$d = \frac{\lambda}{4\pi\sqrt{n_1^2 \sin^2\theta - n_2^2}}$$

Intensity penetration depth of the evanescent wave (λ in vacuum).

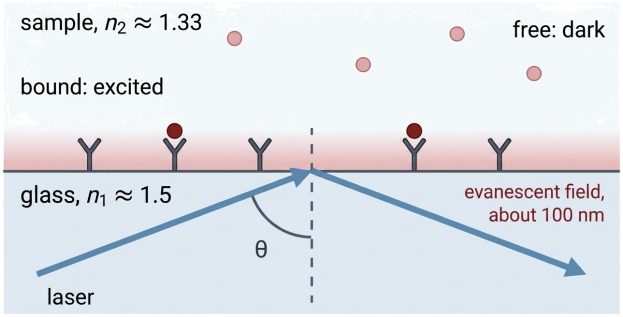


Figure 7.8: TIRF biosensor. A laser hits the glass-sample interface beyond the critical angle and is totally reflected. Its evanescent field reaches only about 100 nm into the sample, so the fluorescent labels held at the surface by the gray capture antibodies glow (dark cardinal dots), while labeled molecules in the bulk (pale dots) stay dark.

TIRF as a biosensor

Figure 7.8 shows the idea. Capture molecules (antibodies, aptamers, or single DNA strands) are fixed to the glass. When the analyte binds, it brings a fluorescent label into the evanescent field, carried either by the analyte itself or by a second labeled antibody that binds the captured analyte (a *sandwich assay*). Only labels within about 100 nm of the surface are excited, so the fluorescence measures what is bound, in real time, without washing away the unbound sample. TIRF microscopes

use the same trick to image only the part of a living cell that touches the glass.

> **WORKED EXAMPLE 7.3** *The penetration depth of a TIRF sensor*
>
> A 455 nm laser beam hits the interface between crystalline quartz ($n_1 = 1.55$; the fused silica of "quartz" slides and cuvettes has 1.46) and a liquid ($n_2 = 1.33$) at an angle of incidence of 75°. (a) Find the intensity penetration depth d. (b) What fraction of the surface intensity reaches a fluorophore bound 10 nm from the surface, and one diffusing 200 nm away?
>
> SOLUTION. (a) First check that the beam is totally reflected: $\theta_c = \arcsin(1.33/1.55) = 59.1° < 75°$. Then $n_1^2 \sin^2\theta - n_2^2 = (1.55)^2(0.9330) - (1.33)^2 = 0.4727$, and Eq. (7.11) gives
>
> $$d = \frac{455\,\text{nm}}{4\pi\sqrt{0.4727}} = 52.7\,\text{nm}.$$
>
> (b) $I(10\,\text{nm})/I_0 = e^{-10/52.7} = 0.83$ and $I(200\,\text{nm})/I_0 = e^{-200/52.7} = 0.022$. A bound label is excited 37 times more strongly than one 200 nm away, and labels microns away are not excited at all. ■

7.6 Surface Plasmon Resonance

SURFACE PLASMON RESONANCE (SPR) detects binding without any label. Chapter 6 sensed an antibody binding its target as a change in the charge-transfer resistance of an electrode; SPR senses it as a change in the refractive index within about 200 nm of a gold surface, which rises when proteins (index about 1.45) displace water (1.33).

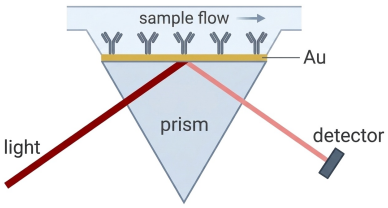


Figure 7.9: Kretschmann geometry. A thin gold film on the prism base carries capture molecules on its sample side. At the resonance angle the p-polarized reflection drops sharply.

Exciting a plasmon with light

The free electrons of a metal can oscillate together. At a metal surface the oscillation travels along the interface as a wave, a *surface plasmon*, whose field peaks at the metal and decays within a few hundred nanometers into the liquid. Light can hand its energy to the plasmon only if the two waves travel along the surface in step, with the same wavelength along the interface. Light arriving from the water side never matches, because the plasmon's wavelength along the surface is always shorter; the fix is to slow the light down in glass.

In the *Kretschmann* configuration (Fig. 7.9) a gold film about 50 nm thick coats the base of a glass prism of index n_p, and the sample flows over the gold. Light enters through the prism beyond the critical angle, and its evanescent field tunnels through the thin gold. Along the

surface the light behaves as if it had the index $n_p \sin\theta$, which grows with the angle; at one angle, the resonance angle θ_{sp}, it matches the plasmon:

$$n_p \sin\theta_{sp} = n_{sp}, \qquad n_{sp} = \mathrm{Re}\sqrt{\frac{\varepsilon_m n_s^2}{\varepsilon_m + n_s^2}}. \tag{7.12}$$

The plasmon's effective index n_{sp} depends on the sample index n_s and on the metal's relative permittivity ε_m, which for gold in the red is negative, about $-11.6 + 1.2i$ at 633 nm (the small imaginary part is its loss). For buffer ($n_s = 1.333$) this gives $n_{sp} = 1.447$, and a glass prism with $n_p = 1.515$ resonates at $\theta_{sp} = \arcsin(0.955) = 72.8°$. At that angle the light's energy flows into the plasmon and is absorbed in the metal, and the reflected intensity drops nearly to zero. Only p-polarized light, whose electric field lies in the plane of incidence and can push the electrons normal to the surface, couples; s-polarized light serves as a reference.

KEY EQUATION

$n_p \sin\theta_{sp} = n_{sp}$

SPR resonance: the light's index along the surface matches the plasmon's effective index.

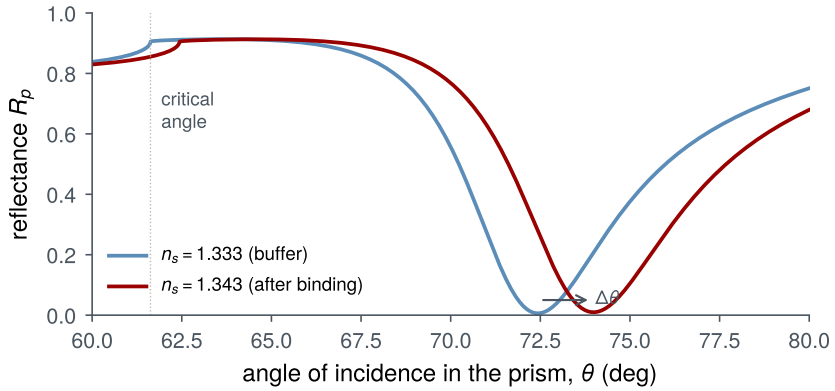


Figure 7.10: Computed p-polarized reflectance of a glass prism ($n_p = 1.515$) with a 50 nm gold film ($\varepsilon_m = -11.6 + 1.2i$, a commonly used value at 633 nm) in contact with buffer and with a sample of slightly higher index (three-layer Fresnel model). The dip moves from 72.4° to 74.0° when the index rises by 0.010.

Reading binding as an angle

When molecules bind and the index near the gold rises, n_{sp} grows and the dip moves to a larger angle (Fig. 7.10). The full optical calculation for this sensor puts the dip at 72.43° in buffer and at 72.58° for $n_s = 1.334$, a *sensitivity* of $0.15°/0.001 = 150°$ per refractive index unit (RIU). Instruments track the dip angle (or, at a fixed angle, the reflected intensity on the steep side of the dip) and resolve index changes of 10^{-6} to 10^{-7} RIU.[2] They report *response units*: 1000 RU is a dip shift of 0.1° and corresponds to about 1 ng/mm^2 of bound protein, so 1 RU is 6.7×10^{-7} RIU on this sensor, or about 1 pg/mm^2.

The plasmon field reaches about 180 nm into water, so SPR also responds to the bulk liquid just above the bound layer. The index of water changes by about 10^{-4} per kelvin, so a small temperature drift

[2] J. Homola, S. S. Yee, G. Gauglitz, *Sens. Actuators B* (1999), a standard review of SPR sensors.

looks like binding; instruments subtract a reference channel without capture molecules.

Binding curves

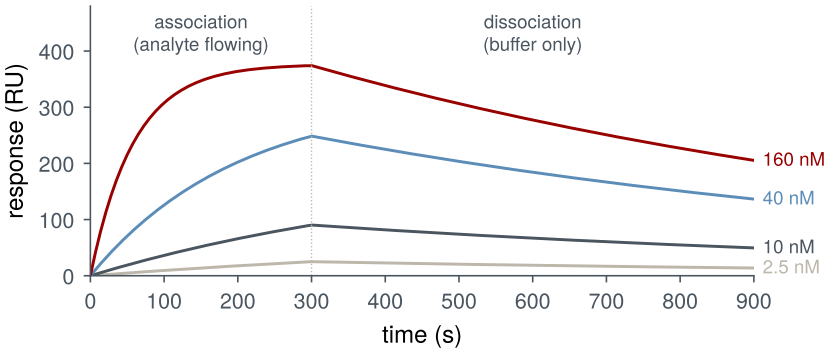


Figure 7.11: Sensorgrams for one-to-one binding with $k_a = 1 \times 10^5\,\mathrm{M^{-1}\,s^{-1}}$, $k_d = 1 \times 10^{-3}\,\mathrm{s^{-1}}$ ($K_D = 10\,\mathrm{nM}$) and $R_{\max} = 400\,\mathrm{RU}$ (illustrative values). Analyte flows for 300 s, then buffer. Higher concentrations rise faster and higher; all curves fall with the same rate k_d.

SPR is valued because it follows binding as it happens. A *sensorgram* (Fig. 7.11) plots the response R against time while the analyte flows over the capture surface and then while buffer washes it off. The response follows the number of occupied sites: with analyte at concentration C it rises as $R(t) = R_{\mathrm{eq}}\left(1 - e^{-k_{\mathrm{obs}}t}\right)$, with

$$k_{\mathrm{obs}} = k_a C + k_d, \qquad R_{\mathrm{eq}} = R_{\max}\frac{C}{C + K_D}, \qquad K_D = \frac{k_d}{k_a}, \tag{7.13}$$

where k_a and k_d are the association and dissociation rate constants, $R_{\max}$ is the response with every site occupied, and K_D is the dissociation constant of the binding isotherm of Chapter 6. In the wash the response decays as $e^{-k_d t}$. For the curves of Fig. 7.11, $K_D = 10\,\mathrm{nM}$; at $C = 40\,\mathrm{nM}$ the response heads for $R_{\mathrm{eq}} = 320\,\mathrm{RU}$ with a time constant $1/k_{\mathrm{obs}} = 200\,\mathrm{s}$, and in the wash the complex has a half-life of $\ln 2/k_d = 693\,\mathrm{s}$, about 12 min. Because k_{obs} grows linearly with C, runs at several concentrations give k_a and k_d separately. As for any affinity sensor (Chapter 4), the isotherm also sets the useful range: the fraction of occupied sites goes from 0.1 to 0.9 as C goes from $K_D/9$ to $9K_D$, about two decades.

7.7 *Pulse Oximetry*

The pulse oximeter is the most widely used optical biosensor: a Beer-Lambert measurement on blood that never leaves the body. Chapter 1 introduced its signal, the photoplethysmogram (PPG). Light that crosses a fingertip is partly absorbed by blood, and each heartbeat

adds a little arterial blood and dims the light a little. The detected light has a large, steady *DC* part from tissue, bone, and venous blood, and a small *AC* part, typically 0.1 to 2 % of the DC, from the pulsing arterial blood. This section turns that signal into oxygen saturation.

Two absorbers, two wavelengths

Hemoglobin carries oxygen, and its oxygenated (HbO_2) and deoxygenated (Hb) forms have different spectra (Fig. 7.12). In the red, near 660 nm, Hb absorbs about ten times more strongly than HbO_2; in the near infrared, beyond the crossing point near 800 nm, HbO_2 absorbs more. The arterial oxygen saturation is the oxygenated fraction, $S = C_{HbO_2}/(C_{HbO_2} + C_{Hb})$, normally 95 % to 100 %. An oximeter estimates it from the PPG at one red and one infrared wavelength.

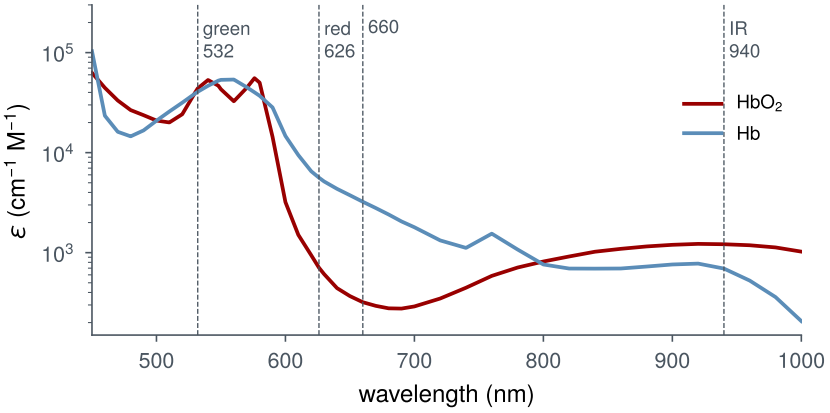


Figure 7.12: Molar extinction coefficients of oxyhemoglobin and deoxyhemoglobin from 450 to 1000 nm, from S. Prahl's compilation (Oregon Medical Laser Center), on a log scale. Dashed lines mark the green and red OLEDs of the organic oximeter (532 and 626 nm) and the 660 and 940 nm LEDs of most clinical probes.

The ratio of ratios

During a heartbeat the arterial blood adds a small absorbance ΔA_λ at each wavelength. Because it is small, $I = I_{DC}\,10^{-\Delta A} \approx I_{DC}(1 - 2.303\,\Delta A)$, so the fractional dip is $AC/DC \approx 2.303\,\Delta A_\lambda$; a pulse of 1 % is an absorbance of only 0.0043. In tissue, light does not travel in a straight line: it scatters many times, and its mean path through the pulsing vessels, L_λ, is longer than the finger is thick and different at each wavelength. The *modified* Beer-Lambert law writes the pulsatile absorbance with this path, $\Delta A_\lambda = [\varepsilon_{HbO_2}(\lambda)\,S + \varepsilon_{Hb}(\lambda)(1-S)]\,C_{tot}\,L_\lambda$, where C_{tot} is the total hemoglobin concentration. Dividing the red result by the infrared one gives the *ratio of ratios*

$$R = \frac{(AC/DC)_{red}}{(AC/DC)_{IR}} = \frac{\varepsilon_{HbO_2,r}\,S + \varepsilon_{Hb,r}(1-S)}{\varepsilon_{HbO_2,ir}\,S + \varepsilon_{Hb,ir}(1-S)} \cdot \frac{L_r}{L_{ir}}. \qquad (7.14)$$

KEY EQUATION

$$R = \frac{(AC/DC)_{red}}{(AC/DC)_{IR}}$$

The ratio of ratios. LED brightness, detector gain, and C_{tot} cancel.

Dividing AC by DC cancels the LED power, the detector gain, and the

static tissue; taking the ratio cancels the finger thickness and C_{tot}. Only the path-length ratio L_r/L_{ir} remains, and it depends on the tissue, the probe, and the person. Solving Eq. (7.14) for S with $R' = R\,L_{ir}/L_r$ gives

$$S = \frac{\varepsilon_{\text{Hb},r} - R'\,\varepsilon_{\text{Hb},ir}}{(\varepsilon_{\text{Hb},r} - \varepsilon_{\text{HbO}_2,r}) + R'\,(\varepsilon_{\text{HbO}_2,ir} - \varepsilon_{\text{Hb},ir})}. \tag{7.15}$$

At 660 and 940 nm, R is about 0.5 in healthy blood and rises as saturation falls.

Calibration and accuracy

Because L_r/L_{ir} is not known, no oximeter relies on Eq. (7.15) alone. Manufacturers calibrate empirically: healthy volunteers breathe gas mixtures with less oxygen, arterial blood is drawn at several plateaus between about 70 and 100 % saturation and analyzed by a CO-oximeter (giving SaO_2), and a curve $\text{SpO}_2 = f(R)$ is fit to the paired data. A straight line such as $\text{SpO}_2 = 110 - 25R$, used in teaching labs, is the simplest version (Fig. 7.13). Accuracy is quoted as the root-mean-square difference between SpO_2 and SaO_2, A_{rms}, which combines the mean bias b and the scatter s about it, $A_{\text{rms}} \approx \sqrt{b^2 + s^2}$. FDA's 2013 guidance for pulse oximeters recommends $A_{\text{rms}} \le 3.0\,\%$ for transmission sensors and $\le 3.5\,\%$ for reflectance sensors over 70 to 100 %. So a reading of 92 % means roughly 89 to 95 %, which matters near the thresholds at which clinicians give oxygen.

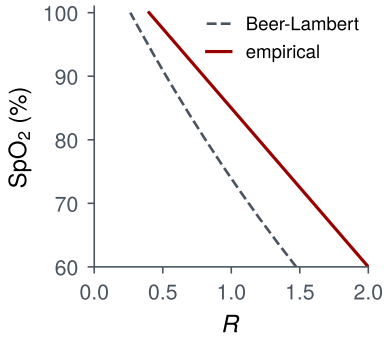


Figure 7.13: Calibration curves. Dashed: Eq. (7.15) at 660 and 940 nm with $L_r = L_{ir}$. Cardinal: the generic teaching-lab line $110 - 25R$. The difference comes mostly from the unequal path lengths.

Worked Example 7.4 *SpO_2 from a recording*

A finger probe at 660 and 940 nm records $(\text{AC}/\text{DC})_{\text{red}} = 0.70\,\%$ and $(\text{AC}/\text{DC})_{\text{IR}} = 1.40\,\%$. (a) Find R and SpO_2 from the line $110 - 25R$. (b) Find S from Eq. (7.15) with $L_r = L_{ir}$, using $\varepsilon_{\text{HbO}_2} = 320$ and $\varepsilon_{\text{Hb}} = 3227$ at 660 nm, and $\varepsilon_{\text{HbO}_2} = 1214$ and $\varepsilon_{\text{Hb}} = 693$ at 940 nm ($\text{cm}^{-1}\,\text{M}^{-1}$). (c) Why do the answers differ?

Solution. (a) $R = 0.70/1.40 = 0.50$ and $\text{SpO}_2 = 110 - 25(0.50) = 97.5\,\%$.

(b) With $R' = R = 0.50$,

$$S = \frac{3227 - 0.50(693)}{(3227 - 320) + 0.50(1214 - 693)} = \frac{2880}{3167} = 0.909.$$

(c) The pure Beer-Lambert model with equal paths says 90.9 %, the calibration line 97.5 %. The model assumes that red and infrared light sample the pulsing blood over the same path. They do not: the measured R is what the model would give at 97.5 % if the red path were about 1.5 times the infrared path (Problem 7.4). The empirical calibration absorbs that ratio for the average person, which is also why an oximeter is only as good as the people it was calibrated on.

■

Transmission, reflectance, motion, and perfusion

In *transmission* mode the LEDs and photodiode sit on opposite sides of a fingertip or earlobe. In *reflectance* mode they sit side by side, a few millimeters apart, and detect light that scatters back out of the tissue; they work on the forehead, wrist, and chest, which makes wearables possible, but their pulsatile signal is weaker. The layout matters: in printed reflectance sensors of red and near-infrared OLEDs and an organic photodiode on the underside of the wrist, the pulsatile signal decayed exponentially as the emitter-detector spacing grew from 2 to 6 mm, and OLEDs that wrapped around the photodiode collected 40 to 50 % more red signal than OLEDs on two of its sides.[3]

[3] Y. Khan, D. Han, J. Ting, M. Ahmed, R. Nagisetty, A. C. Arias, *IEEE Access* (2019).

Three effects limit every oximeter. With *low perfusion*, as in a cold finger, the AC/DC ratio falls well below a percent and the noise in R grows. With *motion*, venous blood and the coupling of the probe to the skin change, and both wavelengths see nearly the same artifact, which pushes R toward 1; with the line $110 - 25R$ that is 85 %, the value toward which noisy oximeters drift. *Other hemoglobins* fool the two-wavelength model: carboxyhemoglobin from carbon monoxide absorbs like HbO_2 in the red, so the oximeter reads falsely high in smoke inhalation. Only a multiwavelength CO-oximeter (Eq. (7.4)) separates them.

Skin pigmentation and bias

Melanin in the outer skin absorbs strongly at short wavelengths and much less in the near infrared. Because it does not pulse, it should cancel in the AC/DC ratios. In practice it does not cancel completely: it reduces the red signal more than the infrared one, changes the effective path lengths, and lowers the signal-to-noise ratio, and the volunteers in many calibration studies had lighter skin. The result is a measurable bias. In 10,789 paired hospital measurements (Sjoding *et al.*, *N. Engl. J. Med.*, 2020), among the pairs with SpO_2 between 92 and 96 %, the arterial saturation was below 88 % ("occult hypoxemia") in 11.7 % of the measurements in Black patients and 3.6 % in White patients. An oximeter that reads two or three points high in one group delays oxygen therapy for that group. FDA issued a safety communication on the problem in 2021, and its draft guidance of January 2025 for clinical testing asks for more participants spread across skin tones and limits how much the bias may differ between skin-tone groups (check its current status before you design a study). Chapter 14 returns to the regulatory and ethical side.

Organic and printed oximeters

Clinical oximeters use inorganic LEDs and a silicon photodiode in a rigid clip. Organic semiconductors (Chapter 11) can be printed over large, flexible areas, which suggests sensors that conform to skin and map oxygenation instead of measuring one point. The first all-organic oximeter used solution-processed green (532 nm) and red (626 nm) organic LEDs (OLEDs) and a printed organic photodiode, in transmission through a finger.[4] Green replaced the usual near infrared because efficient, air-stable near-infrared OLEDs were not available, and it works as the reference channel: the two hemoglobin coefficients at 532 nm differ by less than 10 % (Fig. 7.12), while the red absorption changes eightfold between deoxygenated and oxygenated blood. The sensor measured pulse rate within 1 % and oxygenation within 2 % of a commercial oximeter. A later printed array of red and near-infrared OLEDs and organic photodiodes worked in reflectance mode.[5] On the forehead it measured SpO_2 with a mean error of 1.1 %, and on a forearm whose blood flow was controlled by a pressure cuff it mapped the change in oxygenation in two dimensions, about -9 % during ischemia and $+8$ % just after the cuff was released. Chapter 12 covers the printing methods.

[4] C. M. Lochner, Y. Khan, A. Pierre, A. C. Arias, *Nat. Commun.* (2014).

[5] Y. Khan, D. Han, A. Pierre, J. Ting, X. Wang, C. M. Lochner, *et al.*, A. C. Arias, *Proc. Natl. Acad. Sci. USA* (2018).

7.8 Light Sources, Photodiodes, and the Front End

Every optical sensor in this chapter has the same chain (Fig. 7.14): a source driven by a timing circuit, an optical path through the sample, a detector, a transimpedance amplifier, an analog-to-digital converter (ADC), and processing that removes ambient light and computes the result. Designing one means choosing each block and checking, with a photocurrent budget, that enough signal reaches the ADC above the noise.

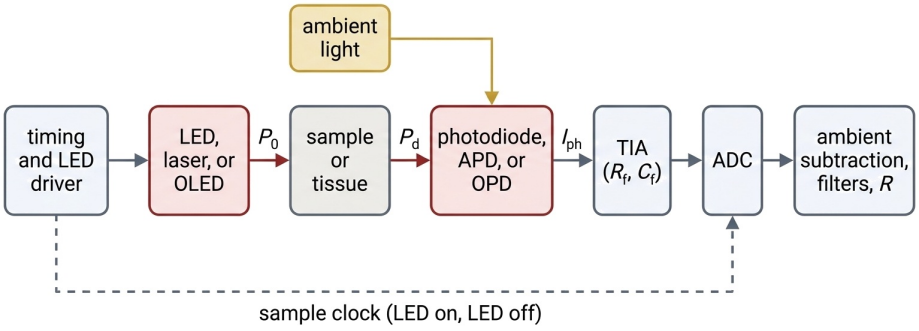


Figure 7.14: The optical sensing chain. The same timing that pulses the source tells the ADC when to sample with the source on and off, so that ambient light, which reaches the detector in both samples, can be subtracted.

Light sources

A *light-emitting diode* (LED) is a *pn junction*, where p-type and n-type semiconductor (Section 5.7) meet. When a current flows forward through it, electrons from the n side and holes from the p side meet at the junction and give up their energy as photons of about the material's *band gap* E_g, the least energy that frees an electron to carry current, so the material sets the color. LEDs are the default for wearables: cheap, efficient, small, easy to pulse in microseconds, with spectral widths of 20 to 50 nm and a wide beam that suits diffuse tissue. *Lasers* give a narrow line and a collimated, intense beam; TIRF and Raman usually need them. *OLEDs* emit from an area and can be printed on plastic for conformal sensors and arrays, but they are dimmer and less stable, especially in the near infrared.

The photodiode

A photodiode runs an LED backward. A photon with more energy than the band gap E_g can give an electron that energy, creating a free electron and a *hole*, the vacancy it leaves, which moves like a positive charge. In a *photodiode* this happens in or near a pn junction, where a built-in electric field pulls the electron to one side and the hole to the other, so each absorbed photon pushes about one electron around the external circuit. The photocurrent is therefore proportional to the optical power P_d on the detector, $I_{ph} = \mathcal{R} P_d$, with the *responsivity*

$$\mathcal{R} = \eta \frac{q\lambda}{hc} = \eta \frac{\lambda\,[\text{nm}]}{1240}\ \text{A/W}, \tag{7.16}$$

where η, the *external quantum efficiency*, is the fraction of incident photons that produce a collected electron (Fig. 7.15). The responsivity grows with wavelength because a watt of red light holds more photons than a watt of blue. Silicon ($E_g = 1.12$ eV) responds from about 300 to 1100 nm and peaks near 0.6 A/W around 900 nm; photons beyond about 1100 nm carry too little energy and pass through. A larger area collects more light but adds capacitance, which slows the amplifier.

Three other detectors appear in biosensors. An *avalanche photodiode* (APD) has an internal gain of tens to hundreds at a high reverse bias, at the cost of extra noise. A *photomultiplier tube* (PMT) multiplies each photoelectron by about 10^6 in a vacuum tube and can count single photons; it suits weak fluorescence on the bench but not a wearable. *Organic photodiodes* (OPDs) can be printed in any shape on flexible plastic, with quantum efficiencies of a few tens of percent and more dark current than silicon.

KEY EQUATION

$\mathcal{R} = \eta\lambda/1240$ A/W

Photodiode responsivity, λ in nm, η the external quantum efficiency.

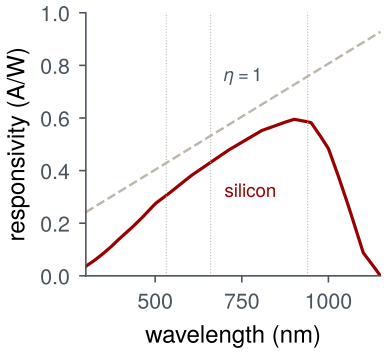


Figure 7.15: Responsivity of an ideal detector ($\eta = 1$) and of a typical silicon photodiode (illustrative curve). Dotted lines: 532, 660, and 940 nm.

The transimpedance front end

The photocurrent goes into the transimpedance amplifier (TIA) of Chapter 4 (Fig. 7.16). The op amp holds its inverting input at virtual ground, so the photodiode sits at zero bias and its whole current flows through the feedback resistor: $v_o = I_{ph} R_f$, positive for the diode drawn here, which pulls the photocurrent out of the summing node. A capacitor C_f across R_f sets the bandwidth, $f_c = 1/(2\pi R_f C_f)$, and keeps the stage stable with the photodiode's capacitance at its input.

Two noise sources from Chapter 4 set the floor: the *shot noise* of the total detector current I (signal, ambient light, and dark current), $i_{\text{sh}} = \sqrt{2qIB}$ in a noise bandwidth B, and the *thermal noise* of the feedback resistor, $i_R = \sqrt{4kTB/R_f}$. A large R_f gives more volts per amp and less noise, but the output must not saturate on the largest DC current, which is usually the ambient light.

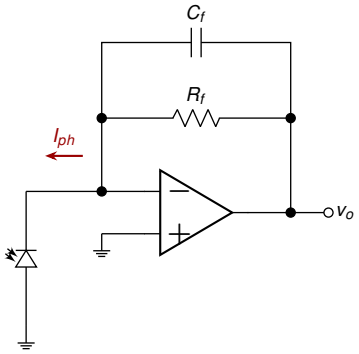


Figure 7.16: Photodiode transimpedance amplifier. The virtual ground holds the photodiode at zero bias, and the photocurrent flows through R_f, so $v_o = I_{ph} R_f$; C_f sets the bandwidth.

Worked Example 7.5 *A photocurrent budget for a finger PPG*

A 660 nm LED emits 5.0 mW into a finger, and 0.050 % of it reaches a silicon photodiode with $\eta = 0.80$. The pulsatile part is 1.0 % of the DC light. (a) Find the responsivity, the DC photocurrent, and the AC photocurrent. (b) Choose R_f for a DC output near 1 V. (c) Estimate the signal-to-noise ratio of the AC signal in a noise bandwidth of 10 Hz, counting shot noise and the thermal noise of R_f at 300 K.

Solution. (a) $\mathcal{R} = 0.80 \times 660/1240 = 0.426\,\text{A/W}$. The power at the detector is $P_d = (5.0\,\text{mW})(5.0 \times 10^{-4}) = 2.5\,\mu\text{W}$, so $I_{DC} = (0.426)(2.5\,\mu\text{W}) = 1.06\,\mu\text{A}$, and the AC part is 10.6 nA peak to peak.

(b) $R_f = 1\,\text{V}/1.06\,\mu\text{A} = 940\,\text{k}\Omega$; a standard 1 MΩ gives 1.06 V DC and 10.6 mV peak to peak of pulse.

(c) The shot noise and the resistor noise are

$$i_{\text{sh}} = \sqrt{2(1.602 \times 10^{-19}\,\text{C})(1.06 \times 10^{-6}\,\text{A})(10\,\text{Hz})} = 1.85\,\text{pA},$$

$$i_R = \sqrt{4(1.381 \times 10^{-23}\,\text{J/K})(300\,\text{K})(10\,\text{Hz})/1\,\text{M}\Omega} = 0.41\,\text{pA},$$

both rms, and together $\sqrt{1.85^2 + 0.41^2} = 1.89\,\text{pA}$. A sinusoidal pulse of 10.6 nA peak to peak has an rms value of $10.6/(2\sqrt{2}) = 3.75\,\text{nA}$, so SNR $= 3750/1.89 = 1980$, or 66 dB. Motion and ambient light, not these noises, limit a real PPG, but a microwatt of light is plenty, so the LED can be pulsed at a low duty cycle to save power (Chapter 13). ■

Rejecting ambient light

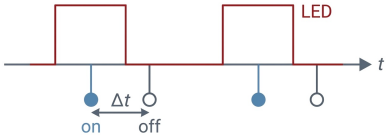


Figure 7.17: Ambient subtraction. The ADC samples once with the LED on and once with it off; the difference keeps the LED signal and cancels slowly varying ambient light.

Room light and sunlight reach the detector, often with more power than the LED signal, and room lights flicker at twice the line frequency, 120 Hz in the United States. An opaque housing, firm skin contact, and a filter that passes only the LED band help first.

Electrically, the LED is pulsed and the ADC takes one sample with the LED on and one with it off, a short time Δt later (Fig. 7.17). The difference removes everything that did not change in Δt. A component at frequency f survives the subtraction with relative amplitude $2\sin(\pi f \Delta t)$, small when $\Delta t \ll 1/f$. With samples 100 µs apart, 120 Hz flicker shrinks to $2\sin(\pi \times 120 \times 10^{-4}) = 0.075$ of its size, 22.5 dB less, and a digital low-pass filter removes the rest; with samples 1 ms apart it shrinks only to 0.74.

Pulsing also saves power. An LED driven at 20 mA but lit for only 100 µs of every 10 ms (a 100 Hz sample rate) has a duty cycle of 0.01 and draws 0.20 mA on average. *Lock-in detection* is the continuous version of the same idea: modulate the LED at a few kilohertz and demodulate at that frequency, so that only light carrying the modulation is kept. Integrated optical front ends do both on one chip.

Summary

- $E = h\nu = hc/\lambda$, $E\,[\text{eV}] = 1240/\lambda\,[\text{nm}]$; visible light is 1.8 to 3.1 eV. Electronic transitions absorb in the visible and UV, vibrations in the infrared.
- Beer-Lambert: $A = -\log_{10} T = \varepsilon c \ell$, absorbed fraction $1 - 10^{-A}$; absorbances add, so n wavelengths resolve n species. Work at A of 0.1 to 1. A solution shows the complement of the color it absorbs.
- Jablonski: absorption, relaxation to the bottom of S_1, then fluorescence (1 to 10 ns, spin allowed) or intersystem crossing and phosphorescence (µs to s, spin forbidden). $\tau = 1/(k_r + k_{nr})$, $\Phi = k_r \tau$, $\tau_0/\tau = 1 + K_{SV}[Q]$, $\tan\phi = 2\pi f \tau$. A fluorometer looks at 90° through an emission filter.
- TIRF: beyond $\sin\theta_c = n_2/n_1$ the evanescent intensity falls as $e^{-z/d}$ with d of about 50 to 100 nm: only bound labels fluoresce.
- SPR: $n_p \sin\theta_{sp} = n_{sp}$; the dip moves about 150°/RIU; 1000 RU $\approx$ 0.1° $\approx$ 1 ng/mm^2.
- Oximetry: $R = (\text{AC/DC})_r/(\text{AC/DC})_{ir}$ depends on S and on L_r/L_{ir}, so calibration is empirical ($A_{\text{rms}} \leq 3$ to 3.5 %) and biased in darker skin.
- Front end: $\mathcal{R} = \eta\lambda/1240$ A/W, $v_o = I_{ph} R_f$; noise $\sqrt{2qIB}$ and

$\sqrt{4kTB/R_f}$; pulse the source and subtract on and off samples taken $\Delta t \ll 1/f$ apart.

The sensors of Chapters 5 to 7 drive a transducer with a current, a voltage, or light and read its response. Chapter 8 records the voltages the body makes itself, the ECG, EEG, and EMG, with the same op amps and a few new circuits.

7.9 *Problems*

7.1 ◦ Photon energies. (a) Find the photon energy in eV for 405, 850, and 1550 nm. (b) Which of these can a silicon photodiode (band gap 1.12 eV) detect? (c) How many 850 nm photons per second are in a 2.0 mW beam?

7.2 • A 1 cm cuvette holds two dyes X and Y. Their coefficients are $\varepsilon_X = 1.2 \times 10^4$ and $\varepsilon_Y = 2.0 \times 10^3$ at 500 nm, and $\varepsilon_X = 3.0 \times 10^3$ and $\varepsilon_Y = 1.8 \times 10^4$ at 600 nm ($\mathrm{M^{-1}\,cm^{-1}}$). The mixture reads $A = 0.360$ at 500 nm and 0.615 at 600 nm. Find both concentrations.

7.3 • A dye has $\tau_0 = 10$ ns and $\Phi_0 = 0.60$. (a) Find k_r and k_{nr}. (b) In the presence of 10 mM of a quencher, the lifetime is 5.0 ns. Find the new quantum yield, the quenching rate constant k_q, and K_{SV}.

7.4 • For the recording of Example 7.4 ($R = 0.50$, 660 and 940 nm):

(a) What value of R' in Eq. (7.15) corresponds to $S = 0.975$?

(b) What path-length ratio L_r/L_{ir} reconciles the Beer-Lambert model with the calibration line?

(c) A second person with the same true saturation has $L_r/L_{ir} = 1.3$. What R is measured, and what does the calibration line report?

7.5 •• *(Design.)* A wrist PPG uses a 530 nm LED emitting 8.0 mW, and 0.010 % of the light reaches a silicon photodiode ($\mathcal{R} = 0.30$ A/W at 530 nm). The pulsatile part is 1.0 % of the DC. In daylight, ambient light adds up to 5.0 µA of photocurrent. (a) Find the DC and AC photocurrents from the LED. (b) Choose R_f so that the TIA output stays below 2.5 V in daylight. (c) Find the shot-noise-limited SNR of the AC signal in a noise bandwidth of 10 Hz with and without the ambient light. What does the ambient light cost?

8
Biopotential Recording

In this chapter you will learn to

- Explain how the heart, brain, muscles, and eyes produce the ECG, EEG, EMG, and EOG at the skin, and give their typical amplitudes and bandwidths.
- Compute the ECG limb leads from electrode potentials, name the P, QRS, and T waves, and find the heart rate from RR intervals.
- Model an electrode on the skin with a half-cell potential and an RC network, and compare wet and dry electrodes.
- Explain how mains wiring puts a common-mode voltage on the body and how electrode mismatch turns part of it into a differential signal.
- Find the gain and CMRR of a three-op-amp instrumentation amplifier and the limit that electrode offset puts on its first stage.
- Calculate how much a right-leg drive lowers the common-mode voltage.
- Choose high-pass, low-pass, and notch filters for ECG and EEG.
- Describe what a delta-sigma converter, an integrated front end, SPI, and Bluetooth Low Energy do, and budget the data rate of a recording.
- Apply the leakage-current limits and the rules that keep a person wired to electrodes safe.

Press three gel electrodes on your chest and snap them to a board the size of a postage stamp. A few seconds later a laptop

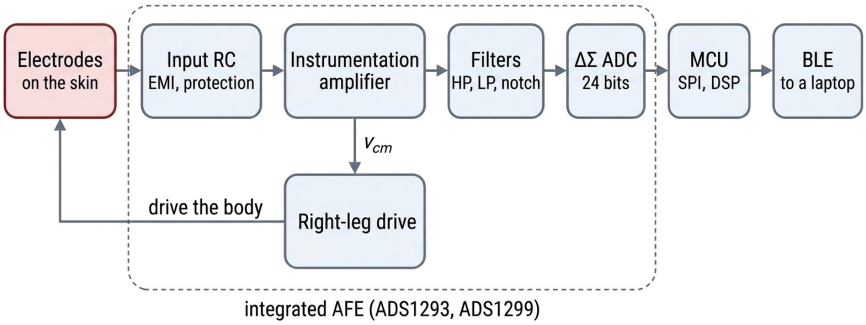


Figure 8.1: The biopotential recording chain. Electrodes turn ionic currents into electron currents; the analog front end (dashed) amplifies the difference between electrodes, filters it, and digitizes it, while a right-leg drive holds the body near the amplifier's reference. A microcontroller reads the samples over SPI and sends them out over Bluetooth Low Energy. The electrodes are treated in Sections 8.1 and 8.2, the instrumentation amplifier in Sections 8.3 and 8.4, the right-leg drive in Section 8.5, the filters in Section 8.6, the converter in Section 8.7, and the microcontroller and the radio in Sections 8.8 and 8.9.

across the room draws your heartbeat: a small bump, a tall spike, a rounded hump, over and over. The spike is about one millivolt at the skin. The same skin also carries a volt or more of 60 Hz hum from the building's wiring, a DC offset of tens to hundreds of millivolts from the electrodes, and the electrical noise of every muscle you tense. Recording a biopotential means keeping that millivolt and rejecting everything else.

This chapter follows the signal from its source to the screen (Fig. 8.1). We start with where biopotentials come from: the heart's electric dipole and the leads that view it, the rhythms of the brain, and the potentials of muscles and eyes. Then we cross the electrode-skin interface, see how mains interference reaches the body, and build the amplifier that rejects it: the instrumentation amplifier, helped by a right-leg drive. Filters and a delta-sigma converter come next; in practice all of these sit on one chip, an integrated analog front end (AFE). The data leave the AFE over SPI and leave the body over Bluetooth Low Energy (BLE), and a few lines of Python turn them into a heart rate or an EEG band power. The chapter closes with electrical safety, which sets the rules for every measurement on a person. It uses the op amps of Chapter 4 and the electrode interface of Chapter 6, and it closes Part II of the book.

Two labs in Part V carry the design to hardware. In Lab 3 a team designs the AFE daughter board that an Arduino Nicla Sense ME plugs onto, and in Lab 4 it writes the firmware that configures the AFE over SPI and streams samples to a laptop. The register settings, the SPI frames, and the decoding of the raw bytes are in those labs; this chapter gives the ideas behind them. Chapter 14 puts the parts together into a design project.

8.1 Where Biopotentials Come From

EVERY EXCITABLE CELL holds its inside about 70 to 90 mV below its outside. Chapter 3 explains why: potassium leaks out through open channels until the field it leaves behind balances its concentration gradient. When a cell fires, sodium (or calcium) channels open and the inside swings positive, for about a millisecond in a nerve and for a few hundred milliseconds in heart muscle. The ions that cross the membrane close their circuit through the salty fluid around the cells. That fluid conducts like a dilute salt solution, so the tissue of the body is a *volume conductor*: the currents of millions of cells that fire together spread through it and set up small potential differences at the skin.

KEY IDEA A biopotential at the skin is the sum of many membrane currents spreading through the body's conducting fluid. Size and synchrony set its amplitude: the heart gives millivolts, the cortex tens of microvolts.

What reaches the skin is much smaller than the membrane potential. A 100 mV cardiac action potential becomes about 1 mV between two chest electrodes, and the synchronized activity of the cortex becomes tens of microvolts on the scalp, after the skull has smeared it out. Table 8.1 lists the four biopotentials of this chapter.

Signal	Source	Amplitude at the skin	Bandwidth	Electrodes
ECG	heart	0.5 to 4 mV	0.05 to 150 Hz; 0.5 to 40 Hz for monitors	chest, limbs
EEG	cortex	5 to 300 µV	0.5 to 40 Hz; gamma to 100 Hz	scalp, ear
EMG	skeletal muscle	10 µV to 5 mV	20 to 500 Hz	over the muscle
EOG	eye	50 µV to 3.5 mV	DC to 50 Hz	beside the eyes

Table 8.1: The biopotentials of this chapter. The ECG is the largest and the EEG the smallest; the EMG is the fastest and the EOG reaches down to DC.

The table gives the first numbers of any design. To bring a signal to about 1 V at a converter, a 1 mV R wave of the ECG needs a voltage gain near 1000 and a 50 µV alpha wave of the EEG a gain near 20 000. To keep the band, the sampling rate must be more than twice the highest frequency (the Nyquist criterion of Chapter 1): more than 300 samples per second for a diagnostic ECG, 1000 for surface EMG, and 80 for a 40 Hz EEG, for which 250 is common. The EEG is the hardest of the four: twenty times smaller than the ECG and about the size of the noise of a poor electrode.

The heart as an electric dipole

The heart has two small upper chambers, the *atria*, and two large lower chambers, the *ventricles*, which do most of the pumping. A heartbeat starts in the *sinoatrial (SA) node*, a patch of pacemaker cells in the right atrium. A wave of depolarization (the inside of each cell swinging positive) spreads across both atria, pauses for about 0.1 s in the *atri-*

oventricular (AV) node, then races down special conducting fibers and spreads through the ventricles. At each instant the boundary between depolarized and resting muscle carries current. Seen from a distance, the whole heart looks like one *current dipole*, a source and a sink of current side by side, whose size and direction change during the beat: the *cardiac vector* $\vec{M}(t)$.

A pair of electrodes reads the component of $\vec{M}$ along the line that joins them. For electrodes far from the heart,

$$v_{\text{lead}} = \vec{a} \cdot \vec{M} = |\vec{a}|\,|\vec{M}| \cos\theta, \tag{8.1}$$

where $\vec{a}$, the *lead vector*, points along the electrode pair and θ is the angle between the two vectors. A lead perpendicular to $\vec{M}$ reads zero; a lead parallel to it reads the most.

Einthoven's triangle. Willem Einthoven treated the two arms and the left leg as wires to three corners of an equilateral triangle centered on the heart (Fig. 8.2). The three *limb leads* are the potential differences

$$\text{I} = \Phi_{\text{LA}} - \Phi_{\text{RA}}, \qquad \text{II} = \Phi_{\text{LL}} - \Phi_{\text{RA}}, \qquad \text{III} = \Phi_{\text{LL}} - \Phi_{\text{LA}}, \tag{8.2}$$

so by Kirchhoff's voltage law $\text{II} = \text{I} + \text{III}$ at every instant. With the angle measured from lead I and positive toward the feet, lead I points at 0°, lead II at 60°, and lead III at 120°, and in the ideal triangle each limb lead reads $M\cos(\theta - \theta_{\text{lead}})$ for a vector of size M at angle θ. Two leads therefore fix the vector in the frontal plane. Its net direction during the QRS complex, the *mean electrical axis*, normally lies between -30° and 90°.

Augmented and precordial leads. Emanuel Goldberger added three *augmented limb leads*, each a limb against the average of the other two:

$$\begin{aligned} \text{aVR} &= \Phi_{\text{RA}} - \tfrac{1}{2}(\Phi_{\text{LA}} + \Phi_{\text{LL}}), \\ \text{aVL} &= \Phi_{\text{LA}} - \tfrac{1}{2}(\Phi_{\text{RA}} + \Phi_{\text{LL}}), \\ \text{aVF} &= \Phi_{\text{LL}} - \tfrac{1}{2}(\Phi_{\text{RA}} + \Phi_{\text{LA}}). \end{aligned} \tag{8.3}$$

They point at -150°, -30°, and 90°, they sum to zero, and they fill the gaps between the limb leads. Six *precordial* leads, V1 to V6, measure electrodes in an arc across the chest against the average of the three limbs (the Wilson central terminal of Section 8.5) and view the heart in the horizontal plane. The standard 12-lead ECG needs ten electrodes: four on the limbs, counting the right leg, and six on the chest (Fig. 8.3). A wearable patch uses two or three closely spaced electrodes and records one or two leads with a smaller amplitude.

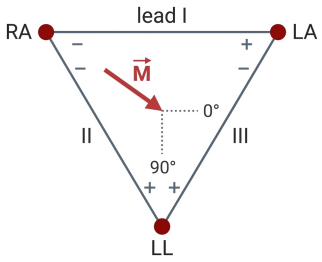


Figure 8.2: Einthoven's triangle. The right arm (RA), left arm (LA), and left leg (LL) sit at the corners. Each lead reads the projection of the cardiac vector $\vec{M}$ on its side; the signs mark the polarity. Angles are measured from lead I, positive toward the feet.

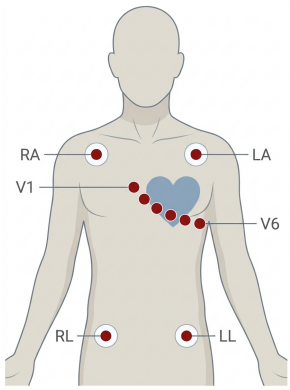


Figure 8.3: Electrode positions for a 12-lead ECG on the torso: the limb electrodes RA, LA, LL, and RL near the shoulders and hips, and the chest electrodes V1 to V6.

WORKED EXAMPLE 8.1 *Leads from three limb potentials*

At the peak of the R wave, the potentials of the limb electrodes against a distant reference are $\Phi_{RA} = -0.30\,\text{mV}$, $\Phi_{LA} = 0.20\,\text{mV}$, and $\Phi_{LL} = 0.70\,\text{mV}$. (a) Find leads I, II, and III and check Einthoven's law. (b) Find aVR, aVL, and aVF. (c) Which way does the cardiac vector point, and how large is it? (d) A smartwatch records an ECG between an electrode on the back of the case (left wrist) and one on the crown, touched by a finger of the right hand. Which lead is that, and what does it read at this instant?

SOLUTION. (a) From Eq. (8.2), $\text{I} = 0.20 - (-0.30) = 0.50\,\text{mV}$, $\text{II} = 0.70 - (-0.30) = 1.00\,\text{mV}$, and $\text{III} = 0.70 - 0.20 = 0.50\,\text{mV}$. Indeed $\text{I} + \text{III} = 1.00\,\text{mV} = \text{II}$.

(b) From Eq. (8.3), $\text{aVR} = -0.30 - 0.45 = -0.75\,\text{mV}$, $\text{aVL} = 0.20 - 0.20 = 0\,\text{mV}$, and $\text{aVF} = 0.70 - (-0.05) = 0.75\,\text{mV}$. They sum to zero.

(c) aVL reads zero, so the vector is perpendicular to aVL's direction of $-30°$: it points at $60°$ (the opposite choice, $-120°$, would make lead II negative). That is the direction of lead II, so lead II reads the whole vector, $M = 1.00\,\text{mV}$. Check: lead I reads $M\cos 60° = 0.50\,\text{mV}$. An axis of $60°$ is normal.

(d) Left arm minus right arm is lead I, here 0.50 mV, half of lead II. One lead is enough to time beats and look for P waves, not to locate damage in the heart as a 12-lead ECG does. ■

The ECG waveform

Figure 8.4 shows two beats as lead II records them. The *P wave* is the depolarization of the atria: about 0.1 to 0.25 mV and less than 0.12 s long. The *PR interval*, from the start of the P wave to the start of the QRS complex, is 0.12 to 0.20 s and is mostly the delay in the AV node. The *QRS complex* is the depolarization of the ventricles: the largest deflection, about 1 mV, and 0.06 to 0.10 s wide; a QRS wider than 0.12 s points to a conduction block. The flat *ST segment* that follows is the plateau of the ventricular action potential. A shift of the ST segment above or below the baseline is a sign of ischemia (heart muscle short of blood), which is why a diagnostic ECG must pass frequencies down to 0.05 Hz (Section 8.6). The *T wave*, 0.1 to 0.5 mV, is the repolarization of the ventricles. The *QT interval*, from the start of the QRS to the end of T, is about 0.35 to 0.44 s and shortens as the heart rate rises.

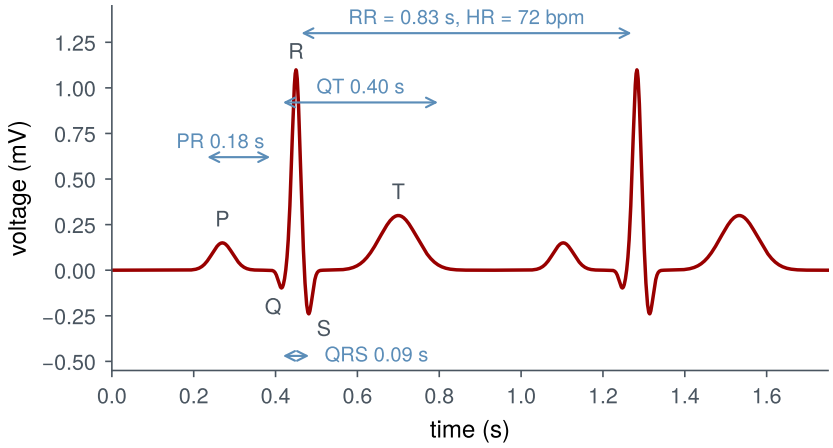


Figure 8.4: Two beats of a lead II ECG at 72 beats per minute (illustrative data, a sum of smooth bumps). The P wave, the QRS complex, and the T wave are labeled with their typical intervals. The RR interval between R peaks sets the heart rate.

Heart rate and its variability. The time between successive R peaks is the *RR interval*. The heart rate in beats per minute (bpm) is

$$\mathrm{HR} = \frac{60}{\overline{RR}}, \qquad \overline{RR} \text{ in seconds}, \tag{8.4}$$

so a mean RR interval of 0.83 s is 72 bpm; 60 to 100 bpm is normal at rest. The RR interval is not constant. The autonomic nervous system adjusts it beat by beat: it shortens while you breathe in and lengthens while you breathe out. This *heart-rate variability* (HRV) reflects the balance of the nervous system and is one of the features wearables use to estimate stress and recovery. Two time-domain measures are standard.[1] SDNN is the standard deviation of the RR intervals, and RMSSD is the root mean square of the differences between successive intervals:

$$\begin{aligned} \mathrm{SDNN} &= \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(RR_i - \overline{RR}\right)^2}, \\ \mathrm{RMSSD} &= \sqrt{\frac{1}{N-1}\sum_{i=1}^{N-1}\left(RR_{i+1} - RR_i\right)^2}. \end{aligned} \tag{8.5}$$

Both are in milliseconds, typically tens of milliseconds in a healthy adult at rest. RMSSD follows the fast, breath-to-breath changes; SDNN also includes slower drifts. One missed beat doubles an RR interval and inflates both, so RR data must be cleaned first. An irregular RR sequence with no P waves is the signature of atrial fibrillation, the most common arrhythmia.

[1] Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, *Circulation* (1996). The task force recommends five-minute recordings for short-term HRV.

The EEG: rhythms of the cortex

The EEG comes mostly from the *pyramidal neurons* of the cerebral cortex. They stand side by side, perpendicular to the surface, and their

synaptic currents flow along the same axis, so when thousands of them receive input together their currents add. Action potentials contribute little: they are too brief and too poorly synchronized. Between the cortex and the scalp lie the cerebrospinal fluid, the skull, which conducts poorly, and the scalp, which spreads the current sideways. A scalp electrode therefore sees the smoothed activity of several square centimeters of cortex, typically 10 to 100 µV. (Electrodes on or in the brain see far more detail; Chapter 9 covers them.)

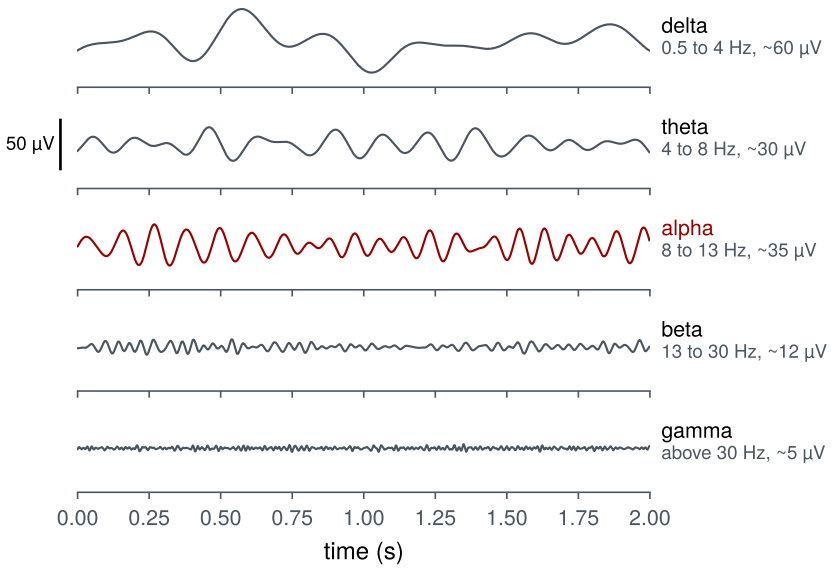


Figure 8.5: The EEG rhythms with typical scalp amplitudes (illustrative data: band-limited noise). Slow rhythms are large and fast ones small. The alpha rhythm grows over the back of the head when the eyes close.

The EEG is described by its rhythms (Fig. 8.5). *Delta* waves (0.5 to 4 Hz) dominate deep sleep. *Theta* (4 to 8 Hz) appears in drowsiness. *Alpha* (8 to 13 Hz) is the rhythm of relaxed wakefulness with closed eyes, strongest over the back of the head and suppressed within a fraction of a second when the eyes open. *Beta* (13 to 30 Hz) accompanies alert, focused, or anxious states, and *gamma* (above 30 Hz) is small and easily swamped by scalp muscle activity. Because alpha appears with relaxation and beta with arousal, the ratio of alpha to beta power is a simple stress index; the wearable EEG project of Chapter 14 can use it (Section 8.9).

The 10-20 system and ear-EEG. Clinical EEG places 21 electrodes by the *10-20 system*: measure from the nasion (the bridge of the nose) to the inion (the bump at the back of the skull) and between the two ears, and put electrodes at 10 % and 20 % steps of those distances (Fig. 8.6); each position is named by a letter for the region and a number (Cz is the top of the head). A wearable cannot carry 21 wet electrodes, so research groups have moved the electrodes to the ear, printed on a

flexible sheet that sticks around it.[2] A student team built such an array: ten printed silver electrodes 3 mm in diameter on a stretchable sheet, eight of them channels and two the reference and the bias electrode. Closely spaced electrodes see smaller differences than electrodes far apart on the scalp, so ear-EEG amplitudes are smaller, but the alpha rhythm is still visible.

[2] S. Debener *et al.*, *Sci. Rep.* (2015); M. G. Bleichner and S. Debener, *Front. Hum. Neurosci.* (2017).

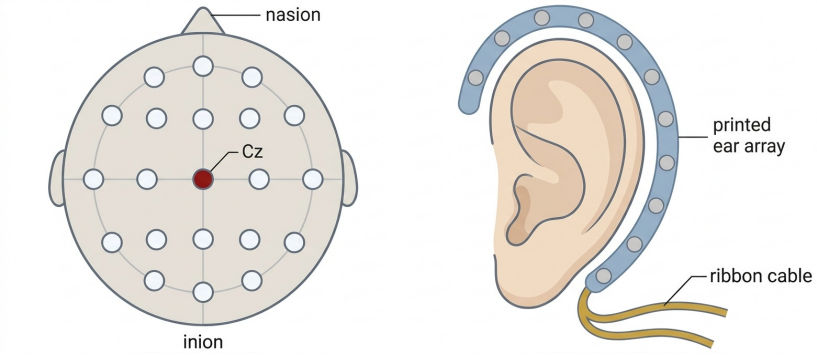


Figure 8.6: EEG electrode placement. Left: the 10-20 system seen from above, with Cz at the top of the head. Right: a printed electrode array around the ear, drawn schematically (the student array described in the text has ten electrodes), connected to its front end by a flat ribbon cable.

EMG and EOG

A muscle contracts when its motor neurons fire. Each motor neuron drives a *motor unit*, from tens to thousands of muscle fibers that fire together. Surface electrodes over the muscle record the sum of many motor units firing out of step, an irregular signal whose amplitude grows with force, from tens of microvolts at rest to a few millivolts in a strong contraction, with energy from about 20 to 500 Hz. Surface EMG drives gesture recognition and prosthetic control; in an EEG or ECG recording it is noise. A screen-printed armband of 64 electrodes, whose classifier was trained and updated on the wearable itself, recognized 13 hand gestures with 97 % accuracy in two participants.[3] The *EOG* comes from the eye itself, which is a dipole with the cornea positive with respect to the retina. Electrodes beside the eyes see a voltage that changes with the angle of gaze and a large spike at every blink. Measuring gaze needs an amplifier that passes DC; in the EEG the EOG is the main artifact of the frontal channels.

[3] A. Moin, A. Zhou, A. Rahimi, *et al.*, J. M. Rabaey, *Nat. Electron.* (2021).

8.2 The Electrode-Skin Interface

In the body, current is carried by ions; in a wire, by electrons. The electrode is where one becomes the other, and it sets the offset, the noise, and much of the motion artifact of every biopotential recording.

The skin

The outer layer of the skin, the *stratum corneum*, is 10 to 20 µm of dead, flattened, keratin-filled cells over most of the body and much thicker on the palms and soles (Fig. 8.7). It is dry, and it is the main electrical barrier of the skin. Below it lie the living *epidermis* and the *dermis*, which holds blood vessels, nerves, and sweat glands and conducts well. Sweat ducts cross the stratum corneum and form parallel conducting paths, which is why the skin's impedance falls when you sweat. Cleaning the skin with alcohol and abrading it lightly removes part of the stratum corneum and lowers the impedance by about an order of magnitude.

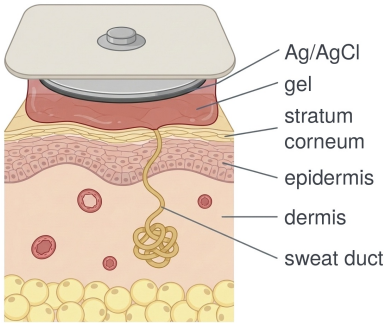


Figure 8.7: A wet gel electrode on the skin. The gel wets the stratum corneum; ions in the gel exchange charge with the Ag/AgCl layer on the metal.

The half-cell potential

A metal in an electrolyte exchanges charge with it until an electric double layer forms at the surface (Chapter 6). The potential across the double layer is the *half-cell potential* E_{hc}. The standard biopotential electrode is silver coated with silver chloride in a gel of chloride ions:

$$\text{AgCl} + \text{e}^- \rightleftharpoons \text{Ag} + \text{Cl}^-, \qquad E^\circ = +0.222\,\text{V vs. SHE}. \tag{8.6}$$

The reaction is fast in both directions, so current crosses the interface with little extra voltage, and the potential is fixed by the chloride concentration of the gel through the Nernst equation. Two identical electrodes in the same gel have nearly the same E_{hc}, and a differential amplifier sees only their difference, a few millivolts for a good pair. Electrodes of different metals, dry electrodes on sweaty skin, or electrodes at different stages of drying can differ by hundreds of millivolts. Standards for ECG monitors therefore require the front end to keep working with ± 300 mV of DC offset between electrodes, a number that limits the gain of the first amplifier stage (Section 8.4).

Polarizable and nonpolarizable electrodes. Chapter 6 met the two ideal electrodes. A *perfectly polarizable* electrode passes no charge across its interface; it behaves like a capacitor. Platinum, gold, and stainless steel come close. Their double layer is easily disturbed, so they drift and produce large motion artifacts. A *perfectly nonpolarizable* electrode passes current freely with no change of its potential. Ag/AgCl comes close, which is why it is the clinical standard: low drift, low noise, and small motion artifacts. A printed silver electrode without chloride behaves more like a polarizable electrode; chloriding it, or printing an Ag/AgCl ink, moves it toward the nonpolarizable end (Chapter 12; Lab 2 prints such electrodes).

Electrode types

Wet gel electrodes,[4] a snap stud on an Ag/AgCl disc in a foam pad with gel and adhesive, are cheap, stable, and low in impedance; the gel dries within a day or two and can irritate the skin. *Dry electrodes* (metal, carbon-loaded rubber, conductive polymer) need no gel and rely on sweat and pressure for contact. Their impedance is ten to a hundred times higher and more variable, so they give a usable heart rate but a less reliable waveform, as in a smartwatch. *Printed electrodes* put silver or Ag/AgCl ink on a thin plastic film, with wavy traces that survive stretching (Chapters 11 and 12), and *textile electrodes* use silver-plated yarn knitted into a garment.

[4] A. Searle and L. Kirkup, *Physiol. Meas.* (2000), compare wet, dry, and insulating electrodes; Y. M. Chi *et al.*, *IEEE Rev. Biomed. Eng.* (2010), review dry and non-contact electrodes.

An equivalent circuit

Figure 8.8 is the simplest useful model of an electrode on the skin. The half-cell potential E_{hc} is a battery. The double layer is a capacitor C_d in parallel with a resistor R_d that represents charge transfer across it (a nonpolarizable electrode has a small R_d, a polarizable one a huge R_d). The gel and the tissue add a series resistance R_s. The impedance is

$$Z(f) = R_s + \frac{R_d}{1 + j2\pi f R_d C_d}, \qquad f_d = \frac{1}{2\pi R_d C_d}, \tag{8.7}$$

which falls from $R_s + R_d$ at low frequency, where the capacitor is an open circuit, to R_s above the corner f_d, where it shorts R_d (Fig. 8.9). The stratum corneum adds a second, similar RC network in series. Biopotentials live at low frequency, where the impedance is largest, so an impedance measured at 1 kHz flatters an electrode.

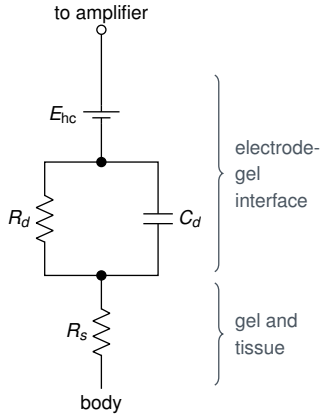


Figure 8.8: Equivalent circuit of an electrode on the skin: the half-cell potential E_{hc}, the double layer $R_d \parallel C_d$, and the series resistance R_s of gel and tissue.

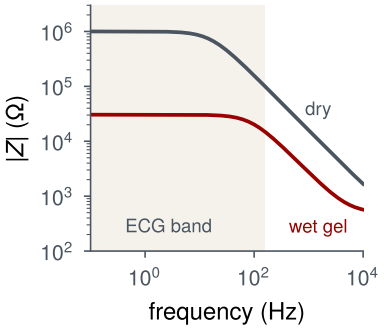


Figure 8.9: Impedance magnitude of Eq. (8.7) for a wet gel electrode (R_d = 30 kΩ, C_d = 60 nF) and a dry one (R_d = 1 MΩ, C_d = 10 nF), both with R_s = 500 Ω (illustrative values).

Motion artifacts. When an electrode moves, the charge of the double layer is disturbed and E_{hc} shifts; when the skin stretches, the potential across the stratum corneum changes. Both produce slow swings of millivolts, with energy from about 0.1 to 10 Hz, right in the ECG band, so filtering cannot remove them. The remedies are mechanical and chemical: Ag/AgCl with gel, a foam pad and adhesive that hold the electrode still, light abrasion, strain relief for the lead wires, and a high amplifier input impedance so that changes of the electrode impedance matter less.

> **Worked Example 8.2** *Impedance of wet and dry electrodes*
>
> Use the illustrative values of Fig. 8.9. (a) Find the corner frequency and $|Z|$ at 0.5, 10, and 1000 Hz for each electrode. (b) Each electrode drives an amplifier input of 10 MΩ. What fraction of a 0.5 Hz signal reaches the input? What if the input is 1 GΩ?

SOLUTION. (a) Wet: $f_d = 1/(2\pi \times 30\,\text{k}\Omega \times 60\,\text{nF}) = 88\,\text{Hz}$. At 0.5 Hz the capacitor is negligible and $|Z| \approx R_s + R_d = 30.5\,\text{k}\Omega$; at 10 Hz, $|Z| = 30.3\,\text{k}\Omega$; at 1 kHz, $|Z| = 2.7\,\text{k}\Omega$. Dry: $f_d = 1/(2\pi \times 1\,\text{M}\Omega \times 10\,\text{nF}) = 15.9\,\text{Hz}$, and $|Z| = 1.00\,\text{M}\Omega$, 847 kΩ, and 15.9 kΩ at the three frequencies. At 1 kHz the dry electrode looks only six times worse than the wet one; at the frequencies of the ECG it is thirty times worse.

(b) The electrode and the input form a voltage divider. Wet: $10/(10 + 0.0305) = 0.997$, a 0.3 % loss. Dry: $10/(10 + 1.0) = 0.91$, a 9 % loss that changes whenever the contact changes. With a 1 GΩ input the dry electrode loses only 0.1 %. Section 8.3 shows that a mismatch between two electrodes does more damage than this loss: it turns hum into signal. ■

8.3 *Interference: Mains, Common Mode, and Mismatch*

YOUR BODY IS A CONDUCTOR in a room full of wires at 120 V and 60 Hz (230 V and 50 Hz in much of the world). It couples to them through the air, and the coupling puts a hum on the whole body that can be a thousand times larger than the ECG.

Electric-field coupling and common-mode voltage

The body and a nearby mains wire form a capacitor, which passes alternating current, with an impedance $1/(\omega C)$ at angular frequency $\omega = 2\pi f$ (Chapter 1). Figure 8.10 shows the model of B. B. Winter and J. G. Webster.[5] A capacitance C_p of a few picofarads couples the body to the mains wiring, and a capacitance C_b of about 200 pF couples it to earth. The changing mains voltage drives a *displacement current*, the current through a capacitor,

$$i_d \approx V_{\text{mains}}\, \omega C_p \tag{8.8}$$

flows through C_p into the body and out through C_b and through any electrode that connects the body to the amplifier's ground. Because $1/\omega C_p$ is hundreds of megohms, much more than anything else in the path, i_d is set by C_p alone: a fraction of a microampere. If the body is connected to the amplifier common through an electrode of impedance Z_G, the body sits at $v_{\text{cm}} = i_d Z_G$ with respect to that common, a few millivolts. If it is not connected at all, C_p and C_b form a capacitive divider and the body floats at $V_{\text{mains}}\, C_p/(C_p + C_b)$, around a volt.

[5] B. B. Winter and J. G. Webster, *IEEE Trans. Biomed. Eng.* (1983). The capacitances are typical; they vary with posture, shoes, and the room.

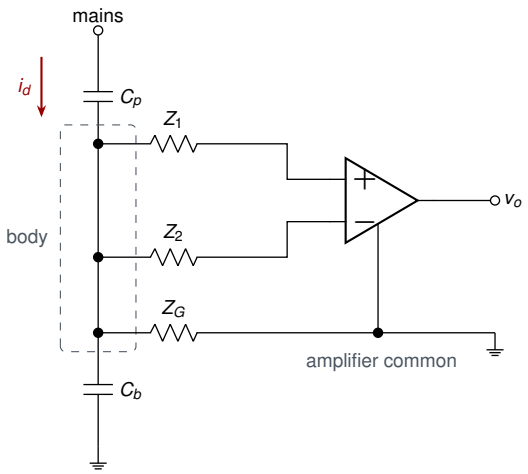


Figure 8.10: How mains hum reaches a biopotential amplifier. The displacement current i_d through C_p flows into the body and returns to earth through C_b and through the ground electrode Z_G and the amplifier common, which here is tied to earth. The body then sits at a common-mode voltage $v_{\text{cm}} \approx i_d Z_G$ above the amplifier common. The recording electrodes Z_1 and Z_2 carry both inputs to that voltage. The triangle stands for a differential amplifier (feedback not shown).

Common-mode and differential signals

Any two input voltages can be split into a part they share and a part that differs:

$$v_{\text{cm}} = \frac{v_1 + v_2}{2}, \qquad v_d = v_1 - v_2. \tag{8.9}$$

The ECG is a differential signal; the hum on the body is common mode. A real amplifier responds to both, $v_o = A_d v_d + A_{\text{cm}} v_{\text{cm}}$, and its *common-mode rejection ratio* (Chapter 4) compares the two gains:

$$\text{CMRR} = \frac{A_d}{A_{\text{cm}}}, \qquad \text{CMRR}_{\text{dB}} = 20 \log_{10} \frac{A_d}{A_{\text{cm}}}. \tag{8.10}$$

Dividing the output by A_d refers it to the input: the common-mode voltage appears as an apparent differential signal $v_{\text{cm}}/\text{CMRR}$.

That sets the CMRR a design needs. If the body sits 10 mV above the amplifier common and the hum referred to the input must stay below 10 µV, the amplifier needs a CMRR of at least 1000, or 60 dB. A body floating at 1.8 V would need 180 000, or 105 dB, at the edge of what the best integrated amplifiers offer at 60 Hz. The one-op-amp difference amplifier of Chapter 4, built from 1 % resistors at a gain of 1, has a worst-case CMRR of only about 50 and would leave 200 µV of hum. This is why every biopotential amplifier connects the body to its common through a third electrode, and why the best ones drive that electrode (Section 8.5).

Magnetic coupling. The lead wires and the body also form a loop. A changing magnetic field B through a loop of area A induces $v = A\,\mathrm{d}B/\mathrm{d}t$, which at frequency f is $2\pi f B A$. A loop of 200 cm^2 in a 0.5 µT field at 60 Hz, possible near a transformer or a power cord, picks up $2\pi(60)(0.5 \times 10^{-6})(0.02) = 3.8$ µV, directly as a differential signal.

CMRR does nothing against it. The remedy is geometric: keep the lead wires together, twist them, and keep power cords away.

Electrode mismatch turns common mode into signal

Each amplifier input has a finite impedance to the amplifier common, Z_{cm}: a large resistance in parallel with a few to hundreds of picofarads of input, cable, and board capacitance. With an electrode in series, each input is a voltage divider (Fig. 8.11). If the two electrodes had equal impedances, both dividers would pass the same fraction of v_{cm} and the amplifier would reject it. They never do: electrodes differ by tens of kilohms. The two inputs then see different fractions of v_{cm}, and the difference is a real differential signal at the amplifier's input:

$$v_d = v_{cm}\left(\frac{Z_{cm}}{Z_1 + Z_{cm}} - \frac{Z_{cm}}{Z_2 + Z_{cm}}\right) \approx v_{cm}\,\frac{Z_2 - Z_1}{Z_{cm}}, \tag{8.11}$$

where the approximation holds for $Z_1, Z_2 \ll Z_{cm}$. This error arises before the amplifier, so no CMRR can reject it. The cures are a common-mode input impedance Z_{cm} much larger than the electrode mismatch (short, low-capacitance input wiring), matched electrodes, and a small v_{cm}, which is the job of the right-leg drive.

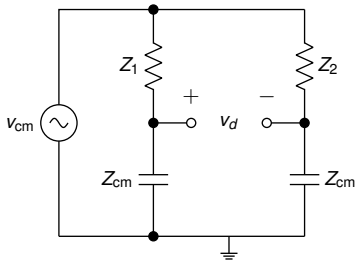


Figure 8.11: The potential divider. The body's common-mode voltage reaches the two amplifier inputs through unequal electrode impedances Z_1 and Z_2, so the inputs differ by $v_d \approx v_{cm}(Z_2 - Z_1)/Z_{cm}$.

KEY IDEA Electrode mismatch converts common mode into a differential signal before the amplifier. No amount of CMRR removes it; only a large input impedance and a small v_{cm} do.

WORKED EXAMPLE 8.3 *Hum from the room*

A person sits near 120 V, 60 Hz wiring with $C_p = 3\,\text{pF}$ to the line and $C_b = 200\,\text{pF}$ to earth. (a) What voltage does the body float at if nothing else is connected? (b) Find the displacement current. A ground electrode of $Z_G = 50\,\text{k}\Omega$ now connects the body to the common of an amplifier whose common is at earth potential. What is v_{cm}? (c) The recording electrodes have $Z_1 = 20\,\text{k}\Omega$ and $Z_2 = 70\,\text{k}\Omega$, and each amplifier input has 100 pF to the common. Find the differential hum at the input and compare it with a 1 mV ECG and a 20 µV alpha wave.

SOLUTION. (a) The capacitive divider gives $120\,\text{V} \times 3/(3 + 200) = 1.77\,\text{V}$ rms.

(b) At 60 Hz, $1/\omega C_p = 1/(377 \times 3\,\text{pF}) = 884\,\text{M}\Omega$, far more than anything else in the path, so $i_d = 120\,\text{V}/884\,\text{M}\Omega = 0.136\,\mu\text{A}$. The path to earth through C_b (13.3 MΩ at 60 Hz) is in parallel with Z_G and barely matters, so $v_{cm} = i_d Z_G = 0.136\,\mu\text{A} \times 50\,\text{k}\Omega = 6.8\,\text{mV}$. Connecting the body lowered the hum from volts to millivolts.

(c) $|Z_{cm}| = 1/(2\pi \times 60 \times 100\,\text{pF}) = 26.5\,\text{M}\Omega$, and Eq. (8.11) gives $v_d \approx 6.8\,\text{mV} \times 50\,\text{k}\Omega/26.5\,\text{M}\Omega = 12.8\,\mu\text{V}$. On a 1 mV ECG that is about 1 %, visible on the flat parts of the trace; on an EEG it is about two thirds of the alpha wave. Ten times less input capacitance would give ten times less hum. ■

8.4 The Instrumentation Amplifier

THE FIRST AMPLIFIER after the electrodes has five jobs. It must present an input impedance far above the electrode impedance (hundreds of megohms or more, with little capacitance); reject common mode (a CMRR of 80 to 100 dB or more at the mains frequency); add little noise (a few microvolts for ECG, under a microvolt for EEG); survive the electrode offset of ±300 mV without saturating; and give its input bias currents a DC path. Chapter 4 built the difference amplifier from one op amp and four resistors. Its input impedance is only its resistors, and its CMRR depends on how well they match. The *instrumentation amplifier* (IA) fixes the first problem with two input buffers and eases the second by amplifying before subtracting.

The three-op-amp instrumentation amplifier

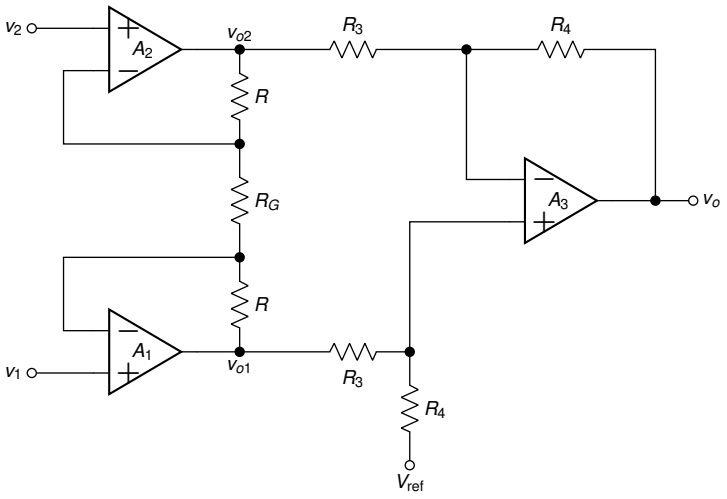


Figure 8.12: The three-op-amp instrumentation amplifier. The input stage amplifies the difference $v_1 - v_2$ by $1 + 2R/R_G$ and passes the common mode with a gain of one; the difference stage subtracts and rejects what is left. Both inputs drive op-amp inputs directly, so they draw almost no current.

Figure 8.12 shows the circuit. The two rules of the ideal op amp (Chapter 4) put the inverting input of A_1 at v_1 and that of A_2 at v_2, so the voltage across R_G is $v_1 - v_2$ and the current $(v_1 - v_2)/R_G$ flows through it. No current enters the op-amp inputs, so the same current flows through both resistors R, and the output difference is

$$v_{o1} - v_{o2} = (v_1 - v_2)\left(1 + \frac{2R}{R_G}\right). \tag{8.12}$$

If both inputs rise together, no current flows in R_G and both outputs simply follow: the common mode passes with a gain of one. The difference stage then forms $v_o = V_{\text{ref}} + (R_4/R_3)(v_{o1} - v_{o2})$, so the overall

differential gain is

$$G = \left(1 + \frac{2R}{R_G}\right)\frac{R_4}{R_3}, \tag{8.13}$$

set by one resistor, R_G. Integrated IAs put the matched resistors on the chip and leave R_G outside.

KEY EQUATION

$G = \left(1 + \frac{2R}{R_G}\right)\frac{R_4}{R_3}$

Gain of the three-op-amp instrumentation amplifier; one resistor, R_G, sets it.

CMRR and resistor matching

The difference stage rejects common mode only as well as its resistors match. Chapter 4 showed that if each of its four resistors is off by up to a fraction ϵ (its tolerance) and the errors add in the worst direction, the stage's rejection ratio is

$$\mathrm{CMRR}_{\mathrm{diff}} \approx \frac{1 + G_d}{4\epsilon}, \qquad G_d = \frac{R_4}{R_3}. \tag{8.14}$$

The input stage multiplies the differential signal by $G_1 = 1 + 2R/R_G$ before the difference stage sees it but leaves the common mode alone, so the whole IA has

$$\mathrm{CMRR} \approx G_1\,\mathrm{CMRR}_{\mathrm{diff}}. \tag{8.15}$$

KEY IDEA Amplify before you subtract: the gain of the input stage multiplies the CMRR of the difference stage.

This is the reason to put gain in the first stage. Laser-trimmed integrated IAs reach 100 to 120 dB at DC; their CMRR falls at higher frequency, where small capacitance mismatches matter, so read the datasheet value at 60 Hz.

Input impedance and bias current. The inputs of the IA are op-amp inputs: with CMOS op amps, picoamperes of bias current and more than 1 GΩ in parallel with a few picofarads. That solves the loading of Example 8.2. But the bias currents need a DC path back to the amplifier common, or the inputs drift to a supply rail. Through the body, that path is the third electrode. If an electrode falls off, its input floats; front ends use this on purpose, injecting a tiny current into each input and flagging any input that drifts toward a rail. This is *lead-off detection*.

Offset sets the first-stage gain

The electrode offset is a DC differential voltage, so the first stage amplifies it along with the signal. A first stage with gain G_1 and an output swing of $\pm V_{\mathrm{sw}}$ about its reference saturates unless

$$G_1\,(V_{\mathrm{offset}} + v_{\mathrm{signal}}) < V_{\mathrm{sw}}. \tag{8.16}$$

With ±300 mV of offset, G_1 is limited to a few or tens. The classic answer is to split the gain: a modest first stage, a high-pass filter that removes the DC (Section 8.6), and a second stage with the rest of the

gain. The modern answer, the one the ADS1293 takes, is to keep the analog gain low (3.5), digitize the offset and the signal together with a 24-bit converter, and remove the DC in software (Section 8.7).

8.5 Right-Leg Drive and the Wilson Central Terminal

A THIRD ELECTRODE connects the body to the amplifier. Connected passively, it holds the body at $v_{cm} = i_d Z_G$. Driven by an amplifier, it does far better. It is traditionally placed on the right leg, which no standard limb lead uses, hence the name; on a patch or a headband it is simply a third electrode.

The driven right leg

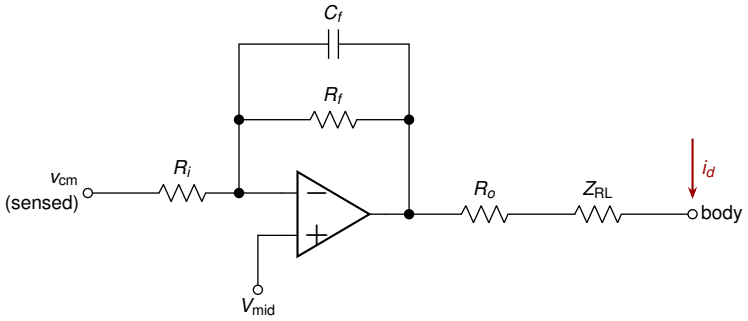


Figure 8.13: A right-leg drive. The amplifier senses the common-mode voltage of the recording electrodes, inverts and amplifies it with gain $G = Z_f/R_i$ ($Z_f = R_f \parallel C_f$), and drives the body through the right-leg electrode Z_{RL} and a protective resistor R_o. The feedback drives the body's common-mode voltage toward V_{mid}.

A *driven right leg* (RLD) senses the common-mode voltage of the recording electrodes (the average of the inputs, or the midpoint of R_G in the IA), inverts and amplifies it, and feeds it back to the body through the right-leg electrode (Fig. 8.13). It works like a thermostat: when the body drifts up, the drive pushes it down. The displacement current i_d must still leave the body through the right-leg path, so the body sits at $i_d(Z_{RL} + R_o)$ plus whatever the amplifier output adds. The amplifier output is $-G\,v_{cm}$ (measured from V_{mid}), so $v_{cm} = i_d(Z_{RL} + R_o) - G\,v_{cm}$, and

$$v_{cm} = \frac{i_d\,(Z_{RL} + R_o)}{1 + G(j\omega)}, \qquad G(j\omega) = \frac{Z_f}{R_i} = \frac{R_f/R_i}{1 + j\omega R_f C_f}. \tag{8.17}$$

The drive divides the common-mode voltage by $1 + G$. The capacitor C_f lowers the gain at higher frequencies so that the loop, which includes the electrode and cable capacitances, stays stable. The resistor R_o protects the subject: if the amplifier output jumps to a rail, no more than V_{rail}/R_o can flow into the body.

KEY EQUATION

$$v_{cm} = \frac{i_d\,(Z_{RL} + R_o)}{1 + G}$$

The right-leg drive divides the common-mode voltage by one plus its loop gain.

Integrated front ends have the drive amplifier on the chip and leave R_i, R_f, and C_f outside. TI's evaluation board for the ADS1293 uses

$R_i = 10\,\text{k}\Omega$, $R_f = 10\,\text{M}\Omega$, and $C_f = 1\,\text{nF}$. The ADS1299 calls the same function the *bias drive*. Lab 3 chooses the parts for your board.

> **Worked Example 8.4** *How much does the right-leg drive help?*
>
> Use the body of Example 8.3: $i_d = 0.136\,\mu\text{A}$ at 60 Hz and a right-leg electrode of $50\,\text{k}\Omega$. (a) Find v_{cm} with a passive ground electrode. (b) Find v_{cm} with the evaluation-board drive ($R_i = 10\,\text{k}\Omega$, $R_f = 10\,\text{M}\Omega$, $C_f = 1\,\text{nF}$), ignoring R_o, and the hum that the electrode mismatch of that example now makes. (c) The drive runs on 3.3 V. What R_o keeps the current into the body below $10\,\mu\text{A}$ if the output sticks at a rail, and what does it cost?
>
> Solution. (a) $v_{\text{cm}} = 0.136\,\mu\text{A} \times 50\,\text{k}\Omega = 6.8\,\text{mV}$.
>
> (b) At 60 Hz, $\omega R_f C_f = 377 \times 10\,\text{M}\Omega \times 1\,\text{nF} = 3.77$, so $|G| = (10^7/10^4)/\sqrt{1+3.77^2} = 1000/3.90 = 256$, with a phase of $-75°$. Then $|1+G| = 257$ and $v_{\text{cm}} = 6.8\,\text{mV}/257 = 26\,\mu\text{V}$. The mismatch ratio of Example 8.3 was $50\,\text{k}\Omega/26.5\,\text{M}\Omega = 1.89 \times 10^{-3}$, so the hum falls from $12.8\,\mu\text{V}$ to $26\,\mu\text{V} \times 1.89 \times 10^{-3} = 0.05\,\mu\text{V}$, negligible even for EEG.
>
> (c) $R_o \geq 3.3\,\text{V}/10\,\mu\text{A} = 330\,\text{k}\Omega$. It adds to Z_{RL} in Eq. (8.17): $v_{\text{cm}} = 0.136\,\mu\text{A} \times 380\,\text{k}\Omega/257 = 0.20\,\text{mV}$, still thirty times better than the passive electrode. ■

The Wilson central terminal

A chest lead needs a reference that does not favor any limb. Frank Wilson joined the three limb electrodes through equal resistors (Fig. 8.14). The junction, the *Wilson central terminal*, sits at the average potential

$$\Phi_{\text{WCT}} = \tfrac{1}{3}\left(\Phi_{\text{RA}} + \Phi_{\text{LA}} + \Phi_{\text{LL}}\right), \qquad V_i = \Phi_{V_i} - \Phi_{\text{WCT}}. \tag{8.18}$$

For the limb potentials of Example 8.1, $\Phi_{\text{WCT}} = (-0.30 + 0.20 + 0.70)/3 = 0.20\,\text{mV}$, and a chest electrode at 1.40 mV reads V2 = 1.20 mV. The resistors load the electrodes, so modern front ends buffer each limb first; the ADS1293 builds the terminal on the chip, which lets one channel record a chest lead (the five-electrode setup of Lab 4). Goldberger's augmented leads (Eq. (8.3)) leave the measured limb out of the average, which makes them 1.5 times larger than a limb measured against the WCT: aVF $= \frac{3}{2}(\Phi_{\text{LL}} - \Phi_{\text{WCT}}) = 1.5 \times 0.50\,\text{mV} = 0.75\,\text{mV}$, as in the example.

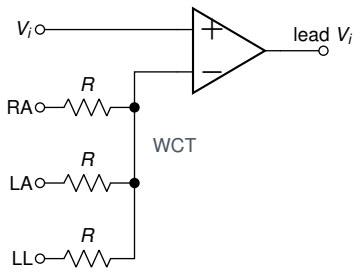


Figure 8.14: The Wilson central terminal (WCT): the average of the three limb potentials through equal resistors (buffered inside an AFE). A precordial lead is a chest electrode measured against it by a differential amplifier (feedback not shown).

8.6 Filtering Biopotentials

A filter keeps the frequencies that carry information and removes the rest. Chapter 1 introduced the RC low-pass and high-pass

filters and the language of cutoffs and decibels; here we choose bands for biopotentials and size the circuits.

Choosing the band

For a diagnostic ECG the American Heart Association recommends a pass band from 0.05 Hz to 150 Hz for adults.[6] The low edge keeps the slow ST segment undistorted and the high edge keeps the sharp corners of the QRS. A monitor or a wearable usually narrows the band to 0.5 to 40 Hz, which cuts baseline wander, EMG, and mains hum and is enough for heart rate and rhythm. EEG uses about 0.5 to 40 Hz, wider if gamma matters, and surface EMG about 20 to 500 Hz. Figure 8.15 compares the two ECG bands.

[6] P. Kligfield *et al.*, *Circulation* (2007). The low cutoff of 0.05 Hz preserves the ST segment; linear digital filters with zero phase distortion may use a cutoff up to 0.67 Hz.

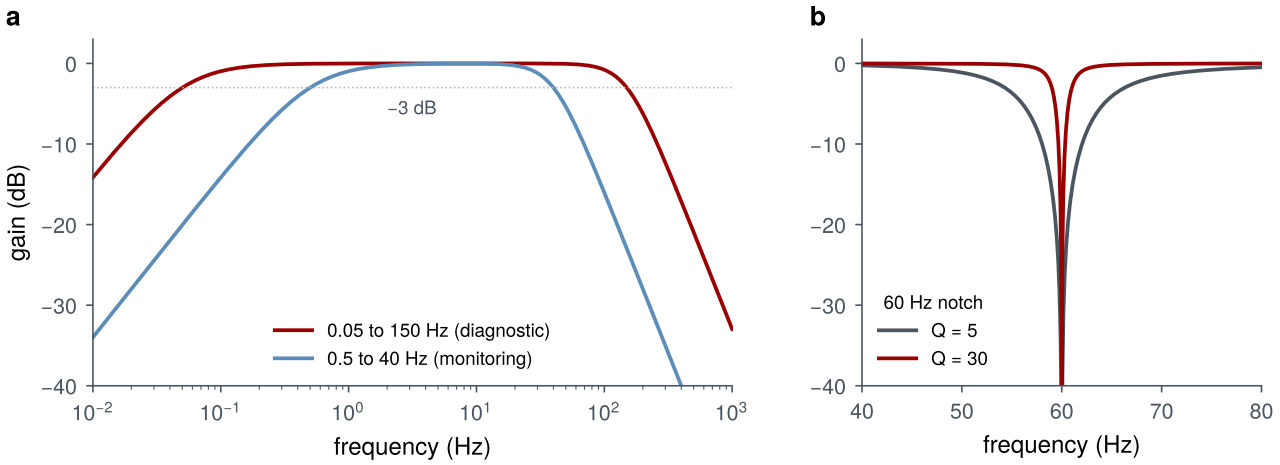


Figure 8.15: (a) Magnitude responses of a diagnostic ECG chain (first-order high-pass at 0.05 Hz, second-order Butterworth low-pass at 150 Hz) and a monitoring chain (0.5 Hz and 40 Hz). (b) Second-order 60 Hz notch filters with $Q = 5$ and $Q = 30$. The higher Q removes a narrower band but rings longer.

AC coupling

A first-order high-pass filter, a series capacitor C and a resistor R to the reference, blocks the electrode offset after the first amplifier stage. This is called *AC coupling*. Its cutoff and time constant are

$$f_c = \frac{1}{2\pi RC}, \qquad \tau = RC = \frac{1}{2\pi f_c}. \tag{8.19}$$

A low cutoff passes the ST segment but recovers slowly: after a motion artifact or an electrode bump drives the stage into saturation, the output returns to the baseline only as $e^{-t/\tau}$. With a 24-bit converter and a low-gain first stage, the analog high-pass can be dropped altogether: the converter digitizes the offset with the signal and a digital high-pass removes it.

Low-pass filters and anti-aliasing

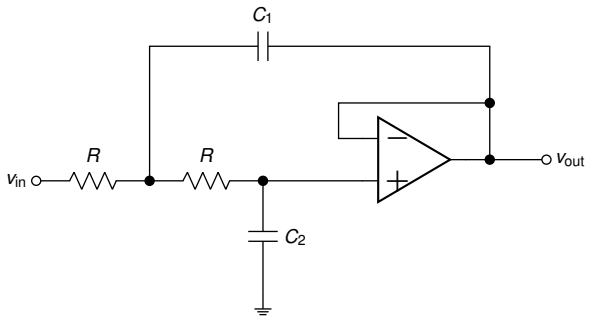


Figure 8.16: Unity-gain Sallen-Key low-pass filter. With equal resistors, $C_1 = 2C_2$ gives a Butterworth response.

The low-pass filter sets the upper edge of the band and, before an ADC, prevents aliasing. A first-order RC falls at 20 dB per decade above its cutoff; a second-order section falls at 40 dB per decade, twice as fast. The unity-gain *Sallen-Key* low-pass (Fig. 8.16) is the usual second-order section: one op amp wired as a follower, two resistors, and two capacitors. With equal resistors R,

$$f_0 = \frac{1}{2\pi R\sqrt{C_1 C_2}}, \qquad Q = \frac{1}{2}\sqrt{\frac{C_1}{C_2}}, \tag{8.20}$$

where C_1 is the capacitor from the middle node to the output and C_2 the one from the op-amp input to ground. The quality factor Q sets the shape near f_0. A *Butterworth* response, flat in the pass band with no peaking, has $Q = 1/\sqrt{2} = 0.707$, so $C_1 = 2C_2$, and its magnitude is $|H| = 1/\sqrt{1 + (f/f_0)^4}$. For $f_0 = 150\,\text{Hz}$ with $C_2 = 10\,\text{nF}$ and $C_1 = 20\,\text{nF}$, $R = 75.0\,\text{k}\Omega$, and the filter passes only 2.2 % (-33 dB) at 1 kHz.

A delta-sigma converter (Section 8.7) samples its input at a rate far above the output data rate and filters digitally, so the analog anti-aliasing filter only has to remove what lies near the converter's internal sampling frequency, hundreds of kilohertz or more. A first-order RC at each input is then enough, and that is what the project boards use.

The mains notch

Even with a right-leg drive some mains hum remains. A *notch* removes a narrow band around $f_0 = 60\,\text{Hz}$ (or 50 Hz). The second-order notch

$$H(s) = \frac{s^2 + \omega_0^2}{s^2 + (\omega_0/Q)\,s + \omega_0^2}, \qquad \text{BW} = \frac{f_0}{Q}, \tag{8.21}$$

with $s = j2\pi f$, is zero at f_0 and passes everything far from it. Its

KEY EQUATION
$\text{BW} = f_0/Q$, $\tau \approx Q/(\pi f_0)$
A sharper notch rings longer.

magnitude is

$$|H(f)| = \frac{|f_0^2 - f^2|}{\sqrt{(f_0^2 - f^2)^2 + (f f_0/Q)^2}}, \tag{8.22}$$

and its -3 dB width is f_0/Q. A high Q removes little of the ECG, but the filter's response to a step or to the onset of hum decays with a time constant of about $Q/(\pi f_0)$, so a sharp notch rings for a long time. A 60 Hz notch with $Q = 30$ is 2 Hz wide, passes 0.90 of the signal at 58 Hz, and rings with $\tau = 159$ ms; with $Q = 5$ it is 12 Hz wide and settles with $\tau = 27$ ms, but passes only 0.32 at 58 Hz. In practice the notch is digital, because an analog notch drifts with its component values; `scipy.signal.iirnotch` builds Eq. (8.21) for a given sampling rate. A notch distorts the ECG near its frequency, so a diagnostic recording keeps it off when the hum is small.

8.7 *Digitizing: Delta-Sigma Converters and Integrated Front Ends*

The ECG must survive a ± 300 mV offset and still be resolved to a microvolt: a ratio of 6×10^5, or about 19.2 bits, before any margin. Biopotential front ends therefore use 24-bit *delta-sigma* ($\Delta\Sigma$) converters, which reach that resolution with very little analog gain in front.

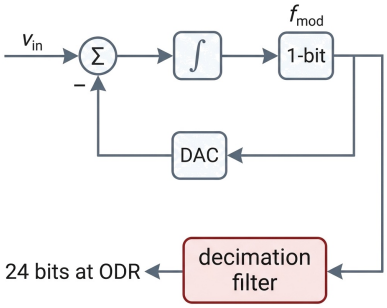


Figure 8.17: A first-order delta-sigma converter: integrator, one-bit quantizer, and DAC in a feedback loop at f_{mod}, followed by a digital decimation filter.

How a delta-sigma converter works

A $\Delta\Sigma$ converter trades speed for precision. Its *modulator* (Fig. 8.17) subtracts from the input the output of a one-bit DAC, which can only be $+V_{ref}$ or $-V_{ref}$, integrates the difference, and decides at each tick of a fast clock f_{mod} whether the integrator is above or below zero. The feedback keeps the integrator near zero, so over many ticks the fraction of ones in the bit stream follows the input: an input of zero gives as many ones as zeros, and an input at three quarters of full scale gives seven ones in eight. Each decision is crude, but the average of thousands of them is precise. The loop also pushes the error of the one-bit decisions toward high frequencies (*noise shaping*). A digital *decimation filter* then averages the stream, removes that high-frequency noise, and delivers multi-bit samples at the output data rate

$$\mathrm{ODR} = \frac{f_{mod}}{\mathrm{OSR}}, \tag{8.23}$$

where OSR, the oversampling ratio, is the number of modulator ticks averaged into each sample. A larger OSR gives a lower rate and less noise. The decimation filter is also an excellent anti-aliasing filter for

everything except frequencies near f_{mod}, which is why the analog filter in front can be a simple RC.

Step size and noise

An N-bit converter with a reference V_{ref} and a programmable-gain amplifier (PGA) of gain G in front accepts inputs from $-V_{\mathrm{ref}}/G$ to $+V_{\mathrm{ref}}/G$ and divides that span into 2^N steps, each

$$\mathrm{LSB} = \frac{2V_{\mathrm{ref}}}{G\,2^N}. \tag{8.24}$$

At 24 bits the step is tens of nanovolts, far below the analog noise of the electrodes and the amplifier, so the noise, not the step, sets the resolution. Datasheets quote the input-referred noise in a stated bandwidth, as an rms value or as a peak-to-peak value; for random (Gaussian) noise the peak-to-peak value seen in a long record is about 6.6 times the rms.

KEY EQUATION

$\mathrm{LSB} = \dfrac{2V_{\mathrm{ref}}}{G\,2^N}$

Step size of a bipolar converter with a gain-G amplifier in front.

Integrated front ends

An integrated analog front end puts the dashed box of Fig. 8.1 on one chip a few millimeters on a side: switches that route any electrode to any channel, an amplifier and a $\Delta\Sigma$ converter per channel, decimation filters, a reference, a clock, the right-leg (bias) drive, lead-off detection, and an SPI port through which a microcontroller sets every option and reads the samples. Table 8.2 compares the two front ends of the design projects. The ADS1293 trades noise for low power and the flexible routing of a 3- or 5-lead ECG; the ADS1299 has eight low-noise channels that sample at the same instant, which is what EEG needs, at forty times the power.

	ADS1293 (wearable ECG)	ADS1299 (wearable EEG)
Channels	3, from 6 inputs routed freely	8 differential, sampled at once
Amplifier gain	fixed, 3.5	programmable, 1 to 24
Input range	±400 mV	$\pm 4.5\,\mathrm{V}/G$ (±187.5 mV at 24)
Input noise	7 µV p-p, 0.1 to 40 Hz (10 µV in low-power mode)	about 1 µV p-p (70 Hz band)
Data rates	853 SPS in TI's examples, 200 in Lab 4	250 to 16 000 SPS
Power	about 0.3 mW per channel	about 39 mW for 8 channels
Extras	RLD, Wilson terminal, lead-off	bias drive, lead-off, test signals
Output code	offset binary	two's complement
SPI mode	0	1

Table 8.2: The two front ends of the design projects, simplified from the datasheets (SPS: samples per second). Lab 3 designs a board around one of them and Lab 4 configures it.

From codes to volts. A converter reports each sample as an integer *code*. Two conventions are common. In *offset binary* (ADS1293) zero volts reads half of the full-scale code, and the code grows with the input.

In *two's complement* (ADS1299) zero volts reads 0, and a negative input comes out as a large number with its top bit set, which the software turns back into a negative count. Either way,

$$v_{\text{in}} = (c - c_0) \times \text{LSB}, \tag{8.25}$$

where c is the code, c_0 the code for zero volts, and LSB the voltage of one step. Lab 4 assembles the bytes of each sample and decodes both formats; mixing them up flips the sign of large inputs.

8.8 Getting the Data Out: SPI and Bluetooth Low Energy

An AFE is a peripheral: it converts, and a microcontroller decides when to read. In the design projects the microcontroller is the Arduino Nicla Sense ME, a board the size of a stamp with an Arm Cortex-M4 processor and a Bluetooth radio. It talks to the AFE over SPI and to the laptop over BLE.

SPI

The *serial peripheral interface* connects one controller (here the microcontroller) to one or more peripherals with four wires. The controller drives the clock SCLK and pulls a chip-select line CS low to talk to one peripheral. Data go out on COPI (controller out, peripheral in; formerly MOSI) and come back on CIPO (controller in, peripheral out; formerly MISO), one bit per clock in both directions at once, most significant bit first. A fifth line, DRDY (data ready), lets the AFE tell the controller that a new sample is waiting. The two sides must also agree on which clock edge samples the data, the SPI *mode*: in mode 0 the clock idles low and data are sampled on rising edges (Fig. 8.18); in mode 1 they are sampled on falling edges. The ADS1293 uses mode 0 and the ADS1299 mode 1. A wrong mode typically shifts every byte by one bit or returns garbage.

A 24-bit sample is three bytes. After DRDY falls, the controller reads every channel of that conversion in one burst, so that no sample mixes bytes of two conversions. For three ECG channels a command byte plus nine data bytes is 80 clocks, 80 µs at an SCLK of 1 MHz, a small fraction of the 5 ms between samples at 200 samples per second. Lab 4 gives the command bytes and registers of both chips. One more rule comes from the wiring: the AFE's digital supply must equal the voltage of the microcontroller's I/O pins, which Lab 3 sets on the board.

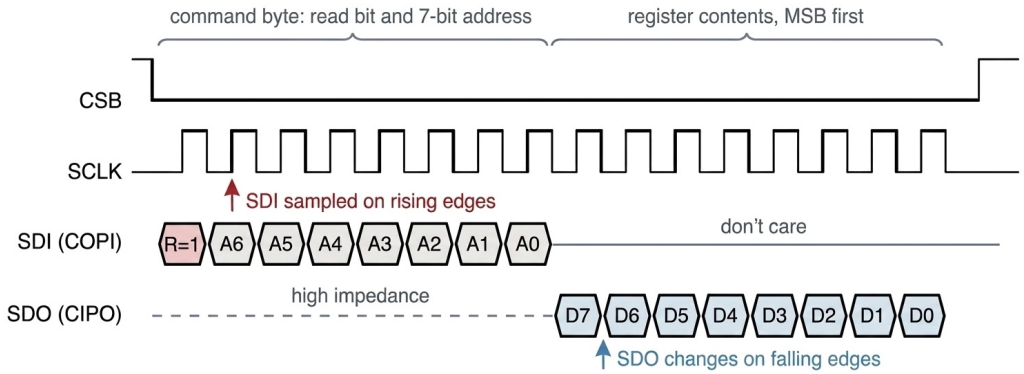


Figure 8.18: Reading one register of a peripheral over SPI in mode 0, as the ADS1293 does it (its pins are CSB, SDI, and SDO). With CSB low, the controller sends a command byte, a read bit and a 7-bit address; during the next eight clocks the peripheral returns the register contents, most significant bit first. Data are sampled on rising edges of SCLK.

Bluetooth Low Energy

BLE splits a link into roles. A *peripheral* (the wearable) advertises its presence a few times per second; a *central* (the phone or laptop) scans, connects, and then sets the rhythm of the connection. Once connected, the two exchange data through the *generic attribute profile* (GATT). The peripheral holds a table of *services*, each identified by a long number (a UUID), and each service holds *characteristics*: a value with properties such as read, write, and notify (Fig. 8.19). When the central subscribes to a characteristic, the peripheral may send *notifications*, new values pushed without a request. Notifications are how a wearable streams samples.

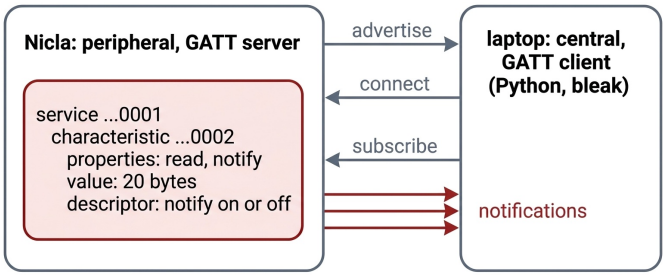


Figure 8.19: A BLE link for streaming. The wearable advertises; the laptop connects and subscribes to a characteristic; from then on each new value is pushed as a notification, at most 20 bytes at a time with the default packet size.

Three numbers set the throughput. The *payload* of a notification is 20 bytes unless the two sides negotiate a larger packet size (the ATT MTU), which both must support. The *connection interval*, between 7.5 ms and 4 s in steps of 1.25 ms, is chosen by the central within limits the peripheral may request. And the number of packets per connection event depends on both radios' firmware. The throughput, and what a stream of N_{ch} channels of b bits at f_s samples per second needs, are

$$\begin{aligned} \text{throughput} &= \frac{n_{\text{per event}} \times \text{payload}}{T_{\text{interval}}}, \\ \text{needed} &= N_{\text{ch}} \times \frac{b}{8} \times f_s, \end{aligned} \tag{8.26}$$

KEY EQUATION

$N_{\text{ch}} \dfrac{b}{8} f_s$ bytes per second

What a stream of N_{ch} channels of b bits at f_s needs.

in bytes per second. A design must leave margin, because the laptop may choose a long interval and the radio loses packets in a crowded room. Measure it: put a sequence number in every notification and count the gaps (Lab 4).

WORKED EXAMPLE 8.5 *An EEG headband from electrode to radio*

An eight-channel EEG headband uses the ADS1299 (Table 8.2) with printed silver electrodes, whose offsets may reach 300 mV, and streams 24-bit samples at 250 samples per second over BLE in 20-byte notifications. (a) Which gain keeps the offset inside the input range, and what is the step size? (b) The input noise is about 1 µV peak to peak. What SNR does a 20 µV peak-to-peak alpha wave get? (c) What data rate does the stream need, and how many notifications per second? (d) The laptop picks a 15 ms connection interval and the link delivers two notifications per event. Does the stream fit? If not, what can the design do?

SOLUTION. (a) The range $\pm 4.5\,\mathrm{V}/G$ must exceed 0.3 V, so $G < 15$: choose 12, with a range of ±375 mV. Eq. (8.24) gives LSB $= 2(4.5)/(12 \times 2^{24}) = 44.7\,\mathrm{nV}$. At gain 24 the range of ±187.5 mV would saturate; gain 24 suits gel Ag/AgCl electrodes, whose offsets stay within tens of millivolts.

(b) The rms noise is about $1\,\mu\mathrm{V}/6.6 = 0.15\,\mu\mathrm{V}$, about three steps. A sine of 20 µV peak to peak is $20/(2\sqrt{2}) = 7.07\,\mu\mathrm{V}$ rms, so SNR $= 20\log_{10}(7.07/0.15) = 33\,\mathrm{dB}$. The ADS1293, with seven times the noise in high-resolution mode and ten in Lab 4's low-power mode, would give 17 to 20 dB less: the EEG needs the ADS1299.

(c) Eq. (8.26): $8 \times 3 \times 250 = 6000$ bytes/s (48 kbit/s), or $6000/20 = 300$ notifications per second. A 15 ms interval gives 66.7 events per second, so 4.5 notifications per event.

(d) Two per event carry $2 \times 20/15\,\mathrm{ms} = 2667$ bytes/s, less than half of what (c) needs. Sending 16-bit samples after removing the offset and halving the rate to 125 samples per second, which still covers 40 Hz, needs $8 \times 2 \times 125 = 2000$ bytes/s and fits. Better still for a stress monitor: compute band powers on the microcontroller and send eight channels times four bands of 2-byte values once a second, 64 bytes/s. Two ECG channels at 200 samples per second need only 1200 bytes/s, about one notification per event. ■

8.9 From Samples to Heart Rate and Band Power

THE SAMPLES THAT ARRIVE ON THE LAPTOP still carry baseline wander, hum, and noise. A few lines of Python remove them

and extract the numbers the projects need. Baseline wander from breathing and motion is removed by a high-pass at about 0.5 Hz, mains hum by a notch, and fast noise by a low-pass at 40 Hz. Offline, `scipy.signal.sosfiltfilt` runs each filter forward and then backward, which cancels its phase shift; a real-time filter on the microcontroller cannot look ahead and delays the signal instead.

Finding R peaks

The classic real-time QRS detector of Pan and Tompkins[7] uses the fact that the QRS complex is the steepest part of the ECG (Fig. 8.20). A band-pass filter from about 5 to 15 Hz keeps the QRS energy and rejects the P and T waves, baseline wander, and hum. A derivative emphasizes the steep slopes, squaring makes every value positive and favors large slopes, and a moving-window average about 150 ms long (the width of a wide QRS) merges each complex into one smooth bump. A threshold that adapts to the heights of recent peaks marks the bumps that are beats, and a *refractory period* of 200 ms prevents counting one beat twice. At 250 samples per second the window is 38 samples and the refractory period 50. The RR intervals then give the heart rate (Eq. (8.4)) and HRV (Eq. (8.5)): R peaks found at samples 101, 309, 516, and 727 are on average 208.7 samples, or 0.835 s, apart, which is 71.9 bpm. Always check a detector against a hand count of the pulse.

[7] J. Pan and W. J. Tompkins, *IEEE Trans. Biomed. Eng.* (1985).

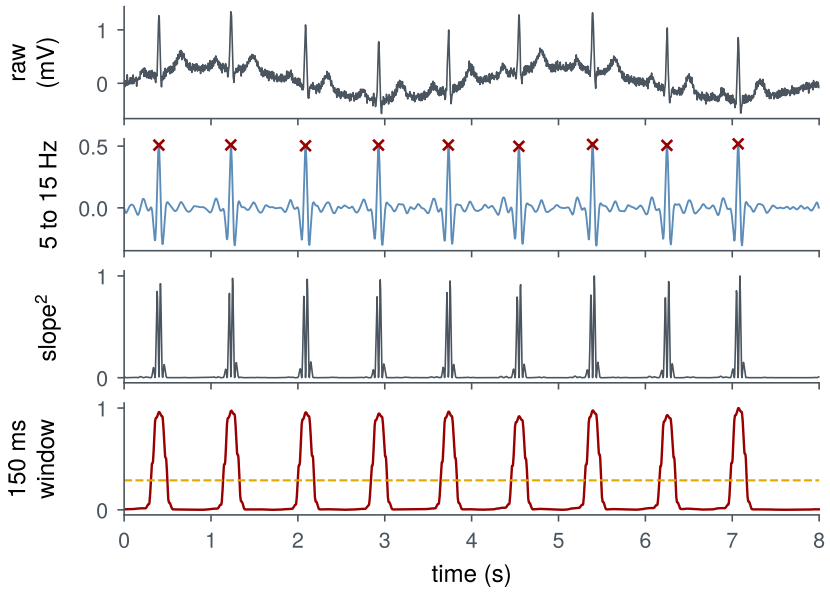


Figure 8.20: A Pan-Tompkins style detector on 8 s of synthetic ECG at 250 samples per second with baseline wander, 60 Hz hum, and noise (illustrative data). From top: the raw signal; the 5 to 15 Hz band-pass, with the R peaks found; the squared slope; the 150 ms moving-window integral with its threshold (dashed). The detector finds all nine beats, 72 bpm.

EEG band power

The power of an EEG band is the area under its *power spectral density* (PSD), the signal power per hertz, between the band edges. *Welch's method* cuts the record into overlapping segments, multiplies each by a smooth window, averages their spectra, and returns the PSD in $\mu V^2/Hz$; with segments T seconds long the frequency bins are $1/T$ apart (`scipy.signal.welch`). Integrating over 8 to 13 Hz gives the alpha power in μV^2, and its square root is the rms amplitude of the alpha rhythm. *Relative* power divides a band by the total over, say, 0.5 to 30 Hz, which cancels differences in electrode contact between sessions.

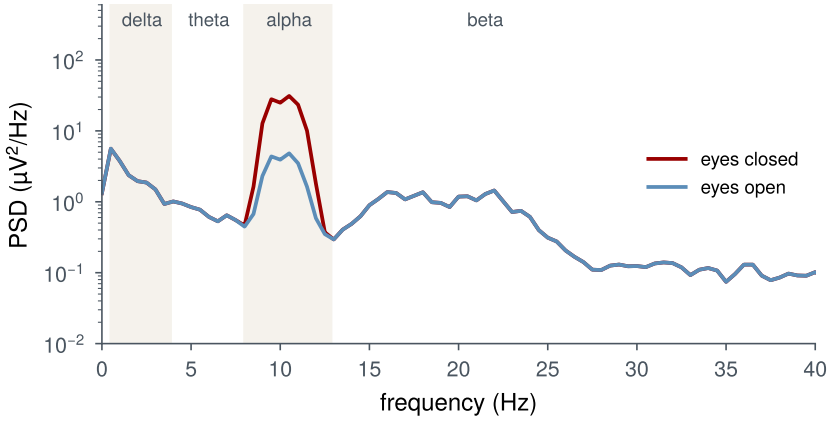


Figure 8.21: Welch PSD (2 s segments, 0.5 Hz bins) of one EEG channel with the eyes open and closed (illustrative data). Closing the eyes raises the alpha band by a factor of six; the other bands are unchanged.

Before trusting any EEG index, check the one effect everyone shows: alpha power rises when the eyes close (Fig. 8.21). In that recording, 2 s segments at 250 samples per second hold 500 samples and give bins 0.5 Hz apart, ten of them in the alpha band. Closing the eyes raises the alpha power from 11 to 67 μV^2 while beta stays at 12 μV^2, so the alpha/beta ratio goes from 0.92 to 5.6, and the rms alpha amplitude with the eyes closed is $\sqrt{67} = 8.2\,\mu V$. The wearable EEG project of Chapter 14, which detects relaxation or stress, can use this ratio: relaxation raises it and mental effort or anxiety lowers it. It is a crude index that varies from person to person, and blinks, jaw clenching, and electrode motion all push power into the bands, which is why such an index is averaged over minutes.

8.10 *Electrical Safety*

A BIOPOTENTIAL AMPLIFIER is wired to a person through gel and low-impedance electrodes. That is what makes it work, and it is also

what makes it dangerous if anything in the system touches the mains.

What current does to the body

The skin's resistance normally protects us; gel electrodes are designed to remove that protection. For 60 Hz current flowing hand to hand for one to three seconds, about 1 mA is the threshold of perception, about 10 mA (6 to 16 mA) the let-go current above which muscles clamp onto the conductor, and roughly 75 to 400 mA can start ventricular fibrillation, the chaotic beating that stops the heart from pumping.[8] These are *macroshock* levels, current entering through the skin and spreading through the body. Current delivered straight to the heart through a catheter (*microshock*) is dangerous at far lower levels, which is why the limit for equipment with a cardiac connection is 10 µA.

[8] J. G. Webster, ed., *Medical Instrumentation: Application and Design* (Wiley, 2009), chapter on electrical safety.

Suppose a mains-powered circuit develops a fault that puts 120 V on the amplifier common, and the person, wired by a chest electrode (10 kΩ), touches a grounded metal frame with a wet hand (1 kΩ) through about 500 Ω of body. The current is $120\,\text{V}/11.5\,\text{k}\Omega = 10\,\text{mA}$, the let-go level. A battery-powered circuit at 3.7 V cannot drive more than $3.7\,\text{V}/11.5\,\text{k}\Omega = 0.3\,\text{mA}$ through the same path, and only if two of its nodes were connected to the person directly.

Leakage-current limits

IEC 60601-1, the standard for medical electrical equipment, classifies the parts that touch the patient (*applied parts*) and limits the current that may flow through them (Table 8.3). Type BF parts touch the body surface, as ECG and EEG electrodes do; type CF parts may contact the heart. *Normal condition* means the equipment is intact; *single fault* means one protective means has failed, such as a broken ground wire. Chapter 14 covers how a device is tested and cleared.

Current	Normal condition	Single fault
Touch current (enclosure)	100 µA	500 µA
Patient leakage, type B and BF (AC)	100 µA	500 µA
Patient leakage, type CF (AC)	10 µA	50 µA
Earth leakage	5 mA	10 mA

Table 8.3: Leakage-current limits of IEC 60601-1 (third edition) for currents up to 1 kHz. The DC limits for patient leakage are lower; see the standard for the full table.

Design strategies

Three strategies keep a biopotential system within those limits. Run it from a battery, so that no path leads to the mains. Isolate anything that must connect to a mains-powered device, with an isolated DC-DC converter for power and digital isolators or an optical or radio link for data; a BLE link is an isolation barrier. And limit the current that

any single fault can drive into the person, with resistors in series with the electrodes and the right-leg drive (Example 8.4). A laptop on its charger is a mains-powered device: its USB ground is coupled to the mains through the charger. That is the reason for the rule of every measurement in this book with electrodes on a person: *never connect a person to a circuit that is connected, by any cable, to the mains*. Run the board from its battery, unplug the laptop's charger, and prefer the BLE link.

The battery also sets the run time, its capacity divided by the average current. The radio dominates an ECG patch, the ADS1299's 39 mW an EEG headband; Chapter 13 builds the power budget.

Summary

- Biopotentials are membrane currents summed in the body's conducting fluid: ECG 0.5 to 4 mV, EEG 5 to 300 µV, EMG 10 µV to 5 mV up to 500 Hz, EOG down to DC. EEG rhythms run from delta (0.5 to 4 Hz) to beta (13 to 30 Hz); alpha (8 to 13 Hz) rises when the eyes close.
- A lead reads the projection of the cardiac vector: $\mathrm{I} = \Phi_{\mathrm{LA}} - \Phi_{\mathrm{RA}}$, $\mathrm{II} = \Phi_{\mathrm{LL}} - \Phi_{\mathrm{RA}}$, $\mathrm{III} = \Phi_{\mathrm{LL}} - \Phi_{\mathrm{LA}}$, and $\mathrm{II} = \mathrm{I} + \mathrm{III}$; augmented leads use the average of two limbs, chest leads the Wilson terminal $\frac{1}{3}(\Phi_{\mathrm{RA}} + \Phi_{\mathrm{LA}} + \Phi_{\mathrm{LL}})$. The ECG shows the P wave (atria), the QRS (ventricles, about 1 mV), and the T wave; $\mathrm{HR} = 60/\overline{RR}$, and HRV is measured by SDNN and RMSSD.
- The electrode is E_{hc} in series with $R_d \parallel C_d$ and R_s; Ag/AgCl is nearly nonpolarizable. Offsets reach ±300 mV; dry electrodes have ten to a hundred times the impedance of wet ones.
- Mains couples $i_d = V\omega C_p$ (a fraction of a microampere) into the body, which sits at $v_{\mathrm{cm}} = i_d Z_G$. Electrode mismatch converts it into $v_d \approx v_{\mathrm{cm}}\,\Delta Z / Z_{\mathrm{cm}}$, which no CMRR can remove.
- Three-op-amp IA: $G = (1 + 2R/R_G)(R_4/R_3)$ and $\mathrm{CMRR} \approx G_1(1 + G_d)/(4\epsilon)$; the electrode offset limits G_1. Right-leg drive: $v_{\mathrm{cm}} = i_d(Z_{\mathrm{RL}} + R_o)/(1 + G)$, with R_o limiting the fault current.
- Bands: diagnostic ECG 0.05 to 150 Hz, monitoring 0.5 to 40 Hz. A high-pass recovers with $\tau = RC$; a Sallen-Key Butterworth has $C_1 = 2C_2$ and $f_0 = 1/(2\pi R\sqrt{C_1 C_2})$; a notch has $\mathrm{BW} = f_0/Q$ and rings with $\tau \approx Q/(\pi f_0)$.
- A ΔΣ converter averages a fast one-bit stream: $\mathrm{ODR} = f_{\mathrm{mod}}/\mathrm{OSR}$ and $\mathrm{LSB} = 2V_{\mathrm{ref}}/(G\,2^N)$, tens of nanovolts at 24 bits, so noise sets the resolution. Integrated front ends (ADS1293, ADS1299) hold the whole analog chain on one chip.

- SPI moves bytes between AFE and microcontroller on four wires plus DRDY, in a mode both sides share. BLE throughput $n \times \text{payload}/T_{\text{interval}}$ (20-byte payloads by default) must exceed $N_{\text{ch}}(b/8)f_s$.
- Processing: band-pass, notch, an R-peak detector (5 to 15 Hz, slope, square, 150 ms window, 200 ms refractory period); Welch band powers and the alpha/beta ratio.
- Safety: about 1 mA perception, 10 mA let-go, 75 to 400 mA fibrillation; type BF limit 100 µA normal, 500 µA single fault. Run on batteries, isolate, limit current, and never connect a person to a mains-connected circuit.

This chapter closes Part II. The biopotentials here are recorded through the skin, smeared by tissue and bone. Chapter 9 opens Part III by moving the electrodes onto and into the nervous system, where they record single neurons and also stimulate them.

8.11 Problems

8.1 ◦ At the peak of the R wave a recording gives lead I $= 0.80$ mV and lead III $= 0.20$ mV.

(a) Find lead II, aVR, aVL, and aVF.

(b) Find the mean electrical axis and the size of the cardiac vector. Is the axis normal?

8.2 • The body sits at $v_{\text{cm}} = 5$ mV at 50 Hz. The electrodes are 15 kΩ and 45 kΩ, and each amplifier input has 1 GΩ in parallel with 20 pF to the common.

(a) Find $|Z_{\text{cm}}|$ and the differential hum at the input.

(b) What input capacitance would bring the hum below 0.1 µV?

8.3 •• Splitting the gain of a three-op-amp IA.

(a) Explain why the input stage multiplies the CMRR of the difference stage by G_1, and show that the whole IA has CMRR $\approx G_1(1+G_d)/(4\epsilon)$ for resistor tolerance ϵ.

(b) The IA needs a total gain of 10 with 0.1 % resistors. Find the CMRR in decibels for $(G_1, G_d) = (10, 1)$, $(5, 2)$, and $(2, 5)$.

(c) On ±5 V supplies (output swing ±4.5 V) with a 300 mV electrode offset, which split do you choose, and why?

8.4 • An ADS1299 channel runs at gain 24 with its 4.5 V reference.

(a) Find the full-scale input range and the step size (LSB).

(b) The modulator runs at $f_{\text{mod}} = 1.024\,\text{MHz}$. What oversampling ratio gives 250 samples per second?

(c) After decoding, a sample reads a signed count of -3000. What is the input voltage?

(d) Why can the ADS1293, at its fixed gain of 3.5, accept a 300 mV electrode offset that saturates this channel?

8.5 ◦ Electrical safety.

(a) Which IEC 60601-1 patient-leakage limits apply to a chest ECG patch?

(b) A student records an ECG with the Nicla on USB to a laptop that is charging. Explain the hazard and give two fixes.

(c) What series resistance keeps the current from any electrode below 10 µA even if a 3.3 V rail is shorted to it?

Part III

Interfacing with the Nervous System

9
Neural Recording and Stimulation

In this chapter you will learn to

- Describe how pacemakers, deep brain stimulators, and brain-machine interfaces stimulate or record, and what a closed-loop device adds.
- Compare EEG, ECoG, local field potentials, and single units by amplitude, bandwidth, spatial scale, and invasiveness.
- Explain why stiff electrodes fail in the brain, and compare the bending stiffness of a silicon needle and a polymer thread.
- Compute the data rate of a many-channel recorder and the savings of spike detection.
- Design an AC-coupled neural amplifier and add up its noise budget.
- Estimate how much power the heat limit allows a small implant.
- Model the electrode-tissue interface and explain capacitive and faradaic charge injection.
- Choose a charge-balanced stimulation waveform, check it against the Shannon limit and the electrode's charge injection capacity, and size the compliance voltage of a current-source stimulator.

In 2012 a woman who had been paralyzed for fifteen years reached for a bottle of coffee with a robotic arm, brought it to her mouth, and drank from it. She controlled the arm with her thoughts. A square of silicon smaller than a fingernail, carrying 96 needle electrodes, sat in the motor cortex of her brain. Each needle picked up the electrical spikes of a few nearby neurons, a cable through her skull carried the signals to a rack of computers, and a decoder turned the

pattern of spikes into arm movements.[1] A second family of devices runs the other way. A deep brain stimulator, a thin lead with four contacts placed deep in the brain and wired to a pulse generator in the chest, sends trains of current pulses that quiet the tremor of Parkinson's disease. Both devices touch neurons through an electrode, and both must do it for years without hurting the tissue or wearing out.

[1] L. R. Hochberg *et al.*, *Nature* (2012).

This chapter opens Part III, on the nervous system. Chapter 8 recorded the summed activity of millions of cells through the skin; here the electrode moves to the brain and the nerves themselves, close enough to hear single neurons, and it can also send current back. The chapter is about that electrode and the circuits behind it. We start with the devices already in clinical use, from the cardiac pacemaker to deep brain stimulators and brain-machine interfaces. Then we follow the recording path: the signals you can pick up from the scalp to the inside of the cortex, the electrodes that reach them, the data that thousands of channels produce, the low-noise amplifier that every channel needs, and the heat limit on how small an implant can be. The second half turns to stimulation: what the electrode looks like to a stimulator, why pulses must be charge balanced, how much charge is safe, and how a current-source stimulator is built. The chapter uses how neurons fire (Chapter 3), op amps and noise (Chapter 4), the electrode-electrolyte interface (Chapter 6), and biopotential amplifiers (Chapter 8). Chapter 10 applies the same ideas to hearing and sight.

9.1 *Neural Interfaces in the Clinic*

A *neural interface* is a device that exchanges information with the nervous system through electrodes. It can *record* (read the electrical activity of neurons), *stimulate* (inject charge to make neurons fire or to block their firing), or both. A device that records, decides, and then stimulates on its own is *closed loop*. Stimulation works because an excitable cell fires when its membrane is pushed past threshold (Chapter 3), and current driven through a nearby electrode can do the pushing. Applied to heart muscle, this idea gave the first electronic implant.

From the pacemaker to deep brain stimulation

The cardiac *pacemaker* of Chapter 1, first fully implanted in 1958, is the model for every stimulator that followed. It listens to the heart's rhythm and, when the heart pauses or beats too slowly, sends a pulse of a few volts lasting a fraction of a millisecond that starts a contraction: it stimulates *on demand*. A pulse that small is enough because the electrode touches the muscle it excites, and the sealed generator runs

for about a decade on one battery before it is replaced.

Deep brain stimulation (DBS), first performed in 1987, is a pacemaker for the brain (Fig. 9.1, right). A lead 1.27 mm in diameter, with four platinum-iridium ring contacts 1.5 mm long, is passed through a small hole in the skull to a deep target such as the subthalamic nucleus. A pulse generator in the chest sends it 130 to 185 pulses per second, each 60 to 90 µs long, at a few volts or milliamperes. DBS treats Parkinson's disease, essential tremor, and dystonia; why high-frequency pulses relieve the symptoms is still debated.[2] A non-rechargeable DBS battery is replaced surgically every 3 to 5 years; a rechargeable one lasts longer but must be charged through the skin (Chapter 13).

Other stimulators work on nerves outside the brain. A *cuff electrode*, a soft sleeve with metal contacts wrapped around a nerve, records or stimulates all of its fibers at once. *Vagus nerve stimulation* (VNS) uses a helical cuff around the vagus nerve in the neck to treat epilepsy and depression, and *spinal cord stimulators* treat chronic pain with leads in the spinal canal. Cochlear and retinal implants stimulate the senses (Chapter 10). A closed-loop device already in clinical use is the responsive neurostimulator for epilepsy, which detects the onset of a seizure in its recordings and stimulates to stop it.[3]

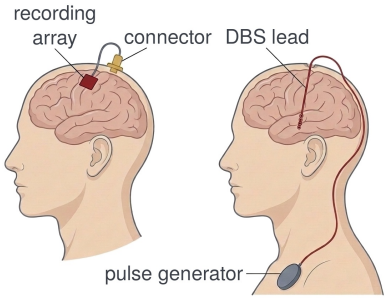


Figure 9.1: Two clinical neural interfaces. Left: a recording array in the motor cortex with a connector on the skull, as used in brain-machine interface trials. Right: a deep brain stimulator, a lead in a deep brain structure tunneled under the skin to a pulse generator in the chest.

[2] A. M. Lozano *et al.*, *Nat. Rev. Neurol.* (2019).

[3] T. L. Skarpaas, B. Jarosiewicz, and M. J. Morrell, *Epilepsy Res.* (2019).

Brain-machine interfaces

A *brain-machine interface* (BMI), also called a brain-computer interface, records the activity of neurons, decodes what it means, and uses it to drive a computer cursor, a robotic arm, a speech synthesizer, or a person's own paralyzed muscles (Fig. 9.1, left). Each neuron in the motor cortex fires fastest for movements in its own preferred direction, so the firing rates of dozens of neurons, read together, predict the intended movement. In the BrainGate trials people with tetraplegia (paralysis of all four limbs) moved cursors and robotic arms with 96-electrode arrays in the motor cortex, and later trials decoded attempted handwriting and speech into text.[4] Noninvasive BMIs that read the EEG from the scalp carry far less information but need no surgery; in 2021 the FDA authorized one that drives a hand orthosis for stroke rehabilitation. An implanted BMI is a class III device, approved only after clinical studies (Chapter 14).

Figure 9.2 shows the building blocks of a fully implanted, wireless interface that both records and stimulates.[5] On the recording path a low-noise amplifier (LNA) boosts the microvolt signals of the electrodes, an analog-to-digital converter (ADC) digitizes them, and a transmitter sends them out through the skin. On the stimulation path a receiver takes commands, a controller sets the pulse timing, and a current digital-to-analog converter (DAC) drives the electrodes. A battery,

[4] L. R. Hochberg *et al.*, *Nature* (2006); F. R. Willett *et al.*, *Nature* (2021) and (2023).

[5] The block diagram and the three trends below follow the tutorial by R. Muller, M. M. Ghanbari, and A. Zhou, "Miniaturized wireless neural interfaces: a tutorial," *IEEE Solid-State Circuits Mag.* (2021).

a receiving coil, or an energy harvester powers everything. In a closed-loop device a processor (the dashed block) classifies the recorded signals on the implant and triggers stimulation without waiting for an outside computer.

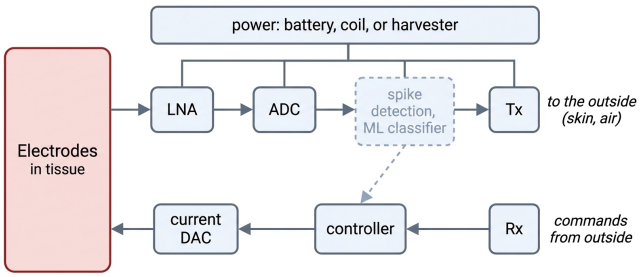


Figure 9.2: Building blocks of an implanted wireless neural interface. The recording path (top) amplifies and digitizes the electrode signals and sends them out; the stimulation path (bottom) turns commands into current pulses. The dashed block (ML, machine learning) closes the loop: the implant decides when to stimulate from what it records. After the tutorial by Muller, Ghanbari, and Zhou.

Three trends drive the engineering. *Scale*: integrate electrodes with chips so the channel count can grow from about a hundred to many thousands. *Closing the loop*: process signals on the implant, so it can act in milliseconds and send out decisions instead of raw data. *Miniaturization*: shrink the whole implant to a few millimeters, so it can be injected or placed through a blood vessel and no wire crosses the skin, since such a wire is a path for infection.

9.2 *Recording by Depth*

The closer the electrode is to a neuron, the larger and the more local the signal it records. Figure 9.3 shows where electrodes can sit, from outside the skull to inside the cortex, and Table 9.1 lists what each one sees.

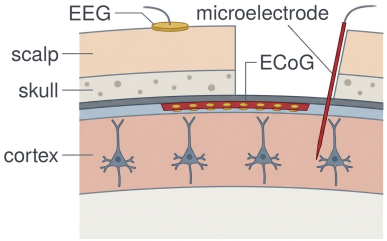


Figure 9.3: Recording depth. EEG electrodes sit on the scalp; ECoG grids lie on the brain under the skull and dura; penetrating microelectrodes record local field potentials and single-neuron spikes inside the cortex.

Electroencephalography (EEG) records from the scalp the summed synaptic currents of millions of neurons, smeared by the poorly conducting skull (Chapter 8 builds an EEG amplifier). It needs no surgery and is safe for daily use, but its signals are small, blurred over centimeters, and mixed with artifacts from muscles, eyes, and motion. *Electrocorticography* (ECoG) places a grid of disk electrodes on the cortex under the dura, the tough membrane around the brain, after a craniotomy (the removal of a piece of skull). Clinical grids with contacts about a centimeter apart map the source of seizures before epilepsy surgery. Without the skull in the way the signal is about ten times larger, sharper in space, and rich in high frequencies. ECoG still sees populations of neurons, not single cells, and the surgery carries risks of infection and bleeding.

Inside the brain a penetrating microelectrode records two signals at once. *Local field potentials* (LFPs) are the low-frequency part of the voltage, from synaptic currents within a few hundred micrometers.

Single units are the spikes of individual neurons within about 50 to 100 µm of the electrode: about a millisecond long and 50 to 500 µV, so they need a bandwidth of roughly 300 Hz to 7 kHz. An electrode that picks up several neurons records *multi-unit activity*, and spike-sorting software assigns the spikes to cells by their shape. Penetrating electrodes give the most information per channel, but they need the largest surgery and they provoke a scar that degrades the recordings over months (Section 9.3). Electrodes in the brain can also record chemistry: a carbon-fiber microelectrode swept quickly in voltage oxidizes dopamine and serotonin and reads their concentration from the current (Chapter 6).

Modality	Amplitude	Bandwidth	Spatial scale	Invasiveness
EEG	5 to 300 µV	0.5 to 100 Hz	centimeters	none (scalp)
ECoG	10 µV to a few mV	0.5 to 500 Hz	1 to 10 mm	craniotomy, on cortex
LFP	0.1 to 1 mV	below 300 Hz	0.1 to 1 mm	penetrating
Single unit	50 to 500 µV	0.3 to 7 kHz	50 to 100 µm	penetrating

Table 9.1: Neural recording modalities, from the outside in. Values are typical ranges, not limits; the EEG and ECoG amplitudes follow the classic table in Webster's *Medical Instrumentation*.

KEY IDEA Moving the electrode inward makes the signal larger and more local, and the surgery larger. Every modality is a point on that trade.

Inside and outside the cell

An electrode *inside* a neuron records the membrane potential of Chapter 3, about 100 mV from rest to peak. An implanted microelectrode sits *outside*, in the extracellular fluid, and sees only the small voltage drop that the membrane currents create as they flow through the well-conducting fluid. Its shape follows the membrane current: near the cell body the inward sodium current comes first, so the extracellular spike starts with a sharp negative trough of about 100 µV lasting a few tenths of a millisecond, followed by a small positive bump (Fig. 9.4). The extracellular spike is a thousand times smaller, 100 mV against 100 µV, or $20 \log_{10} 1000 = 60$ dB: an extracellular amplifier needs about 60 dB more gain to produce the same output.

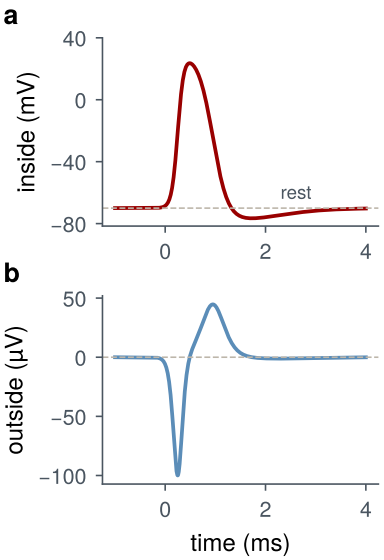


Figure 9.4: One action potential seen two ways (illustrative waveforms). (a) Inside the cell, about 100 mV. (b) Outside, near the cell body, a negative trough of about 100 µV that follows the membrane current. Note the factor of 1000 between the scales.

9.3 *Electrodes Inside the Brain*

RECORDING SINGLE NEURONS IN PEOPLE became routine with one device, the *Utah array*, developed by Richard Normann's group at the University of Utah.[6] It is a 10×10 grid of tapered silicon needles on a 400 µm pitch, 1.0 or 1.5 mm long, on a base about 4 mm square (Fig. 9.5). Each needle is one electrode: it is insulated with parylene except at its sharp tip, which is coated with platinum or iridium oxide and records the neurons around it. In most arrays 96 of the 100 needles are wired through a bundle of fine wires to a connector on the skull. The array is made more by machining than by *lithography*, the

[6] P. K. Campbell, K. E. Jones, and R. A. Normann, *Biomed. Sci. Instrum.* (1990); C. T. Nordhausen, E. M. Maynard, and R. A. Normann, *Brain Res.* (1996).

patterning through a mask with light by which chips are made (Chapter 12 lists its steps): a dicing saw cuts deep grooves into a thick silicon wafer, leaving 100 square posts, and a wet etch sharpens the posts into needles. A pneumatic inserter drives the array into the cortex at high speed, since pushing slowly would dimple the brain.

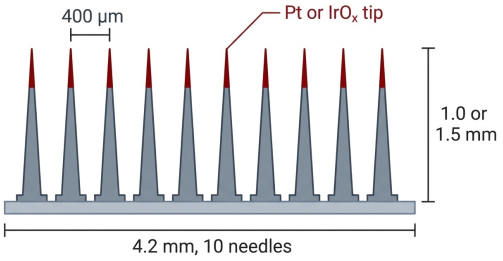


Figure 9.5: Side view of a Utah array (one row of ten needles, to scale in the horizontal direction). The silicon needles are insulated except at the metallized tips (red).

Why rigid electrodes fail

A Utah array can record single units for years in some people, but the number of working channels and the spike amplitudes usually fall over months.[7] The body treats the array as a foreign object. Insertion tears small blood vessels and starts an inflammatory response. The immune cells of the brain, microglia and astrocytes, gather around each needle and over weeks form a dense *glial scar* that insulates the electrode and pushes the nearest neurons away; some neurons near the tip die. The insulation can crack or peel, and the tip metal can corrode, especially if the electrode is also used for stimulation. And the brain moves: it pulses with each heartbeat and shifts with breathing and head motion by several to tens of micrometers relative to the skull, while the array is tethered to the skull by its cable. This *micromotion* rubs the stiff needles against the tissue all the time.

[7] J. C. Barrese *et al.*, *J. Neural Eng.* (2013), analyzed failure modes of silicon arrays in nonhuman primates.

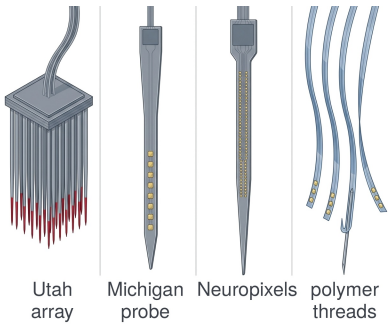


Figure 9.6: Four kinds of intracortical electrodes, from left: a Utah array; a Michigan-style planar shank; a Neuropixels CMOS shank with dense sites; flexible polymer threads.

The root of the mechanical problem is the *modulus mismatch* of Chapter 1: brain tissue has a Young's modulus of order 1 kPa, silicon of order 150 GPa, eight orders of magnitude stiffer. A flexible electrode bends with the brain instead of cutting into it. For a slender probe what matters is its *bending stiffness*, the product EI of the modulus E and the *second moment of area* I of its cross section, which measures how much material sits far from the center of the beam. For a rectangle of width w and thickness t, and for a circle of diameter d,

$$I_{\text{rect}} = \frac{wt^3}{12}, \qquad I_{\text{circ}} = \frac{\pi d^4}{64}. \tag{9.1}$$

A probe of length L clamped at its base, like a diving board (a *cantilever*), deflects by δ under a sideways force F at its tip with

$$F = \frac{3EI}{L^3}\,\delta. \tag{9.2}$$

Because t enters as t^3, the law Chapter 1 found for sheets, making a probe thin helps far more than choosing a softer material. Chapter 11 returns to bending in layered films.

KEY EQUATION

$F = 3EI\delta/L^3$

Tip force of a cantilever. $I = wt^3/12$, so thickness dominates the bending stiffness.

Worked Example 9.1 *Stiff needle, soft thread*

A Utah needle is roughly a silicon cylinder 80 µm in diameter and 1.5 mm long ($E = 160$ GPa). A thread electrode is a polyimide ribbon 20 µm wide and 5 µm thick ($E = 2.5$ GPa) of the same length. (a) Compare their bending stiffness. (b) The brain moves 10 µm sideways relative to the clamped base of each. What force does each exert on the tissue, and what pressure is that over its side?

Solution. (a) From Eq. (9.1), the needle has $I = \pi(80 \times 10^{-6}\,\mathrm{m})^4/64 = 2.01 \times 10^{-18}\,\mathrm{m}^4$, so $EI = 3.22 \times 10^{-7}\,\mathrm{N\,m^2}$. The thread has $I = (20 \times 10^{-6}\,\mathrm{m})(5 \times 10^{-6}\,\mathrm{m})^3/12 = 2.08 \times 10^{-22}\,\mathrm{m}^4$, so $EI = 5.21 \times 10^{-13}\,\mathrm{N\,m^2}$. The needle is about 6×10^5 times stiffer, although silicon is only 64 times stiffer than polyimide: most of the difference comes from the thickness.

(b) From Eq. (9.2) with $L = 1.5$ mm and $\delta = 10$ µm,

$$F_{\text{needle}} = \frac{3(3.22 \times 10^{-7}\,\mathrm{N\,m^2})(1 \times 10^{-5}\,\mathrm{m})}{(1.5 \times 10^{-3}\,\mathrm{m})^3} = 2.9\,\mathrm{mN},$$

and in the same way $F_{\text{thread}} = 4.6$ nN. Spread over the side of the needle (80 µm × 1.5 mm), 2.9 mN is a pressure of about 24 kPa, more than ten times the modulus of the brain: the tissue, not the needle, has to give. The thread pushes with about 0.15 Pa and simply moves with the brain. ■

Planar shanks and Neuropixels

The *Michigan probe*, first developed at the University of Michigan, is the other classic silicon design: a flat shank patterned by lithography, with many sites along its length, so it samples many depths of the cortex at once (Fig. 9.6). *Neuropixels* probes build the amplifiers and switches into the shank itself, in a standard chip (CMOS) process. The first version has 960 sites along a 10 mm shank, any 384 of which can be recorded at once; Neuropixels 2.0 puts 5120 sites on four shanks, each only 70 × 24 µm in cross section.[8] These probes are research tools for animals and for short recordings in people during surgery.

[8] N. A. Steinmetz *et al.*, *Science* (2021).

Flexible threads

If stiffness causes the damage, make the electrodes thin and soft: ultrathin polyimide threads, conformable films for the brain surface, and mesh electronics so flexible that they can be injected through a syringe

(their materials are the subject of Chapter 11). Some groups go further and place electrodes inside a blood vessel on a stent, so that no craniotomy is needed.[9]

A thread a few micrometers thick buckles instead of entering the brain; the company Neuralink solved this with a surgical robot.[10] Each polyimide thread, 4 to 6 µm thick, carried 32 electrode contacts in that 2019 system. A loop at its end hooks onto a fine tungsten needle; the camera-guided robot pushes needle and thread into the cortex between surface blood vessels and withdraws the needle, leaving the thread behind. The clinical implant, the N1, has 64 threads of 16 electrodes each (1024 in all) and a sealed package 23 mm across and 8 mm thick that sits flush with the skull and holds the amplifiers, a battery, and a radio; nothing crosses the skin. The first person received it in January 2024. In the following weeks a number of threads retracted from the brain and the number of useful electrodes fell; the company recovered, and then exceeded, the original cursor-control performance by changing its recording and decoding software.[11] Thin threads solve the stiffness problem, but they must stay where they were put.

[9] Injectable mesh: T. Zhou *et al.*, *Proc. Natl. Acad. Sci. USA* (2017). Stent electrodes: T. J. Oxley *et al.*, *Nat. Biotechnol.* (2016).

[10] E. Musk and Neuralink, "An integrated brain-machine interface platform with thousands of channels," *J. Med. Internet Res.* (2019).

[11] Neuralink, "PRIME study progress update: user experience" (May 2024).

Impedance of a small contact

The impedance of a recording contact matters twice: its thermal noise adds to the signal (Section 9.5), and with the amplifier's input impedance it forms a voltage divider. For a small contact the impedance at 1 kHz, the frequency where it is usually quoted, is set by the double-layer capacitance of the interface (Chapter 6). A capacitance per unit area c_{dl} on a contact of area A gives

$$|Z| \approx \frac{1}{2\pi f \, c_{dl} A}, \tag{9.3}$$

so shrinking the contact raises its impedance in proportion. A 14×24 µm thread contact of smooth metal ($c_{dl} \approx 20\,\mu\text{F/cm}^2$) has $C = 67$ pF and $|Z| \approx 2.4$ MΩ at 1 kHz. Rough or porous coatings fight back by raising the real surface area: iridium oxide stores charge in surface redox reactions, and the conducting polymer PEDOT:PSS (Chapter 11) swells with electrolyte so that ions charge it through its whole volume. Neuralink's PEDOT:PSS contacts measured 37 kΩ, as if they had 64 times the capacitance of smooth metal. The spreading resistance of the tissue in series, $R_s = \rho/(4a)$ for a disk of radius a (Chapter 6), is about 17 kΩ for a disk of the same area in saline ($\rho \approx 70\,\Omega\,\text{cm}$), and no coating can lower it.

9.4 *Scaling Up: Channels and Data*

The human brain has about 86 billion neurons, and the best implants record from a few hundred to a few thousand of them at a time. The step that made scaling possible was putting the electronics next to the electrodes (Fig. 9.7). A 100-electrode Utah array with a matching chip that amplified, digitized, and transmitted the signals wirelessly freed animals from cables. Today Neuralink's chip handles 1024 channels in a sealed package, Neuropixels 2.0 builds 5120 sites over its circuits, and the Argo system of Paradromics bonds more than 65,000 wire electrodes to an amplifier array.[12]

[12] The wireless 100-channel chip: R. R. Harrison *et al.*, *IEEE J. Solid-State Circuits* (2007). Argo: K. Sahasrabuddhe *et al.*, *J. Neural Eng.* (2021).

Data rate

Every channel produces samples whether or not a neuron fires. With N channels sampled at f_s with b bits each, the raw data rate is

$$R_{\text{raw}} = N f_s b. \tag{9.4}$$

KEY EQUATION

$R_{\text{raw}} = N f_s b$

Raw data rate: channels times sample rate times bits per sample.

Spikes need f_s of about 20 kS/s, safely above twice their 7 kHz bandwidth (the Nyquist rate of Chapter 1), and about 10 bits. For a thousand channels that is hundreds of megabits per second, far more than a radio in a small implant can send on its power budget. The implant must reduce the data before it transmits. In increasing order of savings, it can *band-limit* (send only the LFP band, at about 1 kS/s), *detect spikes* (send only short "snippets" of the waveform around each spike, or only the spike times), or *bin* (send the number of spikes per channel in bins of 10 to 50 ms, which is what cursor decoders use).

A spike detector compares each sample with a threshold, usually 4 to 5 times the rms noise of that channel and on the negative side, since extracellular spikes start with a negative trough. Noise alone almost never crosses such a threshold, while a 100 µV spike on a few microvolts of noise crosses it easily.

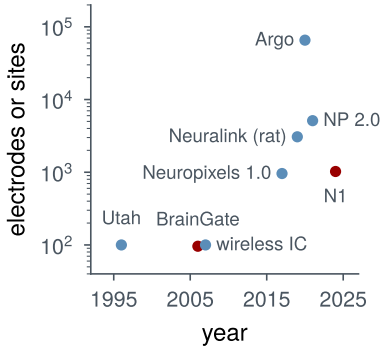


Figure 9.7: Electrode or site counts of selected recording devices, as reported by their developers. Red: implanted in people. A Neuropixels probe records 384 of its sites at a time.

> **Worked Example 9.2** *The data rate of a 1024-channel implant*
>
> A recorder has 1024 channels sampled at 20 kS/s with 10 bits. (a) Find the raw data rate. (b) Its radio spends 1 nJ per bit (an assumed, typical order for a low-power implant radio). How much power would sending the raw data take? (c) Find the data rate and radio power if the implant sends 32-sample snippets for an average of 20 spikes per second per channel, or spike counts in 20 ms bins with 2 bits per bin.
>
> Solution. (a) $R_{\text{raw}} = 1024 \times 20 \times 10^3\,\text{s}^{-1} \times 10 = 204.8\,\text{Mb/s}$.
>
> (b) $204.8 \times 10^6\,\text{b/s} \times 1 \times 10^{-9}\,\text{J/b} = 205\,\text{mW}$, more than the

whole heat budget of a small implant (Section 9.6).

(c) Snippets: $1024 \times 20 \times 32 \times 10 = 6.55\,\text{Mb/s}$, or 6.6 mW, a 31-fold saving. Binned counts: 50 bins per second give $1024 \times 50 \times 2 = 102\,\text{kb/s}$, or 0.10 mW, a 2000-fold saving. The binned stream fits easily; the raw stream does not fit at all. ■

9.5 *The Neural Recording Amplifier*

Every channel needs its own amplifier, so the amplifier sets the noise, power, and area of the whole implant. It must resolve spikes of tens of microvolts, so its noise must be a few microvolts rms. It must pass from below 1 Hz (for LFPs) or 300 Hz (for spikes only) up to about 7 kHz, ignore electrode offsets of tens to hundreds of millivolts, and use only microwatts, since a thousand of them share a budget of milliwatts. The integrated biopotential front ends of Chapter 8 spend 0.3 to 5 mW per channel on the same job; an implant must do it on a small patch of chip per channel.

AC coupling with capacitive feedback

An electrode offset is a DC voltage between two contacts that comes from small differences in their half-cell potentials (Chapter 6). A gain of 100 would turn a 50 mV offset into 5 V and saturate the amplifier. The fix is a series capacitor, which blocks DC and passes the signal band: a high-pass filter (Chapter 1).

The standard circuit is the capacitive-feedback amplifier of Fig. 9.8.[13] It is the inverting amplifier of Chapter 4 with capacitors in place of resistors. There the gain was $-R_f/R_i$, and with any two impedances it is $-Z_f/Z_i$. With $Z = 1/(j\omega C)$ for each capacitor, the frequency cancels and the midband gain is

$$A_M = -\frac{C_{\text{in}}}{C_f}. \tag{9.5}$$

[13] R. R. Harrison and C. Charles, *IEEE J. Solid-State Circuits* (2003).

Capacitors are what a chip makes best, and their ratio is accurate to a fraction of a percent. The amplifier itself is an *operational transconductance amplifier* (OTA), the on-chip cousin of the op amp, whose output is a current rather than a voltage.

The feedback capacitor alone would let the input node drift, since nothing sets its DC voltage, so each C_f has a very large resistance R_p in parallel. Together R_p and C_f set the low corner of the pass band,

$$f_L = \frac{1}{2\pi R_p C_f}. \tag{9.6}$$

For f_L below 1 Hz with C_f of a few hundred femtofarads, R_p must be of order $10^{12}\,\Omega$, which no resistor on a chip can provide. The trick is the *pseudo-resistor*: a pair of transistors (Chapter 5) connected back to back so that, for small voltages across them, neither conducts much and the resistance is huge. For a large step, such as the artifact of a stimulation pulse, one of them turns on and resets the node quickly. The upper corner is set by how fast the OTA can charge the load capacitance C_L, $f_H \approx G_m/(2\pi|A_M|C_L)$, where the transconductance G_m is the OTA's output current per volt of input.

A design in numbers: a gain of 100 (40 dB) with $C_f = 200\,\text{fF}$ needs $C_{\text{in}} = 20\,\text{pF}$; $R_p = 10^{12}\,\Omega$ gives $f_L = 0.80\,\text{Hz}$; and $G_m = 50\,\mu\text{S}$ with $C_L = 10\,\text{pF}$ gives $f_H = 8.0\,\text{kHz}$. The input impedance at 1 kHz, $1/(2\pi f C_{\text{in}}) = 8.0\,\text{M}\Omega$, is 200 times that of a $37\,\text{k}\Omega$ contact, so the divider costs less than 0.5 % of the signal. A 50 mV electrode offset only charges C_{in} and never reaches the output.

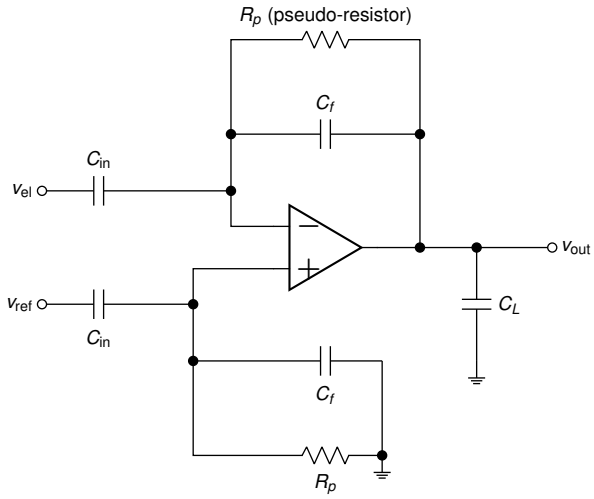


Figure 9.8: The AC-coupled neural amplifier. The input capacitors block the electrode offset; the midband gain is $-C_{\text{in}}/C_f$; the pseudo-resistors R_p in parallel with C_f set a low corner below 1 Hz; the load capacitance C_L sets the high corner. The non-inverting side mirrors the inverting side so that the two inputs see the same impedance.

The noise budget

The input noise of a channel has three main parts, each met in Chapter 4: the thermal noise of the electrode, the noise of the amplifier, and the quantization noise of the ADC referred to the input. Uncorrelated sources add in power:

$$v_{n,\text{tot}} = \sqrt{v_{n,\text{el}}^2 + v_{n,\text{amp}}^2 + v_{n,\text{q}}^2}, \quad v_{n,\text{el}} = \sqrt{4kT\,\text{Re}\{Z\}B}, \quad v_{n,\text{q}} = \frac{\text{LSB}}{\sqrt{12}}. \tag{9.7}$$

Only the resistive part of the electrode impedance produces thermal noise, so using $|Z|$ gives an upper estimate. The brain itself adds the background activity of distant neurons, often several microvolts rms, so a few microvolts of circuit noise is good enough.

Worked Example 9.3 *A noise budget for spike recording*

A PEDOT contact (take $\mathrm{Re}\{Z\} = 37\,\mathrm{k\Omega}$ as an upper bound) feeds an amplifier with 2.2 µV rms of noise over the spike band, 300 Hz to 7 kHz, at 310 K. (a) Find the total analog noise and the SNR of a 100 µV spike. (b) The 10-bit ADC spans 7.2 mV peak to peak referred to the input. Is its quantization noise negligible?

Solution. (a) With $B = 6.7\,\mathrm{kHz}$, $v_{n,\mathrm{el}} = \sqrt{4kT(37\,\mathrm{k\Omega})B} = 2.06\,\mu\mathrm{V}$, and the total is $\sqrt{2.06^2 + 2.2^2} = 3.0\,\mu\mathrm{V}$ rms. The SNR is $100/3.0 = 33$, or 30 dB (spike peak to rms noise).

(b) LSB $= 7.2\,\mathrm{mV}/1024 = 7.0\,\mu\mathrm{V}$, so $v_{n,q} = 7.0/\sqrt{12} = 2.0\,\mu\mathrm{V}$: not negligible, since the total grows to 3.6 µV. More bits, or more analog gain before the ADC, would fix it; designers aim for quantization noise below about a third of the analog noise. ■

Noise costs power

Lower noise costs current. The noise voltage of the amplifier's input transistor falls as one over the square root of its bias current, so for a given design halving the noise takes four times the current, and four times the power. Designers rate amplifiers by their *noise efficiency factor*, which compares an amplifier's noise with that of a single ideal transistor using the same current; good neural amplifiers come within a factor of 2 to 3 of that ideal. Harrison's classic amplifier reached 2.2 µV rms with 80 µW; 1024 of them would dissipate 82 mW, about half the heat budget of the N1 (Section 9.6). The 2019 Neuralink chip's amplifiers accepted 5.9 µV rms and used 5.2 µW each, 5.3 mW for 1024 (about 24 mW with the converters). Accepting about 2.7 times the noise, and running from a lower supply voltage, cut the power fifteenfold.

9.6 *Limits of Miniaturization*

An implant that can be injected through a needle or slid through a blood vessel has to fit its chip, power source, and package in about 2 mm or less. Two limits then press in: heat, and getting power and data through the tissue.

Heat

Brain tissue tolerates only small temperature rises. A steady rise of 1 to 2 °C is the usual design ceiling, and a widely used guideline[14] limits the heat flux from the surface of implanted brain-machine interface hardware to about $40\,\mathrm{mW/cm^2}$. The allowed power of an implant is

[14] The guideline and the 1 °C to 2 °C figure are summarized in A. H. Marblestone *et al.*, "Physical principles for scalable neural recording," *Front. Comput. Neurosci.* (2013).

then

$$P_{\max} \approx q_{\max} A_{\text{surface}}, \qquad q_{\max} \approx 40\,\text{mW/cm}^2, \tag{9.8}$$

so it scales with the *surface* area while the electronics we would like to pack in scale with volume. Halving every dimension of an implant cuts its power budget by four.

Power and data through tissue

Batteries do not shrink well: their energy scales with volume, and they must be recharged or replaced. Small implants therefore receive power wirelessly, by inductive coupling between coils, by radio waves, or by ultrasound (Chapter 13). An antenna or a transducer couples well only when it is not much smaller than the wavelength. Muscle has a relative permittivity $\varepsilon_r \approx 55$, which slows radio waves by $\sqrt{\varepsilon_r}$ and shortens their wavelength, but they still have a wavelength of $c/(f\sqrt{\varepsilon_r}) = 4.0\,\text{cm}$ at 1 GHz, so a millimeter antenna must work at several gigahertz, where tissue absorbs strongly: absorption alone is about 9 dB/cm at 5 GHz, or 27 dB (a factor of 500) over 3 cm.[15] Sound is about five orders of magnitude slower than light, so its wavelength in tissue is already $1540\,\text{m/s}/1\,\text{MHz} = 1.5\,\text{mm}$ at 1 MHz, and brain absorbs it weakly, only about 2 dB over 3 cm at 1.5 MHz. That makes ultrasound attractive for deep, millimeter-size implants. Both kinds of link can carry data as well as power.

[15] For a plane wave in muscle, from its measured permittivity and conductivity; the same calculation gives about 4 dB/cm at 2.4 GHz. The in-body path losses quoted in Chapter 1 also include the spreading of the wave and antenna losses, so they are larger.

9.7 *The Electrode-Tissue Interface in Stimulation*

A NEURON'S MEMBRANE BEHAVES LIKE A CAPACITOR, and an action potential starts when its voltage is pushed past threshold (Chapter 3). Electrical stimulation does this from outside. A *cathode*, an electrode driven negative, draws current toward itself through the tissue. Part of that current leaves the nearby neurons through their membranes, and an outward current through the membrane depolarizes it. That is why cathodic (negative) pulses excite neurons near the electrode most easily. Neurons can also be excited by light: in *optogenetics* a gene delivered to the neurons makes them build a light-gated ion channel, so that a flash of blue light makes them fire. It is mostly a research tool in animals, with first trials in the eye (Chapter 10). Focused ultrasound and magnetic fields can excite neurons too, but electrical stimulation is by far the most common. To design a stimulator, we need the electrical model of what it drives.

The electrode-electrolyte interface of Chapter 6 reappears here as the circuit of Fig. 9.9. The *double-layer capacitance* C_{dl} is the layer of ions that gathers at the metal surface, a capacitor with a gap of molecular

size; for smooth metals it is 10 to 40 $\mu F/cm^2$ of true surface area, somewhat more per unit of drawn area because no metal is perfectly flat. The *charge-transfer resistance* R_{ct}, in parallel with it, carries the current of electrochemical reactions at the surface; it is very large near the rest potential and falls steeply when the interface voltage is driven far from rest. The *spreading resistance* R_s in series is the resistance of the tissue itself, which for a flat disk of radius a facing a distant return electrode is the result of Chapter 6,

$$R_s = \frac{\rho}{4a}, \tag{9.9}$$

with ρ now the resistivity of the tissue. Wires from the stimulator add in series with R_s.

The stimulation examples of this chapter use one electrode: a platinum disk 1 mm in diameter in brain tissue, with model values $R_s = 1.75\,k\Omega$, $C_{dl} = 427\,nF$ ($54\,\mu F/cm^2$ on its $7.85 \times 10^{-3}\,cm^2$), and $R_{ct} = 38.5\,M\Omega$, after the tutorial by Muller *et al.* The spreading resistance implies $\rho = 4aR_s = 4(0.05\,cm)(1750\,\Omega) = 350\,\Omega\,cm$, a typical value for brain. The model has two time constants. The fast one, $R_sC_{dl} = 0.75\,ms$, is how quickly the interface charges or discharges through the tissue. The slow one, $R_{ct}C_{dl} = 16\,s$, is how slowly charge left on the interface leaks away by itself. A 100 µs pulse is far shorter than 16 s, so during a pulse essentially all of the current charges C_{dl}: the electrode acts as a resistor in series with a capacitor.

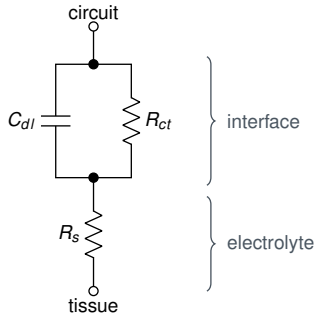


Figure 9.9: Lumped model of a stimulating electrode in tissue. The double-layer capacitance C_{dl} stores charge without chemistry; the charge-transfer resistance R_{ct} carries faradaic current; the spreading resistance R_s is the tissue between the electrode and the distant return.

Capacitive and faradaic charge injection

Charge crosses the interface in two ways. In *capacitive* injection, electrons pile up on the metal side and ions rearrange on the solution side, as in any capacitor; no electrons cross and no chemistry happens, so it is fully reversible. In *faradaic* injection, electrons do cross the interface by oxidizing or reducing something. Some faradaic reactions are *reversible* because their products stay bound to the surface: hydrogen atoms plating onto platinum, or the change of iridium oxide between its Ir^{3+} and Ir^{4+} states. If the opposite charge is returned soon, they run backward. Others are *irreversible*: water reduced to hydrogen gas, water oxidized to oxygen, chloride oxidized, the metal itself dissolving. Their products diffuse away, change the local pH, damage tissue, and corrode the electrode. The range of electrode potentials in which no water is split is the *water window*, roughly $-0.6\,V$ to $+0.8\,V$ versus Ag/AgCl for platinum.

KEY IDEA Stimulate inside the water window, with capacitive or reversible faradaic charge only, and return every charge you inject.

The *charge injection capacity* (CIC) of an electrode material is the charge per unit area it can inject in one phase while staying inside the water window. Table 9.2 lists typical values. Coatings that store charge through reversible surface reactions or through their volume inject ten or more times as much charge as smooth platinum. For the

1-mm disk, the double layer alone holds $Q = C_{dl}\Delta V = 0.26\,\mu C$ within 0.6 V. Platinum's CIC allows 0.39 to 1.2 µC on the same area, more than the pure capacitance, because hydrogen-atom plating adds reversible charge; iridium oxide would allow 8 to 39 µC.

Material	CIC (mC/cm^2)	Mechanism
Platinum, Pt-Ir	0.05 to 0.15	capacitive and surface H atoms
Titanium nitride	about 0.5 to 1	capacitive (porous)
Iridium oxide (activated or sputtered)	1 to 5	reversible Ir redox
PEDOT:PSS	about 1 to 15	volume capacitance

Table 9.2: Approximate charge injection capacities, after the review by S. F. Cogan, *Annu. Rev. Biomed. Eng.* (2008). Values depend on pulse width, potential limits, and how rough the film is; the PEDOT values in the literature vary widely.

9.8 *Safe Stimulation Waveforms*

Stimulators control current, not voltage, in almost every modern design. The charge delivered is the integral of the current, so a current source delivers the same charge whatever the electrode and tissue impedance do, and that impedance changes over months as tissue grows around the electrode. A voltage source delivers a charge that depends on the impedance.

The two basic waveforms are shown in Fig. 9.10. A *monophasic* pulse train sends brief cathodic current pulses and nothing else. Cathodic pulses are the most effective at exciting neurons near the electrode, but each one leaves its charge on the interface. A *biphasic* train follows each cathodic phase with an anodic phase of equal charge, returning the interface to where it started. The *charge per phase*, $Q = IT$ for a rectangular phase of current I and width T, is the number that matters for both excitation and safety. A short *interphase gap* between the phases improves efficacy, because the neuron has time to start its action potential before the reversal begins to undo the depolarization. The amplitude, pulse width, rate, and gap are tuned for each patient.

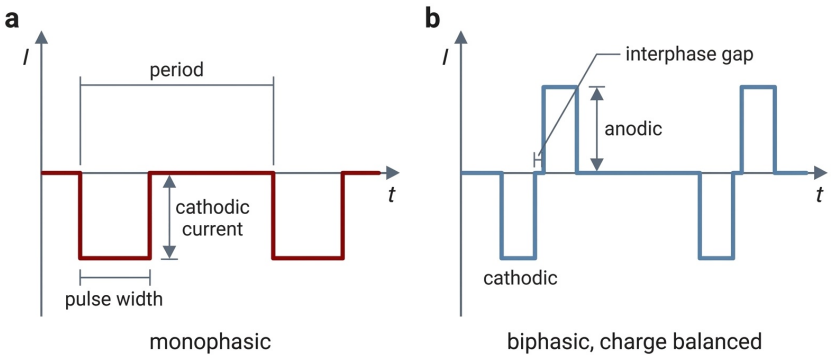


Figure 9.10: Stimulation waveforms. (a) A monophasic cathodic train leaves the charge of every pulse on the electrode. (b) A biphasic train returns it in an anodic phase after a short interphase gap; in a charge-balanced pulse the two rectangles, current times pulse width, have equal areas.

What goes wrong with monophasic pulses

Figure 9.11 simulates the 1-mm platinum model with 1 mA, 100 µs pulses at 130 Hz. Each monophasic pulse adds charge to C_{dl}, and almost none of it leaks away through R_{ct} before the next, so the interface voltage ratchets out of the water window within a few pulses. In a real electrode R_{ct} then collapses as irreversible reactions take over: water splits, the pH shifts, platinum corrodes, and tissue is damaged. The biphasic train returns to zero after every pulse.

Two fixes for residual charge are common. A *passive recharge*, or *shorting*, phase connects the electrode to the return after each pulse and discharges C_{dl} through R_s with time constant $R_s C_{dl}$; after a biphasic pulse it only has to remove the small imbalance left by mismatched current sources. A series *DC-blocking capacitor* guarantees zero net DC current even if the circuit fails.

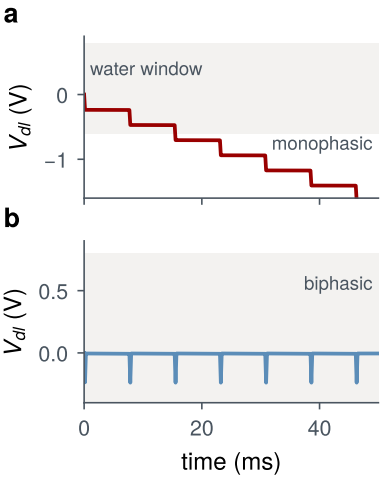


Figure 9.11: Voltage V_{dl} across C_{dl} of the 1-mm platinum model, simulated for 1 mA, 100 µs pulses at 130 Hz. (a) Monophasic pulses ratchet out of the shaded water window (drawn relative to rest); beyond it the linear model fails. (b) Biphasic pulses return to zero.

Worked Example 9.4 *Fixing a failing stimulator*

A start-up's neurostimulator keeps failing in animal experiments. It sends monophasic cathodic pulses of 1 mA and 100 µs at 130 Hz into the 1-mm platinum electrodes modeled in Section 9.7. (a) How much does each pulse charge C_{dl}, and why do the experiments fail? (b) Propose a new waveform. (c) In the new design the anodic current source is 1 % weaker than the cathodic one. With no shorting phase, where does the interface voltage settle? (d) How long a shorting phase removes 99.9 % of the residual voltage after each pulse?

Solution. (a) Each pulse carries $Q = 1\,\text{mA} \times 100\,\mu\text{s} = 0.1\,\mu\text{C}$ and charges C_{dl} by $0.1\,\mu\text{C}/427\,\text{nF} = 0.23\,\text{V}$. In one period, $T = 1/130\,\text{Hz} = 7.7\,\text{ms}$, the leak through R_{ct} removes only the fraction $T/(R_{ct}C_{dl}) \approx 0.05\,\%$ of it. The voltage climbs by 0.23 V per pulse and leaves the water window within three pulses (Fig. 9.11). After that the charge goes into irreversible reactions: gas, pH change, corrosion, dissolved platinum. The waveform, not the biology, causes the failures.

(b) A charge-balanced biphasic waveform (Fig. 9.10b): 1 mA cathodic for 100 µs, a gap of about 50 µs, 1 mA anodic for 100 µs, at 130 Hz, then a shorting phase, with a DC-blocking capacitor in series. A longer, weaker anodic phase of equal charge also works and is less likely to excite neurons itself.

(c) Each pulse now leaves $\Delta Q = 0.01 \times 0.1\,\mu\text{C} = 1\,\text{nC}$, or $\Delta V = 2.34\,\text{mV}$. The voltage climbs until the leak removes as much in each period as each pulse adds. The leak removes the fraction T/τ of the voltage per period, with $\tau = R_{ct}C_{dl} = 16.4\,\text{s}$, so $V_{ss}\,T/\tau = \Delta V$ and $V_{ss} \approx \Delta V\,\tau/T = 5.0\,\text{V}$ (in the linear model). Even a 1 %

mismatch would drive the electrode far out of the water window, so the residual must be removed actively.

(d) Shorting discharges C_{dl} through R_s with $\tau_s = R_s C_{dl} = 0.75\,\text{ms}$, as e^{-t/τ_s}. Leaving 0.1 % takes $\tau_s \ln 1000 = 5.2\,\text{ms}$, which fits in the 7.4 ms left in each period after the pulse. ■

How much charge is safe: the Shannon criterion

Charge balance prevents chemistry; it does not by itself prevent damage from too much stimulation. Excessive stimulation can injure neurons, probably by driving them to fire too hard for too long, and very intense pulses can open pores in nearby cell membranes (the electroporation of Chapter 2). Two quantities matter. The charge per phase Q sets how large a volume of tissue is driven, and the *charge density* Q/A sets how intense the stimulation is at the electrode surface. From animal studies of tissue damage, R. V. Shannon proposed an empirical boundary,[16]

$$\log_{10}\frac{Q}{A} = k - \log_{10} Q, \tag{9.10}$$

with Q in microcoulombs per phase and Q/A in microcoulombs per square centimeter per phase. Experiments with k above 1.85 produced damage, so $k = 1.85$ is used as an absolute maximum; clinical deep brain stimulators use a limit of $30\,\mu\text{C/cm}^2$, which corresponds to the more conservative $k = 1.75$. Equation (9.10) is the same as $Q \cdot (Q/A) = 10^k$, so for an electrode of area A the largest safe charge per phase is

$$Q_{\max} = \sqrt{10^k A}. \tag{9.11}$$

Figure 9.12 plots the boundary with lines of constant electrode area. The criterion was built on data from large electrodes; for microelectrodes it is not established, and the material limit of Table 9.2 often binds first. A safe design must satisfy both.

[16] R. V. Shannon, "A model of safe levels for electrical stimulation," *IEEE Trans. Biomed. Eng.* (1992). Shannon was a hearing scientist who worked on cochlear implants (Chapter 10).

KEY EQUATION

$\log \frac{Q}{A} = k - \log Q$

Shannon criterion, Q in µC per phase, Q/A in $\mu\text{C/cm}^2$. Damage was seen for $k > 1.85$.

WORKED EXAMPLE 9.5 *Shannon and material limits*

Find the largest charge per phase that the Shannon criterion allows, and the charge density at that limit, for (a) the 1-mm disk at $k = 1.85$; (b) a DBS contact 1.27 mm in diameter and 1.5 mm long at $k = 1.75$; (c) a 20 µm disk at $k = 1.85$. Give the current each limit allows at 200 µs per phase for (a) and 90 µs for (b). (d) If each electrode is platinum, which limit binds, Shannon or the charge injection capacity?

SOLUTION. (a) $A = \pi(0.05\,\text{cm})^2 = 7.85 \times 10^{-3}\,\text{cm}^2$ and $Q_{\max} = \sqrt{10^{1.85} \times 7.85 \times 10^{-3}} = 0.75\,\mu\text{C}$, at $Q/A = 95\,\mu\text{C/cm}^2$. At 200 µs per phase that is $I = Q/T = 3.7\,\text{mA}$.

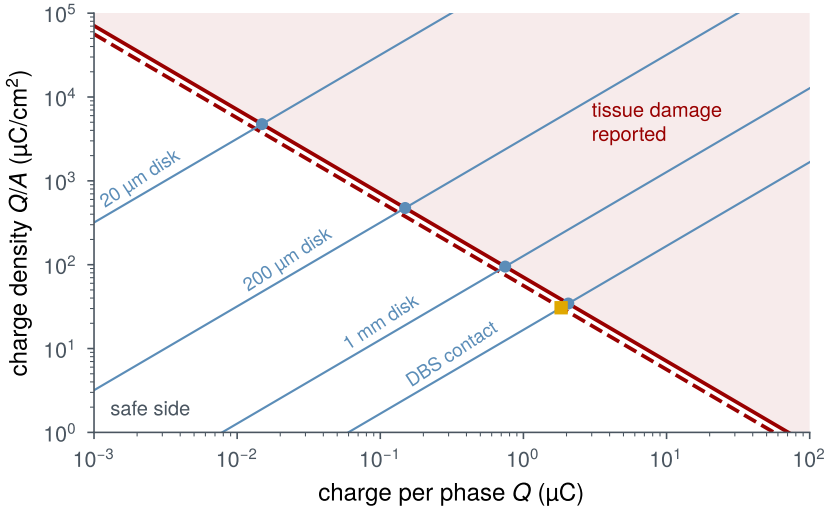


Figure 9.12: The Shannon boundary (solid, $k = 1.85$; dashed, $k = 1.75$) and lines of constant electrode area (blue). Each blue dot is the largest safe charge per phase for that electrode at $k = 1.85$. The gold square is the clinical DBS limit, $30\,\mu C/cm^2$ on a standard DBS contact. Small electrodes reach the boundary at small charge but very high charge density.

(b) The contact is a cylinder, so its area is

$$A = \pi d L = \pi(0.127\,\text{cm})(0.15\,\text{cm}) = 0.0598\,\text{cm}^2.$$

Then $Q_{\max} = \sqrt{10^{1.75} \times 0.0598} = 1.83\,\mu C$ and $Q/A = 30.7\,\mu C/cm^2$: the $30\,\mu C/cm^2$ rule of clinical DBS is the $k = 1.75$ boundary for this contact. At 90 µs it allows 20 mA, far above the few milliamperes used.

(c) $A = 3.14 \times 10^{-6}\,\text{cm}^2$, so $Q_{\max} = 14.9\,\text{nC}$ at $Q/A = 4.7\,\text{mC/cm}^2$.

(d) Platinum injects 0.05 to $0.15\,\text{mC/cm}^2$. The 1-mm disk can then take 0.39 to 1.2 µC, about the Shannon charge, so both limits matter. The DBS contact can take 3.0 to 9.0 µC, more than its 1.83 µC, so Shannon binds. The 20 µm disk can take at most 0.47 nC, thirty times less than its Shannon charge: the material binds, and a microelectrode needs iridium oxide or PEDOT to come near the Shannon limit. ■

9.9 *Stimulator Circuits*

A STIMULATOR IS A SWITCHED CURRENT SOURCE. A *current source* forces a set current through whatever it drives. On a chip it is a transistor biased by a *current mirror*, a matched transistor that copies a reference current, and it holds its current only as long as the voltage across it stays above a minimum, its *headroom*, of a few hundred millivolts. A *current DAC* is a set of such sources weighted 1, 2, 4, 8, and so on, switched on by the bits of a digital word, so the controller sets the

amplitude in steps.

Compliance voltage

A current source can force its current only if it has enough voltage. While a constant current I flows for a time t, the voltage across the electrode of Fig. 9.9 (with R_{ct} negligible during the pulse) is

$$V_{\text{stim}}(t) = IR_s + \frac{It}{C_{dl}} + V_{\text{offset}}, \tag{9.12}$$

a step IR_s across the tissue followed by a ramp as C_{dl} charges, plus any voltage already on the interface. The value at the end of a phase of width T sets the *compliance voltage* the stimulator must provide:

$$V_C \geq I\left(R_s + \frac{T}{C_{dl}}\right) + V_{\text{headroom}}. \tag{9.13}$$

The compliance voltage often sets the supply voltage of the whole stimulator chip (Fig. 9.13).

KEY EQUATION

$V_C \geq I(R_s + T/C_{dl}) + V_{\text{hr}}$

Compliance: the resistive step plus the capacitor ramp plus the current source's headroom.

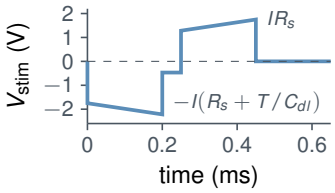


Figure 9.13: Electrode voltage for one biphasic pulse of 1 mA, 200 µs per phase, with a 50 µs gap, into the 1-mm platinum model: a step of $IR_s = 1.75$ V, a ramp of 0.47 V, and the mirror image in the anodic phase.

Push-pull stimulators and H-bridges

Figure 9.14a shows the common *push-pull* stimulator. Two current sources, I_A from a positive rail V_{DDH} and I_C to a negative rail $-V_{\text{DDL}}$, connect to the electrode through switches closed in phases ϕ_{A} (anodic) and ϕ_{C} (cathodic). A third switch, closed in phase ϕ_{short}, connects the electrode to the return to remove residual charge. The return electrode sits at ground, the middle of the supply, so the electrode can swing to $+V_{\text{DDH}}$ or $-V_{\text{DDL}}$ minus the headroom. Matching the anodic and cathodic sources sets the charge balance; the first-generation retinal stimulator chip of Chapter 10, for example, matched them to 1.2 to 1.7 %.

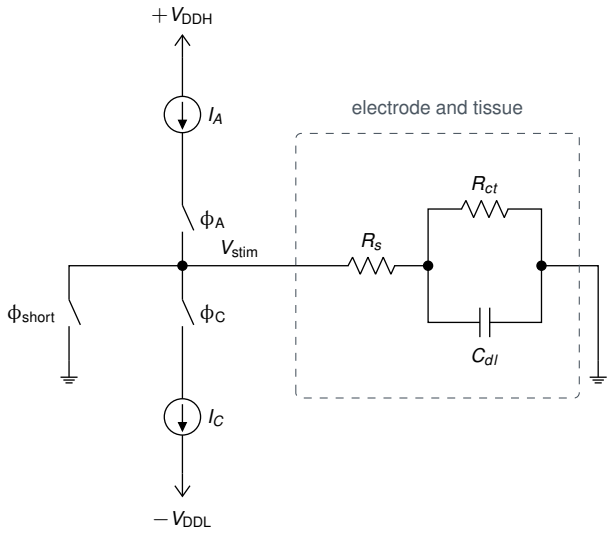
An *H-bridge* (Fig. 9.14b) uses one current source and four switches to reverse the current through a pair of electrodes. Closing S1 and S4 sends the current one way, closing S2 and S3 the other way. It runs from a single supply and puts almost the whole supply across the electrode pair in each direction.

Efficiency and power

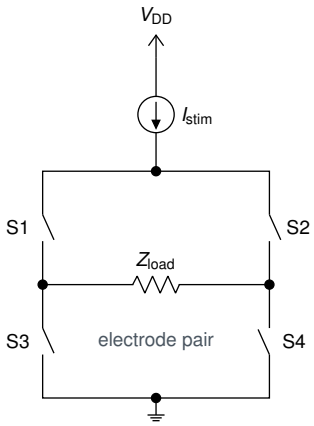
In a power-limited implant the efficiency of stimulation matters:

$$\eta = \frac{P_{\text{stim}}}{P_{\text{in}}}, \qquad P_{\text{stim}} = 2\, V_{\text{stim}} I_{\text{stim}} T_{\text{stim}} f_{\text{stim}}, \tag{9.14}$$

where the factor 2 counts both phases and f_{stim} is the pulse rate. A current source draws its current from the full rail, so the supply delivers $2V_{\text{rail}}ITf$ whatever the electrode needs, and the difference is lost



(a) push-pull stimulator



(b) H-bridge

Figure 9.14: Two current-mode stimulator topologies. (a) Push-pull: a source from $+V_{\mathrm{DDH}}$ for the anodic phase, a sink to $-V_{\mathrm{DDL}}$ for the cathodic phase, and a shorting switch to discharge the electrode; the return electrode is at ground. Model values for a 1-mm platinum disk: $R_s = 1.75\,\mathrm{k\Omega}$, $R_{ct} = 38.5\,\mathrm{M\Omega}$, $C_{dl} = 427\,\mathrm{nF}$. (b) H-bridge: one source and four switches reverse the current through an electrode pair from a single supply. After Muller, Ghanbari, and Zhou.

as heat in the source transistors. Use the lowest rails that meet the compliance of Eq. (9.13); some designs even adjust the supply to each pulse. Averaged over time, a biphasic train draws a current $2ITf$ from the supply, to which the quiescent current of the electronics adds; that average sets the battery size (Chapter 13).

When the rails are too low

Small or scarred electrodes have a high series resistance. Driving 10 µA through 1 MΩ takes 10 V, but from ±3.3 V rails with 0.3 V of headroom each phase can swing only 3.0 V, enough for 3 µA. The current sources leave their working range, the delivered current collapses to a third of the target, and the neurons never reach threshold. There are three fixes. First, check that the 1 MΩ is really the series resistance R_s and not R_{ct}, which a short pulse never sees, and lower it with a larger contact or a better lead; a coating that raises C_{dl} shrinks the ramp. Second, an H-bridge puts the swing of both rails across the electrode pair, about twice what one rail gives. Third, a *charge pump*, a ladder of capacitors and switches that stacks copies of the supply voltage, raises the rails; the high voltage then needs transistors rated for it, which is why stimulator chips often use high-voltage processes. The 256-channel retinal stimulator chip of Chapter 10, for example, used a ±12 V output stage.

Stimulation artifacts and blanking

A stimulation pulse of a few volts appears at nearby recording electrodes as an artifact thousands of times larger than the spikes. It saturates the amplifiers, and a slowly recovering AC coupling keeps them blind afterward. Closed-loop devices therefore *blank*: they disconnect or hold the amplifier inputs during the pulse and a short recovery, reset the amplifier (the pseudo-resistors of Fig. 9.8 conduct strongly for large signals), and subtract the remaining artifact in software. Charge-balanced pulses with a shorting phase shorten the artifact by leaving the electrode near rest. Because charge is what matters, some stimulators also measure the charge they actually deliver, by integrating the electrode current on a capacitor during each phase, and trim the next pulse to cancel any imbalance.

Summary

- Neural interfaces record, stimulate, or both in a closed loop: pacemakers, DBS (130 to 185 Hz, 60 to 90 µs pulses), VNS, spinal cord stimulators, and BMIs that decode the motor cortex.
- Closer is larger and sharper: EEG (5 to 300 µV), ECoG (10 µV to mV), LFPs (below 300 Hz), single units (50 to 500 µV, 0.3 to 7 kHz). Spikes are 100 mV inside the cell and 100 µV outside.
- Stiff arrays fail by glial scarring and micromotion. Bending stiffness EI grows as t^3, so thin polymer threads are nearly a million times more compliant than a silicon needle. Small contacts have $|Z| \approx 1/(2\pi f c_{dl} A)$; IrOx and PEDOT lower it.
- $R_{\text{raw}} = N f_s b$; sending spike snippets cuts it about thirtyfold and binned spike counts about two-thousandfold. Heat limit: about 40 mW/cm^2 of surface.
- AC-coupled amplifier: gain $-C_{\text{in}}/C_f$, low corner $1/(2\pi R_p C_f)$. Noise adds in power; half the noise costs four times the current.
- Interface: $R_s = \rho/(4a)$ in series with $C_{dl} \parallel R_{ct}$. Use current-controlled, charge-balanced biphasic pulses with a shorting phase; keep the charge per phase below the CIC of the material and the Shannon limit $Q_{\max} = \sqrt{10^k A}$, $k \leq 1.85$.
- Compliance $V_C \geq I(R_s + T/C_{dl}) + V_{\text{hr}}$; use the lowest rails that meet it, and an H-bridge or a charge pump when they are too low.

Chapter 10 puts these stimulators to work in the two most successful neural prostheses, the cochlear implant and the retinal implant.

9.10 Problems

9.1 ◦ Choose a recording modality (EEG, ECoG, LFP, or single units) and give one reason for: (a) sleep staging at home; (b) locating the cortical source of seizures before surgery; (c) decoding intended hand-movement direction for a robotic arm; (d) tracking the beta rhythm (13 to 30 Hz) in the subthalamic nucleus through a DBS lead.

9.2 • Design the amplifier of Fig. 9.8 for 46 dB with $C_f = 100\,\text{fF}$: (a) find C_{in}; (b) find R_p for $f_L = 0.5\,\text{Hz}$; (c) find the input impedance at 1 kHz and the fraction of the signal lost with a $1\,\text{M}\Omega$ electrode (treat both as capacitive, so magnitudes add).

9.3 • A 500 µm diameter disk has $c_{dl} = 30\,\mu\text{F/cm}^2$. (a) How much charge can it inject capacitively within 0.6 V? (b) How much as platinum ($0.1\,\text{mC/cm}^2$) or iridium oxide ($2\,\text{mC/cm}^2$)? (c) For each material, is the CIC or the Shannon limit ($k = 1.85$) lower?

9.4 • A stimulator drives 200 µA for 100 µs per phase into $R_s = 5\,\text{k}\Omega$ and $C_{dl} = 10\,\text{nF}$. (a) Find the electrode voltage at the end of the cathodic phase. (b) What rails are needed with 0.3 V of headroom? (c) At 100 Hz, find P_{stim} and the efficiency with ± 5 V and ± 3.3 V rails.

9.5 •• *(Design.)* A 256-channel implant, $4 \times 4 \times 1.5\,\text{mm}$, sends 2-bit spike counts per 20 ms bin with a 2 nJ/bit radio; other logic takes 0.5 mW. Each amplifier needs 5 µV rms over the spike band; a reference design of the same quality reaches 10 µV rms on 0.45 µW. Find (a) the amplifier power and (b) the radio power. (c) Compare with the $40\,\text{mW/cm}^2$ budget of the surface: how many more channels fit?

10
Sensory Prostheses

In this chapter you will learn to

- Describe the chain shared by sensory prostheses, from an external sensor to the neurons it stimulates.
- Explain how the cochlea maps frequency to place, and use the Greenwood function to find the pitch an electrode evokes.
- Design the processor of a cochlear implant: a log-spaced filter bank, envelope extraction, compression, and continuous interleaved sampling.
- Describe the eye and the layers of the retina, and how photoreceptors, bipolar cells, and ganglion cells pass an image on to the brain.
- Convert between visual angle, retinal distance, Snellen acuity, logMAR, and cycles per degree.
- Estimate the best acuity an electrode array can give from its pitch, and the number of electrodes needed to read.
- Compare epiretinal, subretinal, and cortical implants, and explain how a photovoltaic pixel stimulates without wires.

About 737,000 cochlear implants had been implanted worldwide by the end of 2019.[1] Most of their recipients were deaf because the hair cells of their inner ear had died, while the nerve fibers behind those cells survived. A microphone behind the ear, a processor that splits sound into frequency bands, a coil that sends power and data through the skin, and a thin array of electrodes threaded into the cochlea together take the place of the hair cells. Many recipients talk on the telephone. Retinal implants try to do the same for the eye, and their story shows how much harder vision is: the best of them restore a blurry, colorless patch of sight.

[1] U.S. National Institute on Deafness and Other Communication Disorders (NIDCD), cochlear implant fact sheet: about 736,900 registered devices worldwide as of December 2019.

This chapter applies the stimulation of Chapter 9 to two senses. After the chain that every sensory prosthesis shares, it builds the cochlear implant, from the anatomy of the ear to the signal processing that drives the electrodes. It then turns to vision: the eye and the retina, how sharp the vision from an electrode array can be, where a retinal implant can go, and two landmark devices, the epiretinal Argus II, developed with researchers at USC, and the photovoltaic PRIMA, which needs no wires.

10.1 How a Sensory Prosthesis Works

The body has five classical senses: vision, hearing, smell, touch, and taste. Each sense organ contains receptor cells that turn a physical stimulus into the action potentials of sensory nerves (Chapter 3). When the receptor cells die but the nerve and the brain behind them still work, an implant can bypass the receptors and stimulate the nerve directly. That is the idea of a *sensory prosthesis*. Cochlear implants for hearing are standard care; retinal implants for some forms of blindness have reached the market several times. The first implanted stimulator for vision was reported in 1968 by Brindley and Lewin, who placed an array of radio receivers connected to electrodes on the visual cortex of a blind patient; radio signals made the patient see spots of light (*phosphenes*) in the opposite half of the visual field.[2]

[2] G. S. Brindley and W. S. Lewin, *J. Physiol.* (1968).

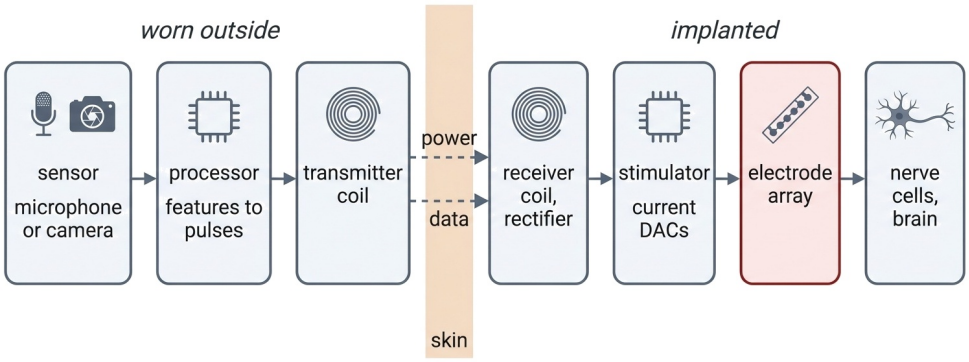


Figure 10.1: The chain shared by sensory prostheses. An external sensor and processor turn the stimulus into a stimulation pattern; an inductive link carries power and data through the skin; an implanted stimulator drives the electrodes, which excite the surviving neurons.

Figure 10.1 shows the chain that cochlear and most retinal implants share. The sensor and the processor are worn outside, where they can be upgraded and run from a replaceable battery. An *inductive link* carries power and commands through the skin: an alternating current in an outer coil induces a voltage in an implanted coil, as in a transformer, so no wire crosses the skin (Chapter 13 designs these links). The implant rectifies that voltage into a DC supply, decodes the commands, and drives the electrodes with current sources whose amplitudes are set by digital words (current DACs). It does no more

processing than it must, because every milliwatt it dissipates warms tissue. The stimulation follows the rules of Chapter 9: controlled current, in charge-balanced biphasic pulses that push charge into the tissue and then pull the same charge back, with the charge per phase below Shannon's safety limit.

10.2 Hearing and the Cochlea

SOUND IS A PRESSURE WAVE IN AIR. Its frequency sets the pitch we hear, from about 20 Hz to 20 kHz for a young listener, and its amplitude sets the loudness, measured in decibels (Chapter 1). The outer ear funnels the wave to the eardrum; three tiny bones of the middle ear, the ossicles, carry the vibration to the oval window of the *cochlea*, a fluid-filled tube coiled about two and a half turns, like a snail shell (Fig. 10.2).

Along the cochlea runs the *basilar membrane*, about 35 mm long in people, which carries the sensory hair cells. The membrane is narrow and stiff at the base, near the oval window, and wide and floppy at the apex. A sound wave travels along it and peaks at the place that resonates at the sound's frequency, much as a short, tight string sounds high and a long, slack one low: high frequencies peak near the base, low frequencies near the apex. This mapping of frequency to place is called *tonotopy*, and the brain keeps it all the way to the auditory cortex.

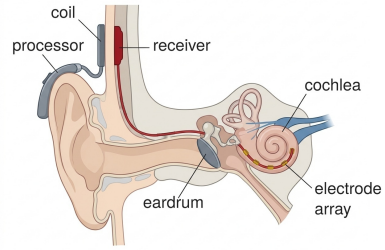


Figure 10.2: A cochlear implant. A microphone and processor behind the ear send power and data through a coil held over the implanted receiver by a magnet; a lead carries the stimulation to an electrode array coiled inside the cochlea.

Where the membrane moves, *inner hair cells* bend their bundles of stereocilia. The bending opens mechanically gated ion channels (Chapter 3), the hair cell depolarizes, and it releases neurotransmitter onto the fibers of the auditory nerve, whose cell bodies form the *spiral ganglion*. Most deafness is *sensorineural*: the hair cells are lost, from genetics, noise, infection, drugs, or age, while many spiral ganglion neurons survive. A cochlear implant stimulates those neurons directly. Because they keep their tonotopic order, an electrode near the base evokes a high pitch and one deeper in evokes a low pitch.

The place-frequency map of the human cochlea is well described by the *Greenwood function*,

$$f = 165.4\left(10^{2.1x} - 0.88\right)\ \text{Hz}, \tag{10.1}$$

where x is the position along the basilar membrane as a fraction of its length, measured from the apex ($x = 0$) to the base ($x = 1$). It spans about 20 Hz to 20.7 kHz, the range of human hearing, and is nearly logarithmic over most of the length (Fig. 10.3): each *octave*, a doubling of frequency, takes roughly the same few millimeters.

KEY EQUATION

$f = 165.4(10^{2.1x} - 0.88)$

Greenwood map; x is the fraction of the cochlear length from the apex.

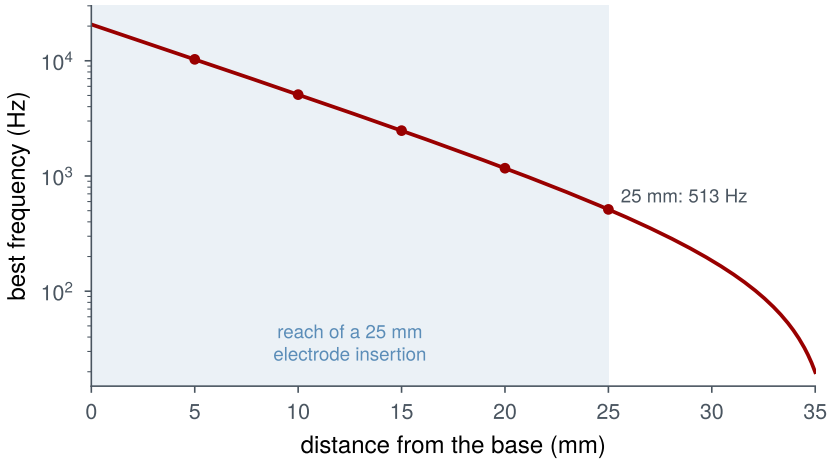


Figure 10.3: The Greenwood map for a 35 mm cochlea, plotted against distance from the base, where an electrode array enters. An array inserted 25 mm (shaded) reaches the place of about 500 Hz; lower frequencies lie deeper in the apex.

Worked Example 10.1 *Where the electrodes sit on the frequency map*

A cochlear implant array is inserted 25 mm into a 35 mm cochlea. (a) What frequency corresponds to its deepest contact? (b) Where along the cochlea do 1 kHz and 4 kHz lie? (c) The processor sends the band around 250 Hz to the deepest contact. What does that mismatch mean for the listener?

Solution. (a) The deepest contact is 10 mm from the apex, $x = 10/35 = 0.286$, so $f = 165.4(10^{0.600} - 0.88) = 165.4(3.98 - 0.88) = 513$ Hz.

(b) Solve Eq. (10.1) for x: $x = \log_{10}(f/165.4 + 0.88)/2.1$. For 1 kHz, $x = 0.400$, which is $35 \times (1 - 0.400) = 21.0$ mm from the base; for 4 kHz, $x = 0.666$, 11.7 mm from the base. The two frequencies, two octaves apart, are about 9 mm apart.

(c) The deepest contact stimulates neurons tuned near 500 Hz, so a 250 Hz sound is heard about an octave too high. Every band is shifted upward in the same way. Listeners adapt to this shift over weeks to months, which is one reason performance keeps improving after the implant is turned on. Longer arrays reach deeper but risk more damage to the delicate apex. ■

10.3 *The Cochlear Implant*

A cochlear implant has two halves. Outside, behind the ear, sit a microphone, a speech processor, a battery, and a transmitting coil held to the skin by a magnet. Under the skin, the receiver-stimulator contains the receiving coil, a rectifier and regulators, a decoder, and the current sources. A lead carries the stimulation into the cochlea,

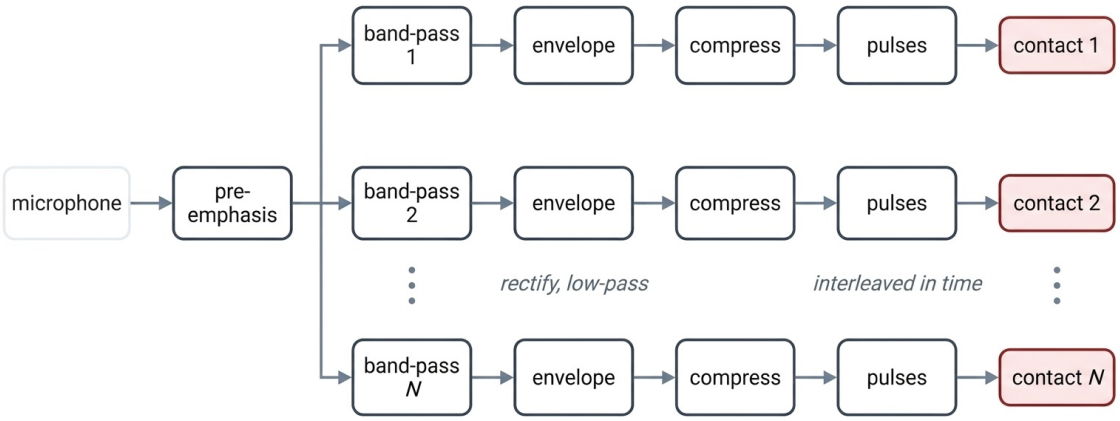


Figure 10.4: Continuous interleaved sampling. A filter bank splits the sound into bands; each band's envelope is extracted, compressed into the narrow electric dynamic range, and used to set the amplitude of biphasic pulses on one contact. The pulses on different contacts never overlap in time.

where an electrode array of typically 12 to 22 contacts lies in the scala tympani, the lower fluid-filled chamber of the cochlea, close to the spiral ganglion. Each contact is a stimulation channel: with a reference electrode outside the cochlea, it delivers current pulses into the nerve fibers nearby.

The processor

The processor turns sound into a pattern of pulses, one channel per contact (Fig. 10.4). A pre-emphasis filter first boosts the high frequencies, which carry many consonants but little of the energy of speech. A *filter bank* then splits the signal into bands, one per channel. Each filter is a *band-pass filter*: it passes a range of frequencies and rejects those above and below, like a high-pass and a low-pass filter of Chapter 1 in series. Because the cochlea maps frequency logarithmically, the bands are spaced logarithmically too: each band edge is a fixed ratio r above the one before. For N bands from f_{lo} to f_{hi},

$$f_k = f_{\text{lo}}\, r^k, \qquad r = \left(\frac{f_{\text{hi}}}{f_{\text{lo}}}\right)^{1/N}, \qquad k = 0, 1, \ldots, N, \tag{10.2}$$

so every band covers the same fraction of an octave, $\log_2 r$, and its center is the geometric mean of its edges, $\sqrt{f_k f_{k+1}}$.

Worked Example 10.2 *A log-spaced filter bank*

Design a 12-band filter bank from 250 Hz to 8 kHz. (a) Find the ratio between band edges and the width of each band in octaves. (b) List the edges. (c) Find the center and bandwidth of the lowest and the highest band.

Solution. (a) $r = (8000/250)^{1/12} = 32^{1/12} = 1.335$. Each band spans $\log_2 1.335 = 5/12 = 0.42$ octave, since the whole bank covers 5 octaves ($2^5 = 32$).

(b) $f_k = 250 \times 1.335^k$: 250, 334, 445, 595, 794, 1060, 1414, 1888, 2520, 3364, 4490, 5993, and 8000 Hz.

(c) Lowest: 250 to 334 Hz, center $\sqrt{250 \times 334} = 289$ Hz, bandwidth 84 Hz. Highest: 5993 to 8000 Hz, center 6924 Hz, bandwidth 2007 Hz. The bandwidths grow 24-fold while the relative bandwidth stays at 29 % of the center, just as the cochlea's own filters do. ■

From each band the processor keeps only the slowly varying *envelope*, the outline of the band's oscillations. It rectifies the band signal (takes its absolute value) and low-pass filters it at a few hundred hertz, the way an AM radio recovers its audio. The fine timing inside each band is discarded. Even so, the envelopes of a handful of bands are enough to understand speech in quiet, as experiments with normal-hearing listeners showed.[3]

[3] R. V. Shannon *et al.*, "Speech recognition with primarily temporal cues," *Science* (1995). Shannon, also the author of the stimulation safety criterion of Chapter 9, worked on cochlear implants.

The envelope must then be *compressed*. Normal hearing spans about 120 dB from the quietest audible sound to discomfort, and speech alone spans several tens of decibels. Electric hearing is far narrower: between the *threshold* current T, where a sound is first heard, and the *most comfortable* current C, a contact often has 20 dB or less. A logarithmic map turns the envelope e into a current,

$$I = T + (C - T)\,\frac{\log(e/e_{\min})}{\log(e_{\max}/e_{\min})}, \tag{10.3}$$

where $e_{\min}$ and $e_{\max}$ bound the acoustic range that is kept. An envelope at $e_{\min}$ gives T, one at $e_{\max}$ gives C, and equal steps in decibels give equal steps in current between them. T and C are measured for each contact of each patient when the device is fitted.

Continuous interleaved sampling

Each channel's current then sets the amplitude of a train of biphasic pulses on its contact. In the *continuous interleaved sampling* (CIS) strategy, the pulses on different contacts are staggered so that no two contacts are ever on at the same time.[4] Simultaneous currents in the conductive fluid of the cochlea would add in the tissue and smear the channels together; interleaving keeps each channel's field its own. The cost is time. With N channels at a rate f_p pulses per second each, the stimulator must deliver $N f_p$ pulses per second, and each pulse, both phases plus the gap between them, must fit in a slot of $1/(N f_p)$ (Fig. 10.5). Pulse rates of hundreds to a few thousand pulses per second per channel are common.

[4] B. S. Wilson *et al.*, "Better speech recognition with cochlear implants," *Nature* (1991).

These pulses carry little charge. At 340 µA for 40 µs, one phase delivers 13.6 nC; on a contact of about 0.3 mm^2 that is 4.5 µC/cm^2, and

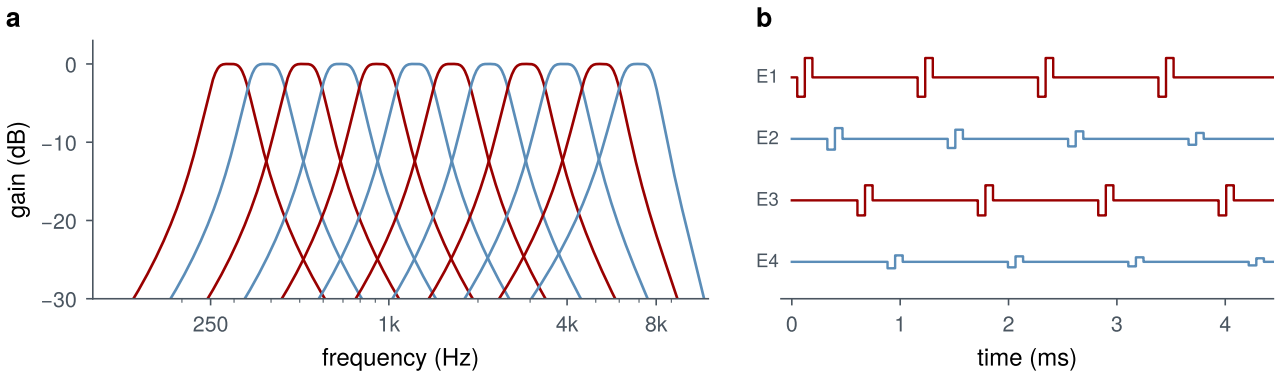


Figure 10.5: (a) A 12-channel filter bank with log-spaced bands from 250 Hz to 8 kHz (computed responses of second-order Butterworth band-pass sections). (b) The interleaved pulse pattern of CIS on four contacts at 900 pulses/s per channel (schematic): each contact gets its biphasic pulse in its own time slot, with an amplitude set by its band's envelope.

Shannon's $k = \log_{10}(Q/A) + \log_{10} Q$ is about -1.2, far below the limit of 1.85 (Chapter 9). Short pulses at modest currents keep cochlear implants well inside the safe region.

The FDA approved cochlear implants for adults in the mid-1980s and, since 2000, for children from 12 months of age (9 months for one device since 2020); implanting deaf children early, while the auditory brain is still developing, gives the best spoken language. Many adult users understand speech in quiet without lip reading. Speech in noise, music, and the pitch of voices remain hard, because a few overlapping channels of envelope carry far less information than thousands of hair cells and the fine timing they encode.

10.4 *The Eye and the Retina*

Light from an object is bent by the cornea and the lens and focused onto the *retina*, the thin layer of neural tissue lining the back of the eye (Fig. 10.6). The cornea does most of the focusing; the lens changes shape to adjust it for near and far objects. The retina converts photons into electrical signals, processes the image, and sends the result to the brain through about a million axons that leave the eye together as the *optic nerve*.

The *fovea*, a small pit about 1.5 mm across in the center of the *macula*, is where vision is sharpest. There the inner layers of the retina are pushed aside so light reaches the photoreceptors directly, and the cones are packed most densely. Everywhere else, light must first pass through the other layers, because the retina is "inverted": its photoreceptors face away from the light, toward the *retinal pigment epithelium* (RPE), a dark cell layer that absorbs stray light and recycles the light-sensitive pigment.

From the back of the retina to the front, five classes of cell process

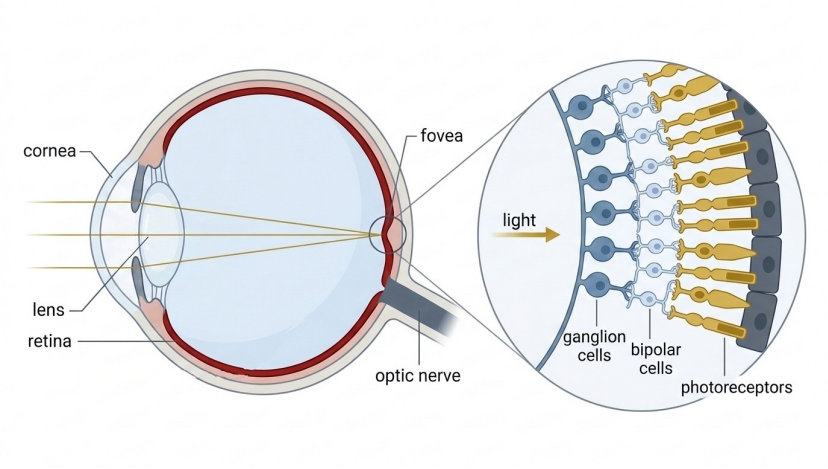


Figure 10.6: The eye. Cornea and lens focus light onto the retina; the fovea, in the center of the macula, has the highest density of cones; the optic nerve leaves at the optic disc. The inset shows the retinal layers as they lie beside the fovea (in the pit itself the inner layers are pushed aside): light crosses the ganglion and bipolar cell layers before it reaches the photoreceptors at the back.

the image (Fig. 10.12 in Section 10.6):

- *Photoreceptors*, rods and cones, convert light into graded changes of membrane potential.
- *Bipolar cells* carry the signals from photoreceptors toward the ganglion cells.
- *Horizontal* and *amacrine cells* connect cells sideways: they let neighbors inhibit each other, which sharpens edges, and shape responses to motion and to changes of light level.
- *Ganglion cells* are the first cells in the chain that fire action potentials; their axons form the optic nerve. In engineering terms they are the retina's analog-to-digital converters.

Everything before the ganglion cells works with *graded potentials*, smooth changes of membrane voltage without spikes (Chapter 3): the retina does analog image processing before it digitizes. That matters for implants. An implant that stimulates the bipolar cells keeps some of this processing; one that stimulates ganglion cells must do all of it in electronics.

Photoreceptors and visual processing

The human retina has two kinds of photoreceptor. About 92 million *rods* handle vision in dim light: they are very sensitive (a rod responds to a single photon), slow, and saturate in daylight. About 4.6 million *cones* work in brighter light; they are less sensitive and faster, and they come in three types with different spectral sensitivities, which is what gives us color.[5] Short-wavelength (S, "blue") cones peak near 420 nm, middle (M, "green") cones near 534 nm, long (L, "red") cones

[5] Counts from C. A. Curcio *et al.*, *J. Comp. Neurol.* (1990).

near 564 nm, and rods near 498 nm (Fig. 10.7). All four fall to almost nothing by 700 nm, so near-infrared light is invisible to them. The center of the fovea contains only cones, packed at about 150 000 per square millimeter with centers about 3 µm apart.

Photoreceptors work backward from what an engineer might expect. In the dark, ion channels in the cell are held open, a steady inward "dark current" of sodium flows, and the cell sits depolarized near −40 mV, releasing the neurotransmitter glutamate all the time. A photon absorbed by the pigment (rhodopsin, in rods) starts a chemical cascade that closes many of these channels, so the cascade is the photoreceptor's amplifier. The cell *hyperpolarizes* toward −70 mV and releases less glutamate: light is signaled by a decrease in transmitter. Bipolar cells read this signal in two ways. *OFF bipolar cells* follow the photoreceptor; *ON bipolar cells* invert it and depolarize in light.

Each ganglion cell collects signals from a small patch of retina, its *receptive field*. Because horizontal and amacrine cells let neighbors inhibit each other, the field has a center and an opposing surround: an ON-center cell fires most for a small spot of light that fills its center, and less if the light also covers the surround. The retina therefore reports edges and local contrast more than absolute brightness. Near the fovea a ganglion cell's center may receive input from a single cone; in the periphery, from hundreds of rods.

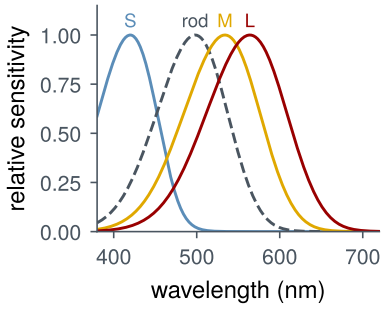


Figure 10.7: Normalized spectral sensitivity of the three cone types and the rods, computed from a standard visual-pigment template (V. I. Govardovskii *et al.*, *Vis. Neurosci.*, 2000) with the peak wavelengths given in the text.

From the eye to the cortex

The optic nerves meet at the optic chiasm, where half the fibers of each eye cross, so that each hemisphere of the brain receives the opposite half of the visual field. After a relay in the thalamus, the signals reach the *primary visual cortex* (V1) at the back of the head. V1 is *retinotopic*: neighboring points in the visual field map to neighboring points on the cortex. The map is far from uniform. The *cortical magnification*, millimeters of cortex per degree of visual field, falls from about 23 mm/deg at the center of gaze to 1.6 mm/deg at 10° (Fig. 10.8), so about a third of the cortex that serves the central 10° serves the central 2°, only 4 % of that field's area. A cortical implant with evenly spaced electrodes therefore produces phosphenes that are crowded near the center of gaze and sparse at the edges.

The visual cortex is also shaped by experience early in life. In classic experiments, Hubel and Wiesel closed one eye of a kitten for a month during a critical period a few weeks after birth; afterward almost no cells in the visual cortex responded to that eye, while closing an adult cat's eye had no such effect. The brain must have learned to see, so retinal implants help people who lost their sight after years of vision, not people blind from birth.

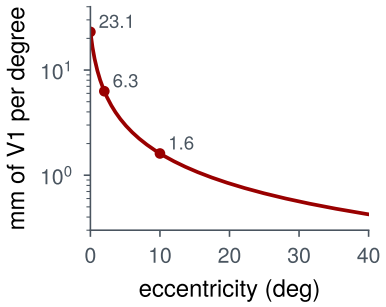


Figure 10.8: Cortical magnification in human V1 against the angle E from the center of gaze (the eccentricity), $M \approx 17.3/(E + 0.75)$ mm/deg, with values at 0, 2, and 10° marked (after J. C. Horton and W. F. Hoyt, *Arch. Ophthalmol.*, 1991). The fovea takes a large share of the cortex.

10.5 *Estimating Spatial Acuity*

ACUITY IS MEASURED IN ANGLES, because the size of an image on the retina depends on both the size of an object and its distance. An object of size h at distance d subtends a *visual angle* $\theta \approx h/d$ (in radians) at the eye, and its image on the retina has a size θ times the distance from the eye's optical nodal point to the retina, about 16.5 mm in an adult eye (Fig. 10.9). One degree is $\pi/180$ radian, so

$$\text{one degree} \approx 16.5\,\text{mm} \times \frac{\pi}{180} = 288\,\mu\text{m of retina.} \tag{10.4}$$

KEY EQUATION

$1^\circ \approx 288\,\mu\text{m}$

Retinal distance per degree of visual angle in the human eye.

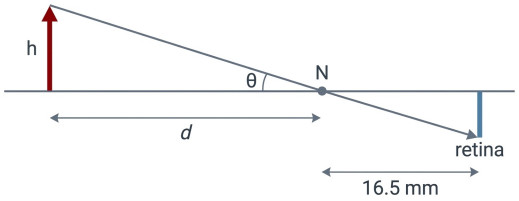


Figure 10.9: Visual angle. Rays through the nodal point N pass undeviated, so an object h at distance d makes an image $\theta \times 16.5$ mm on the retina, with $\theta = h/d$.

Clinicians express acuity as a *Snellen fraction*. Normal acuity, 20/20, means you can read at 20 feet a letter whose strokes and gaps subtend 1 arcmin, a sixtieth of a degree (the whole letter is 5 arcmin tall). On a chart 6 m away such a stroke is 1.75 mm wide, and its image on the retina is $288/60 = 4.8\,\mu$m, less than two cone spacings. A person with 20/200 acuity needs letters ten times larger: at 20 feet they read what a normal eye reads at 200 feet. For an acuity of $20/x$, the *minimum angle of resolution* (MAR), its logarithm (logMAR), and the grating acuity are

$$\text{MAR} = \frac{x}{20}\ \text{arcmin}, \qquad \text{logMAR} = \log_{10}\text{MAR},$$
$$\text{grating acuity} = \frac{30}{\text{MAR}}\ \text{cycles per degree}, \tag{10.5}$$

the last because one cycle of a black-and-white grating contains two bars of width MAR, and a degree has 60 arcminutes. Each 0.1 of logMAR is one line of a standard eye chart. Table 10.1 lists a few values. In the United States, 20/200 or worse in the better eye with correction defines legal blindness.

Snellen	MAR	logMAR	cyc/deg
20/20	1	0.0	30
20/200	10	1.0	3
20/400	20	1.3	1.5
20/1262	63	1.8	0.48

Table 10.1: Acuity in four units, from Eq. (10.5); MAR in arcminutes. The last row is the best grating acuity measured with the Argus II.

The acuity of the cone mosaic

The retina samples the image with its photoreceptors, so their spacing sets a Nyquist limit, just as a sampling rate does in time (Chapter 1). Foveal cones 3 µm apart each cover $3/288 = 0.0104^\circ$, or 0.63 arcmin, and two samples per cycle give a limit of $1/(2 \times 0.0104) = 48$ cycles/deg. Because the cones are packed hexagonally, their rows are closer than one spacing, and the limit rises to about 60 cycles/deg, or 20/10: the best acuity of a human eye.

The acuity of an electrode array

An implant samples the visual field with its electrodes. If each electrode produces one phosphene, the finest detail it can convey is set by the electrode *pitch* p, the distance between electrode centers. One pitch subtends $p/(288\,\mu\text{m})$ degrees, and taking the MAR equal to that angle,

$$\text{MAR} = 60\,\frac{p}{288\,\mu\text{m}}\ \text{arcmin}, \qquad 20/x \text{ with } x = 20\,\text{MAR} \approx 4.17\,p/\mu\text{m}. \tag{10.6}$$

A grating then needs two electrodes per cycle (Fig. 10.10).

This is an upper bound on what the hardware can deliver. Current spreads through the retina, so a phosphene is usually larger than the pitch, and nearby electrodes may evoke merged phosphenes. An electrode on the surface of the retina also excites the axons of ganglion cells that pass beneath it on the way to the optic nerve, which stretches phosphenes into streaks. On the other hand, patients who move their heads to scan a camera across a scene can sometimes resolve finer detail than one static frame allows. Figure 10.11 compares the bound with reported results.

KEY EQUATION

$20/x,\ x \approx 4.17\,p[\mu\text{m}]$

Acuity limit of an electrode pitch p: one pitch per MAR.

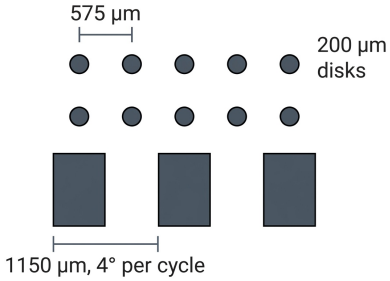


Figure 10.10: Part of the Argus II grid and the finest grating it can represent: one bar per electrode, so one cycle spans two pitches, 1150 µm or 4° of visual field.

WORKED EXAMPLE 10.3 *Electrode pitch and acuity*

Find the acuity bound for (a) the Argus II, 575 µm pitch; (b) the Alpha IMS, 70 µm; (c) PRIMA, 100 µm. (d) Compare with the foveal cone mosaic.

SOLUTION. (a) One pitch is $575/288 = 2.0° = 120$ arcmin, so the bound is about 20/2400. A grating needs 4° per cycle, or 0.25 cycles/deg (Fig. 10.10).

(b) One pitch is $70/288 = 0.243°$, or 14.6 arcmin: about 20/290, or 2.1 cycles per degree.

(c) $100/288 = 0.347° = 20.8$ arcmin: about 20/420.

(d) The cones, 3 µm apart, give 0.63 arcmin, or 20/12.5 by the same rule. The Argus II pitch is about 190 times coarser, and its 0.25 cycles/deg is 240 times worse than the 60 cycles/deg of the best human eye. ■

10.6 *Retinal Disease and Where to Stimulate*

RETINAL IMPLANTS ARE FOR DISEASES that destroy the photoreceptors but leave the inner retina, the optic nerve, and the visual cortex working. *Retinitis pigmentosa* (RP) is a family of inherited disorders, affecting about one person in 4000, in which the rods degenerate first,

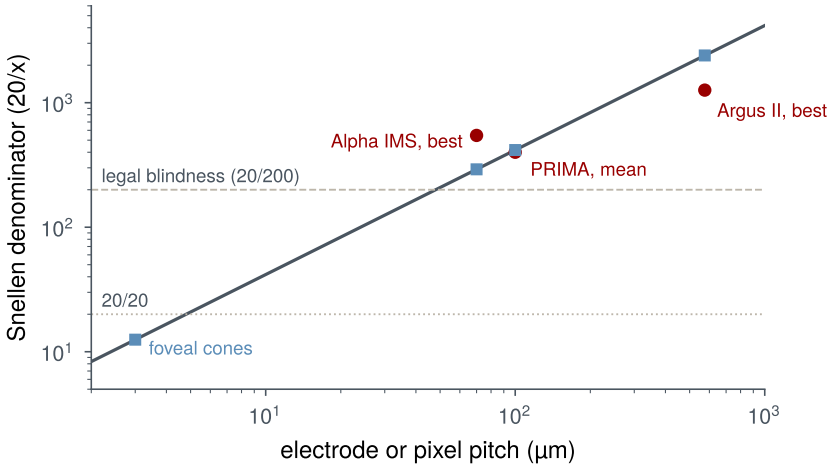


Figure 10.11: The acuity bound of Eq. (10.6) (line, blue squares) and reported acuities (red): the best grating acuity with the Argus II, the best Landolt C acuity with the Alpha IMS, and the mean acuity at one year with PRIMA. The foveal cone spacing is shown for comparison. Below the dashed line is better than legal blindness.

causing night blindness and a shrinking visual field, and then the cones, until some patients have no useful sight left. *Age-related macular degeneration* (AMD) destroys the photoreceptors and RPE of the macula, the region that serves reading and recognizing faces, while peripheral vision survives; its advanced dry form is called *geographic atrophy*. In both diseases many bipolar and ganglion cells survive, although the inner retina rewires as it loses its input. *Glaucoma*, by contrast, kills the ganglion cells and the optic nerve, so no retinal implant can help; only an implant in the brain could.

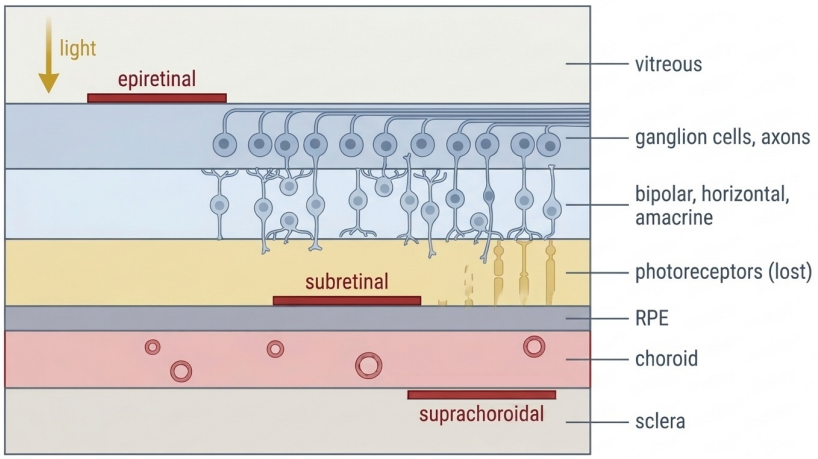


Figure 10.12: The layers of the retina, drawn flat with light arriving from the top, and the three retinal implant sites. Epiretinal arrays lie on the ganglion cells; subretinal arrays replace the lost photoreceptors under the retina; suprachoroidal arrays sit outside the choroid, against the sclera.

Figure 10.12 shows where an implant can go. An *epiretinal* array is tacked onto the inner surface of the retina, against the ganglion cells. The surgery is relatively simple, but the array bypasses all retinal processing, which the camera and computer must then do, and it also

excites the passing ganglion-cell axons. A *subretinal* array is slipped between the retina and the RPE, where the photoreceptors were. It drives the bipolar cells, so the retinal network still processes the signal, and if it carries its own photodiodes it uses the eye's own optics and eye movements; the surgery is harder and the retina can detach. A *suprachoroidal* array, outside the choroid (the blood-vessel layer) against the sclera (the white outer wall), is the safest to implant but the farthest from its targets. Beyond the eye, electrodes in a cuff around the optic nerve sit where the axons from the whole retina pass but are hard to map to the visual field, and *cortical* implants in V1, the approach Brindley started, work even when the eye or the optic nerve is gone.

10.7 The Argus II Epiretinal Prosthesis

The first retinal implant approved in the United States came out of Los Angeles. Second Sight Medical Products, founded in 1998 in Sylmar, California, with the implant engineer and entrepreneur Alfred Mann among its founders, developed the Argus implants with researchers at USC's Doheny Eye Institute led by Mark Humayun. The first-generation Argus I, with 16 electrodes, was implanted in six people starting in 2002. The Argus II, with 60 electrodes, was tested in a 30-patient trial at 10 centers in Europe and the United States. It received the CE mark in Europe in 2011 and FDA approval in 2013 under a *humanitarian device exemption* (HDE), a pathway for devices that treat rare conditions.[6]

[6] M. S. Humayun *et al.*, *Ophthalmology* (2012); Chapter 14 explains the HDE pathway.

The system

A miniature camera on a pair of glasses captures the scene; a *video processing unit* (VPU) worn on the belt converts each frame into stimulation commands; and a coil on the glasses sends power and data to a receiving coil on the eye (Fig. 10.13). The implant's coil and electronics case are fixed to the outside of the eye on a band around the sclera, and a thin cable passes through the eye wall to a platinum array of 60 electrodes, each 200 µm in diameter, in a 6×10 grid with a 575 µm pitch, held on the retina over the macula by a small tack. The six rows span $5 \times 575 = 2.88$ mm between centers and the ten columns 5.18 mm, so at 288 µm per degree the array covers about $10 \times 18^\circ$ of the visual field, 20° on the diagonal, against about 180° across for normal binocular vision.

Because the array sits on the ganglion cells, the VPU must do the image processing that the retina would have done. It crops the camera image to the field the array covers, downsamples it to the 6×10 grid,

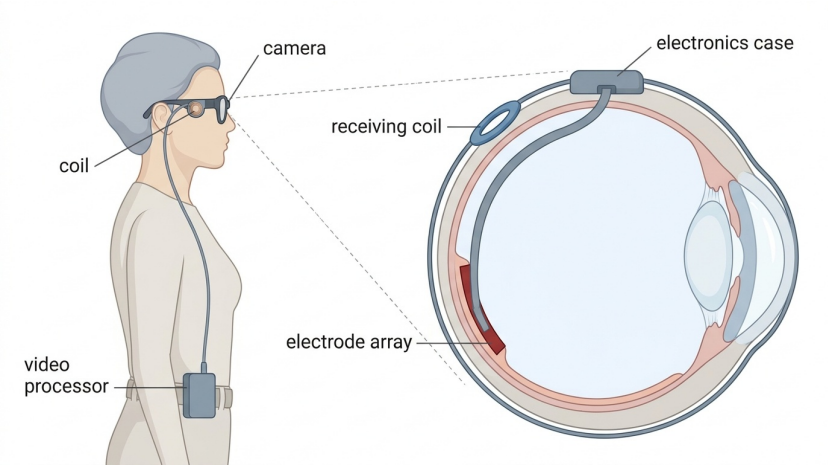


Figure 10.13: The Argus II system: a camera on the glasses, a video processing unit worn on the body, and an external coil on the glasses that powers and commands the implant on the eye.

enhances edges (standing in for the retina's center-surround fields), and maps brightness to current between each electrode's threshold and maximum. A camera of 640×480 pixels at 30 frames per second and 8 bits per pixel produces 73.7 Mb/s; 60 amplitudes of 6 bits sent 20 times per second are only 7.2 kb/s, about 10 000 times less. All of the processing happens outside the body, where power and computing are cheap.

The stimulator chips

Two generations of stimulator chips show the circuit side. A chip designed for the first system had 8 outputs, current DACs with 6-bit resolution and full scales up to 600 µA, supplies of ± 6.5 V, and anodic and cathodic currents matched to 1.2 to 1.7 %. It was tested with 600 µA biphasic pulses of 1 ms per phase into 10 kΩ loads, and its rails were just enough: the load drops $600\,\mu\text{A} \times 10\,\text{k}\Omega = 6.0$ V, and the current source needs 0.5 V more of headroom, the compliance rule of Chapter 9. A later 256-channel epiretinal stimulator chip received power and data on two separate inductive links and used a ± 12 V push-pull output stage; its link was designed to deliver about 100 mW.[7] Higher rails drive small electrodes whose impedance rises as tissue grows around them, but they cost power: a current source draws its current from the full rail whatever the electrode needs, so an electrode that drops 1.5 V on a 12 V rail uses only 12.5 % of the power drawn.

[7] K. Chen *et al.*, "An integrated 256-channel epiretinal prosthesis," *IEEE J. Solid-State Circuits* (2010).

Worked Example 10.4 *Charge density on a 200 micrometer electrode*

Suppose an Argus-type electrode (200 µm diameter) needs 150 µA for 0.45 ms per phase to evoke a phosphene (an assumed

value of the right order for epiretinal stimulation). (a) Find the charge per phase and the charge density. (b) Check the Shannon criterion. (c) Can smooth platinum inject it? (d) What does the test condition of the first-generation chip, 600 µA for 1 ms, mean for this electrode?

SOLUTION. (a) $A = \pi(0.01\,\text{cm})^2 = 3.14 \times 10^{-4}\,\text{cm}^2$; $Q = 150\,\mu\text{A} \times 0.45\,\text{ms} = 67.5\,\text{nC}$; $Q/A = 215\,\mu\text{C/cm}^2$.

(b) $k = \log_{10} 215 + \log_{10} 0.0675 = 2.33 - 1.17 = 1.16$, well below 1.85. At $k = 1.85$ this electrode could take up to $Q_{\max} = \sqrt{10^k A} = 149\,\text{nC}$ per phase (Chapter 9).

(c) No. Smooth platinum's 0.05 to 0.15 mC/cm² allows only 16 to 47 nC on this area. The electrode needs a roughened, high-surface-area platinum or an iridium oxide coating.

(d) $0.6\,\mu\text{C}/3.14 \times 10^{-4}\,\text{cm}^2 = 1.9\,\text{mC/cm}^2$ and $k = 3.06$: far beyond both limits. The full-scale current is for larger electrodes or shorter pulses; firmware must cap the charge per phase for each electrode size. ■

Results

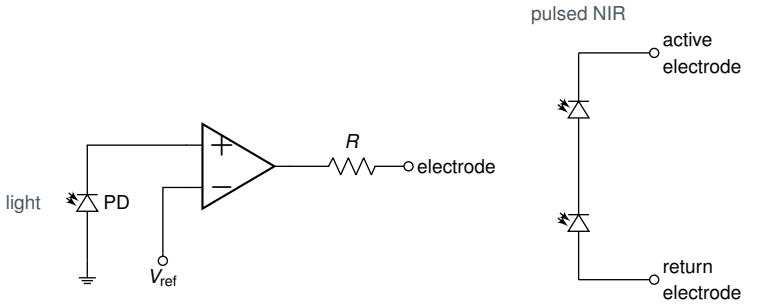
In the 30-patient trial, all subjects did better with the system on than with it off on tasks such as locating a bright square on a screen or walking along a line. Seven of the 30 could score on a grating acuity test, at best 1.8 logMAR (20/1262), about twice as fine as the 20/2400 bound of Example 10.3, which head scanning can allow. With practice, users picked out doors, windows, and large letters from patterns of flashes. Second Sight stopped producing the Argus II in 2019. It showed that an implant could give blind people useful, if crude, vision for years.

10.8 Subretinal and Photovoltaic Implants

A SUBRETINAL IMPLANT PUTS THE LIGHT SENSORS BACK where the photoreceptors were. The *Alpha IMS*, made by the German company Retina Implant AG, was a silicon chip about 3×3 mm carrying 1500 pixels on a 70 µm grid, each with a photodiode, an amplifier, and a stimulating electrode (Fig. 10.14a). Placed under the fovea, it lets the eye's own optics focus the scene onto the photodiodes. Each pixel amplifies its photocurrent and drives its electrode, so the image becomes stimulation right where it falls, the surviving bipolar network still processes it, and the patient looks around with normal eye movements. The chip covers $3/0.288 = 10.4^\circ$ on a side. The amplifiers need power, which comes through a cable from a coil under the skin behind the ear. The best patients read a Landolt C (a ring with a gap whose

direction the patient names) at 20/546, and many located objects better with the implant on.[8] The device received the CE mark in 2013, but the company closed in 2019.

[8] K. Stingl *et al.*, *Proc. R. Soc. B* (2013) and *Vision Res.* (2015).



(a) amplified pixel (Alpha IMS) (b) photovoltaic pixel (PRIMA)

Figure 10.14: Two ways to turn light into stimulation under the retina, simplified. (a) Each Alpha IMS pixel amplifies its photodiode signal against a reference V_{ref} with a differential amplifier whose gain is set from outside the chip (the gain-setting parts are not drawn), so it needs a wired power supply. (b) A PRIMA pixel is a few photodiodes in series that drive current between an active and a return electrode directly from the energy of pulsed near-infrared light; it needs no wires.

Photovoltaic implants: PRIMA

A *photovoltaic* implant removes the wires. A photodiode under light makes a current in proportion to the light intensity (Chapter 7); like a small solar cell, it can also deliver that current into a load at a few tenths of a volt. In PRIMA, developed in Daniel Palanker's laboratory at Stanford, each pixel connects a few photodiodes in series between an active electrode in its center and a return electrode around its edge (Fig. 10.14b), so their voltages add. A camera on a pair of glasses captures the scene, a pocket processor converts it, and a projector in the glasses shines the image onto the implant as pulses of near-infrared light. The light both carries the image and powers the pixels, which turn it into pulses of current. Because near-infrared light is invisible to photoreceptors (Fig. 10.7), the patient's surviving peripheral vision is not disturbed. The chip is 2×2 mm and about 30 µm thick, with 378 pixels 100 µm across, and it covers $2/0.288 = 6.9^\circ$ on a side, enough for the central field used in reading.[9]

PRIMA targets geographic atrophy from AMD. In the PRIMAvera trial, 38 patients in Europe were implanted under the atrophic macula. Of the 32 who completed a year, 26 (81 %) improved by at least 0.2 logMAR (two lines on the chart); the mean gain was 0.51 logMAR, and the mean final acuity of about 20/400 matched the pixel size, as Example 10.3 predicts. Most participants used it to read letters and words at home. About half had surgical complications, among them macular holes and retinal detachment. PRIMA, now owned by Science Corporation, received the CE mark in Europe in July 2026; it still needs FDA approval in the United States.

[9] Trial published by F. G. Holz *et al.*, *N. Engl. J. Med.* (2026); specifications as reported there and by Science Corporation.

10.9 Trends in Visual Prostheses

THE MAIN TREND IS MORE PIXELS. Example 10.5 works out how many are needed to read. More pixels bring the problems of Chapter 9: smaller electrodes with higher impedance and charge density, and more power and heat in a small eye. Current spreading between closely spaced electrodes limits how far shrinking helps; local return electrodes, as in PRIMA, confine the field. For people whose retina or optic nerve is gone, cortical prostheses with penetrating arrays revive Brindley's approach with the electrodes and wireless electronics of Chapter 9. For some forms of RP, gene therapy and optogenetics, which make surviving retinal cells sensitive to light, may replace electrodes altogether.

> **WORKED EXAMPLE 10.5** *How many pixels to read*
>
> (a) What pitch gives an acuity bound of 20/200, 20/100, and 20/80? (b) How many pixels cover a $10 \times 10^\circ$ field at each pitch? (c) If each pixel took 10 µW on average (an assumed value), which designs fit a 10 mW budget?
>
> SOLUTION. (a) From Eq. (10.6), $p = 288 \times x/1200$ in micrometers: 48, 24, and 19 µm.
>
> (b) The field is 2.88 mm on a side, so $(2880/p)^2$ gives 3600, 14 400, and 22 500 pixels.
>
> (c) 36 mW, 144 mW, and 225 mW: none of them. Even the 20/200 design needs less than 3 µW per pixel to fit, which is why photovoltaic pixels that draw their energy from light, and stimulation only of pixels that see an edge, are attractive. ■

Summary

- A sensory prosthesis: external sensor and processor, inductive link, implanted stimulator, electrodes on surviving neurons.
- The cochlea is tonotopic, with high frequencies at the base: $f = 165.4(10^{2.1x} - 0.88)$; a 25 mm insertion reaches about 500 Hz.
- A cochlear implant processor: log-spaced bands, $f_k = f_{\text{lo}} r^k$ with $r = (f_{\text{hi}}/f_{\text{lo}})^{1/N}$; envelopes; log compression between T and C; CIS pulses interleaved in slots of $1/(Nf_p)$.
- In the retina, ganglion cells are the first cells to fire spikes. RP and AMD destroy photoreceptors but spare the inner retina; glaucoma does not. Epiretinal arrays drive ganglion cells, subretinal arrays bipolar cells. Implants help only people who once had sight.

- One degree is about 288 µm of retina. For $20/x$, MAR $= x/20$ arcmin, and a pitch p (in µm) bounds acuity at about $20/(4.17p)$: 20/2400 for the Argus II, 20/420 for PRIMA; 20/200 needs 48 µm.
- Argus II: 60 electrodes at 575 µm, 20° diagonal, processing outside the body; best 20/1262. PRIMA: 378 photovoltaic pixels powered by near-infrared light; mean about 20/400.

Part IV, starting in Chapter 11, turns to how implants are built.

10.10 Problems

10.1 ° For a 35 mm cochlea: (a) find the best frequency 8 mm and 18 mm from the base; (b) find where 500 Hz and 2 kHz lie; (c) how far apart are 2 kHz and 4 kHz?

10.2 • A CIS processor drives 16 channels at 1200 pulses/s each. (a) Find the slot per pulse. (b) With a 10 µs gap, find the longest phase. (c) If the phases must be 50 µs to reach threshold on some contacts, what per-channel rate is possible?

10.3 • A contact has $T = 80$ µA and $C = 400$ µA; the processor keeps 40 dB of acoustic range. (a) Find the electric dynamic range in decibels. (b) Find the current for envelopes at 10, 25, and 40 dB above $e_{\min}$. (c) With 30 µs phases at C on a 0.2 mm^2 contact, find Q, Q/A, and k.

10.4 • A 1024-electrode epiretinal implant is updated 30 times per second with 6-bit amplitudes, plus 16 bits of overhead per frame. (a) Find the data rate. (b) Compare with a camera of 1280×720 pixels, 8 bits, 30 frames per second.

10.5 •• *(Design.)* Design a subretinal photovoltaic array for reading at 20/160 over a $6 \times 6°$ field. (a) Find the pitch and the pixel count. (b) If each pixel needs 20 nC per phase to reach threshold on an electrode of the pixel's area, check the charge density against iridium oxide (3 mC/cm^2) and the Shannon criterion. (c) Comment on what limits further shrinking.

Part IV

Building Devices

11
Materials and Mechanics of Soft Electronics

In this chapter you will learn to

- Explain why bioelectronic devices are made soft and thin, and choose a substrate from its temperature limit, expansion, modulus, and use on or in the body.
- Sort materials into conductors, semiconductors, and insulators by band gap and conductivity.
- Compare evaporation, sputtering, CVD, and ALD, and choose a method for a given film and substrate.
- Find the bending strain in a layered stack from its neutral plane, and place a fragile layer at the neutral plane with an encapsulation.
- Describe how thin films crack, buckle, and peel, and estimate the stress left by a thermal mismatch.
- Compute the resistance of a stretched conductor, and pick a geometry (serpentine, wrinkles, island-bridge) for a stretchable circuit.
- Explain why nanoparticle inks sinter far below the melting point, and compute the resistance of a printed trace from its sheet resistance.
- Explain conjugation and doping in organic materials, extract the mobility of an OTFT, and compute the transconductance of an OECT.
- Estimate how long an encapsulation keeps water away from a device.

Take apart a sweat-sensing patch and look at what it is made of. The base is a polyester film thinner than a sheet of paper. The wiring is silver, printed as an ink of nanoparticles and baked at a temperature

not far above that of boiling water. The electrodes that touch the sweat are carbon, and on top of one of them sits a layer of a blue conducting polymer that swells with water and turns ion concentrations into a current. The only silicon is a chip the size of a grain of rice at one end, and the whole patch bends around the forearm without cracking.

Every device in the first three parts of this book is built from layers like these. The electrochemical sensors of Chapter 6 are thin films of carbon, gold, or platinum under an enzyme or a membrane. The polyimide threads of Chapter 9 are 5 µm thick so that they bend with the brain, and their contacts carry a coating of conducting polymer a fraction of a micrometer thick. The photovoltaic retinal implant of Chapter 10 is a silicon chip only 30 µm thick. Part IV of the book is about how such devices are made.

This chapter follows a soft device from the bottom up. It starts with why devices are made soft and thin, the three classes of electronic materials, and the substrates that carry them. Then it makes thin films in a vacuum chamber, bends them (strain and the neutral plane), breaks them (how films fail), and stretches them (stretchable conductors). The second half turns to the materials of printed and organic electronics: conductive inks and how heat or light turns them into metal; organic conductors and semiconductors, and the two transistors built from them, the OTFT and the OECT; and the encapsulation that keeps water out. Chapter 12 then puts the inks through printers.

11.1 *Why Soft and Thin*

Tissue is soft. Brain is near 1 kPa, about as stiff as a custard; skin is 0.1 to 1 MPa; even cortical bone, at about 20 GPa, is several times softer than silicon (Fig. 11.1). Chapter 1 defined the *Young's modulus E*, the stress (force per area) needed per unit of strain (fractional stretch), and showed what a stiff device does to soft tissue: it presses and rubs, it lifts off the skin and picks up motion noise, and inside the body it provokes the scar tissue that walls off rigid neural probes (Chapter 9).

key idea Make stiff materials thin, or use materials that are soft to begin with. Most soft devices do both.

There are two ways to make an electronic device soft. The first is to use materials that are soft to begin with: elastomers such as silicone rubber, hydrogels, and conducting polymers. The second is to make stiff materials thin. The bending stiffness of a sheet grows as the cube of its thickness ($D = Et^3/[12(1-\nu^2)]$ per unit width, Chapter 1), so a film ten times thinner is a thousand times easier to bend. Gold is thirty times stiffer than polyimide, yet a gold line 100 nm thick on a polyimide sheet a few micrometers thick follows the ridges of a fingertip. Most soft devices combine the two routes.

A *thin film* is a layer of material much thinner than it is wide, from

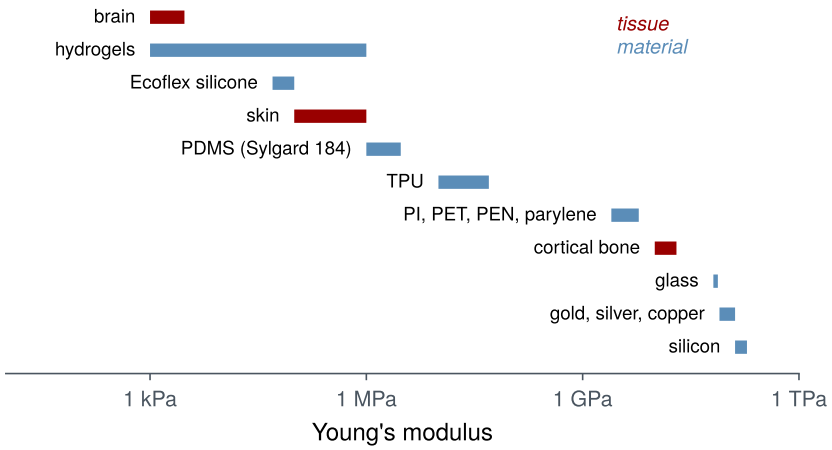


Figure 11.1: Young's modulus of tissues (red) and materials (blue), typical ranges. Plastics are thousands of times stiffer than skin; silicones and hydrogels close the gap.

a single layer of atoms to a few micrometers, held up by a thicker *substrate*. Almost every layer of an integrated circuit is a thin film, and bioelectronic devices add a few of their own: the gold or platinum of an electrode, the iridium oxide that stores charge for stimulation, the parylene that keeps body fluid out, the PEDOT:PSS that lowers an electrode's impedance. Film properties differ from bulk ones, because grain size, defects, and stress depend on how the film was made.

Soft substrates bring one constraint that silicon processing does not have: the substrate may melt, shrink, or decompose at the process temperature. A silicon wafer survives 1000 °C; polyethylene terephthalate (PET) softens near 80 °C and polyimide near 360 °C. Many of the methods in this chapter, such as plasma-enhanced deposition or flash-lamp sintering, exist to put the energy into the film and not into the substrate.

11.2 *Conductors, Semiconductors, and Insulators*

Every electronic system needs three kinds of material. *Conductors* carry current with little loss: the wires, electrodes, and contacts. *Semiconductors* carry current that a voltage or light can switch and modulate: the channels of transistors and diodes, and many sensing layers. *Insulators* carry no current: gate dielectrics, the insulation between wires, the substrate, and the encapsulation.

The difference between them is an energy gap (Fig. 11.2). Electrons in a solid can take only certain energies, which group into *bands*. In a metal a band is only partly filled, so an electric field easily pushes electrons into empty states just above the filled ones. In a semiconductor the filled band is separated from the next empty one by a gap E_g of about 1 to 3 eV (1.12 eV for silicon), small enough that heat, light,

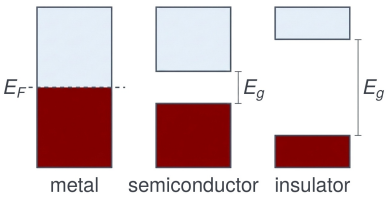


Figure 11.2: Energy bands, with filled states in red. A metal has a partly filled band, filled up to the Fermi energy E_F; a semiconductor has a gap E_g of about 1 to 3 eV that heat, light, or doping can bridge; an insulator's gap is too wide.

doping, or a gate voltage can put carriers into play. In an insulator the gap is so wide (9 eV for SiO_2) that essentially no carriers exist. The conductivity

$$\sigma = nq\mu \tag{11.1}$$

is the product of the carrier density n, the charge q, and the *mobility* μ, the drift velocity of a carrier per unit of electric field. The carrier density runs from about $10^{23}\,\mathrm{cm}^{-3}$ in a metal to almost none in an insulator, so the conductivity spans more than 20 orders of magnitude (Fig. 11.3).

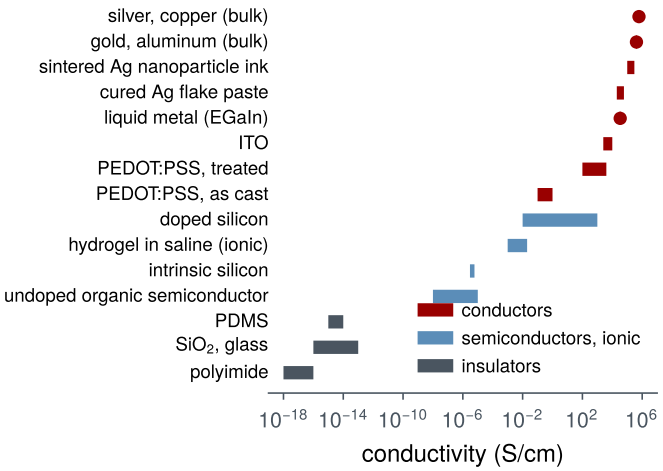


Figure 11.3: Typical conductivities at room temperature. Printed conductors fall short of bulk metal by a factor of 2 to 30; conducting polymers sit between metals and semiconductors; hydrogels conduct by ions, not electrons.

Silicon technology builds all three classes from a few elements: doped silicon for semiconductors, SiO_2, Si_3N_4, and HfO_2 for insulators, and aluminum, copper, and tungsten for conductors. Flexible electronics draws on a wider palette (Table 11.1), because each layer must survive bending, low process temperatures, and often printing from solution.

	Silicon electronics	Flexible and printed electronics
Conductors	Al, Cu, W, doped poly-Si	Ag and Au nanoparticle and flake inks, Ag nanowires, Cu inks, carbon, PEDOT:PSS, liquid metal (EGaIn), ITO, graphene
Semi-conductors	crystalline Si, Ge, GaAs	small molecules (pentacene, C8-BTBT), polymers (P3HT), metal oxides (IGZO), Si nanomembranes
Insulators	SiO_2, Si_3N_4, HfO_2	polymers (parylene, PMMA, PVP, CYTOP, SU-8), ALD Al_2O_3, the plastic substrate itself

Table 11.1: The three classes of electronic materials, in silicon technology and in flexible electronics. The rest of this chapter meets most of them.

11.3 Substrates: From Plastic to Hydrogels

The substrate is the largest part of most flexible devices, and it decides much of what the rest can be. Seven properties matter. The *maximum process temperature*, set by the *glass transition* T_g (above which a polymer turns from glassy to rubbery and starts to soften) or by melting or decomposition, limits every later step: a silver ink that needs 200 °C cannot go on PET. The *thermal expansion* and the *water uptake* decide whether layers printed one after another still line up (see the end of this section). The *Young's modulus* and the thickness set the bending stiffness and how well the device conforms. *Surface roughness* limits how thin a layer can be printed on top without shorting through, and *transparency* matters for optical sensors. Finally, *biocompatibility* decides where on or in the body the device can go. Table 11.2 lists the most common substrates, and Fig. 11.4 shows several of them side by side.

Substrate	Max. process temperature	CTE (ppm/K)	Modulus	Water uptake	Typical use
Paper	about 150 °C	set by humidity	2 to 5 GPa	high, several %	disposable diagnostics that wick fluid
PET	T_g 70 to 80 °C; 150 °C if heat stabilized	15 to 33	2 to 5 GPa	0.1 to 0.4 %	low-cost wearables, printed sensors
PEN	$T_g \approx 120$ °C; about 180 °C	13 to 20	5 to 6 GPa	0.1 to 0.4 %	wearables that need hotter cures
Polyimide (Kapton HN)	about 400 °C (T_g 360 to 410 °C)	20	2.5 GPa	1.8 to 2.8 %	flexible circuits; neural and retinal implants
Parylene C	melts at 290 °C (T_g below 90 °C)	35	2.8 GPa	below 0.1 %	conformal implant coating
Silicon	above 1000 °C	2.6	130 to 190 GPa	none	chips; flexible once thinned
PDMS (Sylgard 184)	about 200 °C	310	1 to 3 MPa	below 0.1 %; gas permeable	stretchable skin devices, microfluidics
Ecoflex silicone	about 200 °C	large	50 to 100 kPa	low	very soft skin patches
TPU	about 100 °C	large	10 to 50 MPa	0.5 to 1.5 %	stretchable printed circuits
Textiles	polyester melts at 250 °C	n/a	very compliant	cotton about 8 %	garments worn all day
Hydrogels	dry out above about 60 °C	n/a (swell)	1 kPa to 1 MPa	70 to 99 % water	skin-electrode gels, tissue-like coatings

Table 11.2: Substrates for bioelectronics, with typical values (CTE is the thermal expansion coefficient, defined at the end of this section). Grades vary widely; use the datasheet of the film you buy. The modulus column spans eight orders of magnitude, from silicon to hydrogels.

Plastic films are the workhorses. PET (the plastic of drink bottles) is cheap, clear, and smooth, but softens near 78 °C; heat-stabilized grades survive brief bakes at 120 to 150 °C. PEN (polyethylene naphthalate) is stiffer and stable to about 180 °C. *Polyimide* survives soldering and 350 °C processing and is the base of flexible circuit boards and of most thin-film neural electrodes, but it is amber, it absorbs water, and it swells. *Parylene C* is deposited from the vapor at room temperature (Section 11.4) as pinhole-free films a few micrometers thick that follow every surface; it is among the best-accepted implant materials. Silicon wafers and thin glass are flat and smooth and survive high temperatures, but they are rigid until thinned (Section 11.5). *Paper* is cheap, disposable, and porous, so it wicks fluid like a microfluidic channel, which paper diagnostics use; its fibers make it rough, and its size

changes with humidity.

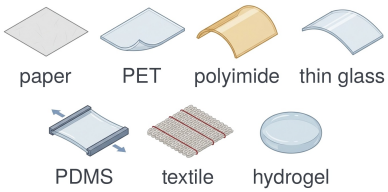


Figure 11.4: Substrates side by side: paper, PET, amber polyimide, thin glass, clear PDMS, a knit textile, and a hydrogel disk.

Stretchable substrates

Skin and muscle are soft, and they also stretch. *Elastomers* stretch by tens to hundreds of percent and spring back. The most common is *polydimethylsiloxane* (PDMS), sold as Sylgard 184: a base and a curing agent mixed 10:1 and cured at 60 to 100 °C into a clear rubber with a modulus near 1.5 MPa, a thousand times softer than PET. It molds features down to nanometers, and it is permeable to gases, which is good for skin and bad for barrier layers. Softer silicones (Ecoflex) reach 100 kPa and stretch nearly tenfold. Thermoplastic polyurethane (TPU) is stiffer and tougher, melts for lamination and 3D printing, and holds printed inks better. The modulus sets how hard the device pulls on the skin: a strip of cross-section A stretched by a strain ε pulls back with a force $F = E\varepsilon A$. The difficulty with all elastomers is adhesion: metal and oxide films do not bond to silicone well, so devices use adhesion layers, plasma activation, or mechanical interlocking.

Electronic textiles weave or knit conductive fibers into fabric, or print and laminate circuits onto it; knits stretch by rearranging their loops, not by straining the fiber. *Hydrogels*, cross-linked polymer networks that hold 70 to 99 % water, match the softness of tissue and conduct ions. They are the gel of a wet ECG electrode (Chapter 8), drug reservoirs, and soft coatings on implants.

Dimensional stability

A printed device is built layer by layer, and each layer must land on the one below to within about 10 µm. Plastic substrates move in three ways. They expand with temperature: the *coefficient of thermal expansion* (CTE) α is the strain per kelvin of a free material, quoted in parts per million per kelvin (ppm/K). They swell with water. And oriented films shrink permanently the first time they are heated.

Take polyimide (Kapton HN, $\alpha = 20\,\text{ppm/K}$). Printed at 25 °C and cured at 125 °C, a 100 mm pattern grows while hot by $\alpha\,\Delta T\,\ell = 20 \times 10^{-6} \times 100\,\text{K} \times 100\,\text{mm} = 200\,\mu\text{m}$, or 0.2 %, twenty times the alignment tolerance. The growth reverses on cooling. So the second electrode of a two-sided polyimide capacitor is printed at 100 % scale if the film has cooled, but at 100.2 % if it is printed on a platen held at 125 °C. Polyimide also grows by about 20 ppm per percent of relative humidity: a swing from 30 to 70 % RH stretches the same pattern by 80 µm. PET and PEN heated above their glass transition shrink for good, by up to a percent or more, unless they are heat stabilized (preshrunk at the factory). The remedies are to bake the film once before

the first print, to print every layer at the same temperature and humidity, and to align each layer on printed fiducial marks (Chapter 12).

11.4 Making Thin Films

A THIN FILM CAN BE LAID DOWN from a liquid or from a vapor. Liquid routes, such as spin coating a polymer or printing an ink, appear in Section 11.8 and Chapter 12. The classical route, and the one that gives the densest and purest films, is *vacuum deposition* (Fig. 11.5). Its two families differ in how the material reaches the surface. In *physical vapor deposition* (PVD), atoms are knocked or boiled off a solid source and fly to the substrate, where they condense; no chemical reaction is needed, so almost any metal can be deposited. In *chemical vapor deposition* (CVD), precursor gases react on the substrate, and the solid product stays behind.

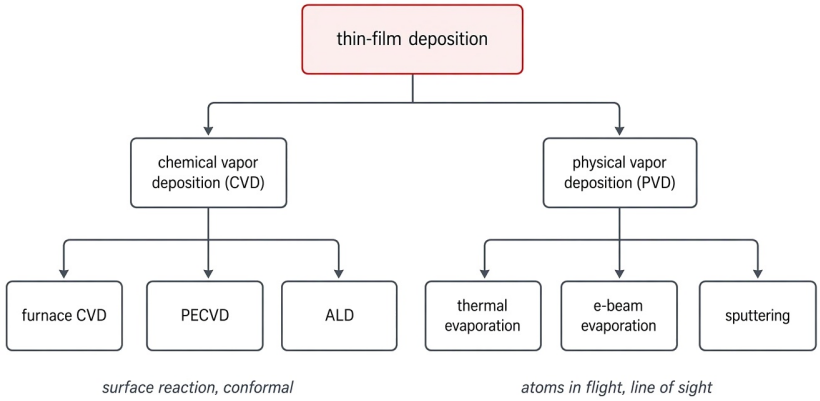


Figure 11.5: The vacuum deposition methods of this chapter. CVD builds a film from a chemical reaction at the surface; ALD is its self-limiting, one-layer-at-a-time form. PVD moves atoms from a solid source to the substrate by heat (evaporation) or by ion bombardment (sputtering).

PVD needs a vacuum because an atom flies in a straight line only until it hits a gas molecule. The average distance between collisions, the *mean free path*, is about 67 nm in room air but about 50 m at 10^{-6} Torr. (Vacuum pressures are still quoted in torr; 1 Torr is 133 Pa, and 760 Torr is one atmosphere.) At that pressure atoms cross the chamber without scattering, and background gas such as water lands on the growing film slowly enough not to spoil it.

When the atoms land, they hop over the surface and gather into small islands that grow and merge. Gold and silver on plastic or oxide grow this way, so a film thinner than about 10 nm is a set of disconnected islands that does not conduct. A few nanometers of titanium or chromium underneath bond to the substrate, help the noble metal cover it, and keep a gold electrode from peeling off in saline.

Physical vapor deposition

The three workhorses of PVD are sputtering, thermal evaporation, and electron-beam evaporation (Fig. 11.6).

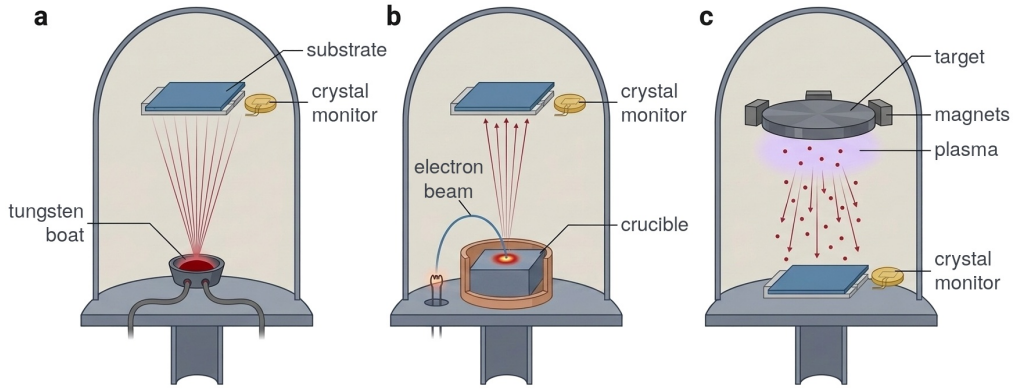


Figure 11.6: Three PVD tools. (a) Thermal evaporation: a current heats a tungsten boat until the metal in it evaporates; the vapor rises in straight lines to the substrate. (b) Electron-beam evaporation: a magnetically bent electron beam melts a small spot of the source in a water-cooled crucible. (c) Magnetron sputtering: argon ions from a plasma knock atoms out of a target, and magnets behind the target keep the plasma dense. In each chamber a quartz crystal monitor beside the substrate weighs the film as it grows.

In *sputtering*, argon gas at 1 to 20 mTorr is turned into a *plasma*, a glowing gas of ions and free electrons, by a high voltage. The source material, the *target*, is the cathode: argon ions accelerate into it and knock atoms off its surface like a cue ball breaking a rack. At these pressures the mean free path is millimeters to a few centimeters, so the ejected atoms scatter on the way and arrive from a spread of angles. Sputtered films therefore coat the sides of steps better than evaporated ones, and they are dense and well bonded. An insulating target would charge up under a steady voltage, so oxides are sputtered with a radio-frequency source at 13.56 MHz, and adding oxygen to the argon sputters compounds such as indium tin oxide (ITO), the transparent conductor of touch screens.

In *thermal evaporation* the source sits in a boat of tungsten or molybdenum, and a current of about a hundred amperes heats it until the material evaporates. At 10^{-6} Torr the vapor flies in straight lines, like light from a lamp. The method is simple and gentle: it deposits low-melting metals (gold, silver, aluminum) and organic molecules that a plasma would damage. At a typical 0.1 nm/s, 100 nm of gold takes about 17 minutes. Line of sight leaves sharp shadows, a weakness on rough surfaces and a strength for *shadow masks* and for *lift-off*, in which metal is evaporated over a patterned *photoresist* (a light-sensitive polymer exposed through a mask) and dissolving the resist removes the metal on top of it, leaving metal only in the openings. An *electron-beam evaporator* instead melts a small spot of the source with a beam of electrons, which reaches the temperatures of titanium, platinum, and oxides; it too deposits by line of sight.

Chemical vapor deposition and ALD

In CVD the film is the solid product of a reaction between gases at the substrate surface; silicon nitride, the standard passivation layer of chips, grows from silane and ammonia. Because the reaction happens wherever the gas touches the hot surface, CVD coats the tops, sidewalls, and bottoms of features alike. The films are often of high purity, since the precursors are distilled gases and the byproducts leave as gas, and rates can reach micrometers per hour. Furnace CVD runs at 600 to 800 °C, which no plastic survives. *Plasma-enhanced CVD* (PECVD) supplies part of the reaction energy with a plasma and grows silicon nitride and oxide at 200 to 350 °C, compatible with polyimide but not with PET. Parylene is a CVD process that works at room temperature: a vapor of reactive monomer polymerizes on every surface it touches and coats an implant with a pinhole-free film.

Atomic layer deposition (ALD) is CVD split into two half-reactions that each stop by themselves (Fig. 11.7). For aluminum oxide, a pulse of trimethylaluminum (TMA) reacts with the hydroxyl groups on the surface until every site is used, and the excess is purged away; a pulse of water then converts the surface back to hydroxyls and completes one layer of oxide, $2\,Al(CH_3)_3 + 3\,H_2O \longrightarrow Al_2O_3 + 6\,CH_4$. Each cycle adds about 0.1 nm.[1] Thickness is set by counting cycles, and because each half-reaction saturates on every exposed surface, the film follows deep trenches and porous electrodes. ALD runs at 80 to 250 °C, so it can coat plastics, and its pinhole-free oxides are among the best moisture barriers for implanted devices. Its weakness is speed: a 20 nm barrier at 0.11 nm per cycle needs 182 cycles, about 24 minutes at 8 s per cycle, where evaporating 20 nm of aluminum takes 20 seconds.

[1] S. M. George, *Chem. Rev.* (2010), a review of ALD chemistry.

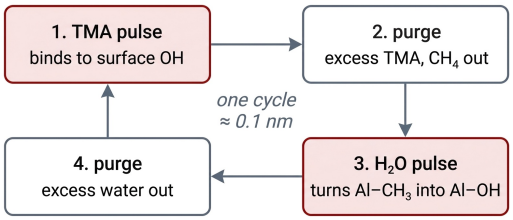


Figure 11.7: One ALD cycle for Al_2O_3. Each pulse reacts only until the surface is saturated, so the growth per cycle is fixed and the film is conformal. Repeating the cycle N times gives a film about $0.1N$ nanometers thick.

Table 11.3 compares the two families. Neither is cheap in general: PVD needs high-vacuum pumps but deposits almost any metal from one solid source near room temperature, while CVD needs a precursor chemistry for each material, toxic-gas handling, and often high temperature. The material, the coverage needed, and the substrate's temperature limit usually decide.

	CVD (and ALD)	PVD
Mechanism	reaction of precursor gases on the surface	atoms evaporated or sputtered from a solid
Coverage	conformal; ALD covers deep trenches	line of sight; sputtering partly conformal
Purity	high; some H, C, or Cl from precursors	set by source and vacuum
Speed	furnace CVD fast; ALD slow	0.1 to 1 nm/s typical
Materials	oxides, nitrides, Si, W, graphene, parylene	almost any metal; oxides by RF sputtering or e-beam
Temperature	furnace 600 to 800 °C; PECVD 200 to 350 °C; ALD 80 to 250 °C; parylene room temperature	near room temperature

Table 11.3: Chemical and physical vapor deposition compared.

Measuring thickness

You cannot see a 50 nm film grow, so the chamber weighs it. A *quartz crystal monitor* beside the substrate is a thin quartz disk that vibrates at its resonant frequency, usually 6 MHz. Film that lands on it adds mass and lowers the frequency in proportion, by about 85 Hz per nanometer of silver. Given the film's density, the controller converts the frequency drop into thickness and rate and adjusts the source power to hold, say, 0.1 nm/s. The crystal sits beside the substrate and receives a slightly different flux, so a *tooling factor*, found once from a measured film, scales one to the other.

After deposition, a *stylus profilometer* drags a fine diamond tip across a step at the film's edge and reads its height to a few nanometers. Transparent films such as oxides and parylene are measured with light reflected from their top and bottom surfaces (*reflectometry* and *ellipsometry*), to a fraction of a nanometer. Other tools look inside the film: X-ray diffraction reads the spacing of atomic planes, and so the strain and grain size, and nanoindentation presses a diamond tip into the film to measure its modulus.

11.5 Bending and the Neutral Plane

BEND A THICK BOOK and its pages slide past each other. Glue the pages together and bend again: the outer pages must stretch, the inner ones must compress, and somewhere in the middle one page keeps its length. A film bonded to a substrate behaves like the glued book (Fig. 11.8). The surface that keeps its length is the *neutral plane*, and a layer a distance y above it, in a stack bent to radius R, is stretched by

$$\varepsilon(y) = \frac{y}{R}. \tag{11.2}$$

The reason is geometry. A layer at radius $R + y$ is an arc of length $(R + y)\theta$, while the neutral plane has length $R\theta$; the difference divided

KEY EQUATION

$$\varepsilon = \frac{y}{R}, \quad \varepsilon_{\text{surface}} = \frac{t}{2R}$$

Bending strain at a distance y from the neutral plane, and at the surface of a sheet of thickness t. Halve the thickness and you halve the strain.

by $R\theta$ is y/R. Above the neutral plane, on the outside of the bend, the strain is tensile; below it, compressive.

A single sheet

For a uniform sheet of thickness t the neutral plane is in the middle, so the surface, $t/2$ away, is strained by

$$\varepsilon_{\text{surface}} = \frac{t}{2R}. \tag{11.3}$$

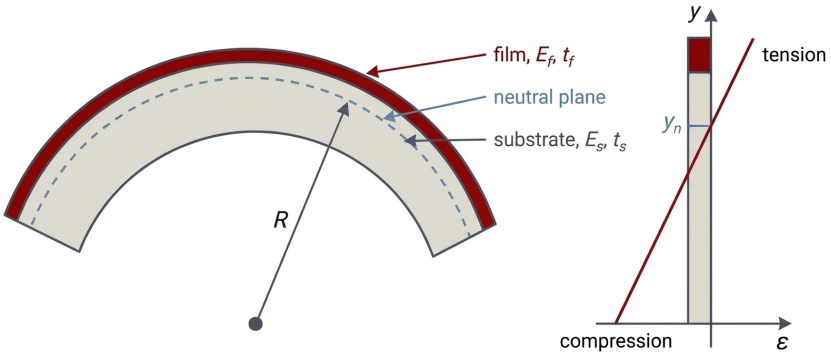


Figure 11.8: A film on a substrate bent to radius R (schematic, thicknesses exaggerated). The strain grows linearly with distance from the neutral plane (dashed), tensile above and compressive below. A stiff film pulls the neutral plane toward itself.

This is the result Chapter 1 used. Adding up the bending stresses $E\varepsilon = Ey/R$ across the thickness, each weighted by its distance from the neutral plane, gives the moment needed to bend the sheet and with it the bending stiffness of Chapter 1, $D = Et^3/[12(1-\nu^2)]$ per unit width; for a narrow strip the factor $1-\nu^2$ drops out, and D times the width is the EI of the probes of Chapter 9. Halving the thickness halves the surface strain and divides the stiffness by eight.

A film much thinner than its substrate sits at the substrate surface and sees the same strain, whatever its own thickness. A 125 µm PET sheet bent to $R = 5$ mm has a surface strain of $125/(2\times 5000) = 1.25\,\%$, and many oxide and metal films start to crack near 1 % (Section 11.6).

The same formula says how thin silicon must be to bend. Polished silicon fractures near 1 % strain, less with surface flaws, so take 0.5 % as a safe design value; then $R_{\text{min}} = t/(2\times 0.005) = 100\,t$. A 500 µm wafer bends only to 50 mm, the curve of a large bottle. Ground and polished to 20 µm, the same chip bends to 2 mm and wraps a fingertip (Fig. 11.9), and it is $(500/20)^3 \approx 16\,000$ times less stiff, so it hardly stresses the adhesive and the tissue under it. A 1 µm silicon membrane bends to 0.1 mm.

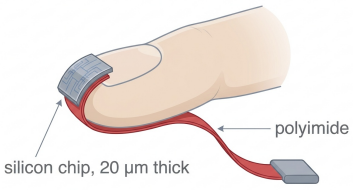


Figure 11.9: A silicon chip thinned to about 20 µm bends around a fingertip. At the full wafer thickness of 500 µm it would break at a radius of a few centimeters.

The neutral plane of a stack

When the film is not thin compared with the substrate, or is much stiffer, the neutral plane moves. Its position follows from force balance: in pure bending, the tensile forces above the neutral plane must cancel the compressive forces below it. With $\varepsilon = (y - y_n)/R$ in every layer, this gives

$$y_n = \frac{\sum_i E_i t_i \bar{y}_i}{\sum_i E_i t_i}, \tag{11.4}$$

where layer i has modulus E_i, thickness t_i, and mid-height $\bar{y}_i$, all measured from the bottom of the stack. The neutral plane is the modulus-weighted center of the stack, as a center of mass is the mass-weighted average position: a stiff layer pulls it toward itself.

KEY EQUATION

$y_n = \dfrac{\sum E_i t_i \bar{y}_i}{\sum E_i t_i}$

The neutral plane is the modulus-weighted average height of the layers.

WORKED EXAMPLE 11.1 *Thin and thick silver on plastic*

On a 250 µm PET substrate you deposit 10 nm of silver; on another, 10 µm. Both are bent to $R = 5$ mm with the silver outside. Which film sees the larger strain? Use $E_{\text{Ag}} = 80$ GPa and $E_{\text{PET}} = 3$ GPa.

SOLUTION. *First estimate, neutral plane in the substrate middle.* The film surface is $(t_s + t_f)/2$ from it, so the strain is $(t_s + t_f)/2R$: $250.01/10\,000 = 2.50\,\%$ for the 10 nm film and $260/10\,000 = 2.60\,\%$ for the 10 µm film. The thick film sits slightly farther out, so it would see slightly more strain.

With the moduli. The estimate assumes the film does not move the neutral plane. Check with Eq. (11.4), measuring from the bottom of the PET (thicknesses in micrometers, moduli in gigapascals). For the 10 µm film,

$$\begin{aligned} y_n &= \frac{(3)(250)(125) + (80)(10)(255)}{(3)(250) + (80)(10)} \\ &= \frac{93\,750 + 204\,000}{1550} = 192.1\ \mu\text{m}. \end{aligned}$$

The stiff silver carries more than half of the stack's Et (800 against 750) and pulls the neutral plane 67 µm above the PET middle. The top of the film is $260 - 192.1 = 67.9$ µm above it, so

$$\varepsilon_{\text{top}} = \frac{67.9\ \mu\text{m}}{5000\ \mu\text{m}} = 1.36\,\%.$$

For the 10 nm film, $E_f t_f = 0.8$ against 750, the neutral plane moves only 0.13 µm, and the film sees 2.50 %.

Answer. The thin film sees the larger strain, 2.50 % against 1.36 % (Fig. 11.10). The geometric estimate, which says the opposite, holds only when the film is thin or compliant enough not to move the

neutral plane. Both strains exceed the roughly 1 % at which evaporated metal films begin to crack, so both samples would degrade under repeated bending to 5 mm. ■

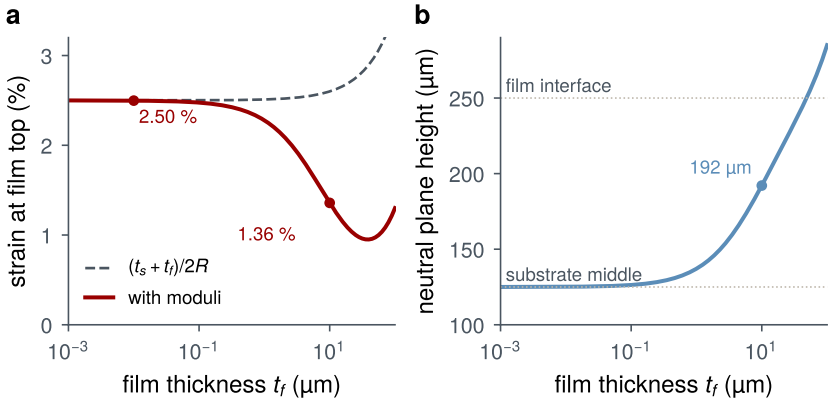


Figure 11.10: Silver on 250 µm PET bent to 5 mm. (a) Strain at the top of the film. The geometric estimate $(t_s + t_f)/2R$ (dashed) rises with film thickness, but once the silver's $E_f t_f$ approaches that of the PET, the neutral plane moves up (b) and the film strain falls. Beyond about 30 µm the silver dominates and the strain rises again, now as half the silver thickness over R.

Neutral-plane engineering

If the neutral plane can be moved, it can be put where the fragile layer is. The standard trick is to cover the device with an *encapsulation* layer whose stiffness balances that of the substrate (Fig. 11.11). A symmetric stack, the same material and thickness above and below, puts the neutral plane exactly at the device, which then sees almost no strain at any bending radius. With different materials, Eq. (11.4) for a thin device layer at height t_s gives the condition $E_s t_s^2 = E_e t_e^2$, or

$$t_e = t_s \sqrt{E_s / E_e}. \tag{11.5}$$

Take a 50 nm gold line on 25 µm polyimide (2.5 GPa) bent to $R = 2$ mm. It sits at the polyimide surface and sees about $t_s/2R = 0.6\,\%$, more than a thin gold line survives for millions of bending cycles. A layer of parylene C (2.8 GPa) of $25\,\mu\text{m} \times \sqrt{2.5/2.8} = 23.6\,\mu\text{m}$ on top puts the neutral plane in the gold, which then sees only the strain across its own half-thickness, $0.025\,\mu\text{m}/2\,\text{mm} \approx 1 \times 10^{-5}$. The parylene does double duty: it also keeps sweat and body fluid away from the metal (Section 11.11).

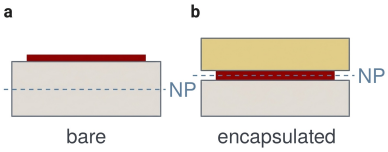


Figure 11.11: Neutral-plane (NP) engineering. (a) A thin device layer (red) on a substrate sits $t_s/2$ above the neutral plane and sees $t_s/2R$. (b) An encapsulation (gold) with $E_e t_e^2 = E_s t_s^2$ moves the neutral plane into the device layer, which then sees almost no strain.

11.6 How Thin Films Fail

A BENT OR STRETCHED FILM FAILS in a few recognizable ways (Fig. 11.12). In tension, a brittle film such as ITO, silicon nitride, or

a sintered nanoparticle film first develops *microcracks* at flaws. As the strain increases, cracks run across the whole film perpendicular to the load; these *channel cracks* cut the film into strips, and current must detour or stop. In compression the film can *buckle*: it lifts off the substrate in a blister or a wavy ridge. Cracks can also turn at the interface and run under the film, peeling it off, which is *delamination*.

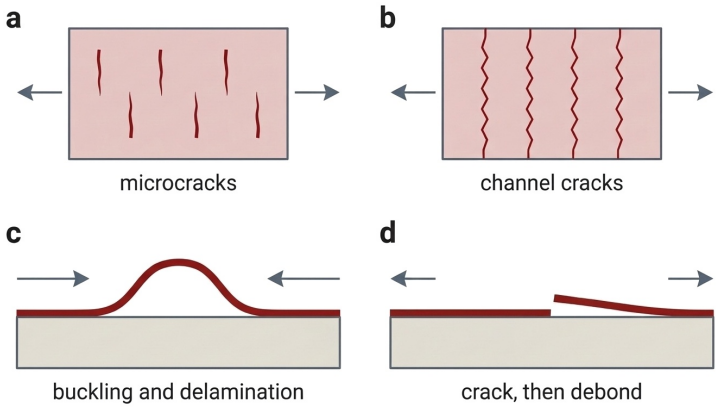


Figure 11.12: Failure modes of thin films under load (arrows). (a) and (b) are top views of a film in tension: microcracks start at flaws and grow into channel cracks across the film. (c) and (d) are cross sections: a film in compression buckles off the substrate; a crack in tension turns along a weak interface and peels the film.

Brittle films (oxides, nitrides, ITO, amorphous silicon) crack with little plastic deformation, and their resistance jumps when channel cracks form. The most useful single number for a film on a substrate is its *crack onset strain*, the strain at which the resistance starts to rise or cracks appear under the microscope. For ITO on PET it is about 1 %, so a 100 nm ITO electrode on 125 µm PET survives bending only to $R = t_s/(2 \times 0.01) = 6.25\,\text{mm}$, and on 25 µm polyimide to 1.25 mm; many evaporated metal films crack at strains of the same order. Repeated bending below the crack onset strain still causes *fatigue*: the resistance creeps up over thousands of cycles as small cracks grow, faster for larger strains. *Ductile* films (gold, copper, aluminum) neck and form voids before they break, and a ductile metal film well bonded to a polymer can be stretched far beyond its free-standing failure strain, because the substrate stops any one neck from growing.[2]

[2] N. Lu, X. Wang, Z. Suo, J. Vlassak, *Appl. Phys. Lett.* (2007): copper on polyimide stretched beyond 50 %.

Residual stress

A film can be stressed with nothing pushing on it. *Residual stress* is built in during fabrication and stays after the device cools. *Intrinsic* stress comes from growth: energetic sputtered atoms hammer extra atoms into the film and leave it in compression, porous evaporated films that densify end up in tension, and polymer films such as the epoxy SU-8 shrink as they cure. *Thermal* stress comes from cooling a film and a substrate that expand at different rates. A film deposited stress-free at a temperature ΔT above room temperature is forced to

shrink with its thick substrate by $\alpha_s \Delta T$ on cooling, while on its own it would shrink by $\alpha_f \Delta T$. The difference is a *mismatch strain*

$$\varepsilon_{\text{mismatch}} = (\alpha_f - \alpha_s)\,\Delta T, \tag{11.6}$$

the same in both in-plane directions, and tensile when the film has the larger CTE. For this *biaxial* strain the stress is

$$\sigma = \frac{E_f}{1 - \nu_f}\,\varepsilon_{\text{mismatch}}, \tag{11.7}$$

where ν is the *Poisson ratio*, the sideways shrinkage per unit of stretch (about 0.3 for metals, 0.5 for rubber).

KEY EQUATION

$$\sigma = \frac{E_f}{1 - \nu_f}(\alpha_f - \alpha_s)\Delta T$$

Thermal mismatch stress of a film cooled by ΔT on a thick substrate.

Copper ($\alpha = 16.5\,\text{ppm/K}$, $E = 120\,\text{GPa}$, $\nu = 0.34$) sputtered onto silicon (2.6 ppm/K) at 250 °C and cooled to 25 °C carries a tensile strain of $(16.5 - 2.6) \times 10^{-6} \times 225 = 3.1 \times 10^{-3}$, or about 570 MPa, more than most copper films can hold without yielding. Silver (18.9 ppm/K) sintered at 150 °C on polyimide (20 ppm/K) is nearly matched and carries only about −18 MPa, a small compression. On a stiff wafer, residual stress bows the wafer, and measuring the curvature with a laser or a profilometer gives the stress (Stoney's equation).[3] On thin plastic the same stress curls the whole sheet.

[3] G. G. Stoney, *Proc. R. Soc. Lond. A* (1909).

Cracking and peeling

A crack grows when growing pays: extending it releases elastic energy stored in the film and costs the energy of the two new surfaces. The *fracture energy* G_c is the energy needed per unit area of new crack, in J/m^2. When the crack runs through a layer the failure is *cohesive*, and G_c belongs to that material; when it runs along the boundary between two layers the failure is *adhesive*, and G_c belongs to the interface. The chemical bonds across a clean interface give only a fraction of a joule per square meter; plastic flow near the crack tip and rough interfaces raise it by orders of magnitude, and a thin titanium or chromium layer that bonds chemically to an oxide is what makes gold stick. Weak organic layers fail at 0.1 to $1.6\,\text{J/m}^2$, which is why flexible organic devices delaminate; well-bonded metal-oxide interfaces reach $10\,\text{J/m}^2$ and more, as measured by bending or prying apart a stack bonded between two beams.

A film under residual tension σ stores elastic energy in proportion to its thickness h, and a channel crack across it releases[4]

$$G = Z\,\frac{\sigma^2 h}{\bar{E}_f} \tag{11.8}$$

per unit of crack area, where $\bar{E}_f = E_f/(1 - \nu_f^2)$ and Z is a number near 2 ($Z = 1.98$ when film and substrate have the same modulus,

[4] J. L. Beuth, *Int. J. Solids Struct.* (1992); J. W. Hutchinson and Z. Suo, *Adv. Appl. Mech.* (1991).

larger on soft substrates). The film cracks by itself when G reaches G_c, so above a critical thickness $h_c = G_c \bar{E}_f / (Z\sigma^2)$. A silicon nitride barrier with 300 MPa of tension, $\bar{E}_f = 213\,\text{GPa}$, and $G_c = 5\,\text{J/m}^2$ cracks spontaneously above about $h_c = 6\,\mu\text{m}$. Thin films survive stresses that would shatter thick ones.

11.7 *Stretchable Conductors*

Bending strains a film by a percent or so. Skin stretches by 10 to 30 % at the joints, and a beating heart or a moving muscle by similar amounts. Devices that follow them need conductors that stretch. Start with what stretching does to a resistor.

A conductor of resistivity ρ, length ℓ, and cross-section A has the resistance $R = \rho\ell/A$. Stretch it to $\ell = \ell_0(1+\varepsilon)$. A *liquid metal* such as eutectic gallium-indium (EGaIn),[5] sealed in an elastomer channel, keeps its volume ($A\ell = A_0\ell_0$) and its resistivity, so

$$\frac{R}{R_0} = \frac{\ell}{\ell_0}\frac{A_0}{A} = \left(\frac{\ell}{\ell_0}\right)^2 = (1+\varepsilon)^2. \tag{11.9}$$

A solid shrinks sideways by its Poisson ratio ν: each transverse dimension scales as about $(1+\varepsilon)^{-\nu}$, so $A = A_0(1+\varepsilon)^{-2\nu}$ and

$$\frac{R}{R_0} = (1+\varepsilon)^{1+2\nu} \approx 1 + (1+2\nu)\,\varepsilon. \tag{11.10}$$

For an incompressible material ($\nu = 0.5$) the two agree. The small-strain slope $1+2\nu$, about 1.6 for metals, is the geometric part of the *gauge factor* $(\Delta R/R_0)/\varepsilon$ of the strain gauges of Chapter 5; metal foil gauges reach about 2. Liquid metal does not crack, so it follows Eq. (11.9) to strains of 100 % and more (Fig. 11.13); a solid metal wire yields or breaks after a few percent.

[5] M. D. Dickey *et al.*, *Adv. Funct. Mater.* (2008).

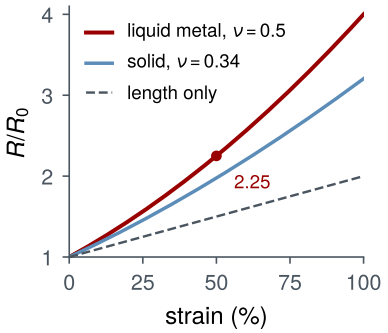


Figure 11.13: Resistance of a stretched conductor from geometry alone (Eqs. 11.9 and 11.10). A solid metal breaks long before the right side of the plot.

Worked Example 11.2 *A liquid-metal interconnect*

An EGaIn line ($\rho = 29.4 \times 10^{-8}\,\Omega\,\text{m}$) fills a 5 cm channel with a 0.5 mm × 0.5 mm cross-section in an elastomer. Find its resistance relaxed and at 50 % and 100 % strain, and compare the change at 1 % strain with that of a copper wire ($\nu = 0.34$).

Solution. $R_0 = \rho\ell/A = (29.4 \times 10^{-8}\,\Omega\,\text{m})(0.05\,\text{m})/(2.5 \times 10^{-7}\,\text{m}^2) = 58.8\,\text{m}\Omega$. At 50 %, $R = 2.25R_0 = 132\,\text{m}\Omega$; at 100 %, $4R_0 = 235\,\text{m}\Omega$. Even stretched to twice its length the line is a fraction of an ohm, fine for carrying power or a biopotential to an amplifier. At 1 %, the liquid metal changes by $(1.01)^2 - 1 = 2.01\,\%$ and copper by $(1.01)^{1.68} - 1 = 1.69\,\%$, similar. The difference is that copper cannot survive much more strain. EGaIn is about

17 times more resistive than copper ($1.7 \times 10^{-8}\,\Omega\,\mathrm{m}$), and a thin skin of gallium oxide on its surface lets it hold its shape in open channels. ■

Geometry instead of material

Solid metals and silicon cannot stretch much, but a structure made of them can, if it is shaped so that it bends or unfolds instead of stretching (Fig. 11.15). Four strategies cover most designs.

Serpentines. A line drawn as a meander of arcs is longer along its contour than from end to end. Pulling on it first rotates and bends the arcs, which strains the metal only a little, and the line straightens before the metal itself is stretched. For semicircular arcs of radius r (Fig. 11.14), each period of length $4r$ along the axis contains a contour length of $2\pi r$, so the line straightens at

$$\varepsilon_{\max} = \frac{2\pi r}{4r} - 1 = \frac{\pi}{2} - 1 = 57\,\%. \tag{11.11}$$

Straight legs between the arcs, or horseshoes of more than half a circle, raise the limit above 100 %. Serpentine copper or gold traces are the standard interconnects of skin-mounted electronics; a team in a past semester printed silver electrodes with serpentine interconnects on thermoplastic polyurethane for an around-the-ear EEG headset.

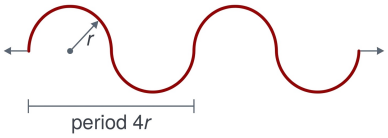


Figure 11.14: A serpentine of semicircles. Its contour length is $\pi/2$ times its end-to-end length, so it can stretch 57 % before the metal itself is pulled straight.

Wrinkles. Bond a thin silicon ribbon to a pre-stretched elastomer and release it: the ribbon buckles into waves without detaching, with a wavelength of about 16 µm for 100 nm of silicon on PDMS. The waves flatten when the elastomer is stretched again, up to the pre-strain, while the silicon itself sees only the small bending strain of the waves. Khang and coworkers built working transistors this way from "wavy" single-crystal silicon.[6]

Island-bridge. A rigid chip smaller than about a millimeter on a soft substrate barely feels the substrate stretch. Put the chips on stiff islands and connect them with serpentine or buckled bridges, and the bridges take nearly all the strain.[7] Stretchable circuit boards use the same idea with packaged components on polyimide islands in polyurethane. The edge of a rigid part is where such boards fail.[8]

Kirigami. A pattern of slits lets a stiff sheet open into a mesh whose ligaments rotate and bend out of plane, so a polyimide sheet that would yield at a few percent strain stretches by tens of percent.

So thinning makes silicon flexible (Section 11.5), and geometry makes it stretchable: with wrinkles or islands and bridges, the circuit stretches by tens of percent while the silicon never sees more than a fraction of a percent. Organic and nanomaterial conductors (Sections 11.8 and 11.9) stretch more on their own, but none matches the mobility of crystalline silicon.

[6] D.-Y. Khang, H. Jiang, Y. Huang, J. A. Rogers, *Science* (2006).

[7] D.-H. Kim *et al.*, *Science* (2008).

[8] In a flexible ECG monitor, copper traces cracked at the edge of the solder joints of the front-end chip, where the stiffness changes abruptly; in bending tests, 50 µm polyimide with 6 µm copper, a low-temperature tin-bismuth solder, and pads with rounded corners held up best. V. Soman, Y. Khan, M. Zabran, *et al.*, *IEEE Trans. Compon. Packag. Manuf. Technol.* (2019).

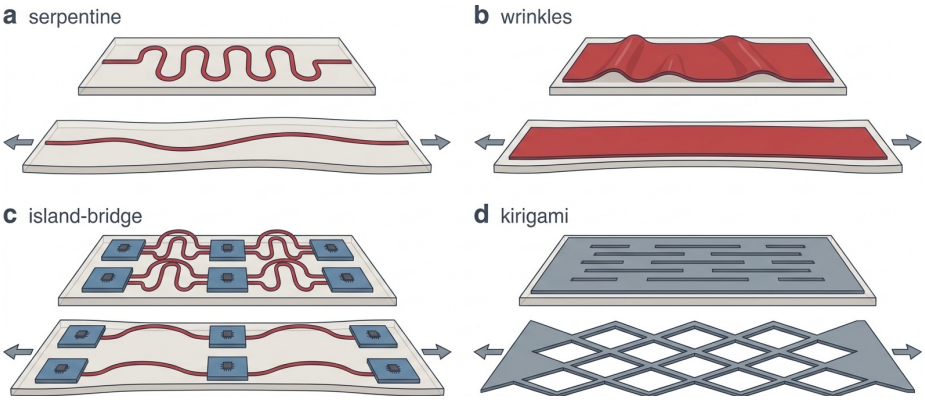


Figure 11.15: Geometric strategies for stretchable electronics. (a) Serpentine interconnects straighten as they stretch. (b) A film bonded to a pre-stretched elastomer wrinkles when the elastomer relaxes, and the wrinkles flatten again on stretching. (c) Island-bridge: rigid islands carry the chips, and stretchable bridges between them take the strain. (d) Kirigami: a pattern of cuts lets a stiff sheet open into a stretchable mesh.

11.8 Conductive Inks and Sintering

A PRINTED CONDUCTOR STARTS AS AN INK: metal particles dispersed in a solvent. The obvious choice, metal powder, does not work, because micrometer grains settle, clog nozzles, and fuse only near the melting point (962 °C for silver). Nanoparticles, tens of nanometers across, stay suspended, pass through inkjet nozzles, and, most important, sinter into a conductor at 100 to 250 °C, which plastic substrates survive.

Why small particles sinter

The reason is surface. A sphere of diameter d with an outer shell one atomic layer thick ($a \approx 0.25\,\text{nm}$) has a fraction

$$f_{\text{surface}} = 1 - \left(1 - \frac{2a}{d}\right)^3 \approx \frac{6a}{d} \tag{11.12}$$

of its atoms at the surface: $1 - 0.9^3 = 27\,\%$ for a 5 nm particle and 3 % at 50 nm. Surface atoms have fewer neighbors, higher energy, and move easily. Very small particles even melt early: gold particles 5 nm across melt near 1150 K, against 1337 K in bulk.[9] But the 10 to 50 nm particles of silver inks melt within about a hundred kelvin of the bulk value, so their sintering at 150 °C is not melting. It is *surface diffusion*: atoms hop over the particle surfaces far below the melting point, and the sharp curvature where two small particles touch draws them into a growing neck.

Silver nanoparticles are made by reducing a silver salt to metal in a solution that also holds a polymer such as polyvinylpyrrolidone (PVP), which coats each particle as it forms. Such *ligands* keep the particles of an ink from sticking together. The ligand is a compromise: it must hold on during storage and printing but let go during sintering, because a molecular shell between particles blocks both conduction and necking.

[9] P. Buffat and J.-P. Borel, *Phys. Rev. A* (1976).

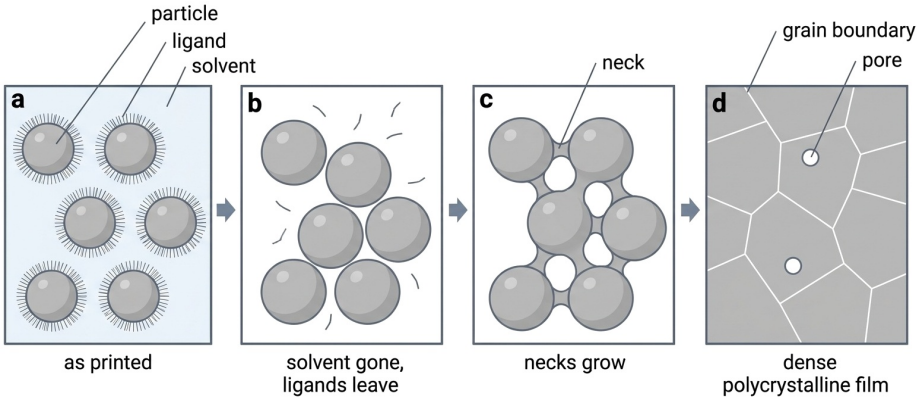


Figure 11.16: Sintering a nanoparticle film. (a) As printed: particles capped with ligands in solvent. (b) Drying and heating drive off the solvent and then the ligands, and the bare particles touch. (c) Atoms diffuse into the contacts and necks grow; the resistivity drops by orders of magnitude here. (d) The film densifies into grains with a few pores, typically two to five times the resistivity of bulk metal.

From particles to metal

A freshly printed nanoparticle line does not conduct. It is a packed bed of particles, each wrapped in ligand and wet with solvent. Heating turns it into metal in three overlapping steps (Fig. 11.16): the solvent evaporates; the ligands desorb or decompose and leave the film; and the bare particles touch and *sinter*, growing necks that weld them into a continuous network. As the necks grow, the resistivity falls by several orders of magnitude and the film densifies. The ligands, not the metal, often set the temperature: silver nanoparticle films conduct after 30 to 60 minutes at 120 to 200 °C.

Thermal sintering in an oven or on a hot plate is simple but slow, and most of the energy heats the substrate, which limits PET to about 150 °C. *Photonic sintering* uses a xenon flash lamp to deliver millisecond pulses of light (Fig. 11.17). The dark particle film absorbs visible light strongly, while clear PET absorbs mainly in the ultraviolet, so the film heats far more than the substrate. Timing does the rest. Heat spreads by diffusion, like the molecules of Chapter 2: in a time t it reaches a depth

$$L_{\text{th}} = \sqrt{D_{\text{th}}\, t}, \tag{11.13}$$

where $D_{\text{th}} = k/(\rho c)$, the *thermal diffusivity*, is the thermal conductivity k divided by the heat capacity per volume (density ρ times specific heat c). For PET, $D_{\text{th}} \approx 1 \times 10^{-7}\,\text{m}^2/\text{s}$, so during a 1 ms pulse the heat reaches only $\sqrt{10^{-7} \times 10^{-3}} = 10\,\mu\text{m}$ into a substrate that is 125 µm thick, while in an hour-long oven bake it reaches 2 cm, through everything. Heating a 1 µm silver film by 300 K takes only about $0.07\,\text{J}/\text{cm}^2$, less than the $0.23\,\text{J}/\text{cm}^2$ that warms the top 10 µm of PET by 150 K, so a pulse of a fraction of a joule per square centimeter absorbed in the film sinters it in milliseconds, on a moving web.

Two other methods avoid heat. In *chemical sintering*, silver particles capped with PVP sinter at room temperature when dipped in dilute

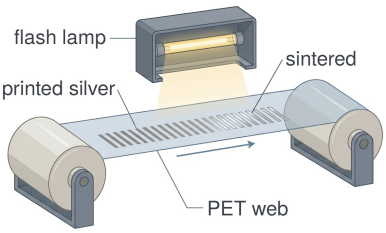


Figure 11.17: Photonic sintering of printed silver lines on a moving PET web.

chloride: Cl^- binds to silver more strongly than PVP, displaces it, and the bare particles coalesce. *Plasma sintering* strips the organic shell with an argon or oxygen plasma. Lasers, or a current passed through the line itself, heat only the printed pattern.

Sheet resistance

Printed films are characterized by their *sheet resistance*

$$R_s = \frac{\rho}{t} \quad (\Omega \text{ per square}), \qquad R = R_s \frac{\ell}{w}, \tag{11.14}$$

where t is the thickness. A square of any size has the same resistance, so a line of length ℓ and width w counts ℓ/w squares. Because thickness is hard to measure on rough printed lines, R_s (measured with a *four-point probe*, which keeps contact resistance out of the reading) is the usual figure of merit, and ρ is quoted as a multiple of the bulk value, 1.59 μΩ cm for silver. Sintered silver nanoparticle films reach 2 to 5 times bulk.

KEY EQUATION

$R = R_s \ell/w, \quad R_s = \rho/t$

A line ℓ/w squares long has resistance R_s per square.

Worked Example 11.3 *Resistance of a printed trace*

A silver trace 100 μm wide, 1 μm thick, and 5 cm long is printed from nanoparticle ink and sintered to three times the bulk resistivity. (a) Find its sheet resistance and resistance. (b) What would bulk silver give? (c) A screen-printed flake paste at 25 times bulk reaches 25 μm thickness. Compare.

Solution. (a) $\rho = 3 \times 1.59\,\mu\Omega\,\text{cm} = 4.77 \times 10^{-6}\,\Omega\,\text{cm}$, so $R_s = \rho/t = 4.77 \times 10^{-6}\,\Omega\,\text{cm}/1 \times 10^{-4}\,\text{cm} = 47.7\,\text{m}\Omega$ per square. The line is $5\,\text{cm}/100\,\mu\text{m} = 500$ squares long, so $R = 500 \times 47.7\,\text{m}\Omega = 23.9\,\Omega$.

(b) Bulk silver of the same size gives a third of that, 7.95 Ω.

(c) $R_s = 25 \times 1.59 \times 10^{-6}\,\Omega\,\text{cm}/25 \times 10^{-4}\,\text{cm} = 15.9\,\text{m}\Omega$ per square, and $R = 7.95\,\Omega$. The flake paste is the more resistive material but makes the lower-resistance line, because it is 25 times thicker. For a biopotential electrode lead either is fine; for a heater or an antenna, the thick paste wins. Only the nanoparticle ink can make the line 1 μm thick, though: flakes are as large as the whole film. ■

Nanowires, flakes, and other inks

Silver nanowires are grown in hot ethylene glycol with PVP, which binds to the side facets of the growing crystal so that silver adds only at the ends. The wires reach lengths of tens of micrometers at diameters of tens of nanometers. A film of them is a network, not a continuous

sheet: current flows only if a chain of touching wires connects the electrodes, which happens above a critical density. For randomly placed sticks of length L in a plane, *percolation theory* gives the threshold

$$N_c = \frac{5.64}{L^2} \tag{11.15}$$

wires per unit area, and above it the conductance rises steeply with density (Fig. 11.18). For wires 10 µm long, $N_c = 0.0564\,\mu m^{-2}$, one wire per $18\,\mu m^2$. A network at five times that density, 30 nm wires, shades only about 8.5 % of the surface and uses about 21 mg of silver per square meter, the metal of a film 2 nm thick, which as a continuous film would not even conduct. Such films reach tens of ohms per square at 90 % optical transmittance, as good as ITO, and the junctions between wires dominate their resistance; pressing, heating, or flashing the network welds them.

KEY EQUATION

$N_c = 5.64/L^2$

Percolation threshold of random sticks of length L; longer wires connect at lower density.

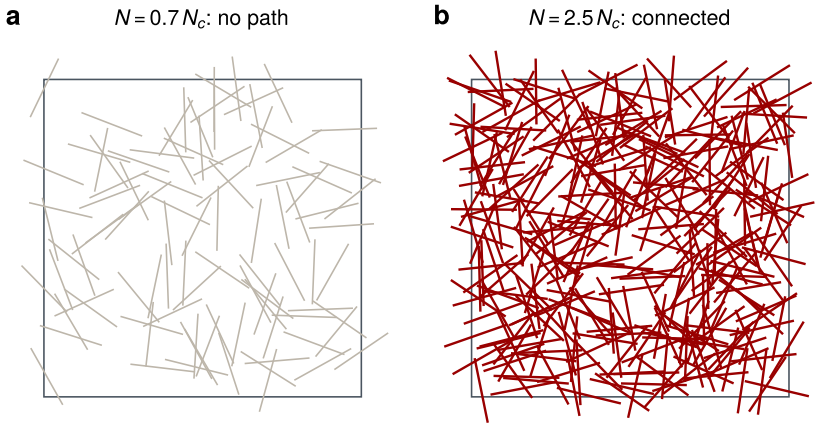


Figure 11.18: Random nanowire networks in a square five wire lengths wide, computed for this figure. (a) Below the threshold, clusters are isolated. (b) At 2.5 times the threshold, the cluster that spans the square (red) includes almost every wire.

Networks also survive strain (Fig. 11.19). Stretched by 10 % on PDMS, a sintered nanoparticle film, a porous and brittle metal film, develops channel cracks, and its resistance rises by orders of magnitude or the line opens. A nanowire network well above threshold has no continuous film to crack: wires many crack spacings long slide and rotate at their junctions instead of breaking, and the resistance rises by tens of percent, close to the geometric $(1+\varepsilon)^2 = 1.21$.

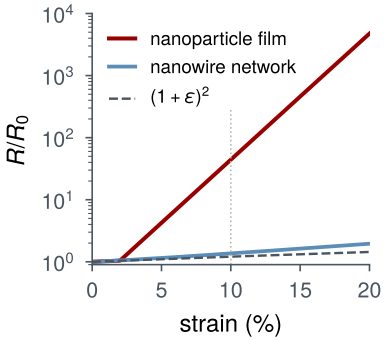


Figure 11.19: Resistance under strain for films on an elastomer (illustrative curves). A sintered nanoparticle film cracks after a few percent; a nanowire network stays connected.

Flake pastes use silver flakes a few micrometers across in a polymer binder, for screen printing. The flakes do not sinter at the cure temperature (80 to 150 °C); they conduct through the contacts of a percolating network held together by the binder, at 10 to 30 times bulk silver, in rough films 5 to 25 µm thick that cannot make fine features. *Copper* conducts almost as well as silver at about a hundredth of the price, but its particles grow an oxide shell in air that blocks sintering, so copper inks are sintered in nitrogen or a reducing gas, or by a flash

fast enough that the metal has no time to oxidize. *Particle-free* inks carry silver as a dissolved compound that decomposes to metal near 90 °C,[10] giving smooth films. *Carbon* inks conduct far worse than silver, but they are cheap and chemically inert, and they form the working electrodes of most screen-printed electrochemical sensors (Chapter 6). Chapter 12 takes up how each ink is formulated for its printer.

[10] S. B. Walker and J. A. Lewis, *J. Am. Chem. Soc.* (2012).

11.9 Organic Semiconductors and Conductors

Organic materials are built on carbon, bonded to itself and to hydrogen, oxygen, nitrogen, and sulfur. They come as *polymers*, long chains of repeating units, and as *small molecules* of a few dozen atoms that can pack into crystals. Most organic materials are insulators: the polyethylene of a bag, and the polyimide and parylene of Section 11.3. A few are semiconductors or conductors, and what makes them so is *conjugation*.

Conjugation, HOMO, and LUMO

In a conjugated molecule, carbon atoms are joined by alternating single and double bonds. Each carbon uses three of its electrons in bonds within the plane of the molecule and has one more in a p_z orbital sticking out of it. The p_z orbitals of neighbors overlap side by side into π orbitals that extend over the whole conjugated segment, so a π electron is not tied to one bond. With N conjugated carbons, N π orbitals form, and the lower half are filled. The highest filled one is the *HOMO* (highest occupied molecular orbital) and the lowest empty one the *LUMO*; they play the roles of the edges of the filled and empty bands, and their difference is the gap (Fig. 11.20). Longer conjugation crowds the levels together and narrows the gap, but real polymers keep a gap of about 1.5 to 3 eV, right in the semiconductor range, because their single and double bonds have different lengths.

Doping and PEDOT:PSS

An organic semiconductor becomes a conductor when it is *doped*: oxidized (electrons removed from the HOMO, *p-doping*) or reduced (electrons added to the LUMO, *n-doping*). Unlike silicon, where a dopant atom replaces a host atom, the dopant here is a counter-ion that sits beside the chain and balances the charge left on it, and doping levels reach one charge per few monomers.

The most important organic conductor is *PEDOT:PSS*, poly(3,4-ethylenedioxythiophene) mixed with polystyrene sulfonate. PEDOT is the conjugated polymer, held in its oxidized (p-doped) state; it

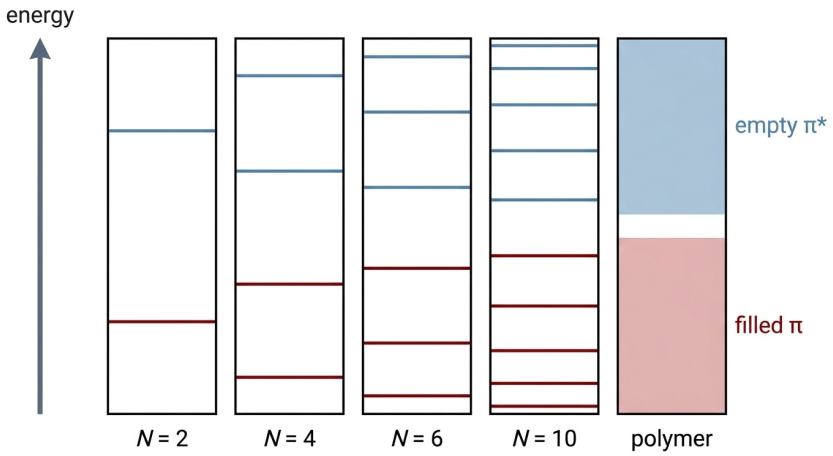


Figure 11.20: π orbital energies of conjugated chains of N carbons in the simplest (Hückel) model. Longer conjugation crowds the levels into bands and narrows the gap between the HOMO (top red line) and the LUMO (bottom blue line). Real polymers keep a gap of 1.5 to 3 eV because the single and double bonds have different lengths.

carries the holes. PSS is a polyanion: its sulfonate groups provide the negative counter-charge for the holes on PEDOT, and its excess acid groups make the complex dispersible in water. PSS does not conduct electronically; it balances charge and provides solubility. As cast from water, PEDOT:PSS conducts about 0.1 to 1 S/cm, because conducting PEDOT-rich grains are wrapped in insulating PSS. Adding a few percent of a polar solvent such as dimethyl sulfoxide or ethylene glycol, called *secondary doping*, adds no charge but lets the PEDOT chains straighten and the excess PSS separate, raising the conductivity to about 1000 S/cm. PEDOT:PSS is transparent in thin layers, flexible, and conducts both holes and ions, which makes it the standard electrode coating for recording from tissue and the channel of electrochemical transistors (Section 11.10).

Semiconducting polymers and small molecules

Poly(3-hexylthiophene) (P3HT) is the classic semiconducting polymer: a polythiophene backbone with hexyl side chains that make it soluble in organic solvents. Its chains stack face to face about 0.38 nm apart, and holes move along chains and hop between stacks, with thin-film mobilities of about 0.01 to 0.1 $\mathrm{cm^2\,V^{-1}\,s^{-1}}$; newer polymers reach 1 $\mathrm{cm^2\,V^{-1}\,s^{-1}}$ and above. *Small molecules* such as pentacene (five fused benzene rings) pack into crystals and reach tens of $\mathrm{cm^2\,V^{-1}\,s^{-1}}$ in carefully grown single crystals; added side groups make soluble versions such as TIPS-pentacene.

Most organic semiconductors conduct holes (*p-type*). Electron-conducting (*n-type*) ones are rarer, because injecting electrons needs contacts with a low *work function* (the energy needed to pull an electron out of a metal), such as calcium, aluminum, or magnesium, which

oxidize easily, and because an electron in the LUMO is easily captured by oxygen and water unless the LUMO lies deep. *Metal oxides* such as indium gallium zinc oxide (IGZO) fill the gap: they are n-type, transparent, and reach about $10\,\mathrm{cm^2\,V^{-1}\,s^{-1}}$, but they need sputtering or hot solution processing.

Why organic mobility is low

In crystalline silicon an electron moves as a wave through a periodic lattice and travels about 47 nm, some 90 lattice spacings, between collisions. In an organic film the molecules are held together only by weak van der Waals forces, the overlap between neighbors is small, and the energy of each site varies with the local disorder. A carrier sits on one molecule or chain segment and *hops* to the next with the help of heat. It is a random walk, so its diffusion coefficient (Chapter 2) is $D \approx \Gamma a^2$ for a hop distance a and a hop rate Γ, and the Einstein relation $\mu = qD/k_BT$ turns diffusion into mobility:

$$\mu \approx \frac{qa^2}{k_BT}\,\Gamma, \qquad \Gamma = \nu_0\, e^{-E_a/k_BT}, \tag{11.16}$$

where k_B is Boltzmann's constant, ν_0 an attempt frequency, and E_a an activation energy. A hop of 0.4 nm at $\Gamma = 10^{12}\,\mathrm{s^{-1}}$ gives $\mu = 0.062\,\mathrm{cm^2\,V^{-1}\,s^{-1}}$ at 300 K, typical of a polymer film; reaching $1\,\mathrm{cm^2\,V^{-1}\,s^{-1}}$ needs a hop every 62 fs, about as fast as molecules vibrate, which is why hopping cannot match the $1400\,\mathrm{cm^2\,V^{-1}\,s^{-1}}$ of electrons in silicon. Because the hop is activated, organic mobility rises with temperature, while in silicon more lattice vibration means more scattering and the mobility falls.

KEY EQUATION

$\mu \approx \frac{qa^2}{k_BT}\,\Gamma$

Hopping mobility. It rises with temperature, the opposite of band transport in silicon.

Grain boundaries lower the mobility further in polycrystalline films: between ordered grains, carriers must cross a disordered region that acts as a barrier and a trap. A material with $1\,\mathrm{cm^2\,V^{-1}\,s^{-1}}$ inside its grains can drop to about $0.05\,\mathrm{cm^2\,V^{-1}\,s^{-1}}$ with 100 nm grains, so crystallization methods (slow drying, solvent annealing, edge casting) aim to grow large, well-oriented grains.

11.10 Organic Transistors: The OTFT and the OECT

The organic thin-film transistor

An organic thin-film transistor (OTFT) is a field-effect transistor whose channel is a thin organic semiconductor film. It works like the MOSFET of Chapter 5: a gate voltage across a thin dielectric of capacitance C_i per area (the C_{ox} of the MOSFET) induces a sheet of carriers at the semiconductor-dielectric interface, and a drain voltage

drives them from the source to the drain. One difference: a silicon MOSFET works by *inversion*, drawing carriers of the opposite type to the doped substrate to its surface, while an OTFT works by *accumulation*: a p-type OTFT turns on when the gate is made negative enough to pile up holes in the undoped organic film. The common geometries differ in the order of the layers (Fig. 11.21). Contacts on top of the semiconductor (*top contact*) usually inject better; contacts underneath (*bottom contact*) can be patterned before the fragile organic layer goes down. A *top gate* buries the channel under the dielectric, which also encapsulates it.

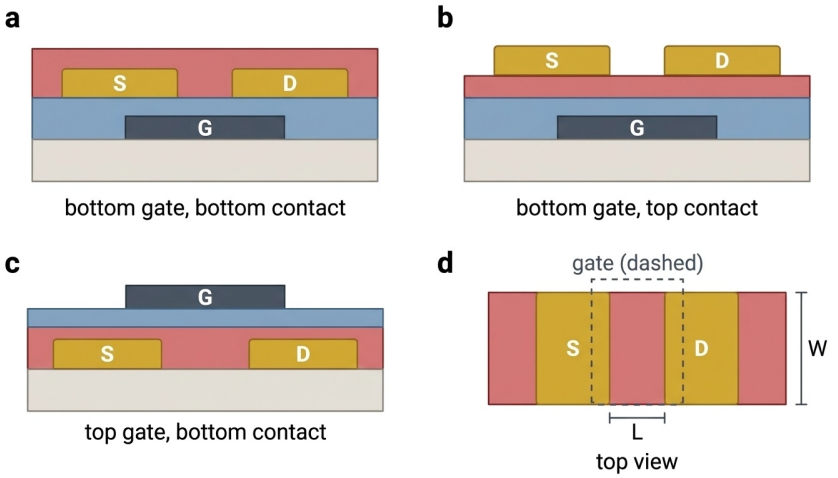


Figure 11.21: OTFT geometries in cross section (substrate light gray, gate dark gray, dielectric blue, gold source and drain, organic semiconductor red) and in top view. The channel of length L and width W lies between source and drain, at the semiconductor-dielectric interface.

The current follows the same square law as the MOSFET. With an overdrive $V_{GS} - V_T$ and drain voltage V_{DS} (both negative for p-type, for which the equations give $|I_D|$),

$$\begin{aligned} I_D &= \frac{W}{L}\,\mu C_i \left(V_{GS} - V_T - \frac{V_{DS}}{2}\right) V_{DS} \quad \text{(linear)}, \\ I_D &= \frac{W}{2L}\,\mu C_i\,(V_{GS} - V_T)^2 \qquad\qquad\quad \text{(saturation)}. \end{aligned} \tag{11.17}$$

In saturation, $\sqrt{|I_D|}$ is a straight line in V_{GS} whose slope gives the mobility and whose intercept gives V_T (Fig. 11.22):

$$\mu = \frac{2L}{WC_i}\left(\frac{\mathrm{d}\sqrt{|I_D|}}{\mathrm{d}V_{GS}}\right)^2. \tag{11.18}$$

Real organic transistors often bend this line, through contact resistance, traps, and a mobility that depends on the gate voltage, and a mobility taken from a short steep segment can be badly overestimated.

KEY EQUATION

$I_D = \frac{W}{2L}\,\mu C_i (V_{GS} - V_T)^2$

OTFT in saturation: the slope of $\sqrt{I_D}$ against V_{GS} gives μ.

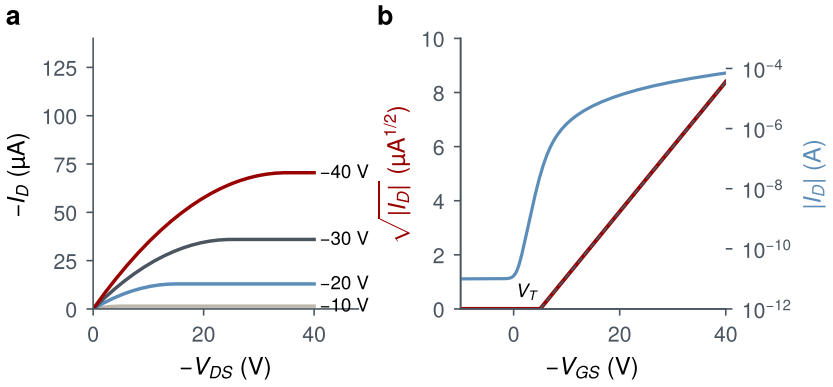


Figure 11.22: A p-type OTFT computed from Eq. (11.17) with the numbers of Example 11.4. (a) Output curves for V_{GS} from $-10\,\text{V}$ to $-40\,\text{V}$. (b) Transfer curve in saturation; $\sqrt{|I_D|}$ (red) is a straight line from $V_T = -5\,\text{V}$, and $|I_D|$ (blue, log scale) spans seven decades from the off current.

Worked Example 11.4 *Extracting mobility from the saturation current*

A p-type OTFT has $W = 1\,\text{mm}$, $L = 50\,\mu\text{m}$, and a 300 nm SiO_2 gate dielectric ($\varepsilon_r = 3.9$). In saturation ($V_{DS} = -40\,\text{V}$) it carries $I_D = -36.0\,\mu\text{A}$ at $V_{GS} = -30\,\text{V}$ and $-70.5\,\mu\text{A}$ at $-40\,\text{V}$. Find μ and V_T.

Solution. The dielectric is a parallel-plate capacitor: $C_i = \varepsilon_r\varepsilon_0/t = 3.9 \times 8.854 \times 10^{-12}\,\text{F/m}/300 \times 10^{-9}\,\text{m} = 1.151 \times 10^{-4}\,\text{F/m}^2 = 11.5\,\text{nF/cm}^2$. The square roots of the currents are $6.00 \times 10^{-3}\,\text{A}^{1/2}$ and $8.40 \times 10^{-3}\,\text{A}^{1/2}$, so the slope is $(8.40 - 6.00) \times 10^{-3}/10\,\text{V} = 2.40 \times 10^{-4}\,\text{A}^{1/2}/\text{V}$. Then

$$\mu = \frac{2(50 \times 10^{-6}\,\text{m})}{(1 \times 10^{-3}\,\text{m})(1.151 \times 10^{-4}\,\text{F/m}^2)}(2.40 \times 10^{-4}\,\text{A}^{1/2}/\text{V})^2$$
$$= 5.0 \times 10^{-5}\,\text{m}^2\,\text{V}^{-1}\,\text{s}^{-1} = 0.50\,\text{cm}^2\,\text{V}^{-1}\,\text{s}^{-1}.$$

The line reaches zero current 25.0 V (= 6.00/0.240) before $V_{GS} = -30\,\text{V}$, so $V_T = -5.0\,\text{V}$. These are the values used for Fig. 11.22. ■

The mobility sets the speed. In saturation the *transconductance*, the change of drain current per volt at the gate, is $g_m = \mu C_i (W/L)|V_{GS} - V_T|$, and the gate capacitance must be charged through it. Taking the whole gate capacitance, channel plus the overlaps L_{ov} of the gate under each contact, $C_G = C_i W(L + 2L_{ov})$, the frequency at which the transistor's current gain falls to one, the *transit frequency*, is

$$f_T = \frac{g_m}{2\pi C_G} = \frac{\mu\,|V_{GS} - V_T|}{2\pi L\,(L + 2L_{ov})}. \tag{11.19}$$

A digital gate cannot switch much faster than f_T, and an amplifier provides gain only well below it.

Building an OTFT from solution raises one more problem: each

KEY EQUATION

$$f_T = \frac{\mu|V_{GS} - V_T|}{2\pi L(L + 2L_{ov})}$$

Transit frequency: proportional to mobility, falling as the square of the channel length.

layer is printed or spun from a solvent, and that solvent must not dissolve or swell the layers already down. The rule is *like dissolves like*, and Hansen made it quantitative by describing each solvent and polymer with three numbers, for dispersion forces, polar forces, and hydrogen bonding. Two materials far apart in this three-dimensional *Hansen space* do not dissolve each other, so the next layer is printed from an *orthogonal* solvent, far from the layer below: for example, a polythiophene from an aromatic solvent, then a fluoropolymer dielectric from a fluorinated solvent. Cross-linking a layer makes it insoluble as well.

The organic electrochemical transistor

An OTFT gates its channel through a solid dielectric, which charges only the interface. An *organic electrochemical transistor* (OECT) gates it through an electrolyte, and the ions enter the whole channel (Fig. 11.23), just as they enter the PEDOT:PSS electrode coatings of Chapter 9. The channel is a film of a mixed conductor, almost always PEDOT:PSS, between gold source and drain contacts; an electrolyte (saline, sweat, a hydrogel, or tissue fluid) covers it, and a gate electrode, often silver/silver chloride, sits in the electrolyte. At zero gate voltage the PEDOT is doped and conducts: the OECT is a *depletion-mode* device, on until the gate turns it off. A positive gate voltage pushes cations from the electrolyte into the film, where they pair with the PSS^- sites; each cation that enters releases a hole, which leaves through the drain, so the channel de-dopes and the current falls.

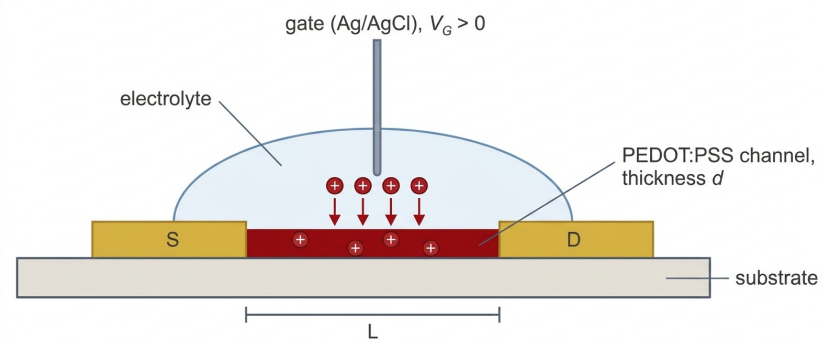


Figure 11.23: An organic electrochemical transistor. A positive gate voltage drives cations from the electrolyte into the whole volume of the PEDOT:PSS channel, where they pair with the sulfonate groups and remove holes, so the drain current falls. Because ions charge the volume, not just a surface, the effective capacitance is enormous.

Because the ions charge the whole volume, the relevant capacitance is a *volumetric capacitance* C^* (farads per cubic centimeter), about $39\,\mathrm{F/cm^3}$ for PEDOT:PSS, and the gate capacitance is C^* times the channel volume WLd. In the Bernards-Malliaras model the saturation

transconductance becomes

$$g_m = \frac{Wd}{L}\,\mu C^*\,|V_T - V_G|, \tag{11.20}$$

the OTFT expression with C_i replaced by C^*d.[11] OECTs reach millisiemens at gate voltages below a volt, inside the water window of Chapter 9, and in direct contact with tissue; OECTs recording brain activity in living animals[12] showed a higher signal-to-noise ratio than surface electrodes. Thicker channels give more transconductance but respond more slowly, because the ions must travel farther. An ion-selective membrane (Chapter 6) makes an OECT respond to one ion: a printed OECT sensor for lithium, the drug many patients with bipolar disorder take, detected 0.1 mM in artificial sweat and was tested on healthy volunteers and on patients.[13]

KEY EQUATION

$g_m = \frac{Wd}{L}\,\mu C^*|V_T - V_G|$

OECT transconductance.

[11] J. Rivnay *et al.*, *Sci. Adv.* (2015); D. A. Bernards and G. G. Malliaras, *Adv. Funct. Mater.* (2007).

[12] D. Khodagholy *et al.*, *Nat. Commun.* (2013).

[13] M. S. Islam, B. P. Kunnel, M. Ferdoushi, *et al.*, Y. Khan, *Device* (2025).

Worked Example 11.5 *Why OECTs amplify so well*

A PEDOT:PSS OECT has $W = 100\,\mu m$, $L = 10\,\mu m$, $d = 100\,nm$, $\mu = 2\,cm^2\,V^{-1}\,s^{-1}$, and $C^* = 39\,F/cm^3$, biased 0.5 V from threshold. (a) Find g_m. (b) Compare with an OTFT of the same W/L on 300 nm SiO_2, with $\mu = 1\,cm^2\,V^{-1}\,s^{-1}$ at a 10 V overdrive. (c) What drain current change does a 100 µV signal at the gate produce in the OECT?

Solution. (a) In centimeters, $Wd/L = (0.01)(1\times10^{-5})/(0.001) = 1\times10^{-4}$ cm, so

$$g_m = (1\times10^{-4}\,cm)(2\,cm^2\,V^{-1}\,s^{-1})(39\,F/cm^3)(0.5\,V) = 3.9\,mS.$$

(b) $g_m = (W/L)\mu C_i V_{ov} = 10\times(1\times10^{-4}\,m^2\,V^{-1}\,s^{-1})(1.151\times10^{-4}\,F/m^2)(10\,V) = 1.15\,\mu S$, about 3400 times less at 20 times the voltage. The reason is the capacitance: $C^*d = 390\,\mu F/cm^2$ against $C_i = 11.5\,nF/cm^2$, a factor of 34 000.

(c) $\Delta I = g_m \Delta V = 3.9\,mS \times 100\,\mu V = 0.39\,\mu A$. Through a 1 kΩ drain resistor that is a 0.39 mV output, a local voltage gain of about 4 right at the tissue. ■

11.11 *Encapsulation and Stability*

Soft devices fail from the outside in. Oxygen and water trap electrons in organic semiconductors and oxidize them, light drives photo-oxidation, and on the body, sweat and tissue fluid bring salts and proteins. Metals corrode, low-work-function contacts oxidize, PEDOT:PSS swells, and polyimide slowly takes up water at body temperature. Lifetime is therefore often set by the *encapsulation*, the barrier over the device, rated by its *water vapor transmission rate* (WVTR)

in grams of water per square meter per day. A single polymer film passes on the order of a gram per square meter per day; ALD oxides and stacks of alternating organic and inorganic layers reach 10^{-3} or better; and an OLED display needs about 10^{-6}. For thin implants, parylene C over an ALD oxide is a common stack, and at the right thickness the same layer also puts the metal at the neutral plane (Section 11.5). Implants meant to last decades put their electronics in a hermetic metal or ceramic case and use polymers only on the leads.

WVTR is measured with the *calcium test*. A thin calcium film under the barrier stays opaque and conductive until water reaches it, then turns into transparent, insulating calcium hydroxide, $\mathrm{Ca} + 2\,\mathrm{H_2O} \longrightarrow \mathrm{Ca(OH)_2} + \mathrm{H_2}$. Each gram of calcium consumes $2 \times 18.0/40.1 = 0.90$ grams of water. A 100 nm calcium film ($1.55\,\mathrm{g/cm^3}$) holds $0.155\,\mathrm{g/m^2}$ of calcium, which consumes $0.139\,\mathrm{g/m^2}$ of water: under a barrier of $1\,\mathrm{g\,m^{-2}\,day^{-1}}$ it is gone in about 3.3 h, under $10^{-3}\,\mathrm{g\,m^{-2}\,day^{-1}}$ in about 140 days, and at the OLED requirement in centuries. Timing the change in the film's resistance or transparency gives the WVTR. The same arithmetic estimates the life of any moisture-sensitive device, once you know how much water it tolerates.

Summary

- Soft devices use soft materials, or stiff materials made thin: bending stiffness goes as Et^3, so a film ten times thinner bends a thousand times more easily. Plastic substrates limit processing: PET softens at its glass transition near 78 °C (heat-stabilized grades take brief bakes to 150 °C), PEN is stable to about 180 °C, and polyimide survives 350 °C.
- Conductors, semiconductors (gap about 1 to 3 eV), and insulators; $\sigma = nq\mu$ spans more than 20 orders of magnitude. PDMS (about 1.5 MPa) is the common elastomer; hydrogels are 70 to 99 % water. Polyimide grows 0.2 % from 25 to 125 °C and 20 ppm per %RH.
- PVD: sputtering at mTorr (partly conformal), evaporation at 10^{-6} Torr (line of sight). CVD is conformal; PECVD runs at 200 to 350 °C, parylene at room temperature, ALD at about 0.1 nm per cycle. Noble metals need about 10 nm or an adhesion layer to be continuous. A quartz crystal monitor weighs the film as it grows.
- Bending: $\varepsilon = y/R$; sheet surface $t/2R$; neutral plane $y_n = \sum E_i t_i \bar{y}_i / \sum E_i t_i$. On 250 µm PET at 5 mm, 10 nm silver sees 2.50 % and 10 µm silver 1.36 %. Encapsulating with $E_e t_e^2 = E_s t_s^2$ centers the neutral plane on the device.
- Films fail by microcracks, channel cracks, buckling, and delamina-

tion, often from about 1 % strain. Thermal stress $[E_f/(1-\nu_f)](\alpha_f - \alpha_s)\Delta T$; a stressed film cracks by itself above $h_c = G_c \bar{E}_f/(Z\sigma^2)$.

- Stretching: liquid metal $R/R_0 = (1+\varepsilon)^2$, solids $(1+\varepsilon)^{1+2\nu}$; serpentines (57 % for semicircles), wrinkles, island-bridge, and kirigami.
- Nanoparticles have many surface atoms ($\approx 6a/d$) and sinter by surface diffusion at 100 to 250 °C once the ligands leave. A 1 ms flash heats only $L_{\text{th}} = \sqrt{D_{\text{th}} t} \approx 10\,\mu\text{m}$ of PET. $R = R_s \ell/w$ with $R_s = \rho/t$; sintered silver is 2 to 5 times bulk, flake pastes 10 to 30 times.
- Nanowires percolate at $N_c = 5.64/L^2$ and stay connected under strain; sintered nanoparticle films crack.
- Conjugation spreads π orbitals along the chain; the HOMO and LUMO bound a gap of 1.5 to 3 eV. In PEDOT:PSS, PEDOT carries holes and PSS balances charge; secondary doping raises the conductivity to about 1000 S/cm. Hopping, $\mu \approx (qa^2/k_B T)\Gamma$, keeps organic mobilities near 0.01 to 10 $\text{cm}^2\,\text{V}^{-1}\,\text{s}^{-1}$.
- OTFT: $I_D = (W/2L)\mu C_i (V_{GS} - V_T)^2$ in saturation; $f_T = \mu |V_{GS} - V_T|/[2\pi L(L + 2L_{\text{ov}})]$, kilohertz to megahertz for organics, tens of gigahertz for silicon. OECT: $g_m = (Wd/L)\mu C^* |V_T - V_G|$ with $C^* \approx 39\,\text{F/cm}^3$, millisiemens below 1 V.
- Encapsulation is rated by WVTR, measured with a calcium film; displays need about $10^{-6}\,\text{g}\,\text{m}^{-2}\,\text{day}^{-1}$.

Chapter 12 puts the inks of this chapter through printers.

11.12 *Problems*

11.1 ° A patch 20 mm wide and 0.2 mm thick is stretched by 20 % with the skin. Find the force for Ecoflex (70 kPa), PDMS (1.5 MPa), TPU (20 MPa), and PET (4 GPa, ignoring that it would yield). Which would the skin hardly notice?

11.2 •• A 1 µm silicon membrane ($E = 130$ GPa) sits on 12.5 µm polyimide ($E = 2.5$ GPa) and is bent to $R = 1$ mm, silicon outside.

(a) Locate the neutral plane and find the strain at the top and bottom of the silicon.

(b) Compare with the estimate $(t_s + t_f)/2R$. Why is it so far off?

(c) What thickness of polyimide on top of the silicon puts the neutral plane at the silicon mid-plane?

11.3 • An EGaIn strain sensor ($\rho = 29.4 \times 10^{-8}\,\Omega\,\text{m}$) is a channel 20 mm long, 200 µm wide, and 50 µm deep. Find its resistance re-

laxed and at 30 % strain, and its gauge factor $(\Delta R/R_0)/\varepsilon$ at 30 % and at small strain.

11.4 • An inkjet-printed silver line is 60 µm wide, 0.5 µm thick, and 20 mm long, at 2.5 times the bulk resistivity. Find R_s and R. How thick would it have to be for 1 Ω, and why might you print several passes instead?

11.5 • An OECT has $W = L = 50\,\mu\text{m}$, $d = 200\,\text{nm}$, $\mu = 1.5\,\text{cm}^2\,\text{V}^{-1}\,\text{s}^{-1}$, and $C^* = 39\,\text{F/cm}^3$. Find g_m 0.3 V from threshold. What change of geometry doubles g_m, and what does it cost?

12
Printed Bioelectronics

In this chapter you will learn to

- Explain why printing suits bioelectronics and sort printing methods by contact, pattern carrier, and format.
- Use Young's equation and the polar and dispersive parts of surface energy to predict whether an ink wets a substrate, and choose a surface treatment that changes the answer.
- Describe how an ink flows with viscosity and shear thinning, and match an ink to a printer.
- Explain the coffee-ring effect and how cosolvents and heat counter it.
- Check an inkjet ink with the Ohnesorge number and predict the width of a printed line.
- Estimate the film thickness left by screen, stencil, gravure, and slot-die printing.
- Choose a printing method from feature size, film thickness, ink, number of copies, and substrate.
- Explain how rough edges and misaligned layers make printed devices differ.
- Describe a printed Ag/AgCl electrode and size its chloride layer.

Tape a sheet of clear plastic to the bench, lay a vinyl stencil on it, and drag a plastic card across a spoonful of silver paste. When you peel the stencil away, sixteen square pads and their wiring sit on the plastic. Fifteen minutes in an oven at 120 °C turn the paste into a conductor with a resistance of a few ohms from end to end. Fold a second print over the first with a slice of foam between them and you

have a pressure sensor array that you can sew into a glove. Nothing in this process needed a vacuum chamber, a mask aligner, or a cleanroom.

That is the promise of printed electronics: functional materials put down only where they are needed, at low temperature, on large and soft substrates, straight from a drawing. Chapter 11 made the inks; this chapter puts them through printers. After a look at why printing is worth the trouble, three sections explain the physics that decides every print, starting each time from everyday experience: how a liquid wets a surface, how an ink flows, and how a wet film dries. Then we tour the printers (inkjet, screen and stencil, roll-based presses, and direct writing) and end with a guide to choosing among them. The last two sections look at why two printed devices are never quite the same and at the printed electrodes and sensors that touch the body. Lab 2 in Part V prints electrodes and a 4×4 capacitive pressure array.

12.1 *Why Print?*

A conventional thin-film process is *subtractive*. To pattern one metal layer you deposit metal everywhere (Chapter 11), coat it with photoresist, expose the resist through a mask, develop it, etch away the unprotected metal, and strip the resist (Fig. 12.1). Each layer repeats the cycle, and most of the metal ends up in the etch bath. Printing is *additive*: deposition and patterning happen in one step, and the material lands only where the drawing puts it. A $1\,\mathrm{cm}^2$ electrode on a $10\,\mathrm{cm} \times 10\,\mathrm{cm}$ sheet uses about a hundredth of the metal of a blanket film.

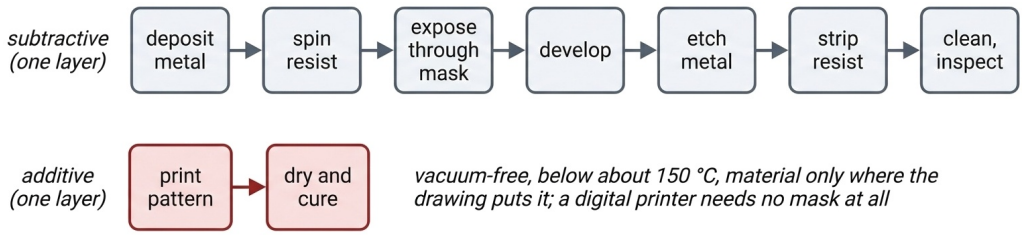


Figure 12.1: Subtractive and additive patterning. Photolithography needs about seven steps, a mask, and a vacuum deposition for every patterned layer, and etches away most of the material it deposits. Printing deposits and patterns in one step and then dries or cures the ink.

Four properties make printing attractive for bioelectronics. *Additive patterning* saves steps and material. *Large area*: presses built for newspapers run webs a meter or more wide, so a patch that covers a chest costs little more than a small one. *Low temperature*: inks dry and cure below about 150 °C, which suits plastics, elastomers, paper, and textiles. *Rapid prototyping*: a digital printer goes from a drawing to a device in an hour. The price is resolution and performance. A printed silver line is tens of micrometers wide and conducts several times worse than bulk silver, and a printed transistor is slow. So printing does not replace the integrated circuit. It makes the parts of a

system that must be large, soft, cheap, or custom (electrodes, interconnects, antennas, sensors, and the film that carries a few silicon chips), while the chips amplify and transmit (Chapter 1).

Three questions sort the methods. *Does the printing element touch the substrate?* Gravure, flexography, and offset press an inked surface against it (*contact*); inkjet, aerosol jet, spray, and slot-die coating do not (*non-contact*); screen printing touches only along the line of the squeegee (*off-contact*). Non-contact methods can print on rough, curved, or fragile surfaces. *Where does the pattern live?* A *digital* method (inkjet, aerosol jet, dispenser; Fig. 12.2) keeps it in a file; a *pattern-carrier* method keeps it in a screen, stencil, plate, or engraved roll that must be made first. *Sheets or rolls?* Roll-to-roll presses print a continuous web and give the highest throughput; digital methods give the easiest design changes.

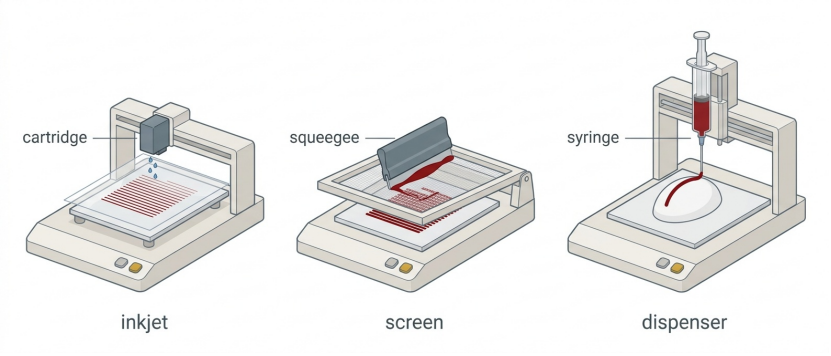


Figure 12.2: Three printers found in a printed-electronics lab. A drop-on-demand inkjet printer jets picoliter drops from a moving cartridge. A screen printer pushes viscous paste through a patterned mesh with a squeegee. A dispenser (direct-write) printer extrudes paste from a syringe needle along a programmed path.

Every method prints an *ink*: a functional material (silver nanoparticles or flakes, carbon, PEDOT:PSS, a dielectric, an enzyme) in a solvent, with additives that set how it flows and wets. Each printer needs its own kind of ink: the same silver is sold as a 10 mPa s inkjet ink and as a screen-printing paste a thousand times thicker. The next three sections explain why.

12.2 *Wetting and Surface Energy*

THE FIRST THING AN INK DOES after leaving the printer is meet the substrate. Water beads up on a freshly waxed car and spreads into a thin sheet on clean glass. An ink does the same, and whether it spreads, beads up, or stays where it was put decides the shape of every printed feature.

Surface tension and the contact angle

A molecule inside a liquid is pulled equally in all directions by its neighbors. A molecule at the surface has neighbors on one side only, so it sits at a higher energy, and the liquid keeps its surface as small as it can; that is why small drops are round. The extra energy per area is the *surface tension* γ, in $\mathrm{J/m^2}$ or, the same thing, $\mathrm{N/m}$. Water has $72.8\,\mathrm{mN/m}$ at $20\,^\circ\mathrm{C}$, among the highest of common liquids; most organic solvents have 20 to $50\,\mathrm{mN/m}$. Solids have a surface energy too. Clean glass and metal oxides have high surface energies, and water spreads on them. Most plastics have low ones, and PDMS and Teflon, at about $20\,\mathrm{mN/m}$, are among the lowest. A rule of thumb from the printing industry follows: an ink wets a substrate well when the substrate's surface energy exceeds the ink's surface tension by about $10\,\mathrm{mN/m}$.

Where solid, liquid, and gas meet, three tensions pull on the edge of the drop, the *contact line* (Fig. 12.3). The solid-gas tension pulls it outward, because covering dry solid with liquid removes solid-gas surface. The solid-liquid tension pulls it inward, and the liquid surface pulls along itself. The drop settles where the horizontal pulls balance:

$$\gamma_{SG} = \gamma_{SL} + \gamma_{LG}\cos\theta_c \quad\Longrightarrow\quad \cos\theta_c = \frac{\gamma_{SG} - \gamma_{SL}}{\gamma_{LG}}. \tag{12.1}$$

This is *Young's equation*, and θ_c, measured through the liquid, is the *contact angle*. A high-energy solid makes θ_c small; a liquid of high surface tension makes it large. Three regimes follow (Fig. 12.4). At $\theta_c = 0$ the liquid wets *completely* and spreads into a film without limit. For $0 < \theta_c \le 90^\circ$ it wets *partially* and settles as a cap of finite size (for water, the surface is called *hydrophilic*). For $\theta_c > 90^\circ$ it does not wet and beads up (*hydrophobic*); a printed line breaks into separate drops.

Printing wants partial wetting, and in practice its lower half: roughly 20 to 60° for inkjet. The ink must spread enough to join neighboring drops and cover a pad, and then stop firmly so the edges stay where they were drawn. With complete wetting, lines widen without limit, merge, and run along scratches; above 90°, drops never join and a line has no continuous path. Silver printed on untreated PDMS stays as isolated drops, and it forms continuous lines only after about half an hour of UV-ozone treatment.[1]

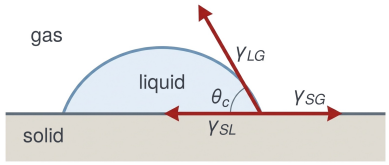


Figure 12.3: Young's equation. At the contact line the solid-gas tension pulls outward, the solid-liquid tension pulls inward, and the liquid-gas tension pulls along the liquid surface at the contact angle θ_c.

KEY EQUATION

$$\cos\theta_c = \frac{\gamma_{SG} - \gamma_{SL}}{\gamma_{LG}}$$

Young's equation: the contact angle balances three surface tensions.

[1] S. Jagannathan, Ph.D. thesis, University of California, Berkeley (2012).

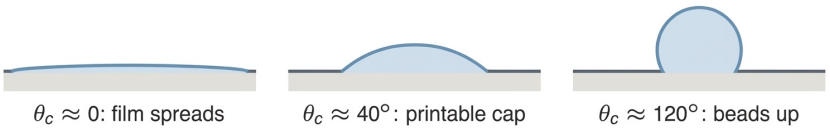


Figure 12.4: Contact-angle regimes for a drop of the same volume. Complete wetting spreads the ink without limit; partial wetting gives a cap whose size you can predict; non-wetting gives beads that do not join into a continuous line.

Polar and dispersive surface energy

Molecules attract each other in two ways. *Dispersion* forces, from fluctuating electron clouds, act between all molecules. *Polar* interactions, between permanent dipoles and through hydrogen bonds, act only between polar molecules. Water is mostly polar; oils, most plastics, and the probe liquid diiodomethane are mostly or entirely dispersive. Owens, Wendt, Rabel, and Kaelble split each surface energy into the two parts, $\gamma = \gamma^d + \gamma^p$, and assumed that a liquid and a solid attract only through the parts they share. Young's equation then becomes the *Owens-Wendt relation*

$$\gamma_L\,(1+\cos\theta) = 2\left(\sqrt{\gamma_S^d\gamma_L^d} + \sqrt{\gamma_S^p\gamma_L^p}\right). \tag{12.2}$$

It explains why water beads on a plastic of fairly high total surface energy: most of what holds water together is polar, and it finds no polar partner on the surface.

Liquid	γ_L	γ_L^d	γ_L^p
water	72.8	21.8	51.0
diiodomethane	50.8	50.8	0
ethylene glycol	48.0	29.0	19.0

Table 12.1: Probe liquids (mN/m, 20 °C), with commonly used dispersive and polar parts. Tabulations differ by a few mN/m.

You measure a solid's two parts with two probe liquids of known components (Table 12.1). Diiodomethane has no polar part, so its angle alone gives the dispersive part of the solid, $\gamma_S^d = [\gamma_L(1+\cos\theta)/2]^2/\gamma_L^d$; the water angle then gives the polar part. On clean PET, water makes 75° and diiodomethane 30° (illustrative angles), which gives $\gamma_S^d = 44.2\,\mathrm{mN/m}$ and $\gamma_S^p = 4.3\,\mathrm{mN/m}$: untreated PET is almost purely dispersive. After a short oxygen plasma the angles are 30° and 28°, and the same steps give $\gamma_S^d = 45.0\,\mathrm{mN/m}$ and $\gamma_S^p = 26.3\,\mathrm{mN/m}$. The plasma left the dispersive part alone and raised the polar part sixfold. With the two parts known, Eq. (12.2) predicts the angle of any ink.

WORKED EXAMPLE 12.1 *Will the ink wet the film?*

A silver nanoparticle ink based on ethylene glycol has the surface tension components of Table 12.1. Predict its contact angle on clean PET ($\gamma_S^d = 44.2$, $\gamma_S^p = 4.3\,\mathrm{mN/m}$) and on plasma-treated PET (45.0 and 26.3 mN/m). Which would you print on?

SOLUTION. Solve Eq. (12.2) for the cosine: $\cos\theta = 2(\sqrt{\gamma_S^d\gamma_L^d} + \sqrt{\gamma_S^p\gamma_L^p})/\gamma_L - 1$. Clean PET: $2(\sqrt{44.2\times29.0} + \sqrt{4.3\times19.0})/48.0 - 1 = 2(35.80+9.04)/48.0 - 1 = 0.868$, so $\theta = 30°$. Treated: $2(36.12+22.35)/48.0 - 1 = 1.44$. A cosine above 1 has no angle; it means complete wetting. Print on the clean film, where drops stop at a

predictable size. The plasma made the surface too easy to wet, and lines would flood sideways; a gentler treatment would land in between. Plasma is the cure for the opposite problem, an ink that beads up, as on PDMS. ■

Changing surface energy

Oxygen plasma, *UV-ozone*, and *corona* discharge make reactive oxygen that burns off adsorbed hydrocarbons and leaves oxygen-containing groups on the surface, which raise its polar surface energy. Plasma takes seconds in a vacuum chamber, UV-ozone minutes in air, and corona treats moving plastic webs at atmospheric pressure. On PDMS a plasma drops the water angle from above 100° to below 10°, but short polymer chains migrate back to the surface within hours to days (*hydrophobic recovery*), so print soon after treating. To *lower* the surface energy, a *self-assembled monolayer* bonds a single layer of molecules to the surface with their tails outward; a fluorinated silane gives a water angle near 107°. *Surface-energy patterning* combines the two: a low-energy coating covers the regions where ink must not go, and ink coated over the whole sheet collects on the rest. One printed oximeter array patterned the layers of its pixels this way.[2]

[2] Y. Khan, D. Han, A. Pierre, *et al.*, *Proc. Natl. Acad. Sci. USA* (2018).

Real surfaces also hold on to the contact line. Add liquid to a drop and its angle rises until the edge jumps outward; remove liquid and the angle falls until the edge retreats. In between, roughness and chemical patchiness *pin* the edge in place. Pinning keeps printed edges where they were drawn while they dry, and it also causes the coffee ring of Section 12.4.

12.3 *Ink Rheology*

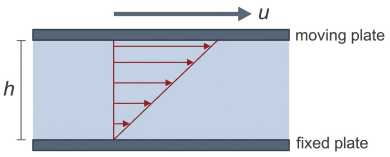


Figure 12.5: Shearing a liquid between parallel plates. The top plate moves at speed u, the fluid velocity grows linearly across the gap h, and the shear rate is u/h everywhere.

Rheology is the science of flow. It answers two questions about an ink: will it move through the printer, and will it stop moving once it lands?

Viscosity is a liquid's resistance to flowing: the reason honey pours slowly and water quickly. To define it, put the liquid between two plates a gap h apart and slide the top plate at speed u (Fig. 12.5). The liquid touching each plate moves with it, so the velocity changes steadily across the gap. The *shear rate* is that velocity gradient, $\dot{\gamma} = u/h$ (in s^{-1}), the *shear stress* is the force per plate area, $\tau = F/A$, and the viscosity is their ratio:

$$\eta = \frac{\tau}{\dot{\gamma}}, \qquad F = \eta A \frac{u}{h}, \tag{12.3}$$

in Pa s. Printers still quote centipoise, 1 cP = 1 mPa s. Water is 1.0 mPa s at 20 °C (0.65 mPa s at 40 °C, which is why inkjet heads are often heated), olive oil about 80, honey 2000 to 10 000, and peanut butter 10^4 to 10^6 mPa s. Solvents are near 1 mPa s; an ink's viscosity comes from its dissolved polymer and suspended particles, and it climbs steeply with their concentration.

For water and simple solvents η is a constant, whatever the shear rate: they are *Newtonian*. Concentrated inks are *non-Newtonian* (Fig. 12.6). Ketchup is the everyday example: it sits in the bottle until you shake or squeeze it, then pours. Such *shear-thinning* fluids get thinner as $\dot{\gamma}$ rises, because the particle networks and tangled polymer chains that resist slow flow break up and align in fast flow. *Shear-thickening* fluids get thicker, because dense particles jam into clusters (cornstarch in water). A *Bingham plastic* does not flow at all until the stress exceeds a *yield stress* (toothpaste). A *thixotropic* fluid thins while sheared and takes seconds to minutes to thicken again. The *power law*

$$\eta = K\dot{\gamma}^{\,n-1} \tag{12.4}$$

describes the first two with two numbers: $n < 1$ for thinning, $n > 1$ for thickening, and $n = 1$ (Newtonian) with $\eta = K$.

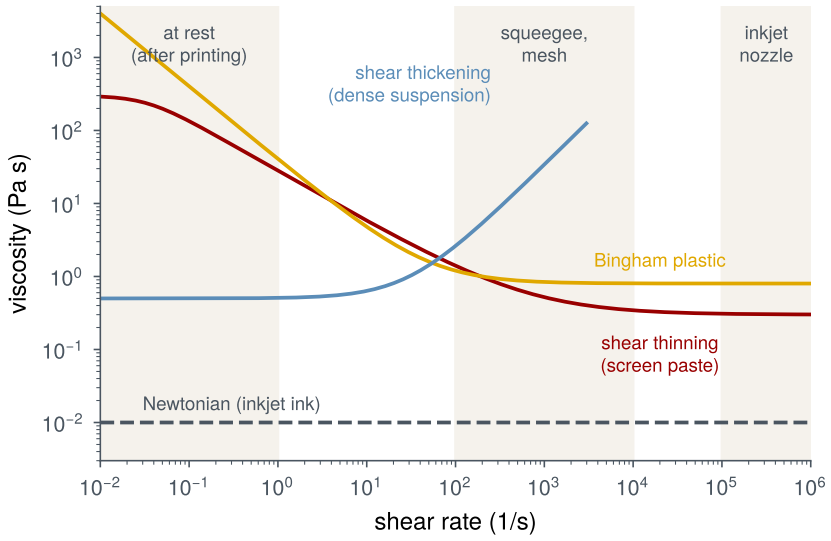


Figure 12.6: Viscosity versus shear rate for the four kinds of fluid (illustrative curves). The shaded bands are typical shear rates at rest after printing, under a squeegee and in a screen mesh, and in an inkjet nozzle. A shear-thinning paste is a thousand times thinner in the mesh than on the substrate.

The shear rate in any printing step is about a speed divided by a gap, $\dot{\gamma} \approx U/h$. A squeegee at 100 mm/s pushing paste through 50 µm mesh openings gives about 2000 s^{-1}; ink leaving a 21.5 µm inkjet nozzle at 6 m/s, about 3×10^5 s^{-1}. Once the ink lands, only surface tension and gravity move it, and $\dot{\gamma}$ falls below 1 s^{-1}. A *rheometer* measures η over this whole range.

Which forces matter?

Gravity, viscosity, and surface tension compete in a printed drop, and their ratios tell which ones matter without solving the flow. A puddle is flat because gravity wins; a dew drop is round because surface tension wins. The crossover size is the *capillary length* ℓ_c,

$$\ell_c = \sqrt{\frac{\gamma}{\rho g}}, \qquad Ca = \frac{\eta U}{\gamma}, \tag{12.5}$$

2.7 mm for water and about 1.6 mm for an ink of 25 mN/m. Printed features are a hundred times smaller, so gravity is negligible and each drop takes the shape surface tension gives it, a slice of a sphere called a *spherical cap*. The *capillary number* Ca compares viscous forces with surface tension for liquid moving at speed U. For 10 mPa s, 25 mN/m, and 1 m/s, $Ca = 0.4$: neither force dominates, which is why Ca appears in almost every printing process. Inertia matters only when ink is jetted (Section 12.5).

12.4 Drying and the Coffee-Ring Effect

A printed feature is still a liquid. As its solvent evaporates the ink thickens until it is solid; sintering or curing then makes it a conductor or a dielectric (Chapter 11). The evaporation rate is the control knob. It rises with temperature and with the solvent's vapor pressure (a low boiling point usually means a high vapor pressure). Slow drying gives the ink time to level, and also to spread; fast drying freezes it in place, defects included.

A spilled drop of coffee dries into a dark ring, and so does a drop of silver ink.[3] Two facts cause it. The contact line is pinned, so the drop cannot shrink inward as it dries; it can only get flatter. And evaporation is fastest near the thin edge. Liquid must flow outward to replace what evaporates at the edge, and it carries the particles with it, which pile up in a ridge (Fig. 12.7a). For a device this matters: a conductor with a thick rim and a thin center has more resistance than its average thickness suggests, and a dielectric printed over the ridge can be thin enough to short.

[3] R. D. Deegan, O. Bakajin, T. F. Dupont, *et al.*, *Nature* (1997).

The cures follow from the mechanism. *Raise the viscosity*, within what the printer allows, so the outward flow is slow. *Dry faster* on a heated substrate, so the film gels before the particles reach the edge. Or *set up a Marangoni flow*. A liquid surface pulls toward wherever its surface tension is highest and drags the liquid below with it. You see this in the "tears" that climb the inside of a wine glass: alcohol evaporates from the thin film on the glass, the film's surface tension rises, and it pulls wine upward. In a drop whose edge has a lower

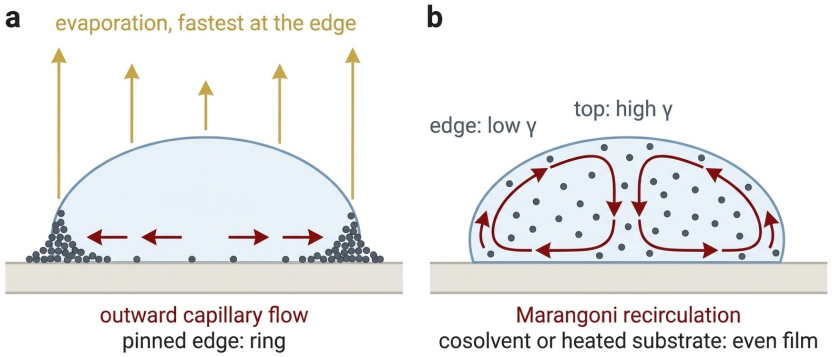


Figure 12.7: (a) The coffee-ring effect. With the edge pinned and evaporation fastest there, liquid flows outward and carries particles into a ridge at the edge. (b) A surface-tension gradient along the drop surface, lower at the edge than at the top, drives a Marangoni flow along the surface toward the top and a return flow near the substrate, which keeps particles circulating instead of piling up at the edge.

surface tension than its top, the surface flows from the edge toward the top and the liquid returns along the substrate, so the particles circulate instead of collecting (Fig. 12.7b).

A *cosolvent* with a *higher* boiling point and a *lower* surface tension than the main solvent creates this gradient. Add 10 to 20 % ethylene glycol (boiling point 197 °C, 48 mN/m) to a water-based ink (100 °C, 72.8 mN/m). Water leaves the edge first, the edge becomes glycol-rich with a lower surface tension than the top, and the Marangoni flow recirculates the particles. A warm substrate helps the same way, because surface tension falls with temperature (about -0.15 mN/m per kelvin for water) and the edge, close to the substrate, is warmer than the top. Traces of surfactant suppress Marangoni flows in water, which makes water-based inks harder to fix this way.

12.5 *Inkjet Printing*

An inkjet printer is the most familiar digital printer: the pattern is a bitmap, and a moving head fires drops wherever the bitmap says. Printed electronics uses *drop-on-demand* heads, which fire a drop only when told to. In a *piezoelectric* head a voltage pulse flexes a wall of a small ink chamber, and the pressure wave pushes a drop of a few picoliters out of the nozzle (Fig. 12.8). In a *thermal* head, the kind in office printers, a heater boils a thin layer of ink and the bubble pushes the drop. Thermal heads suit mostly water-based inks; piezo heads accept most solvents and are the standard for functional inks. The drop leaves with a thin tail, which may break into *satellite* drops that land beside it, so the drive *waveform* is tuned while a strobe camera photographs drops in flight. Lab printers jet at up to about 10 kHz at 5 to 10 m/s and print lines down to about 30 µm. A drop of volume V has diameter $d_0 = (6V/\pi)^{1/3}$: 12.4 µm for 1 pL and 26.7 µm for 10 pL, close to the nozzle diameter. Smaller nozzles print finer but clog more

easily, which keeps them at 10 to 50 µm.

Jettability and the Ohnesorge number

A good ink forms one clean drop per pulse. Three properties decide whether it does. Viscosity resists ejection and damps the tail; surface tension pulls the tail into drops; inertia carries the drop away. For ink of density ρ leaving a nozzle of diameter d at speed U, two ratios compare them: the *Reynolds number* $Re = \rho U d/\eta$ (inertia over viscosity) and the *Weber number* $We = \rho U^2 d/\gamma$ (inertia over surface tension). Their combination $\sqrt{We}/Re$ no longer contains the speed. It is a property of the ink and the nozzle alone, the *Ohnesorge number*:

$$Oh = \frac{\sqrt{We}}{Re} = \frac{\eta}{\sqrt{\rho\gamma d}}, \qquad Z = \frac{1}{Oh} = \frac{\sqrt{\rho\gamma d}}{\eta}. \tag{12.6}$$

A widely used rule says an ink jets well when $1 < Z < 10$. Below $Z = 1$ viscosity wins and the pulse cannot push out a drop; above $Z = 10$ surface tension wins over viscosity and the long tail breaks into satellites. Two more limits close the window (Fig. 12.9): the drop must be fast enough to break free of the nozzle (about $We > 4$), and slow enough not to splash when it lands.[4] In practice inkjet inks have 2 to 20 mPa s and 25 to 50 mN/m at the jetting temperature (one common lab cartridge asks for 10 to 12 mPa s and 28 to 33 mN/m), particles smaller than about a fiftieth of the nozzle, a filter before loading, and a high-boiling cosolvent that keeps the nozzle from drying out.

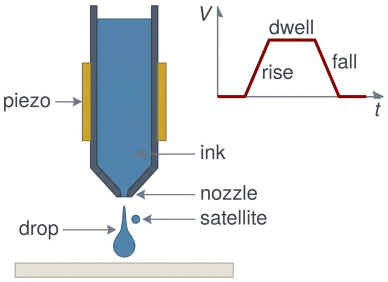


Figure 12.8: A piezoelectric drop-on-demand nozzle and its drive waveform. The rise draws ink in, the dwell lets the pressure wave travel, and the fall ejects the drop; a badly tuned pulse leaves a tail that breaks into satellites.

KEY EQUATION

$$1 < Z = \frac{\sqrt{\rho\gamma d}}{\eta} < 10$$

The printable window of drop-on-demand inkjet.

[4] B. Derby, *Annu. Rev. Mater. Res.* (2010). Other studies find somewhat different bounds; treat the window as a guide.

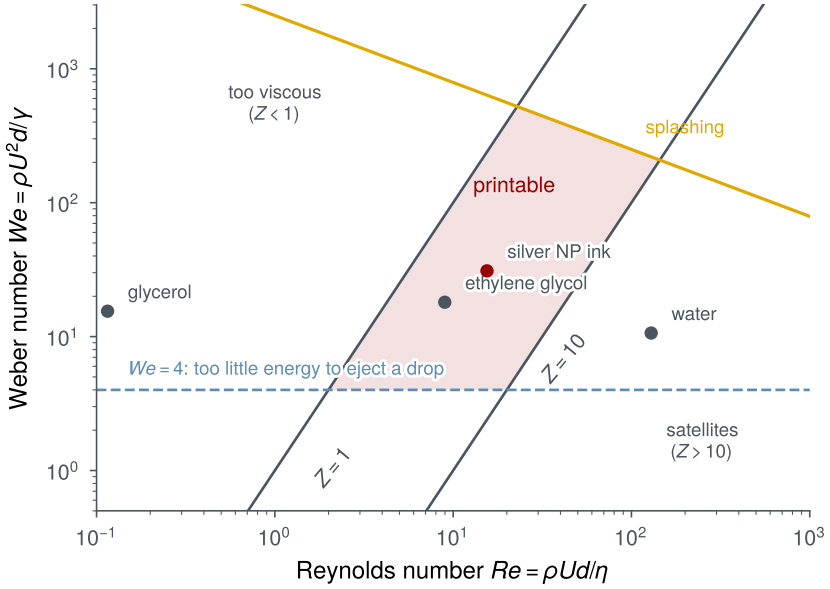


Figure 12.9: The printable window of drop-on-demand inkjet in the plane of Reynolds and Weber numbers, bounded by $Z = 1$, $Z = 10$, the ejection limit $We = 4$, and the onset of splashing. The points are the inks of Worked Example 12.2 and ethylene glycol, jetted at 6 m/s from a 21.5 µm nozzle: the silver ink and ethylene glycol fall inside, water makes satellites, and glycerol is far too viscous.

WORKED EXAMPLE 12.2 *Is this ink jettable?*

A silver nanoparticle ink has $\eta = 10\,\text{mPa s}$, $\gamma = 30\,\text{mN/m}$, and $\rho = 1200\,\text{kg/m}^3$, and is printed from a $21.5\,\mu\text{m}$ nozzle. (a) Find Z. (b) Find Re and We at $6\,\text{m/s}$. (c) Find Z for water ($1.0\,\text{mPa s}$, $72.8\,\text{mN/m}$, $1000\,\text{kg/m}^3$) and glycerol ($1.41\,\text{Pa s}$, $63\,\text{mN/m}$, $1260\,\text{kg/m}^3$).

SOLUTION. (a) $\sqrt{\rho\gamma d} = \sqrt{1200 \times 0.030 \times 21.5 \times 10^{-6}} = 0.0278\,\text{Pa s}$, so $Z = 0.0278/0.010 = 2.78$ ($Oh = 0.36$), inside the window. (b) $Re = 1200 \times 6 \times 21.5 \times 10^{-6}/0.010 = 15.5$ and $We = 1200 \times 36 \times 21.5 \times 10^{-6}/0.030 = 31$, well above the ejection limit of 4; as a check, $\sqrt{31}/15.5 = 0.36 = Oh$. (c) Water: $Z = 39.6$, so it jets with satellites, which is why inks are not plain water. Glycerol: $Z = 0.029$, far too viscous to jet at all. Diluting the silver ink to $4\,\text{mPa s}$ would give $Z \approx 7$, still printable but nearer the satellite side. ■

From drops to lines

When a drop lands it spreads until it reaches its contact angle, and because it is far smaller than the capillary length it becomes a spherical cap. Keeping its volume, a drop of diameter d_0 in flight lands with a *footprint* of diameter

$$b = d_0 \left[\frac{8}{\tan(\theta/2)\,\left(3 + \tan^2(\theta/2)\right)} \right]^{1/3}, \tag{12.7}$$

which comes from setting the cap's volume equal to the drop's. A drop spreads to 2.1 times its diameter at 30° and to 1.26 times at 90°.

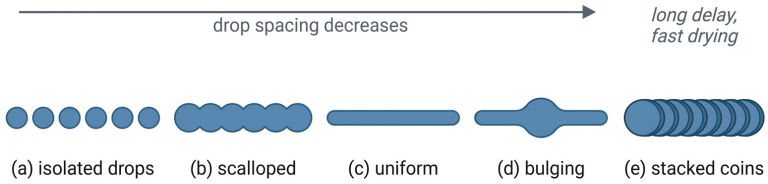


Figure 12.10: Inkjet line shapes, after D. Soltman and V. Subramanian, *Langmuir* (2008). At large drop spacing the drops stay apart; closer spacing first gives a scalloped edge, then a uniform line, and at very small spacing a line with bulges. A long delay between drops on a hot substrate gives a stack of separately dried disks.

The printer places drops at a *drop spacing* p, often quoted as dpi $= 25.4\,\text{mm}/p$ ($25\,\mu\text{m}$ is 1016 dpi). The spacing sets how much liquid each length of line gets, and so the line's shape (Fig. 12.10). Drops farther apart than one footprint stay *isolated* and do not conduct. Closer, they merge, but until there is enough liquid to fill a line one footprint wide the edge stays *scalloped*. Closer still, the line is *uniform*, with parallel edges. At very small spacing the line holds more liquid than it can at a stable angle, and the excess gathers in *bulges* that can short a neighbor. A long delay on a warm substrate lets each drop dry before the next lands, giving *stacked coins*, each with its own coffee ring.

The width of a uniform line follows from volume again.[5] One drop, $\pi d_0^3/6$, fills a length p of line whose cross section is a circular segment of width w and angle θ (in radians):

$$w = \sqrt{\frac{2\pi d_0^3}{3p\,(\theta/\sin^2\theta - \cot\theta)}}. \tag{12.8}$$

The width grows as $1/\sqrt{p}$: halving the spacing doubles the ink per length and widens the line by $\sqrt{2}$. A pinned line cannot be narrower than one footprint, so the largest spacing that still gives parallel edges is the one at which $w = b$:

$$p_{\max} = \frac{2\pi d_0^3}{3b^2\,(\theta/\sin^2\theta - \cot\theta)}. \tag{12.9}$$

Between $p_{\max}$ and b the line is scalloped. The bulging limit $p_{\min}$ depends on printing speed and on how far the contact line can advance, so it is found by experiment (Problem 12.3). In practice you print test lines at several spacings, measure their widths, and choose a spacing in the middle of the uniform range.

[5] J. Stringer and B. Derby, *Langmuir* (2010).

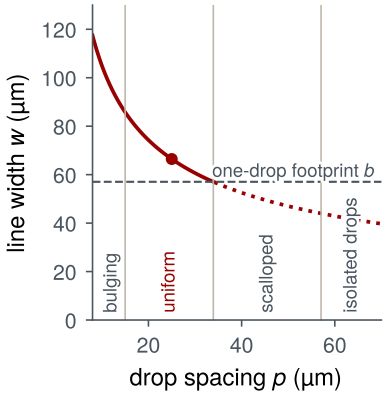


Figure 12.11: Line width versus drop spacing for 10 pL drops at $\theta = 30°$, from Eq. (12.8). The dot is the 25 µm spacing of Example 12.3; the bulging limit (15 µm) is illustrative.

Worked Example 12.3 *From drops to a silver line*

A cartridge jets 10 pL drops of the ink of Example 12.1 onto clean PET ($\theta = 30°$). The ink is 20 wt% silver with a density of 1.5 g/cm^3. (a) Find the drop diameter, its footprint, and $p_{\max}$. (b) Find the line width at $p = 25$ µm. (c) How thick is the silver the line leaves?

Solution. (a) $d_0 = (6 \times 10 \times 10^{-15}\,\mathrm{m}^3/\pi)^{1/3} = 26.7$ µm. With $\tan 15° = 0.268$, Eq. (12.7) gives $b = 26.7 \times [8/(0.268 \times 3.072)]^{1/3} = 57.0$ µm. For $\theta = 0.524$ rad, $\theta/\sin^2\theta - \cot\theta = 2.094 - 1.732 = 0.362$, and Eq. (12.9) gives $p_{\max} = 2\pi(26.7)^3/[3\,(57.0)^2(0.362)] = 33.9$ µm. (b) $w = \sqrt{2\pi(26.7)^3/(3 \times 25 \times 0.362)} = 66.4$ µm, wider than one footprint, as it must be below $p_{\max}$ (the dot in Fig. 12.11). (c) A drop is 15 ng of ink and 3.0 ng of silver, $3.0/10.49 = 286$ µm^3, spread over $pw = 1660$ µm^2 of line: 0.17 µm of dense silver. Inkjet prints thin films; a thick conductor needs several passes or another method. ■

12.6 *Screen and Stencil Printing*

Screen printing is the workhorse of printed sensors: most glucose test strips, many ECG electrodes, and the heater lines in car rear windows are screen printed. It is simple, it prints thick, highly conductive films from viscous pastes, and a screen costs little.

A screen is a woven polyester or steel mesh stretched on a frame and coated with a light-sensitive *emulsion*, which is exposed through a

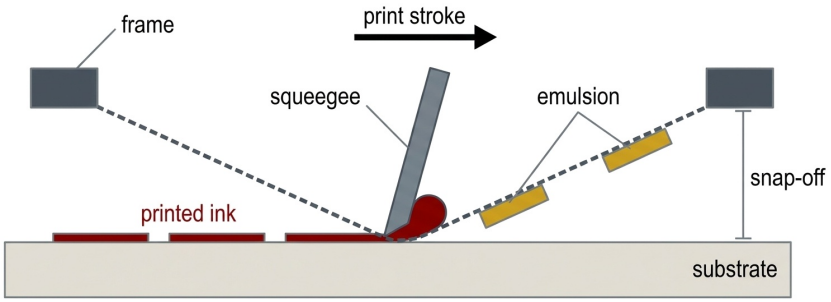


Figure 12.12: Screen printing. The squeegee presses the mesh down to the substrate and pushes a roll of paste ahead of it. Paste passes through the open mesh where the emulsion has been removed and stays on the substrate when the mesh snaps back up behind the squeegee.

film of the pattern and washed out where ink should pass. The screen is held 1 to 3 mm above the substrate (the *snap-off*, Fig. 12.12). A rubber squeegee presses the mesh onto the substrate and drags a rolling bead of paste along. The paste thins under the shear (Section 12.3), fills the open mesh, and touches the substrate; behind the squeegee the mesh peels away and the paste stays. Pastes run from about 1 to 100 Pa s; features go down to about 50 µm on good presses and 100 µm in a typical lab, in films 3 to 25 µm thick per pass. *Rotary* screens, cylinders with the squeegee inside, print roll to roll.

How thick is the print? A mesh is named by its *mesh count* M (threads per inch) and its thread diameter d (Fig. 12.13). The opening is $w = 25.4\,\text{mm}/M - d$, the fraction of the screen that is open is the *open area* $\text{OA} = [w/(w+d)]^2$, and a plain weave is about $2d$ thick. Paste fills the open part of the mesh, so the mesh holds about $\text{OA} \times t_{\text{mesh}}$ of paste per area, and any emulsion standing above the mesh (the *emulsion over mesh*, EOM) adds its own thickness. When the solvent evaporates, only the solids remain:

$$t_{\text{wet}} \approx \text{OA}\, t_{\text{mesh}} + t_{\text{EOM}}, \qquad t_{\text{dry}} = \phi_v\, t_{\text{wet}}, \tag{12.10}$$

with ϕ_v the volume fraction of solids in the paste. Some paste stays in the mesh, so the real print is a little thinner. A line must span at least two or three mesh openings to print reliably, so fine features need fine meshes, which give thinner films.

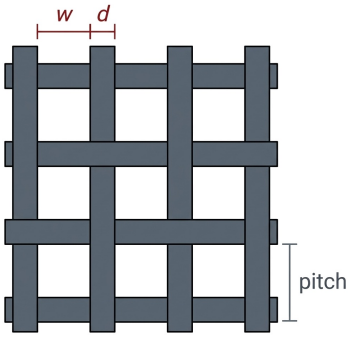


Figure 12.13: A woven mesh seen from above. The thread diameter d and the opening w add up to the pitch, $25.4\,\text{mm}/M$ for a mesh count M in threads per inch.

Worked Example 12.4 *Wet and dry thickness of a screen-printed silver film*

A silver flake paste (65 wt% silver at 10.49 g/cm^3, 10 wt% binder at 1.2 g/cm^3, 25 wt% solvent at 1.0 g/cm^3) is printed through a 325-mesh steel screen with 28 µm wire and 10 µm of emulsion over mesh. (a) Find the open area and the wet thickness. (b) Find the dry thickness. (c) For a cured resistivity of 30 µΩ cm, find the sheet resistance.

SOLUTION. (a) The pitch is $25.4/325 = 78.2\,\mu m$, so $w = 50.2\,\mu m$, $OA = (50.2/78.2)^2 = 0.412$, and with $t_{mesh} \approx 56\,\mu m$, $t_{wet} \approx 0.412 \times 56 + 10 = 33.1\,\mu m$. (b) In 100 g of paste the volumes are 6.20 (silver), 8.33 (binder), and 25.0 cm^3 (solvent), so $\phi_v = 14.5/39.5 = 0.368$ and $t_{dry} = 12.2\,\mu m$. Silver is two-thirds of the paste by mass but only a sixth of it by volume. (c) The sheet resistance, the resistance of any square of the film (Chapter 11), is $R_s = \rho/t = 30 \times 10^{-8}/12.2 \times 10^{-6} = 0.0246\,\Omega$, about $25\,m\Omega$ per square. The resistivity is about 20 times that of bulk silver ($1.59\,\mu\Omega\,cm$), typical of flakes that touch through thin polymer gaps. ■

A *stencil* is a solid sheet with the pattern cut through it: laser-cut steel, polyimide, or the adhesive vinyl of Lab 2. Paste wiped across it fills the openings and stays when the stencil lifts. With no mesh in the way the print is as thick as the stencil, 25 to 200 µm, with sharp walls. Two design rules follow. A stencil cannot hold an *island*, such as the middle of an O or the gap around the center disk of a ring electrode, which falls out when the stencil is cut. And paste releases only if it sticks to the substrate more than to the opening's walls; solder-paste practice asks for an *area ratio* (opening area over wall area, $W/(2t)$ for a long slot of width W in a sheet t thick) above 0.66 and an *aspect ratio* W/t above 1.5.

12.7 *Roll-Based Printing and Coating*

ROLL-BASED PRESSES print a moving web of plastic or paper, up to meters wide, at up to about 10 m/s. They are the presses of newspapers and food packaging, and they make sense for electronics once a design is fixed and thousands of copies are needed.

In *gravure* the pattern is a field of tiny *cells* engraved in a chrome-plated copper roll, which resists wear and solvents and repeats the pattern millions of times. The roll turns through an ink supply, a steel *doctor blade* wipes its surface so ink stays only in the cells, and in the *nip*, the line where the web is pressed against the roll by a rubber-covered *impression roll*, the web pulls ink out of the cells (Fig. 12.14). The dots from neighboring cells then spread and merge into lines and areas, and the web passes a dryer before the next station prints the next layer. Inks run about 10 to 1000 mPa s, and rolls made with microfabrication print lines a few micrometers wide.[6]

Every step is a contest between viscous forces and surface tension, so each depends on the capillary number $Ca = \eta U/\gamma$ at the press speed U (Eq. (12.5)). As Ca rises, cells fill less completely and the doctor blade leaves a thicker residue on the *lands* between cells, but

[6] G. Grau, *et al.*, *Flex. Print. Electron.* (2016).

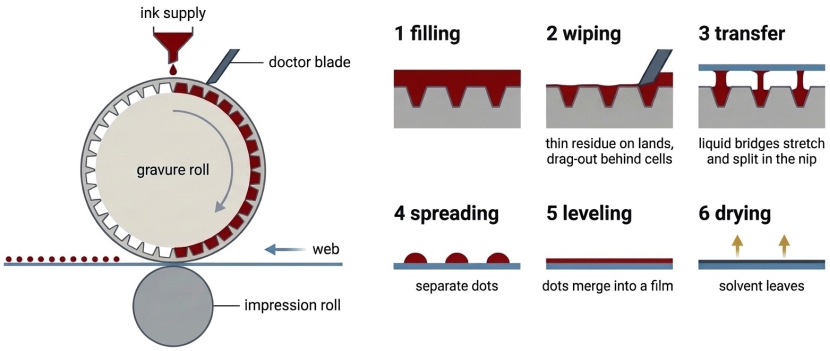


Figure 12.14: Gravure printing. Left: the engraved roll picks up ink, the doctor blade wipes its surface, and in the nip with the impression roll the moving web pulls ink out of the cells. Right: the six steps. Each depends on the capillary number $Ca = \eta U/\gamma$ (after R. Kitsomboonloha, *et al.*, *Langmuir* (2012)).

it drags less ink out of each cell. In the nip a liquid bridge between cell and web stretches and splits, and the fraction that goes to the web, often about half, depends on Ca and on how well the ink wets the web compared with the roll. So ink and press speed are chosen together. The film thickness follows from the cells. An inverted-pyramid cell 20 µm wide and 8 µm deep holds $\frac{1}{3} \times 20^2 \times 8 = 1067\,\mu\text{m}^3$; on a 25 µm grid that is the equivalent of a 1.71 µm layer, and if half transfers the wet film is 0.85 µm, or 85 nm dry at 10 % solids. Gravure prints thin. Two other presses complete the family. *Flexography* prints from raised relief on a soft plate, like a rubber stamp, with features of about 50 µm. *Offset* printing transfers the pattern from a plate to a rubber *blanket* and then to the substrate.

Many layers need no pattern at all: a dielectric over an array, the active layer of a photodiode, an encapsulation. *Slot-die coating* pumps ink at a volume rate Q through a slit of width W just above a web moving at U. Everything pumped lands, so

$$t_{\text{wet}} = \frac{Q}{WU}. \tag{12.11}$$

A 50 mm slot die fed 0.50 mL/min onto a web moving at 1.0 m/min lays a 10 µm wet film; at 1 % solids, as for a PEDOT:PSS dispersion, that dries to 100 nm. *Blade coating* drags a blade over a pool of ink at a fixed gap (the wet film is roughly half the gap); it made the polymer LEDs and photodiodes of the printed oximeter array of Section 12.2. Spin coating, the lab standard, wastes most of the ink and does not scale to rolls.

12.8 Direct Writing: Dispensers, Aerosol Jet, and Liquid Metal

Some features are too thick for inkjet, too curved for a screen, or made of a paste no other printer can move. Direct-write printers

extrude or spray ink from a nozzle that a three-axis stage moves along a programmed path. Like inkjet they are digital.

A *dispenser* pushes paste of thousands to millions of mPa s out of a syringe needle as a continuous bead. Needles of 100 to 400 µm limit features to above about 100 µm. In return the beads are tens to hundreds of micrometers thick, the needle follows curved surfaces, and the same machine prints hydrogels. Past project teams printed silver electrodes on thermoplastic polyurethane (TPU) this way (Chapter 14). Whatever flow rate Q the syringe delivers, the bead's cross section at stage speed v is

$$A = \frac{Q}{v} \approx \tfrac{2}{3}\, wh, \tag{12.12}$$

where the last form is for a flattened bead of width w and height h. A bead 400 µm wide and 60 µm high at 5 mm/s needs about 0.08 mm^3/s. If the stage speeds up and the flow does not, the bead thins and finally breaks up into separate drops, so speed and flow change together.

An *aerosol jet* turns ink into a mist of droplets 1 to 5 µm across and focuses it with a surrounding *sheath* of gas into a jet that writes lines about 10 µm wide from millimeters away, over steps and curves.

Liquid metal is the most stretchable conductor. The gallium-indium alloy EGaIn melts at about 16 °C, flows like water, and conducts at 3.4×10^6 S/m, about a seventeenth of copper. In air it grows an oxide skin a few nanometers thick that holds its shape, so it can be dispensed into lines or injected into silicone channels, where it stays. A stretched trace keeps its volume, so, as Section 11.7 showed, $R/R_0 = (1+\varepsilon)^2$. An EGaIn trace 0.50 mm × 50 µm in cross section and 100 mm long has 1.18 Ω, and 2.65 Ω at 50 % strain; a cured silver-flake trace of the same shape has about 1.2 Ω but cracks at a few percent strain. Keep gallium away from aluminum, which it attacks.

12.9 *Choosing a Printing Method*

NO METHOD WINS ON EVERY COUNT. Table 12.2 gathers typical numbers; sources disagree by up to a factor of ten, because a lab printer and an industrial line are different machines. Roll-based presses cover square meters per second, while digital nozzle printers are slower by orders of magnitude but need no mask. To choose, ask five questions:

1. How small is the smallest feature, and how thick must the film be?
2. How viscous is the ink? Viscosity alone often decides: a 10 Pa s paste is a thousand times too thick for an inkjet head.
3. How many copies? A few, with the design still changing: a digital method or a stencil. Thousands of meters of one design: a press.

4. Is the substrate rough, curved, or fragile? Then choose a non-contact method.
5. How many layers must align, and will each solvent spare the layer below?

Method	Smallest feature (µm)	Ink viscosity (mPa s)	Dry thickness per pass (µm)	Speed (m/min)	Pattern carrier and its cost	Best for
Inkjet	20 to 100	2 to 20	0.01 to 1	0.02 to 5	none (a file)	prototypes, thin films, many materials on one sheet
Aerosol jet	10 to 100	1 to 1000	0.1 to 5	slow	none (a file)	fine lines on curved or stepped parts
Dispenser	100 to 1000	10^3 to 10^6	5 to 500	slow	none (a file)	thick traces, pastes, 3D parts, liquid metal
Screen	50 to 200	10^3 to 10^5	3 to 30	0.6 to 100	mesh screen, low	electrodes, sensors, thick conductors
Stencil	100 to 500	10^4 to 10^6	25 to 200	batch	foil or vinyl, very low	lab electrodes, solder paste
Gravure	5 to 200	10 to 1000	0.02 to 12	8 to 100	engraved roll, high	high-volume thin layers
Flexography	30 to 80	50 to 500	0.2 to 8	5 to 180	relief plate, low to medium	medium runs, conductors, dielectrics
Offset	10 to 50	10^4 to 10^5	0.5 to 2	0.6 to 15	plate and blanket, medium	fine lines in high volume
Slot-die, blade	no pattern (stripes)	2 to 5000	0.1 to 60	0.6 to 5	none	uniform layers over large areas

Table 12.2: Typical ranges for printing methods (after S. Khan, L. Lorenzelli, and R. S. Dahiya, *IEEE Sens. J.* (2015), and M. Hambsch, *et al.*, *Mater. Sci. Eng. B* (2010)). Digital methods need no pattern carrier and suit prototyping; roll-based methods need an expensive master and suit production.

The answers point to a short list of typical choices. For a prototype whose design still changes, print thin, fine lines by inkjet and thick pastes through a stencil or from a dispenser. For thick, low-resistance electrodes in volume, such as ECG electrodes and glucose strips, use a screen. For a uniform layer with no pattern, use a slot die or a blade. For thin patterned layers by the kilometer, use gravure or flexography. For curved or stepped surfaces, use an aerosol jet or a dispenser, and for stretchable wiring, a dispenser with a stretchable paste or liquid metal. Thickness often settles the rest: a screen lays down about 12 µm of dry silver per pass (Worked Example 12.4), an inkjet line less than a fifth of a micrometer (Example 12.3).

12.10 *Edges, Variation, and Registration*

DEVICE GEOMETRY SETS DEVICE PERFORMANCE: an electrode's area sets its impedance, an antenna's length its resonance, a transistor's channel length its current. Printing sets the geometry, so it decides how much two nominally identical devices differ. Two effects dominate: rough edges and misaligned layers.

A printed edge is never straight. Its *line edge roughness* (LER) is the standard deviation of the edge's position about its mean line (Fig. 12.15), and the wiggles have a typical length, the *correlation length* ξ: about one drop for inkjet, one mesh opening for a screen. A printed transistor shows how roughness becomes variation. Its current goes as $1/L$, the gap between the printed source and drain (Chapter 11). With two rough edges the gap varies along the channel by $\sigma_L = \sqrt{2}\,\mathrm{LER}$. A transistor of width W acts like $N = W/\xi$ short pieces side by side whose errors partly cancel, so its current varies by about

$$\frac{\sigma_I}{I} \approx \frac{\sigma_L}{L}\,\frac{1}{\sqrt{N}}. \tag{12.13}$$

For $L = 20\,\mu\mathrm{m}$, an LER of $1.5\,\mu\mathrm{m}$, and $\xi = 10\,\mu\mathrm{m}$, $\sigma_L/L = 10.6\,\%$; a transistor $50\,\mu\mathrm{m}$ wide ($N = 5$) varies by about 4.7 %, one 1 mm wide ($N = 100$) by about 1.1 %. Short, narrow devices suffer most. Printed sensor arrays, whose pixels never match exactly, need per-pixel calibration or a differential design (Chapter 4).

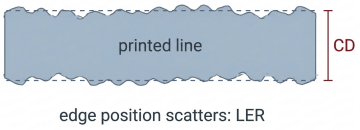


Figure 12.15: Line edge roughness. The critical dimension (CD) is the nominal width of the smallest feature; LER is the standard deviation of each edge's position about its mean line.

A device built from several printed layers works only if each lands on the last; the placement error is the *registration* error. On plastic the substrate itself moves between layers: it expands with temperature and humidity, and oriented films shrink permanently the first time they are heated (Section 11.3). A 100 mm PET sheet that shrinks 0.15 % when its first silver layer is cured moves its far edge by $150\,\mu\mathrm{m}$, and printing the second layer on a platen 35 K warmer (PET expands by about $17\,\mathrm{ppm/K}$) moves it back by $60\,\mu\mathrm{m}$. That is harmless for the 10 mm pads of Lab 2 and fatal for a $50\,\mu\mathrm{m}$ transistor channel. The cures: preshrink the film with a bake before the first print, print every layer at the same temperature and humidity, align each layer on printed *fiducial marks* (the corner crosses in the stencil artwork of Lab 2), and scale later layers by any measured change. Each solvent must also spare the layers below; a solvent that does not dissolve the layer under it is called *orthogonal* (Chapter 11). Hansen's solubility parameters can find one before any trial.[7] For printed transistors with a PMMA gate dielectric over the polymer semiconductor PBTTT, one study used them to pick solvents that dissolve PMMA and spare PBTTT, while solvents that were not orthogonal swelled the semiconductor and roughened its surface.

[7] A. M. Gaikwad, Y. Khan, A. E. Ostfeld, *et al.*, A. C. Arias, *Org. Electron.* (2016).

12.11 *Printed Electrodes and Sensors on the Body*

The printed device that touches the body most often is an electrode. Disposable ECG electrodes, glucose test strips, and many research patches for EEG and EMG are printed. The standard skin

electrode is silver coated with silver chloride, whose reaction $Ag + Cl^- \rightleftharpoons AgCl + e^-$ passes current into the body's chloride-rich fluids with a stable potential and low noise (Chapter 6 explained the interface, Chapter 8 the recording). A printed electrode is a stack (Fig. 12.16): a silver trace and pad, an Ag/AgCl layer on the pad, a printed dielectric that covers the trace, and a gel or adhesive hydrogel against the skin.

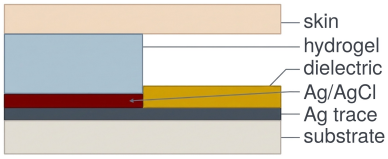


Figure 12.16: A printed biopotential electrode in cross section (not to scale). The dielectric leaves only the Ag/AgCl pad exposed, and the hydrogel connects the pad to the skin.

The Ag/AgCl layer can be printed from an ink of silver flakes and silver chloride particles in a binder. Or a printed silver pad can be chlorided chemically (a brief dip in bleach) or electrochemically, as the anode in a chloride solution. Then every electron that flows converts one silver atom into AgCl, so a charge density q (in C/cm^2) grows a layer

$$t_{AgCl} = \frac{q\,V_m}{F} \tag{12.14}$$

thick, with $F = 96\,485\,C/mol$ the charge of a mole of electrons and $V_m = 25.8\,cm^3/mol$ the molar volume of AgCl ($143.3\,g/mol$ at $5.56\,g/cm^3$): $2.67\,\mu m$ per C/cm^2, if all the charge makes a dense layer.

Worked Example 12.5 *Chloriding a printed silver electrode*

A $2.0\,cm^2$ screen-printed silver pad is made the anode in 0.9 % NaCl at $1.0\,mA/cm^2$ for 5.0 min. (a) How thick is the AgCl? (b) How much silver does it consume? (c) The pad's trace is $1.0\,mm \times 150\,mm$ and $12\,\mu m$ thick at $30\,\mu\Omega\,cm$. Does its resistance matter next to a skin contact of $10\,k\Omega$?

Solution. (a) $q = 1.0\,mA/cm^2 \times 300\,s = 0.30\,C/cm^2$, so $t = 0.30 \times 25.8/96485 = 8.0 \times 10^{-5}\,cm = 0.80\,\mu m$, at a total current of 2.0 mA. (b) Silver's molar volume is $107.87/10.49 = 10.3\,cm^3/mol$, so each micrometer of AgCl consumes $10.3/25.8 = 0.40\,\mu m$ of silver: $0.32\,\mu m$ of the $12\,\mu m$ film. (c) $R = \rho L/(wt) = 3.0 \times 10^{-7} \times 0.150/(1.0 \times 10^{-3} \times 12 \times 10^{-6}) = 3.75\,\Omega$, less than a thousandth of the skin contact: irrelevant for an ECG electrode, though not for a heater or a strain gauge. ■

Printing lets an electrode take any shape on any substrate, and prints on TPU or textiles stretch with the body. Smaller electrodes have higher impedance, but a printed trace of a few ohms is negligible next to a skin contact of kilohms, as Example 12.5 showed. Printing also makes sensors that are more than electrodes. A pulse oximeter made of solution-processed organic LEDs and photodiodes (the photodiodes blade-coated on plastic) measured heart rate and oxygen saturation within 1 and 2 percent of a commercial probe.[8] Inkjet-printed gold electrode arrays measured the impedance of the skin and detected the tissue damage that leads to pressure ulcers before it was visible.[9] Fig-

[8] C. M. Lochner, Y. Khan, A. Pierre, and A. C. Arias, *Nat. Commun.* (2014).

[9] S. L. Swisher, M. C. Lin, A. Liao, *et al.*, *Nat. Commun.* (2015); Y. Khan, F. J. Pavinatto, M. C. Lin, *et al.*, *Adv. Funct. Mater.* (2016).

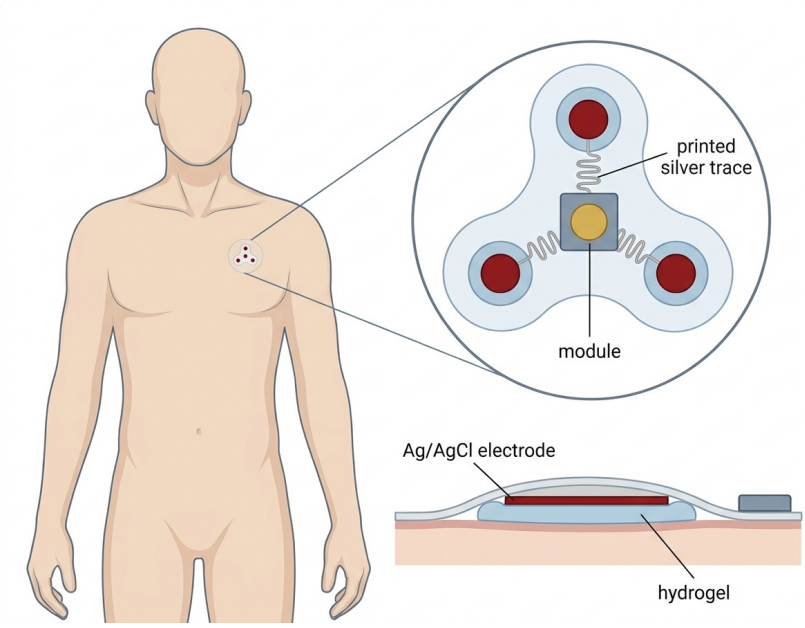


Figure 12.17: A printed ECG patch on the chest. Silver traces on a thin stretchable film connect three Ag/AgCl electrode pads, each covered with hydrogel, to a small rigid module that holds the amplifier and radio. The circle enlarges the patch; the strip below is a section through one electrode.

ure 12.17 shows the system these pieces build: printed electrodes and interconnects on a soft film, with a few silicon chips for amplification and radio.

Lab 2 puts this chapter on the bench: you stencil-print silver and Ag/AgCl, measure your lines and electrodes, and build a 4×4 capacitive pressure array like the one in the project's sensor glove.

Summary

- Printing deposits and patterns in one additive step, over large areas, below about 150 °C, and (for digital methods) straight from a file.
- Young's equation, $\cos\theta_c = (\gamma_{SG} - \gamma_{SL})/\gamma_{LG}$, sets the contact angle; print at roughly 20 to 60°. Owens-Wendt, $\gamma_L(1+\cos\theta) = 2(\sqrt{\gamma_S^d \gamma_L^d} + \sqrt{\gamma_S^p \gamma_L^p})$, predicts it from the dispersive and polar parts; plasma raises the polar part.
- Viscosity $\eta = \tau/\dot{\gamma}$, with $\dot{\gamma} \approx U/h$. Inks should be shear thinning, $\eta = K\dot{\gamma}^{n-1}$ with $n < 1$. Typical viscosities: inkjet 2 to 20 mPa s, screen 1 to 100 Pa s. Printed drops are smaller than the capillary length, so they are spherical caps; $Ca = \eta U/\gamma$ balances viscosity and surface tension.
- A pinned edge that evaporates fastest makes the coffee ring; a cosolvent with a higher boiling point and lower surface tension, or a heated substrate, counters it.

- Inkjet: $Z = \sqrt{\rho\gamma d}/\eta$ between about 1 and 10; a uniform line has width $w = \sqrt{2\pi d_0^3/(3p(\theta/\sin^2\theta - \cot\theta))}$ and is a fraction of a micrometer thick.
- Screen: $t_{\text{wet}} \approx \text{OA}\, t_{\text{mesh}} + t_{\text{EOM}}$, $t_{\text{dry}} = \phi_v t_{\text{wet}}$, about 12 µm of silver per pass. Stencils print thicker and need an area ratio above 0.66. Gravure: every step set by Ca; slot die: $t_{\text{wet}} = Q/(WU)$; dispenser: $A = Q/v$.
- Choose by feature size and thickness, ink viscosity, number of copies, substrate, and layer count: inkjet or stencil for prototypes, screen for thick electrodes in volume, slot die for uniform layers, gravure for volume.
- Rough edges spread device behavior as $(\sqrt{2}\,\text{LER}/L)/\sqrt{W/\xi}$; plastic substrates move between layers, so pre-shrink, print at one temperature, and align on fiducials.
- A printed Ag/AgCl electrode is a silver trace, an Ag/AgCl pad, a dielectric, and a gel; chloriding grows 2.67 µm of AgCl per C/cm^2.

Chapter 13 asks where a printed patch gets its energy.

12.12 *Problems*

12.1 • A screen-printing paste follows the power law with $K = 80\,\text{Pa s}^n$ and $n = 0.35$.

(a) Find its viscosity at rest ($\dot\gamma = 0.05\,\text{s}^{-1}$) and under a squeegee moving at 150 mm/s that forces it through 40 µm mesh openings.

(b) Find the shear stress in the mesh.

(c) The paste is also thixotropic. What goes wrong if it recovers its viscosity too slowly after printing? Too quickly?

12.2 ◦ Three inks are printed from a 30 µm nozzle: (i) a PEDOT:PSS ink with $\eta = 12\,\text{mPa s}$, $\gamma = 35\,\text{mN/m}$, $\rho = 1050\,\text{kg/m}^3$; (ii) a dielectric ink with 60 mPa s, 30 mN/m, $1100\,\text{kg/m}^3$; (iii) an alcohol-based ink with 2.0 mPa s, 22 mN/m, $800\,\text{kg/m}^3$.

(a) Find Z for each and predict its behavior.

(b) Heating the head lowers the viscosity of ink (ii) to 25 mPa s. Is it now jettable?

12.3 •• A 1 pL drop lands at $\theta = 45°$.

(a) Find its footprint and $p_{\max}$.

(b) Find the uniform-line width at $p = 10\,\mu\text{m}$, and that spacing in dpi.

(c) How long does a 20 mm line take at 5 kHz?

(d) Describe an experiment that finds $p_{\min}$, and what you would measure.

12.4 • Stencils.

(a) A 90 µm vinyl stencil has 0.5 mm slots. Find the area ratio and the aspect ratio. Will the paste release?

(b) A 150 µm steel foil has 0.20 mm slots. Repeat.

(c) A ring electrode (a 10 mm disk inside a ring 2 mm wide) is to be stencil-printed with the disk and ring separate. What goes wrong, and how do you fix the stencil?

12.5 • Choose a printing method for each device and give one reason: (a) 10 000 disposable ECG electrodes; (b) a prototype with 30 µm interdigitated gold fingers; (c) an antenna on a curved 3D-printed earpiece; (d) a uniform 100 nm PEDOT:PSS layer over $1\,\text{m}^2$; (e) stretchable interconnects for a glove prototype.

13
Powering Bioelectronics

In this chapter you will learn to

- Explain energy, power, battery capacity, and efficiency, and convert between joules, watt-hours, and milliamp-hours.
- Build the power budget of a wearable from peak currents and duty cycles, and estimate its battery life.
- Compare light, body heat, motion, and radio waves as energy sources for a wearable, and size a solar cell for a load.
- Describe how a lithium-ion cell stores charge, and compute its capacity, energy, and C-rate.
- Choose a battery for an implant and compute the battery a neural stimulator needs.
- Explain inductive power transfer by magnetic coupling and resonance, and compute the efficiency of a coil link from k and Q.
- Compute the specific absorption rate and the heating it causes, and check it against the safety limits.
- Explain when ultrasound powers an implant better than coils do, and how ingestible capsules are powered.

A pacemaker smaller than a matchbox keeps a heart beating for about ten years on one battery that no one ever touches. A cochlear implant has no battery at all: a coil held to the scalp by a magnet sends it power through the skin, all day. A smart patch on the chest, built with the same chips as a fitness band, may need a recharge every night. The difference lies less in the electronics than in the energy: how much each device needs, where it comes from, and how safely it can be delivered through tissue.

Chapters 11 and 12 showed how a soft device is made; this chapter

powers it. Chapter 1 met battery capacity and the duty-cycled power budget in a few lines each. Here we follow the energy from source to load in detail. Section 13.1 reviews energy, power, and capacity and adds efficiency, and Section 13.2 uses them to build a full power budget. Section 13.3 asks how much a wearable can harvest from light, heat, and motion. Sections 13.4 and 13.5 open the lithium-ion cell and the batteries of implants. Sections 13.6 to 13.8 send power through the skin, with coils and with ultrasound, under the limit that tissue heating sets. The last section powers a pill.

13.1 Energy, Power, and Efficiency

ENERGY is what a battery stores and a circuit uses up. It is measured in joules (J). *Power* is the rate at which energy is used, in watts: $1\,\mathrm{W} = 1\,\mathrm{J/s}$. A load at voltage V that draws a current I uses the power $P = VI$ (Chapter 1), and a load that draws a steady power P for a time t uses the energy $E = Pt$. Power says how hard a device works; energy says how long a battery can keep it working.

KEY IDEA Power is a rate, energy a total. Battery life is the energy stored divided by the average power, or the capacity divided by the average current.

Biology gives a useful scale. A resting adult turns about 2000 kilocalories of food a day into heat and work, which is $(2000\,\mathrm{kcal})(4184\,\mathrm{J/kcal})/(86\,400\,\mathrm{s}) = 97\,\mathrm{W}$, about 100 W. A phone in use draws around a watt, a wearable patch about a milliwatt, and a pacemaker a few tens of microwatts, less than a millionth of the power of the body that carries it.

Battery capacity

As Chapter 1 showed, a battery is rated by the charge it can deliver, its *capacity*, in milliamp-hours (mAh): a 100 mAh cell can supply 1 mA for 100 hours, or 10 mA for 10 hours. Since $1\,\mathrm{A} = 1\,\mathrm{C/s}$, one milliamp-hour is $(10^{-3}\,\mathrm{A})(3600\,\mathrm{s}) = 3.6\,\mathrm{C}$. The energy is the charge times the voltage at which it is delivered,

$$E\ [\mathrm{mWh}] = C_{\mathrm{bat}}\ [\mathrm{mAh}] \times V\ [\mathrm{V}], \qquad 1\,\mathrm{mWh} = 3.6\,\mathrm{J}. \tag{13.1}$$

The CR2032 lithium coin cell of many wearables, about 225 mAh at 3.0 V, stores 675 mWh, or 2.4 kJ. That is the food energy of about a seventh of a gram of sugar. Batteries store far less energy per gram than food does, which is why the rest of this chapter counts microamps.

Efficiency

No conversion is perfect. The *efficiency* of a stage is the fraction of its input power that reaches its output, $\eta = P_{\mathrm{out}}/P_{\mathrm{in}}$; the rest becomes

heat. When energy passes through several stages, the efficiencies multiply:

$$\eta_{\text{total}} = \eta_1 \, \eta_2 \, \eta_3 \cdots . \tag{13.2}$$

A wireless link that is 30 % efficient, feeding a rectifier at 85 % and a charger at 90 %, delivers only 23 % of the transmitted power to the battery. In an implant the lost power heats the tissue around it (Section 13.7).

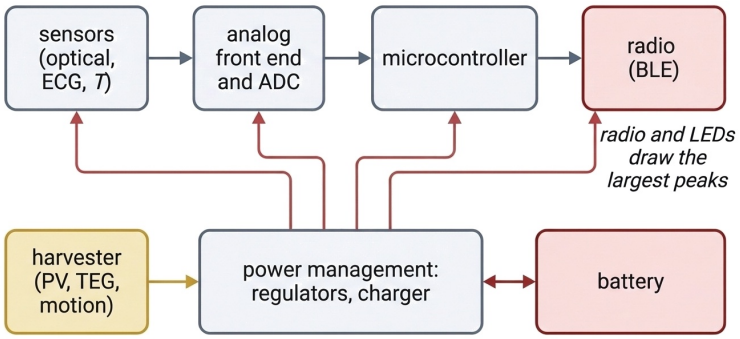


Figure 13.1: Blocks of a wearable sensing system and the flow of power (cardinal). The harvester, when there is one, charges the battery through the power management circuit; the battery covers the peaks that the harvester cannot.

The first stage after any battery is a voltage regulator: a lithium-ion cell falls from 4.2 V to about 3.0 V as it discharges, while the chips need a steady 1.8 V or 3.0 V. A *linear regulator* drops the extra voltage across a transistor that acts as a variable series resistor. The battery current equals the load current, so the efficiency is $V_{\text{out}}/V_{\text{in}}$: 1.8 V made from 3.7 V is only 49 % efficient. A *switching converter* turns a transistor switch on and off at hundreds of kilohertz to a few megahertz and smooths the result with an inductor and a capacitor. It trades voltage for current at 80 to 95 % efficiency, so the battery supplies less current than the load draws:

$$I_{\text{bat}} = \frac{V_{\text{out}} \, I_{\text{load}}}{\eta \, V_{\text{bat}}}. \tag{13.3}$$

A 1.8 V, 1 mA load takes 1 mA from a 3.7 V cell through a linear regulator but only $(1.8)(1)/[(0.90)(3.7)] = 0.54$ mA through a 90 % converter.

13.2 *The Power Budget of a Wearable*

Every sensing system of this book has the same blocks (Fig. 13.1): sensors and their analog front end (Chapters 4 and 8), a microcontroller that processes and stores data, and a radio that sends it. Power reaches them from a battery through a power management circuit: the regulators above and, if there is a harvester, a charger that moves the harvested energy into the battery. A *power budget* lists how much current each block draws and for how long. It is the first calculation of any design, and the design project of Chapter 14 starts with one.

Block	Peak (mA)	Duty cycle	Average (µA)
Red and IR LEDs	20	0.5 %	100
Optical AFE (TIA, ADC)	0.6	10 %	60
Temperature sensor	0.015	1 %	0.15
MCU, active	3.0	2 %	60
MCU, sleep	0.002	98 %	1.96
BLE radio	6.0	0.5 %	30
Regulator and clock	0.003	100 %	3
Total			255

Table 13.1: Power budget of an illustrative wearable PPG and temperature patch on a 3.0 V supply. The peak currents and duty cycles are typical of current low-power parts, not of one product.

Average current and duty cycle

Most blocks are on only part of the time. As in Chapter 1, a block that draws a peak current I_i for a fraction D_i of the time (its *duty cycle*) and nearly nothing otherwise contributes $D_i I_i$ to the average. The battery sees

$$I_{\text{avg}} = \sum_i D_i I_i, \qquad t_{\text{life}} = \frac{C_{\text{bat}}}{I_{\text{avg}}}, \tag{13.4}$$

where C_{bat} is the capacity. Table 13.1 is the budget of a typical patch. Across wearable vital-sign monitors the pattern is the same: the optical probes and the radios take most of the power.[1]

KEY EQUATION

$I_{\text{avg}} = \sum_i D_i I_i, \quad t = C_{\text{bat}} / I_{\text{avg}}$

Average current and battery life; C_{bat} in mAh.

[1] Y. Khan, A. E. Ostfeld, C. M. Lochner, A. Pierre, A. C. Arias, *Adv. Mater.* (2016), a review that tabulates the power needs of wearable monitors and the harvesting options.

Worked Example 13.1 *Budget and battery life of a patch*

(a) Find the average current and power of the patch of Table 13.1. (b) How long does a CR2032 coin cell of about 225 mAh last? (c) Which change helps more: halving the LED duty cycle, or halving the radio duty cycle?

Solution. (a) Summing $D_i I_i$: $100 + 60 + 0.15 + 60 + 1.96 + 30 + 3 = 255\,\mu\text{A}$, or $(255\,\mu\text{A})(3.0\,\text{V}) = 0.77\,\text{mW}$. (b) $t = 225\,\text{mAh}/0.255\,\text{mA} = 882\,\text{h}$, about 37 days (less in practice: a coin cell delivers less than its rated capacity at high pulse currents and low temperature). (c) Halving the LED duty cycle saves 50 µA, halving the radio's saves 15 µA. The LEDs are the larger target here; the photocurrent budget of Chapter 7 shows that they can usually be pulsed even shorter. ■

Life is inversely proportional to the duty cycle only while the active current dominates. A small wireless sensor whose battery lasts 9 h with its radio on all the time lasts 18 h at a 50 % duty cycle and 36 h at 25 %. If the sleep current is a fraction r of the active current, the life becomes $t(D) = t(1)/[D + (1 - D)r]$, which levels off at $t(1)/r$ for small duty cycles (Fig. 13.2). Below that point, cutting the sleep current helps more than shortening the radio's bursts.

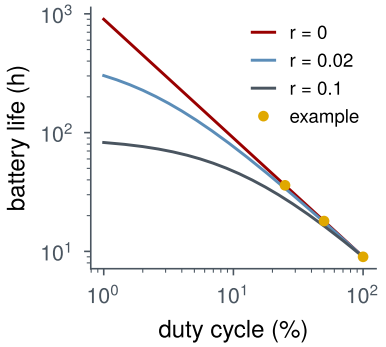


Figure 13.2: Battery life versus duty cycle for sleep-to-active current ratios r, scaled to 9 h at 100 %. Gold dots: 9, 18, and 36 h at 100, 50, and 25 %.

Peaks and the buffer capacitor

A battery also has an internal resistance, and a peak current drops the voltage across it. A fresh CR2032 has roughly 15 Ω, more when cold or nearly empty, so the 20 mA LED pulse of Table 13.1 pulls its voltage down by 0.30 V, enough to reset a microcontroller running near its minimum voltage. A capacitor across the cell supplies the peaks instead and is recharged slowly between them. From $Q = CV$ (Chapter 1), a capacitor that delivers a current I for a time t sags by $\Delta V = It/C$: a 100 µF capacitor sags 0.18 V during a 3 ms, 6 mA radio burst. The buffer is sized for the longest burst.

13.3 *Energy Harvesting for Wearables*

A HARVESTER turns energy that is already present at the body into electricity. Four sources are practical for wearables: light, body heat, body motion, and ambient radio waves. Figure 13.3 compares them. The spread is enormous: sunlight on a solar cell gives a thousand times more power per square centimeter than body heat, but only when the wearer is outside and the cell faces the sky. Light and heat give DC; motion and radio waves give AC that must be rectified.

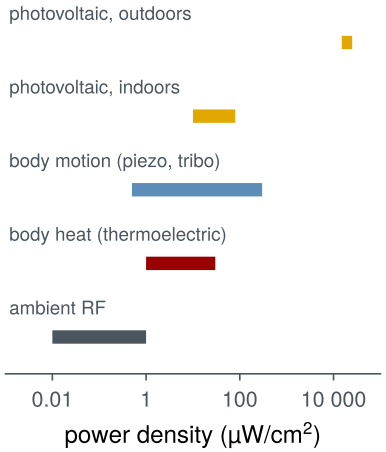


Figure 13.3: Orders of magnitude of the power density that wearable harvesters deliver, after the review of Khan *et al.* (2016), with the photovoltaic values converted from incident to harvested power.

Light

A solar cell under full sun receives about 100 mW/cm² (1000 W/m²). A cell that converts 15 to 25 % of it delivers 15 to 25 mW/cm². Indoors the light is 360 to 1800 times weaker: white light at 200 to 1000 lux carries about 56 to 280 µW/cm²,[2] and cells made for indoor spectra convert 10 to 30 % of it. Flexible amorphous silicon, organic, and perovskite cells can be laminated onto a band or a patch.

> **WORKED EXAMPLE 13.2** *Sizing an indoor solar cell*
>
> A wearable draws 50 µW on average, day and night. Its wearer spends 10 h a day under 300 lux (about 85 µW/cm²), and a flexible cell converts 15 % of that light; the power management circuit passes 80 % of the harvested power to the battery. (a) What cell area is needed? (b) How much battery capacity must carry the 14 h without light, at 3.7 V? (c) How much would an hour outdoors in sunlight give on the same cell?
>
> SOLUTION. (a) The device needs $(50\,\mu W)(24\,h) = 1.2\,mWh$ a day. Each square centimeter delivers $(85\,\mu W)(0.15)(0.80)(10\,h) = 0.102\,mWh$ a day, so the area is $1.2/0.102 = 11.8\,cm^2$, about the size of a large watch strap. (b) $(50\,\mu W)(14\,h) = 0.70\,mWh$, or

[2] M. Jahandar, S. Kim, D. C. Lim, *ChemSusChem* (2021), a review of indoor organic photovoltaics.

$0.70\,\mathrm{mWh}/3.7\,\mathrm{V} = 0.19\,\mathrm{mAh}$; with margin for dark days, a few milliamp-hours. (c) $(100\,\mathrm{mW/cm^2})(0.15)(0.80)(11.8\,\mathrm{cm^2})(1\,\mathrm{h}) = 142\,\mathrm{mWh}$, enough for four months of this load. Outdoors the battery, not the cell, is the limit. ■

Body heat

A *thermoelectric generator* (TEG) turns a temperature difference directly into a voltage, with no moving parts. Heat flowing along a semiconductor leg pushes charge carriers from its hot end to its cold end (the Seebeck effect). A TEG contains N couples of p-type and n-type legs, typically bismuth telluride, connected electrically in series and thermally in parallel between a hot plate and a cold plate. A temperature difference ΔT_{TEG} across the legs gives

$$V_{\mathrm{oc}} = NS\,\Delta T_{\mathrm{TEG}}, \qquad P_{\max} = \frac{V_{\mathrm{oc}}^2}{4R_{\mathrm{el}}}, \tag{13.5}$$

where S is the Seebeck coefficient of one couple (about $400\,\mu\mathrm{V/K}$ for bismuth telluride). $P_{\max}$ is delivered to a load equal to the internal resistance R_{el}, which receives half the open-circuit voltage.

The catch is ΔT_{TEG}. Heat must flow from the body core through tissue, through the TEG, and from a heat sink into the air. Each step has a *thermal resistance*, in kelvin per watt (the temperature difference it takes to push one watt of heat through), and the temperature difference divides among them as a voltage does across resistors in series (Fig. 13.4); the step into still air is the largest. With $R_{\mathrm{body}} = 400\,\mathrm{K/W}$, $R_{\mathrm{TEG}} = 100\,\mathrm{K/W}$, $R_{\mathrm{sink}} = 500\,\mathrm{K/W}$, and 15 K from the core to the air, $\dot{Q} = 15\,\mathrm{K}/1000\,\mathrm{K/W} = 15\,\mathrm{mW}$ of heat flows and the TEG sees only a tenth of the difference, 1.5 K. A $1\,\mathrm{cm^2}$ module with 100 couples of $400\,\mu\mathrm{V/K}$ and $50\,\Omega$ then gives $V_{\mathrm{oc}} = 60\,\mathrm{mV}$ and $P_{\max} = (60\,\mathrm{mV})^2/(200\,\Omega) = 18\,\mu\mathrm{W}$, about 0.1 % of the heat. Walking lowers R_{sink} and raises the power, and a boost converter that starts from tens of millivolts raises the output to a usable voltage.

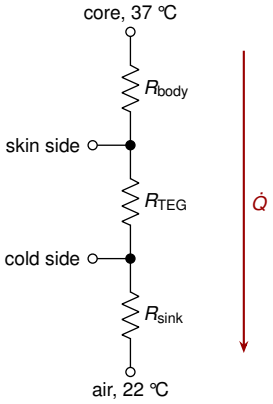


Figure 13.4: Thermal circuit of a TEG on the skin. Heat $\dot{Q}$ flows from the body core to the air through three thermal resistances in series; the temperature difference divides among them like a voltage across a resistor divider, and the TEG gets only its share.

Body motion and radio waves

A *piezoelectric* material, such as the ceramic lead zirconate titanate (PZT) or the polymer PVDF, develops charge on its faces when it is squeezed. A PZT disk of $1\,\mathrm{cm^2}$ in a shoe heel collects about 35 µJ per heel strike at about 200 V: at most 35 µW at one step per second, before conversion losses. A *triboelectric* generator uses the charge that two different materials exchange when they touch (the static charge of a balloon rubbed on hair) and drives a current through a load as they separate. Both give high voltage at tiny charge, so their output

impedance is very high and their output is AC; they need a rectifier and a converter matched to them.

Ambient radio waves from Wi-Fi, cellular, and broadcast transmitters carry well under $1\,\mu W/cm^2$ where people stand, and a rectifier needs a few hundred millivolts to turn on. Ambient RF powers only the most frugal sensors. Power sent on purpose, from a transmitter aimed at the device, is the subject of Section 13.6.

13.4 *Batteries and the Lithium-Ion Cell*

A BATTERY converts chemical energy into electrical energy at the two electrodes of a cell. Each electrode reaction has its own potential (Chapter 6), and the cell voltage is the difference. *Primary* cells are used once; *secondary* cells are rechargeable. The lithium-ion cell is the rechargeable cell of phones, wearables, cars, and rechargeable implants (Fig. 13.5).

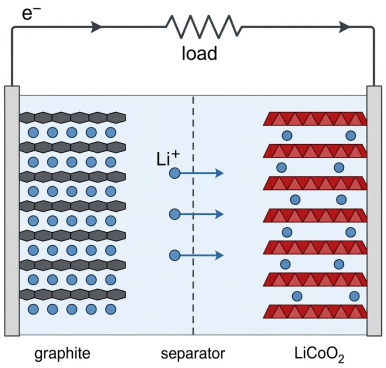


Figure 13.5: A lithium-ion cell during discharge. Lithium ions leave the graphite anode, cross the electrolyte and the porous separator, and slip into the layers of the cobalt oxide cathode; electrons travel the same way through the external circuit. Charging reverses both.

Structure and chemistry

The negative electrode is graphite coated on copper foil, the positive electrode a layered oxide such as lithium cobalt oxide coated on aluminum foil, and between them a porous polymer separator soaked in a lithium salt dissolved in an organic solvent. Neither electrode is consumed: lithium ions shuttle between them and slide into (*intercalate*) the layers of each host, like cards slipped into a deck. On discharge,

$$\begin{aligned} &\text{anode:} && \mathrm{LiC_6} \longrightarrow \mathrm{C_6} + \mathrm{Li^+} + \mathrm{e^-}, \\ &\text{cathode:} && \mathrm{Li_{0.5}CoO_2} + 0.5\,\mathrm{Li^+} + 0.5\,\mathrm{e^-} \longrightarrow \mathrm{LiCoO_2}. \end{aligned} \tag{13.6}$$

The cell voltage reflects how much more strongly the oxide holds lithium than graphite does: in graphite lithium sits only about 0.1 V above lithium metal, in the oxide near 4 V. Lithium-ion cells are therefore rated at 3.6 to 3.7 V nominal and are charged to 4.2 V.

Capacity, energy, and rate

Each lithium ion that moves sends one electron through the external circuit. A mole of electrons carries Faraday's constant, $F = 96\,485\,\mathrm{C/mol}$, so a host of molar mass M that exchanges n electrons per formula unit can store at most

$$q = \frac{nF}{3.6\,M}\ \mathrm{mAh/g}, \tag{13.7}$$

the factor 3.6 converting coulombs to milliamp-hours. A cell's capacity times its average voltage gives its energy (Eq. (13.1)), and per unit mass

or volume its *energy density*. The *C-rate* measures current relative to capacity: a 40 mAh cell at 1C supplies 40 mA for an hour, at C/2 20 mA for two hours, at 2C 80 mA for half an hour. *Cycle life* is the number of full charge and discharge cycles before the capacity falls to 80 % of its initial value, several hundred to a few thousand for lithium ion. Capacity fades because a protective film that forms on the graphite keeps growing and consumes lithium, and because the electrodes crack as they swell and shrink with every cycle.

Safety and flexible cells

A lithium-ion cell stores its energy next to a flammable electrolyte. Overcharge, short circuit, crushing, or heating can start reactions that heat the cell further, *thermal runaway*, with venting and fire. Every cell therefore has a protection circuit that limits voltage, current, and temperature, and the charger supplies a constant current until the cell reaches 4.2 V, then holds that voltage while the current tapers. Flexible cells for wearables are printed from slurries of the same active materials (Chapter 12) or built as small rigid cells joined by serpentine interconnects (Chapter 11), and their capacity is quoted per square centimeter. In one flexible system, a printed lithium-ion battery that stored 6.98 mWh/cm^2 and survived 600 flexing cycles was charged by a flexible solar module and powered a pulse oximeter.[3]

[3] A. E. Ostfeld, A. M. Gaikwad, Y. Khan, A. C. Arias, *Sci. Rep.* (2016).

13.5 Batteries for Implants

Most active implants run on batteries. Batteries are reliable, have decades of history with regulators, and need nothing outside the body. The price is size and replacement: the battery is about half the volume of a pacemaker, and when a primary battery is empty the whole device is replaced in surgery.

Primary implant chemistries

The *lithium-iodine* cell powers most pacemakers. Its cathode is iodine mixed with a polymer, and its electrolyte is a thin layer of lithium iodide that forms by itself where lithium meets iodine. The cell gives 2.8 V, stores about 1 Wh/cm^3, loses almost nothing to self-discharge, and was first implanted in 1972.[4] The solid electrolyte thickens as the cell discharges, so its internal resistance rises steadily. The cell can therefore supply only microamps, but its voltage falls smoothly and predictably near the end, which lets the pacemaker warn of replacement months ahead.

[4] C. F. Holmes, *Electrochem. Soc. Interface* (1999); V. S. Mallela, V. Ilankumaran, N. S. Rao, *Indian Pacing Electrophysiol. J.* (2004).

Devices that need more current use other lithium chemistries. An implantable cardioverter-defibrillator (ICD) must charge its capacitors to deliver a shock of about 30 J within seconds, which takes amps; the *lithium/silver vanadium oxide* (Li/SVO) cell delivers those pulses. *Lithium/carbon monofluoride* (Li/CF_x) has the highest energy density of the group, about 440 mWh/g for a complete cell, and powers neurostimulators and drug pumps. Equation (13.7) gives the theoretical values: CF_x with $x = 1$ stores $96485/(3.6 \times 31.01) = 864$ mAh/g, and at 3.0 V that is 2.6 Wh/g of cathode. Complete cells, with their anode, electrolyte, and case, reach a sixth to a half of the theoretical value: about 0.25 Wh/g for lithium-iodine and Li/SVO cells.

Longevity

The longevity of a battery-powered implant is its usable capacity divided by its average current, exactly as in Eq. (13.4). A pacemaker's drain splits between the pacing pulses and housekeeping: sensing, telemetry, and memory. A pacing pulse of 2.5 V into a 500 Ω lead for 0.40 ms carries $(5\,\text{mA})(0.40\,\text{ms}) = 2.0\,\mu\text{C}$. Pacing two chambers at 70 beats per minute averages 4.7 µA; an output stage that draws twice the delivered charge from the battery makes that 9.3 µA, and with 5 µA of housekeeping the drain is 14.3 µA, about 40 µW at 2.8 V. A 1.2 Ah lithium-iodine cell then lasts 1.2 Ah/14.3 µA = 9.6 years, the decade that pacemaker batteries are known for.

A neural stimulator (Chapter 9) drives larger currents, and its battery is sized the same way: from the charge of each pulse, the pulse rate, the supply voltage, the converter efficiency, and the standby drain.

> **Worked Example 13.3** *The battery for a neural stimulator*
>
> A deep brain stimulator delivers charge-balanced biphasic pulses of 3.0 mA, 60 µs per phase, at 130 Hz, continuously. The output stage runs from a 4.5 V supply (enough to drive 3 mA through about 1 kΩ of electrode and tissue, with headroom) that a converter makes from a 3.0 V battery with 80 % efficiency. The rest of the circuit draws 15 µA. What battery capacity lasts four years?
>
> Solution. Find the average power drawn from the battery, convert it to current, multiply by the lifetime, and add a margin.
>
> *Stimulation.* Each biphasic pulse draws its current for two phases from the supply, so the average stimulation current is
>
> $$I_{\text{stim}} = 2\,I\,t_{\text{ph}}\,f = 2(3.0\,\text{mA})(60\,\mu\text{s})(130\,\text{Hz}) = 46.8\,\mu\text{A},$$
>
> or $(46.8\,\mu\text{A})(4.5\,\text{V}) = 211\,\mu\text{W}$ from the supply. The battery provides $211/0.80 = 263\,\mu\text{W}$ for it. *Everything else*: $(15\,\mu\text{A})(3.0\,\text{V}) =$

45 µW. *Total*: 308 µW, or $308\,\mu W / 3.0\,V = 103\,\mu A$ from the battery.

Four years is 35 064 h, so the battery must deliver

$$(103\,\mu A)(35\,064\,h) = 3.6\,Ah,$$

and with a 20 % margin for aging and for higher settings, about 4.3 Ah, or 13 Wh. At the 440 mWh/g of a Li/CF_x cell that is about 30 g of battery. Most of the energy goes to the pulses: halving the pulse width or the amplitude cuts the battery by about 40 %. This is why primary DBS batteries are replaced every three to five years. ■

For a device that stimulates only a few minutes a day, the balance flips. Pulses of 0.1 mA and 100 µs per phase at 50 Hz, six 5-minute sessions a day for two weeks, use only 0.007 mAh, while a standby current of 2 µA uses $(2\,\mu A)(336\,h) = 0.67\,mAh$. Never leave out the standby drain.

Rechargeable implants

A rechargeable lithium-ion cell trades replacement surgery for a regular chore: rechargeable DBS systems are rated for 9 to 15 years, with the patient charging through a wearable charger for about 30 minutes a day or a few hours every week or two.[5] Rechargeable cells self-discharge faster than primary ones and depend on the patient, which is why pacemakers still use primary cells.

[5] The 15-year rating: Medtronic announcement of FDA approval of the Percept RC neurostimulator, January 2024.

The alternative is wireless power, which keeps the energy source outside, where it can be recharged or replaced. The implant can then be much smaller and can draw tens of milliwatts day and night, as a cochlear implant does. The price is a transmitter the patient must wear, a link that fails when it moves away, and tissue heating that caps the power (Section 13.7). Many devices combine the two: a small rechargeable battery bridges gaps in a wireless link.

13.6 *Inductive Power Transfer*

A COCHLEAR IMPLANT receives its power and data through two coils, one on the scalp and one under the skin, coupled by their magnetic field (Fig. 13.6). The same idea charges electric toothbrushes and phones, and it is the most common way to power an implant from outside. This section builds it in three steps: how one coil induces a voltage in another, why both coils are tuned to resonance, and what sets the efficiency.

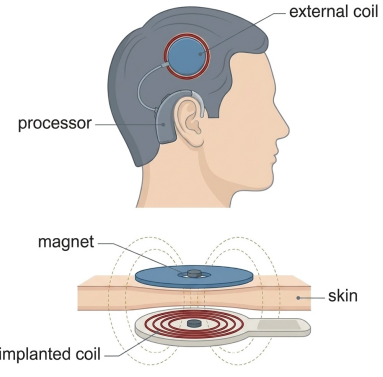


Figure 13.6: A transcutaneous inductive link. The external coil, held over the implant by a magnet, sits a few millimeters above the implanted coil; the processor behind the ear drives it. Below, a side view of the two coils on either side of the skin shows the field they share.

Magnetic coupling

A current through a coil of wire makes a magnetic field that loops through the coil and around it. The field stores energy, and it resists change: when the current changes, the changing field induces a voltage across the coil that opposes the change (Faraday's law of induction). The coil is an *inductor*, and its *inductance* L, in henries (H), sets that voltage, $v = L\,\mathrm{d}i/\mathrm{d}t$. For a sinusoidal current of angular frequency $\omega = 2\pi f$, the voltage amplitude is ωL times the current amplitude. This *reactance* ωL grows with frequency, the opposite of a capacitor's $1/(\omega C)$ (Chapter 1). The coils of implant links have inductances of about a microhenry.

Now bring a second coil near the first. Part of the first coil's field passes through the second, and as that field changes it induces a voltage in the second coil too. The *mutual inductance* M sets this voltage: a current of amplitude I_1 in the transmitter coil induces an open-circuit voltage of amplitude

$$V_2 = \omega M I_1, \qquad M = k\sqrt{L_1 L_2}, \tag{13.8}$$

in the receiver coil. The *coupling coefficient* k, between 0 and 1, is the fraction of the transmitter's magnetic flux that passes through the receiver. A transformer, whose two windings share an iron core, has k close to 1. Coils separated by skin and fat have k of a few percent or less: most field lines of the outer coil close on themselves without reaching the implant.

How fast does the coupling fall with distance? For two coaxial loops of radii a (transmitter) and b (receiver), with N_1 and N_2 turns, a distance d apart, and $b \ll a$,

$$M \approx \frac{\mu_0 \pi N_1 N_2\, a^2 b^2}{2\,(a^2 + d^2)^{3/2}}, \tag{13.9}$$

which is the transmitter's field on its axis times the receiver's area ($\mu_0 = 4\pi \times 10^{-7}\,\mathrm{H/m}$). Once the distance exceeds the transmitter radius, M falls as d^{-3}: doubling the distance cuts the induced voltage about eightfold. Three design rules follow. Inductive links work over distances comparable to the coil size. For a given distance, the transmitter radius that couples best to a small receiver is $a = \sqrt{2}\,d$, so a deeper implant wants a larger outer coil. And a small receiver couples weakly to any transmitter, because it catches little flux.

Resonance

With k near 0.01 the induced voltage is small, and the receiver coil's own reactance ωL_2 blocks most of the current that voltage could drive.

On the outside the same problem appears in reverse: driving a large current through L_1 takes a large voltage. Resonance solves both.

Put a capacitor in series with a coil. Energy now sloshes back and forth between the magnetic field of the inductor and the electric field of the capacitor, the way a swing trades height for speed. At one frequency, the *resonant frequency*

$$f_0 = \frac{1}{2\pi\sqrt{LC}}, \tag{13.10}$$

the two reactances are equal, $\omega_0 L = 1/(\omega_0 C)$, and because they act in opposite directions they cancel. Only the resistance R of the wire is left to limit the current, the way a swing pushed in time with its motion goes higher with every push. The *quality factor*

$$Q = \frac{\omega_0 L}{R} \tag{13.11}$$

compares the reactance with the resistance. It also says how long the circuit rings: struck once, a resonator with $Q = 100$ oscillates for about $Q/\pi \approx 30$ cycles before its amplitude falls to a third. At resonance the voltage across the coil or the capacitor is Q times the drive voltage. Coils for implant links reach Q of 30 to 100 at a few megahertz; a small implanted coil wound with thin wire sits at the low end.

Efficiency of a resonant link

Figure 13.7 shows the circuit. A source drives the transmitter coil L_1 through a tuning capacitor C_1; the receiver coil L_2 is tuned by C_2 and feeds a load R_L, which stands for the rectifier and the implant behind it. R_1 and R_2 are the resistances of the coil wires, where the losses happen.

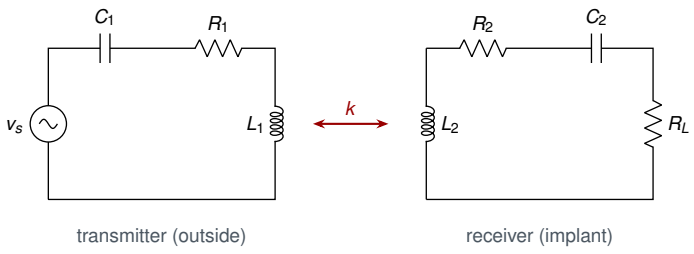


Figure 13.7: A series-resonant inductive link. The coils are coupled with coefficient k; the capacitors tune both to the same frequency. R_1 and R_2 are the coil losses, and R_L stands for the rectifier and the implant. After M. Kiani and M. Ghovanloo, *IEEE Trans. Circuits Syst. I* (2012).

Follow the transmitter current I_1. It does two things: it heats R_1, and it pushes energy into the receiver. To the source, the receiver's pull on the shared field looks like an extra resistance in series with R_1,

$$R_{\text{refl}} = \frac{(\omega_0 M)^2}{R_2 + R_L}. \tag{13.12}$$

The fraction of the transmitter's power that crosses the gap is $R_{\text{refl}}/(R_1 + R_{\text{refl}})$, and of the power that crosses, the fraction $R_L/(R_2 + R_L)$ reaches the load instead of heating R_2. The link efficiency is the product of the two fractions. Its largest value, reached with $R_L = R_2\sqrt{1+k^2Q_1Q_2}$, is

$$\eta_{\max} = \frac{k^2Q_1Q_2}{\left(1+\sqrt{1+k^2Q_1Q_2}\right)^2}. \tag{13.13}$$

Only the product $x = k^2Q_1Q_2$ appears, and it has a plain meaning. With the load shorted ($R_L = 0$), $R_{\text{refl}}/R_1 = (\omega_0 M)^2/(R_1R_2) = k^2Q_1Q_2$: the figure of merit is the resistance the receiver reflects into the transmitter, measured against the transmitter's own loss. When x is small the efficiency is about $x/4$; at $x = 8$ it is 50 %; when x is large it approaches 1. Weak coupling can be made up with high-Q coils: a link with $k = 0.01$ works as well as one with $k = 0.1$ if Q_1Q_2 is 100 times larger. This is what "magnetic resonance" wireless power means.

KEY EQUATION

$\eta_{\max} = \dfrac{x}{(1+\sqrt{1+x})^2}, \quad x = k^2Q_1Q_2$

Best efficiency of a two-coil resonant link.

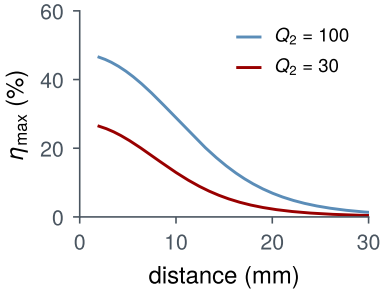


Figure 13.8: Maximum efficiency of the coil pair of Worked Example 13.4 versus distance, for $Q_1 = 100$.

Worked Example 13.4 *A coil link through the skin*

A five-turn transmitter coil of radius 15 mm ($L_1 = 1.97\,\mu$H, $Q_1 = 100$) powers a ten-turn implanted coil of radius 2.5 mm ($L_2 = 1.25\,\mu$H, $Q_2 = 30$) at 13.56 MHz. (a) Find M, k, and $\eta_{\max}$ when the coils are 10 mm apart and when they are 20 mm apart. (b) What capacitors tune the coils? (c) What open-circuit voltage does a 0.5 A transmitter current induce at 10 mm?

Solution. (a) At 10 mm, Eq. (13.9) gives

$$M = \frac{(4\pi\times10^{-7})\pi(5)(10)(0.015)^2(0.0025)^2}{2\,[(0.015)^2+(0.010)^2]^{3/2}} = 23.7\,\text{nH},$$

so $k = M/\sqrt{L_1L_2} = 23.7/1569 = 0.0151$ and $k^2Q_1Q_2 = 0.683$. Then $\eta_{\max} = 0.683/(1+1.297)^2 = 13\,\%$. At 20 mm the denominator grows by $(6.25/3.25)^{3/2} = 2.67$, so $k = 0.0057$, $k^2Q_1Q_2 = 0.096$, and $\eta_{\max} = 2.3\,\%$. Doubling the distance cut the efficiency more than fivefold (Fig. 13.8). (b) $C = 1/(\omega_0^2 L)$ with $\omega_0 = 2\pi(13.56\,\text{MHz})$: $C_1 = 70\,$pF and $C_2 = 110\,$pF. (c) $V_2 = \omega_0 M I_1 = (8.52\times10^7)(23.7\times10^{-9}\,\text{H})(0.5\,\text{A}) = 1.0\,\text{V}$ (amplitude). Resonance in the receiver raises the voltage across L_2 or C_2 by up to Q_2 when it is lightly loaded. ■

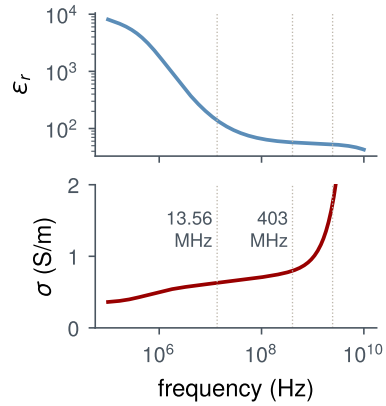


Figure 13.9: Relative permittivity and conductivity of muscle from the four-term Cole-Cole model of Gabriel *et al.* (1996). Dotted lines: 13.56 MHz, 403 MHz (the medical implant band), and 2.45 GHz.

Choosing the frequency

A higher carrier frequency seems better: the induced voltage $\omega M I_1$ grows with ω, and so does each coil's Q. But tissue is salty water, a weak conductor: the changing magnetic field drives currents in it that

become heat, and its conductivity rises with frequency (Fig. 13.9), so the tissue absorbs more as the frequency rises. The best frequency balances the two. For centimeter coils it lies between a few and a few tens of megahertz: cochlear implant links run in this range, and wireless chargers use the industrial, scientific, and medical (ISM) bands at 6.78 and 13.56 MHz. For millimeter receivers deeper in tissue the optimum moves up, to hundreds of megahertz or the low gigahertz.[6]

[6] A. S. Y. Poon, S. O'Driscoll, T. H. Meng, *IEEE Trans. Antennas Propag.* (2010).

13.7 *Tissue Heating and the Specific Absorption Rate*

Every wireless link warms the tissue it crosses. Tissue is mostly salt water, a weak conductor: muscle has a conductivity σ near 1 S/m, copper 6×10^7 S/m. An electric field E in tissue drives a current density $J = \sigma E$, and the current heats the tissue as a current heats a resistor, with a power per unit volume $JE = \sigma E^2$. The *specific absorption rate* (SAR) is this power per kilogram of tissue,

$$\mathrm{SAR} = \frac{\sigma E^2}{\rho} = c \left.\frac{\mathrm{d}T}{\mathrm{d}t}\right|_{t=0}, \tag{13.14}$$

key equation

$\mathrm{SAR} = \sigma E^2/\rho$

Specific absorption rate in W/kg; E is the rms field in the tissue.

in watts per kilogram, where E is the rms field and ρ the density (about 1000 kg/m^3). The second form says what SAR does: at first, before blood flow and conduction carry heat away, the tissue warms at the rate SAR/c, where $c \approx 3500$ J/(kg K) is its specific heat. That is also how SAR is measured, from the initial rate of temperature rise in a phantom, a model of the body filled with a tissue-like liquid or gel.

The body sets its own scale. Its resting 100 W spread over 70 kg is about 1.4 W/kg. The safety limits (Table 13.2) allow a local SAR of 1.6 to 2 W/kg, comparable to the body's own resting heat production, and a whole-body average of 0.08 W/kg, about 6 % of it. Blood flow removes this much heat easily, and the limits keep the temperature rise well below 1 °C. The same limits govern the transmit power of cell phones.

Rule (general public)	Whole body	Local
FCC (US)	0.08 W/kg	1.6 W/kg over any 1 g; 4 W/kg over 10 g for hands, wrists, feet, ankles, and ears
ICNIRP (2020) and IEEE C95.1-2019	0.08 W/kg	2 W/kg over 10 g; 4 W/kg for limbs

Table 13.2: SAR limits for the general public, 100 kHz to 6 GHz. Limits for workers are five times higher.

The US limit of 1.6 W/kg averaged over 1 g of tissue is stricter for small hot spots than 2 W/kg over 10 g: averaging over a smaller mass captures a sharper peak. An implant also heats tissue directly with the power its rectifier and circuits dissipate, and the standard for active

implants (ISO 14708-1) allows its surface to run at most 2 °C above body temperature, as for the brain implants of Chapter 9.

> **WORKED EXAMPLE 13.5** *Heating by an implant link*
>
> An implant link at 403 MHz produces an rms field of 30 V/m in muscle ($\sigma = 0.80\,\mathrm{S/m}$, $\rho = 1090\,\mathrm{kg/m^3}$, $c = 3500\,\mathrm{J/(kg\,K)}$). (a) Find the SAR. (b) What is the largest field the FCC limit allows, and how much more power could the transmitter send? (c) How fast does the tissue warm at that limit?
>
> SOLUTION. (a) SAR $= (0.80)(30)^2/1090 = 0.66\,\mathrm{W/kg}$, within the FCC limit of 1.6 W/kg. (b) $E = \sqrt{(1.6)(1090)/0.80} = 46.7\,\mathrm{V/m}$ rms. Because SAR grows as E^2, and so as the transmitted power, the transmitter can send only $(46.7/30)^2 = 2.4$ times more power before reaching the limit. (c) At 1.6 W/kg the tissue initially warms at $1.6/3500 = 4.6 \times 10^{-4}\,\mathrm{K/s}$, 1.6 K per hour if no heat were carried away. Blood flow and conduction keep the real rise far smaller. ■

SAR caps what an electromagnetic link can deliver. The limit sets a ceiling on the transmitted power, and the link efficiency sets how much of that ceiling reaches the implant. Both work against small, deep implants: a millimeter receiver couples weakly (Eq. (13.9)), and the higher frequency it prefers is absorbed more by the tissue on the way. Small, deep implants are where electromagnetic power runs out, and where ultrasound takes over.

13.8 *Ultrasound Power*

ULTRASOUND is a pressure wave, typically 0.5 to 10 MHz, that soft tissue carries well. An external transducer sends a beam to an implant whose piezoelectric receiver turns it back into electricity, and the implant can talk back by changing how strongly it reflects the wave (backscatter), as in the neural dust motes of Fig. 13.10.

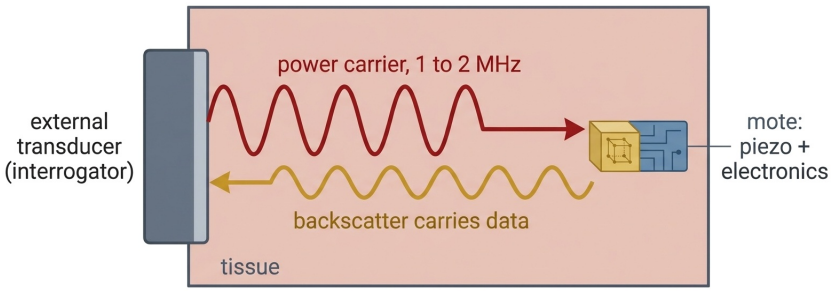


Figure 13.10: Ultrasonic power and data. The interrogator sends ultrasound into the tissue; a piezoelectric crystal in the implant converts it into electrical power, and the implant modulates the reflected wave to send data back.

Waves in tissue

Sound travels in soft tissue at about $c = 1540\,\mathrm{m/s}$, so the wavelength $\lambda = c/f$ is 1.54 mm at 1 MHz. Each medium has an *acoustic impedance* $Z = \rho c$, its density times its sound speed: about 1.6 MRayl for soft tissue ($1\,\mathrm{MRayl} = 1 \times 10^6\,\mathrm{kg/(m^2\,s)}$), 7.8 MRayl for bone, about 30 MRayl for a PZT crystal, and only 0.0004 MRayl for air. At a boundary the fraction of the intensity reflected is

$$R_I = \left(\frac{Z_2 - Z_1}{Z_2 + Z_1} \right)^2 . \tag{13.15}$$

Ultrasound passes between soft tissues almost freely, but a tissue-air boundary reflects 99.9 % of it (hence the gel under an ultrasound probe) and tissue-bone 43 %. A bare PZT receiver reflects most of the wave too, so motes add a matching layer. Soft tissue also absorbs ultrasound, by roughly $0.5\,\mathrm{dB/(cm\,MHz)}$: the loss in decibels grows in proportion to both depth and frequency (Fig. 13.11), and a loss of L decibels leaves $10^{-L/10}$ of the intensity.

The reference for a safe intensity comes from diagnostic imaging.[7] FDA allows a time-average intensity, at the hottest point of the beam and corrected for tissue loss, of up to $720\,\mathrm{mW/cm^2}$ everywhere but the eye, and power links use this number as their reference.

[7] FDA, *Marketing Clearance of Diagnostic Ultrasound Systems and Transducers*, guidance revised February 2023.

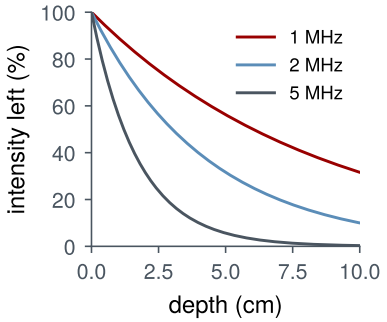


Figure 13.11: Intensity remaining after a depth z of soft tissue at $0.5\,\mathrm{dB/(cm\,MHz)}$. Higher frequency gives a shorter wavelength but a steeper loss.

Why ultrasound wins for small, deep implants

A receiver captures a wave's power efficiently only if it is not much smaller than the wavelength, the argument Chapter 9 met when it shrank an implant. In muscle at 1 GHz, the relative permittivity of about 55 shrinks the electromagnetic wavelength to $\lambda = c_0/(f\sqrt{\varepsilon_r}) \approx 4\,\mathrm{cm}$, so a millimeter antenna is forty times smaller than its wavelength and captures little; at lower frequencies the mismatch is worse. Ultrasound at 1 MHz has a 1.5 mm wavelength, matched to a millimeter receiver, and loses little on the way: 5 cm of tissue costs $(0.5)(1)(5) = 2.5\,\mathrm{dB}$, leaving $10^{-0.25} = 56\,\%$ of the intensity. (At 5 MHz the same path costs 12.5 dB and leaves 5.6 %, so power links stay near 1 to 2 MHz.) At the reference intensity a $1\,\mathrm{mm^2}$ receiver intercepts $(720\,\mathrm{mW/cm^2})(0.01\,\mathrm{cm^2}) = 7.2\,\mathrm{mW}$; even at 25 % conversion, 1.8 mW reaches the circuit, far more than an RF link to a millimeter antenna at that depth can deliver under the SAR limit. A mote 3 mm long, powered by ultrasound, recorded from nerve and muscle in rats.[8]

Ultrasound has its own costs: it does not pass easily through the skull or the lungs, it needs acoustic contact with the skin, and the transmitter must be aimed at the implant. The choice is therefore by size and depth. Ultrasound suits millimeter implants centimeters deep

[8] D. Seo, R. M. Neely, K. Shen, U. Singhal, E. Alon, J. M. Rabaey, J. M. Carmena, M. M. Maharbiz, *Neuron* (2016).

in soft tissue, with no bone or air in the path; coils suit larger, shallow implants, where they couple well and need not be aimed.

13.9 Powering Ingestibles

An ingestible capsule has to fit in a pill about 11 mm across and run for the hours or days it spends in the gut. Batteries remain the usual choice, and silver-oxide button cells (1.55 V) are the common ones because they are small, hold a steady voltage as they discharge, and contain no mercury (Fig. 13.12). A capsule endoscope that photographs the small intestine, 26.2 mm long and 11.4 mm across, runs on two silver-oxide cells for at least 8 hours.[9] If each cell holds about 28 mAh, the pair stores $2(28\,\text{mAh})(1.55\,\text{V}) = 86.8\,\text{mWh}$ (312 J), an average of 10.9 mW over 8 hours. The illumination LEDs and the radio take most of it, which is why the capsule takes more images per second when it moves fast and fewer when it sits still.

Lithium coin cells give 3 V and more energy but are a hazard if their casing fails in the gut. Lithium-ion cells can also overheat in thermal runaway; silver-oxide cells are not prone to it, and they are the only batteries approved to power capsule endoscopes.[10] The alternatives harvest or receive power. Two dissimilar metals in gastric fluid form a galvanic cell (Chapter 6), a "biobattery" that runs on stomach acid. It is the only in-body harvester in commercial use, and only for short-term ingestibles: an ingestible event marker, a tiny sensor attached to a pill, is powered by contact with stomach fluid and signals that the pill was swallowed. FDA granted it a De Novo classification in 2012 (Chapter 14). In pigs, a capsule with galvanic electrodes harvested an average of 0.23 µW per square millimeter of electrode and reported temperature for an average of 6.1 days as it passed through the gut.[11] Power can also be sent from outside, through a coil worn around the abdomen or by ultrasound.

[9] Medtronic product data and FDA 510(k) summary K211684 for the PillCam SB 3 capsule (2021).

[10] A. Abdigazy, M. Arfan, G. Lazzi, C. Sideris, A. Abramson, Y. Khan, *Nat. Electron.* (2024), a review of ingestible design. It also gives the in-body path loss of a radio link from the intestine, about 3 dB/cm at 400 MHz and 8 dB/cm at 2.4 GHz. Path loss counts all the loss between the two antennas, not only absorption: absorption alone, from the muscle properties of Fig. 13.9, is about 1.7 and 3.8 dB/cm.

[11] P. Nadeau, ..., G. Traverso, *Nat. Biomed. Eng.* (2017).

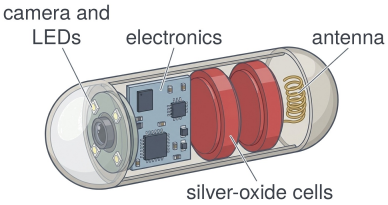


Figure 13.12: Inside an ingestible capsule: two silver-oxide button cells take a large share of the volume, next to the camera or sensor, the electronics, and the antenna. A receiving coil can replace the cells when power is sent from outside.

Harvesting inside the body

The body offers other sources, none yet used in a long-term implant. Thin PZT ribbons sutured to the heart of cows and sheep produced about $1.2\,\mu\text{W/cm}^2$ from a stack of five layers.[12] Glucose and oxygen are present in every body fluid, and sugar stores 20 to 30 times more energy per gram than a lithium-ion cell; a fuel cell with a platinum anode that oxidizes glucose, designed for the cerebrospinal fluid around the brain, gave about $3.4\,\mu\text{W/cm}^2$ in steady state.[13] The cochlea of Chapter 10 keeps a potential of 80 to 100 mV across its fluid compartments, a natural battery that powers its hair cells, and a chip drew

[12] C. Dagdeviren *et al.*, *Proc. Natl. Acad. Sci. USA* (2014).

[13] B. I. Rapoport, J. T. Kedzierski, R. Sarpeshkar, *PLoS ONE* (2012).

about a nanowatt from it in a guinea pig.[14]

These numbers set the scale. A sensor that needs 10 µW would need 43 mm^2 of gastric electrode, which fits on a capsule, 2.9 cm^2 of glucose fuel cell, or 8.3 cm^2 of heart harvester, while a lithium-iodine cell delivers 10 µW for ten years (0.88 Wh) from about 1 cm^3. Harvesting inside the body pays off for loads of a microwatt or less, or for devices that cannot be reached at all.

[14] P. P. Mercier, A. C. Lysaght, S. Bandyopadhyay, A. P. Chandrakasan, K. M. Stankovic, *Nat. Biotechnol.* (2012).

Summary

- Power is the rate of energy use, $P = VI$, and $E = Pt$. Capacity in mAh times voltage is energy in mWh (1 mWh = 3.6 J); a CR2032 stores about 2.4 kJ. Efficiencies multiply; a linear regulator has $\eta = V_{\text{out}}/V_{\text{in}}$, a switching converter 80 to 95 %.
- $I_{\text{avg}} = \sum D_i I_i$ and $t = C_{\text{bat}}/I_{\text{avg}}$; LEDs and radios dominate a wearable budget, until the sleep current takes over. A buffer capacitor sags $\Delta V = It/C$ during a peak.
- Harvesters: sunlight 15 to 25 mW/cm^2, indoor light 10 to 80 µW/cm^2, body heat 1 to 30 µW/cm^2 ($V_{\text{oc}} = NS\Delta T$ with only a kelvin or two across the TEG), ambient RF below 1 µW/cm^2.
- Lithium ion: $q = nF/(3.6M)$, graphite 372 mAh/g, $LiCoO_2$ about half of 274 mAh/g usable; 3.7 V nominal, 4.2 V charged. Implants: Li/I_2 for pacemakers (about ten years), Li/SVO for ICD shocks, Li/CF_x for neurostimulators. Stimulation draws $2It_{\text{ph}}f$; add the converter loss, the standby drain, and a margin.
- Inductive links: $V_2 = \omega M I_1$, $M = k\sqrt{L_1 L_2}$, $M \propto d^{-3}$ beyond the coil radius; resonance at $f_0 = 1/(2\pi\sqrt{LC})$ cancels the reactance, $Q = \omega_0 L/R$; $\eta_{\max} = x/(1+\sqrt{1+x})^2$ with $x = k^2 Q_1 Q_2$.
- SAR $= \sigma E^2/\rho = c\,\mathrm{d}T/\mathrm{d}t$; local limits 1.6 to 2 W/kg, whole body 0.08 W/kg, against a resting body heat of about 1.4 W/kg.
- Ultrasound: $\lambda = 1.5$ mm at 1 MHz, about 0.5 dB/(cm MHz), reference intensity 720 mW/cm^2; it wins for millimeter, deep implants in soft tissue and loses at bone and air.
- Capsules run on silver-oxide cells; in-body harvesters give microwatts per square centimeter or less.

With materials, printing, and power in hand, Chapter 14 puts a whole device together as a design project and follows it toward a patient.

13.10 Problems

13.1 ◦ A sensor node on a 100 mAh battery has a microcontroller drawing 4 mA for 1 % of the time, a radio drawing 15 mA for 0.1 %, a sensor drawing 0.5 mA for 5 %, and a sleep current of 3 µA the rest of the time.

(a) Find the average current.

(b) Find the battery life in days.

(c) Which block would you optimize first?

13.2 •• A TEG with 120 couples ($S = 380\,\mu V/K$ each), $R_{TEG} = 150\,K/W$, and $R_{el} = 60\,\Omega$ sits on skin with $R_{body} = 300\,K/W$ and a heat sink of $R_{sink} = 350\,K/W$. The core-to-air difference is 14 K.

(a) Find ΔT_{TEG}, V_{oc}, and the matched-load power.

(b) Walking halves R_{sink}. Repeat (a).

(c) A boost converter needs 20 mV to start. Does it start in still air?

13.3 • A 150 mAh, 3.7 V lithium-polymer cell (a lithium-ion cell in a flat pouch) powers a smartwatch.

(a) Find its energy in Wh and J.

(b) What current is 2C? How long does a 2C discharge last?

(c) The cell keeps 80 % of its capacity after 500 cycles, and the watch is charged every two days. When does the capacity reach 80 %?

13.4 • At 6.78 MHz, two coils have $L_1 = 2.0\,\mu H$, $L_2 = 0.80\,\mu H$, and $M = 40\,nH$, with $Q_1 = 80$ and $Q_2 = 25$.

(a) Find k and η_{max}.

(b) Find R_2 and the optimal load resistance.

(c) What would Q_2 have to be for 50 % efficiency?

13.5 • An ultrasonic link at 2 MHz powers a mote 4 cm deep.

(a) Find the wavelength in tissue and the fraction of intensity left at the mote (0.5 dB/(cm MHz)).

(b) At the reference intensity of $720\,mW/cm^2$ at the mote, what receiver area gives 1 mW of electrical power with 30 % conversion efficiency?

14
From Prototype to Patient, and the Design Project

In this chapter you will learn to

- Turn a device idea into a one-sentence goal and a specification in which every entry has a target, a unit, and a reason.
- Plan a team project around its riskiest assumption, with an owner for each block and a known critical path.
- Check a design on paper with signal, noise, data-rate, and battery budgets.
- Bring up a sensor board in stages with known test signals before it touches a person.
- Test on yourself and classmates safely: consent, skin contact, battery power, and leakage current.
- Collect, store, and report data that can be trusted, with honest numbers.
- Decide whether a product is a medical device, assign its class, and choose a premarket pathway: 510(k), De Novo, or PMA.
- Plan the clinical evidence (IRB, IDE, study risk, sample size) and name the parts of a quality system.
- Apply the Belmont principles, informed consent, and privacy to wearable and neural devices.

In the last week of the semester a team of four stands at the front of the room. One of them wears a soft strip of printed electrodes curled behind the ear, wired to a small board clipped to their collar. On the screen eight traces scroll by, each a few microvolts tall. The wearer closes their eyes, and within a second a ten-hertz rhythm swells in the traces: the alpha rhythm of a relaxed brain, picked up by a device the team specified, printed, soldered, and programmed themselves.

That demonstration is the end of a project, and it could be the start

of a product. Reaching it takes work that no earlier chapter covers on its own: a specification with numbers, budgets checked on paper, a board brought up one stage at a time, tests on people that are safe, and data that can be trusted. Going from it to a patient takes more. Someone must decide whether the device is a medical device and which rules apply, what evidence shows that it is safe and that it works, how it is designed and made so that every unit is the same, and what we owe the people who test and wear it.

This last chapter follows a device idea along that path, the six steps of Fig. 1.26. The first half turns an idea into a project: the specification (Section 14.1), the plan (Section 14.2), the budgets (Section 14.3), the hardware and its bring-up (Section 14.4), safe tests on people (Section 14.5), the data, demonstration, and report (Section 14.6), and lessons from past projects (Section 14.7). The second half turns a project into a medical device: what a device is (Section 14.8), its class and route to market (Section 14.9), clinical evidence (Section 14.10), quality systems and standards (Section 14.11), and ethics (Section 14.12). The project draws on the whole book: the printed electrodes of Chapter 12, the amplifiers and filters of Chapter 8, the power of Chapter 13, and the four labs that follow this chapter. Regulations change: the rules here were checked in late 2026, and you should check the current version before you rely on one.

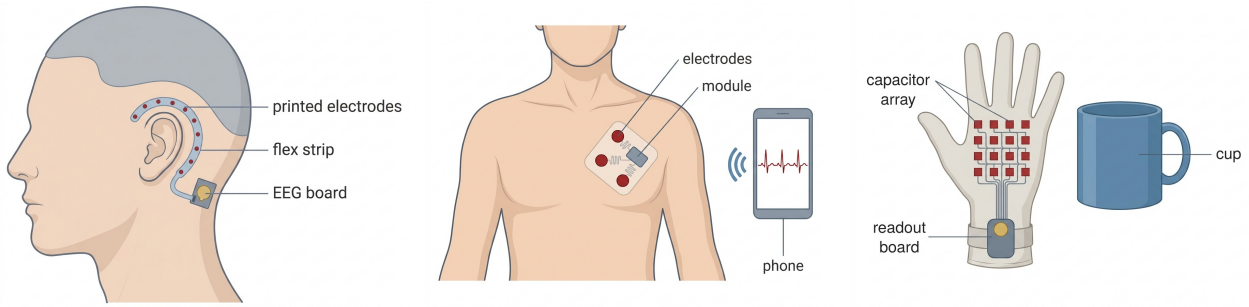
14.1 *From an Idea to a Specification*

A SEMESTER DESIGN PROJECT asks a team, typically of three to five students, to design and build a complete bioelectronic system to a specification of its own. Three projects recur in this book, and each exercises a different part of it (Fig. 14.1):

- a *wearable EEG*: printed electrodes around the ear and an eight-channel biopotential front end (TI ADS1299), to detect a mental state such as relaxation or stress;
- a *wearable ECG*: a chest patch with printed Ag/AgCl electrodes and a three-channel, low-power front end (TI ADS1293) that streams the heart's signal to a phone;
- a *capacitive sensor glove*: a printed array of pressure-sensing capacitors on the palm, read by a capacitance-to-digital converter (TI FDC2214), that classifies the objects it grasps.

Each has every block of the bioelectronic system of Chapter 1. A printed, flexible sensor joins a rigid board through a flat flexible cable (FFC). ECG and EEG electrodes plug straight into the front-end

board of Lab 3; the sensor breakout board of Lab 1 carries the supply on two of its pins, so it suits sensors that do not touch the skin, such as the glove's capacitive array, and never an electrode tail. A front-end chip amplifies and digitizes the signal (Lab 3); a microcontroller sends it over Bluetooth Low Energy (BLE) to a phone or laptop (Lab 4); and a small battery powers it all. Projects of the same scope work as well: a sweat sensor, a printed oximeter, an EMG armband.



(a) wearable EEG (b) wearable ECG (c) capacitive glove

Figure 14.1: Three typical design projects. (a) A strip of printed electrodes around the ear feeds an eight-channel EEG front end behind it. (b) A chest patch with three printed electrodes and a small amplifier and radio module. (c) A glove with a printed row-column capacitor array on the palm and a readout board on the wrist.

Start from a measurement

A good project starts from a measurement, not from a part. "We will use the ADS1299" names a part. "We will detect eyes-closed relaxation from EEG recorded around the ear, with no cap and no gel in the hair" is a goal: it tells you which parts to look for and how you will know you succeeded. Write the goal as one sentence that says for whom, what need, what system, and how you will know it works.

The specification

A *specification* turns that sentence into numbers you can test. For a biopotential or biosensing wearable the same entries come up every time: the signal amplitude and bandwidth (Chapters 1 and 8), the noise floor and the interference the system may add, the channels, sampling rate, and resolution, the electrode impedance, the power and battery life, the size, mass, and wear time, the wireless link, safety (Section 14.5), and *validation*: the test that shows the system works, and against what reference. Each entry gets a target, a unit, and a reason. A requirement without a number cannot be tested, and one without a reason cannot be traded off when it turns out to be expensive.

KEY IDEA Every row of a specification has a target, a unit, and a reason. The budgets check the rows on paper; the validation row is checked on people.

Table 14.1 is a specification for the around-the-ear EEG. Three choices drive the rest. The *signal* is small, a few to a few tens of microvolts, so the total input-referred noise must stay near 1 µV rms and the power-line hum must be pushed below that. The *bandwidth*,

0.5 to 40 Hz, covers the delta to beta bands, including alpha (8 to 13 Hz), and 250 samples per second per channel is six times the top frequency. And the *wearable* constraints (battery, size, radio) are set by a working day, eight hours, and by what fits behind an ear. The budgets of Section 14.3 check the rows they can on paper.

Parameter	Target	Reason
Channels	8 EEG, plus reference and bias (10 electrodes)	wearable EEG systems in the literature average about 8 channels; a C-shaped array around the ear holds 10
Signal	5 to 50 µV; alpha 8 to 13 Hz	around-the-ear EEG is smaller than scalp EEG (5 to 300 µV, typically tens of microvolts)
Bandwidth	0.5 to 40 Hz	delta through beta; above 40 Hz muscle artifacts dominate
DC offset tolerated	±150 mV per electrode pair	gel on Ag/AgCl stays within tens of millivolts (bare silver can drift by hundreds); fits the ±187.5 mV input range at gain 24
Sampling rate	250 samples/s per channel	six times the top of the band
Resolution	24 bits; LSB below 0.1 µV at the input	quantization far below the noise floor
Input-referred noise	$\leq$ 1 µV rms, 0.5 to 40 Hz, amplifier plus electrodes	the alpha band holds an eighth of it, 15 dB below a 2 µV rms ear alpha rhythm (Example 14.1)
Power-line residual	$<$ 1 µV at 60 Hz after the bias drive and filtering	with 20 kΩ of electrode imbalance (Section 14.3)
Electrode impedance	$\leq$ 50 kΩ at 30 Hz, imbalance $\leq$ 20 kΩ	gel or hydrogel on printed Ag/AgCl; measured as in Lab 2
Wireless	BLE, $\geq$ 60 kbit/s of data, latency $<$ 0.5 s	8 channels of 24 bits at 250 samples/s plus margin (Example 14.2)
Battery life	$\geq$ 8 h streaming from a lithium-ion polymer (LiPo) cell of at most 300 mAh	one working day (Example 14.2)
Size and mass	electronics $\leq$ 40 × 25 × 10 mm, total $\leq$ 15 g	sits behind the ear on a flexible strip
Wear time	1-hour sessions; skin checked after each	adhesives and gel are rated for short wear
Safety	battery only; patient leakage $\leq$ 10 µA; bias current $\leq$ 10 µA in a fault	Section 14.5
Validation	eyes-closed alpha power at least twice eyes-open in 4 of 5 blocks	a known physiological effect any working EEG must show (Section 14.6)

Table 14.1: A specification for an around-the-ear EEG. Each row has a target with a unit and a reason; the budgets of this chapter check the rows that can be checked on paper.

14.2 *Planning the Work*

DIVIDE THE WORK BY BLOCKS of the system, so that each person owns a block and the interfaces at its edges (Fig. 14.2). For a team of four a common split is: *sensor and fabrication* (electrode or array design, printing, impedance tests), *analog front end and PCB* (schematic, layout, assembly, bring-up), *firmware and wireless* (register settings, data path, Bluetooth, the logging program), and *data, algorithms, and documentation* (protocol, analysis, classifier, the report's figures). One person also keeps the whole chain running end to end each week.

Write the interfaces down: the connector pinout and cable between electrodes and board, the SPI register settings between front end and microcontroller, and the packet format between microcontroller and phone. One past team's planning document, otherwise a good model,

named one front-end chip in its system overview and a different, unrelated converter in its PCB plan. A written interface that the whole team reviews catches that kind of slip in the first weeks instead of the last.

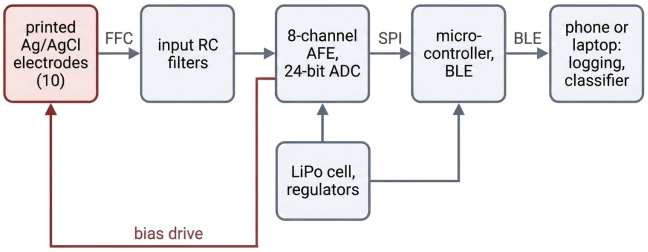


Figure 14.2: Block diagram of an around-the-ear EEG. Each arrow is an interface the team defines in writing; the bias electrode closes a loop from the front end back to the body.

Risks first

List the assumptions the project rests on and ask which one, if false, would sink it. Score each risk by its likelihood and its impact, each from 1 to 5, and multiply. For an ear-EEG team, "the printed electrodes are too noisy or too high in impedance" might score $4 \times 4 = 16$ and "the first PCB does not work" $3 \times 4 = 12$. Test the highest-scoring risk first, with the crudest possible experiment: print a few electrodes, measure their impedance in saline (Lab 2), and record from your own skin with the front end's evaluation board on battery power, before any PCB exists. Then build the thinnest complete path, every block in its simplest form passing real data from electrode to screen, and improve the block that limits the result. A team that perfects its PCB for two months and only then connects it to the electrodes learns at the end of the term that the electrodes do not work.

The schedule

A *Gantt chart* draws each task as a bar on a time axis, with milestones marked. Its most useful property is that it shows dependencies: the PCB must be laid out before it is ordered, ordered before it is assembled, and assembled before it is tested. The longest chain of dependent tasks is the *critical path*; any delay on it delays the end of the project, while tasks off the path have *slack*, time they can slip without delaying anything. In a semester project the hardware chain (specification, layout, fabrication, assembly, bring-up) is almost always the critical path and uses most of the term, while firmware written on an evaluation board and electrodes printed in parallel have a few weeks of slack. Plan backward from the end: reserve the last third of the term for data, analysis, and the report, and collect first data on the evaluation

board with the printed electrodes, so that the data do not wait for the board.

The literature review

The literature review answers two questions: what has been done, and how. Search with a few precise terms ("ear EEG printed electrodes") in Google Scholar, IEEE Xplore, and PubMed, and follow citations in both directions: the references of a good paper and the papers that cite it. Read review papers first; a systematic review of wearable EEG, for example, found that the systems it surveyed used about eight channels on average.[1] Keep a *comparison matrix*, one row per paper and one column per design choice (electrodes, front end, sampling rate, power, battery life, wireless link, what was measured, the result), and write the review by theme, not paper by paper. End with where your project fits and which numbers you take as targets.

[1] C. Anders and B. Arnrich, *Comput. Biol. Med.* (2022).

Reference designs count as literature. Front-end makers publish application notes and evaluation boards for EEG and ECG, and a team that starts from a reference schematic and changes it deliberately does better than one that starts from a blank page. AI tools help with searching, code, and revising your own text, but they also invent plausible references and numbers: treat anything they produce as wrong until you have checked it in a datasheet, a paper, or a measurement, open every reference before you cite it, and say in the report how you used them.

14.3 *Budgets: Checking the Design on Paper*

BEFORE YOU ORDER A BOARD, add up the quantities that most often sink a wearable: the signal against the converter, the noise against the signal, the data against the radio, and the current against the battery. Each budget takes a few minutes and catches mistakes that would otherwise cost a new PCB.

Signal, gain, and resolution

The converter must see the whole signal, including the electrode's DC offset, without clipping, and one least significant bit (LSB), referred back to the input, must be small compared with the smallest feature of interest (Chapter 1 introduced ADC resolution, and Chapter 8 the step size of a front end with gain). For a converter with reference V_{ref}, N bits, and a full-scale range of $\pm V_{\text{ref}}$ after a gain G,

$$\text{range at the input} = \pm\frac{V_{\text{ref}}}{G}, \qquad \text{LSB at the input} = \frac{2V_{\text{ref}}}{G\,2^{N}}. \tag{14.1}$$

KEY EQUATION

$$\text{LSB}_{\text{in}} = \frac{2V_{\text{ref}}}{G\,2^{N}}$$

One converter step referred to the electrodes; the offset must fit in $\pm V_{\text{ref}}/G$.

Skin electrodes have DC offsets of up to a few hundred millivolts, a hundred to a thousand times the signal. Take an ECG patch that must hold a $\pm 300\,\text{mV}$ offset and resolve $5\,\mu\text{V}$. A DC-coupled amplifier into a 12-bit, 3.3 V converter centered on 1.65 V can have a gain of only $(1.65 - 0.15)/0.3 = 5$ (keeping 0.15 V of headroom), so one step is $3.3\,\text{V}/(4096 \times 5) = 161\,\mu\text{V}$ at the input, and a 1 mV R wave spans six codes. A 24-bit front end with a 2.4 V reference at a gain of 6 holds $\pm 0.40\,\text{V}$ and has steps of $2 \times 2.4\,\text{V}/(6 \times 2^{24}) = 48\,\text{nV}$. That is why integrated biopotential front ends use many-bit converters with modest gain (Chapter 8); the other cure is a high-pass filter that removes the offset before most of the gain.

Noise and interference

Independent noise sources add as powers, so the total input-referred noise is the root sum of squares of the amplifier noise, the thermal noise $\sqrt{4kTRB}$ of the electrode resistance R at temperature T in a bandwidth B (k is Boltzmann's constant; Chapter 4), the electrode-skin interface noise, and the quantization noise $\text{LSB}/\sqrt{12}$. The largest term dominates, so find it first.

Worked Example 14.1 *A noise budget for the ear EEG*

The ear EEG's front end, the ADS1299, specifies $1\,\mu\text{V}$ peak to peak of input noise at a gain of 24 (about $0.15\,\mu\text{V}$ rms, taking peak to peak as 6.6 times rms), and its 24-bit converter has $V_{\text{ref}} = 4.5\,\text{V}$. Each electrode has a real impedance of $20\,\text{k}\Omega$ at body temperature (310 K), and a printed gel electrode adds an assumed $0.5\,\mu\text{V}$ rms of interface noise (to be measured). Find the total in the 0.5 to 40 Hz band and compare it with the specification.

Solution. A channel is the difference of two electrodes (active and reference), so each electrode term counts twice. The thermal noise of their $40\,\text{k}\Omega$ in series, in about 40 Hz, is

$$\sqrt{4kTRB} = \sqrt{4\,(1.38 \times 10^{-23})(310)(4 \times 10^{4})(40)}\ \text{V} = 0.165\,\mu\text{V rms},$$

and the two interface terms add to $\sqrt{2} \times 0.5\,\mu\text{V} = 0.71\,\mu\text{V}$ rms. One converter step is $2 \times 4.5/(24 \times 2^{24}) = 22.4\,\text{nV}$, so quantization adds $22.4/\sqrt{12} = 6.5\,\text{nV}$ rms. The total is

$$\sqrt{0.15^2 + 0.165^2 + 0.71^2 + 0.0065^2}\ \mu\text{V} = 0.74\,\mu\text{V rms},$$

within the $1\,\mu\text{V}$ target. Around the ear an eyes-closed alpha rhythm is only about $2\,\mu\text{V}$ rms (an alpha power near $4\,\mu\text{V}^2$, about three times the eyes-open value of Section 14.6), so over the whole band it stands just $20\log_{10}(2/0.74) = 9\,\text{dB}$ above the noise. But the alpha

band, 8 to 13 Hz, holds only 5/40 of the noise power, and within it the rhythm stands $10\log_{10}[4/(0.74^2 \times 5/40)] = 18$ dB above the noise. The interface noise dominates: the electrode, not the chip, decides whether the system meets its specification, which is why the electrodes are tested first. ■

Interference is different from noise. Power-line voltage on the body appears at both inputs as a *common-mode* voltage, and any imbalance ΔZ between the two electrode impedances turns part of it into a differential signal through the amplifier's common-mode input impedance Z_{cm}, about $V_{cm}\,\Delta Z / Z_{cm}$ (Chapter 8). With 0.5 V of 60 Hz on the body, 20 kΩ of imbalance, and 10 pF of input capacitance ($Z_{cm} = 265$ MΩ at 60 Hz), that is 38 µV of hum, larger than the EEG itself, and no common-mode rejection inside the chip can remove it. The bias electrode must lower the common-mode voltage about 38 times, to some 13 mV, and matched electrodes, short leads, and a digital notch filter do the rest.

Data rate and battery life

Multiply out the raw data rate, then compare it with what the link can actually carry. In BLE (Chapter 8) the two ends exchange packets once per *connection interval*; in each interval the device can send one or a few *notifications*, and each carries at most the negotiated packet size (the ATT MTU) minus 3 bytes, only 20 bytes by default. So

$$\text{throughput} \approx \frac{(\text{MTU} - 3) \times (\text{notifications per interval})}{\text{connection interval}}. \tag{14.2}$$

For power, add the average current of each block at the battery, including regulator losses, and divide the usable capacity by it:

$$t_{\text{battery}} = \frac{\eta_{\text{use}}\, Q_{\text{battery}}}{I_{\text{avg}}}, \tag{14.3}$$

where η_{use}, about 0.8, keeps the cell out of deep discharge and covers aging. A linear regulator passes the load current straight through, so the battery current equals the load current; a switching converter trades voltage for current at 80 to 95 % efficiency. Chapter 13 treats batteries and converters.

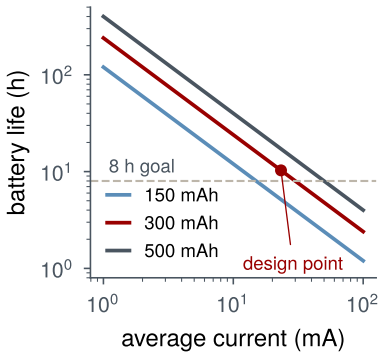


Figure 14.3: Battery life against average current, Eq. (14.3) with $\eta_{\text{use}} = 0.8$. The dot is the ear-EEG design of Worked Example 14.2: 23 mA from 300 mAh.

Worked Example 14.2 *Data and battery budgets for the ear EEG*

Eight channels of 24-bit samples at 250 samples per second stream to a phone. (a) Find the raw data rate. The firmware packs ten samples of all eight channels and a two-byte counter into each notification. Can a link that sends one notification per 15 ms carry it with the default MTU? With an MTU of 247 bytes? (b) Assume

datasheet-level numbers (check your own parts): the front end draws about 5 mW per channel from analog supplies made from the 3.3 V rail at 70 % efficiency; the microcontroller and radio average 6 mA at 3.3 V while streaming; a linear regulator makes 3.3 V from a 3.7 V LiPo cell. Find the battery current and the life from 300 mAh and from 150 mAh.

SOLUTION. (a) $8 \times 24 \times 250 = 48$ kbit/s. A notification holds $10 \times 8 \times 3 + 2 = 242$ bytes and goes out 25 times per second, 6050 B/s. With 20 bytes per notification the same data would need about 303 notifications per second, far more than the 66.7 that one per 15 ms allows. A 247-byte MTU carries 244 bytes, so each packet fits in one notification, and Eq. (14.2) gives $244 \times 66.7 = 16.3$ kB/s, 2.7 times the need. The firmware must request the larger MTU, and the phone must accept it; check the MTU the link actually negotiated, as Lab 4 does.

(b) The front end uses 8×5 mW $= 40$ mW, which costs $40/0.7 = 57.1$ mW from the 3.3 V rail. The microcontroller and radio use $3.3 \times 6 = 19.8$ mW. The rail carries 76.9 mW, or $76.9/3.3 = 23.3$ mA, and the linear regulator draws the same current from the battery. From 300 mAh, $0.8 \times 300/23.3 = 10.3$ h, which meets the eight-hour goal (Fig. 14.3); from 150 mAh, 5.1 h, which does not. The front end is three quarters of the power, so powering down unused channels buys more life than tuning the firmware. ■

For the glove the same kind of budget sets the frame rate: a frame lasts the time of each capacitance conversion times the number of pixels, plus the settling time after every multiplexer switch. At the settings of Lab 2 (3.3 ms per conversion) a 16-pixel frame takes about 0.13 s, 7 to 8 frames per second, close to the rate at which the knitted glove of Section 14.7 recorded grasps; to follow a fast grasp, aim for about 20 frames per second, which needs conversions four times shorter (Lab 2 gives the settings).

14.4 *Building and Bringing Up the Hardware*

THE DESIGN FILES ARE A DELIVERABLE, not a by-product: the sensor fabrication files (stencil artwork, electrode map), the block diagram, the schematic, the PCB layout and its manufacturing files, the firmware, and the analysis code, all in version control. Lab 1 teaches PCB layout with a sensor breakout board, and Lab 3 lays out the front-end board that the electrodes plug into. A few rules matter most for a biopotential board. Start from the front-end maker's reference

schematic. Keep analog inputs short and away from clocks and digital lines, over one continuous ground plane; put a decoupling capacitor at every supply pin; and filter each input with a small series resistor and capacitor against radio-frequency interference. Two past teams lost signal quality to late layout changes: one removed the ground pours in its second revision by mistake, and one routed SPI lines under the front end, where their edges couple into the inputs.

Bring-up: from known signals to people

Bring a board up in stages, each with a known input and an expected output, and do not move on until the current stage passes (Fig. 14.4). Staging separates hardware faults from firmware faults from electrode faults. Firmware for an SPI front end follows the same pattern whatever the chip (Lab 4): reset the chip, read its identification register to prove the link, write the configuration registers and read them back, start conversions, and read a frame each time the data-ready line signals one.

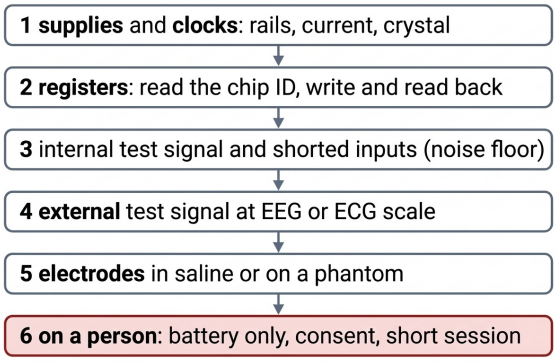


Figure 14.4: Bring-up in stages. Each stage has a known input and an expected output; only a board that has passed stages 1 to 5 touches a person.

Worked Example 14.3 *A test signal at EEG scale*

Make (a) a 100 µV peak-to-peak, 10 Hz test signal from a function generator that cannot go below 1 V peak to peak, and (b) a 100 µV DC level from a 1.5 V cell. (c) A past team tested its EEG board with 60 to 70 mV at a gain of 1. What does that test miss?

Solution. (a) A 10 000:1 divider: $R_1 = 100\,\text{k}\Omega$ in series and $R_2 = 10\,\Omega$ across the input give $1\,\text{V} \times 10/100\,010 = 100\,\mu\text{V}$ peak to peak, from a source impedance of only $10\,\Omega$. Use 1 % resistors and put the divider at the board's input pins so the leads do not pick up hum. The generator is mains powered, which is fine on the bench with no person connected. (b) $R_1 = 1.5\,\text{M}\Omega$, $R_2 = 100\,\Omega$: $1.5\,\text{V} \times 100/1\,500\,100 = 99.99\,\mu\text{V}$. (c) It checks the data path and the scaling

> at gain 1, but not the gain, noise, and saturation at the gain used for EEG (24), where the real signal lives, and a 70 mV signal, 700 times the test level of (a), hides the noise floor. Test at the real gain with a signal of the real size, then with the inputs shorted to measure the noise floor. ■

Design reviews

Hold a design review at three points: when the specification and block diagram are done, before the PCB is ordered, and before the first test on a person. In each, someone who did not do the work goes through a checklist: every requirement has a number and a test; every interface is written down; the budgets of Section 14.3 use the chosen parts' datasheets; the schematic matches the reference design or each difference has a reason; the layout keeps the rules above; the bring-up and test plans exist; and the safety checklist of Section 14.5 is complete. Write down the findings and who will fix each. A review that finds nothing was not a review.

14.5 *Testing on People Safely*

YOUR FIRST PARTICIPANT WILL BE YOU, and your second a teammate. Testing on people is a privilege with rules. This section covers what a course project must do; Sections 14.10 and 14.12 cover the rules for clinical studies and the ethics behind them.

Human-subject rules

United States federal rules (the "Common Rule") define human-subject *research* as a systematic investigation designed to contribute to generalizable knowledge, involving living people from whom investigators obtain data through intervention or interaction. Research must be approved by an *institutional review board* (IRB), the committee that reviews research on people at each institution, before any data are collected. A course project whose only purpose is learning often falls outside that definition, but the team does not decide that: the instructor and the university's IRB office do. If there is any chance you will publish or present the results outside the course, ask for IRB review *before* you collect data; approval cannot be granted afterward. Within a course the usual rules are: test on yourself and on teammates or classmates who agree in writing, after you explain what will be done, for how long, and that they may stop at any time; never tie

participation to a grade; no minors and no patients; no names in data files; and stop at the first sign of discomfort or skin irritation.

Skin contact

Use medical-grade adhesives, gels, and hydrogels within their rated wear time, and fully cure printed inks, since residual solvent irritates skin. Ask participants about allergies to adhesives, silver, and nickel (snap connectors are often nickel plated). Never place electrodes on broken skin, use single-use gel electrodes on one person only, and look at the skin after every session. Check during a long bench run that nothing worn on the body gets warm: the battery, the regulators, the radio.

Electrical safety

A biopotential system connects conductors to a person through low-impedance contacts, sometimes across the chest. Chapter 8 explains what current does to the body and the leakage limits of the medical electrical-equipment standard, IEC 60601-1: in normal conditions, 100 µA of patient leakage for a body-floating (type BF) applied part such as an ECG electrode, and 10 µA for one that may contact the heart (type CF). A course project should stay well below the stricter limit, and the rules follow from that:

- *battery power only* for anything connected to a person; the laptop that logs the data runs on its own battery with its charger unplugged, or receives the data by Bluetooth;
- *no mains-powered instrument* (oscilloscope, bench supply, function generator, LCR meter, a desktop computer over USB) connected to the circuit while electrodes are on a person;
- *limit the current* any electrode can drive, with series resistors on the bias electrode and on any output that touches the skin;
- *no stimulation*: the project records and does not inject current;
- *never wear a device while it charges*.

The bias electrode needs its own limit. If the bias amplifier runs between supplies 5 V apart, a fault could put the full 5 V between two electrodes, so a series resistor of at least $5\,\mathrm{V}/10\,\mu\mathrm{A} = 500\,\mathrm{k}\Omega$ is needed; $1\,\mathrm{M}\Omega$ limits the fault current to 5 µA. In normal use the bias loop carries only the mains current coupled onto the body, about $2\pi \times 60\,\mathrm{Hz} \times 2\,\mathrm{pF} \times 120\,\mathrm{V} = 90\,\mathrm{nA}$, which drops just 90 mV across the resistor, so the loop still works.

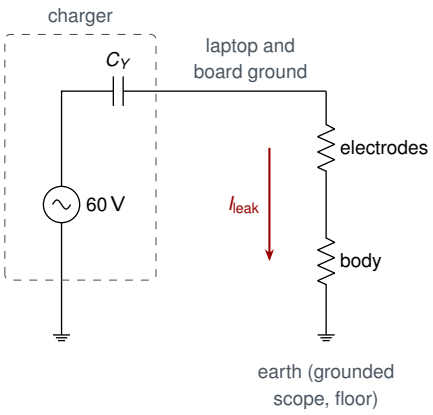


Figure 14.5: How a mains-powered laptop puts current through a person. The charger's isolation (Y) capacitance C_Y couples its primary side, which swings at about half the 120 V line, to the laptop's ground and through the USB cable to the board; electrodes and body close the path to earth.

14.6 *Data, Demonstration, and Report*

DATA YOU CANNOT TRUST ARE WORSE THAN NO DATA, because they lead to confident wrong conclusions. Plan data collection like an experiment. Write a *protocol* before the first session: what is recorded, from where, for how long, under which conditions, in what order, with which reference, and what the participant does at each moment. Mark events (eyes closed, a grasp, a cough) in the data stream itself, with a button or a software marker, rather than on paper. Record a reference whenever one exists: a commercial chest strap for heart rate, a lab EEG system for EEG, labeled objects and video for a glove. Repeat each condition in several short blocks in random order, so that slow drift and fatigue do not masquerade as an effect.

Did it work? Validation

The validation row of the specification is a test whose answer is known in advance. Closing the eyes in a quiet room raises the power of the EEG alpha rhythm (8 to 13 Hz) over the back of the head, and the effect is visible around the ear too (Fig. 14.6).

Suppose a team records five 30-second blocks of each condition and finds closed-to-open ratios of alpha power of 4.0, 2.6, 2.8, 1.7, and 3.4. Four of five exceed 2, so the system passes the last row of Table 14.1. A second check matters as much: if the noise floor near 35 Hz is $0.02\,\mu V^2/Hz$, the 5 Hz alpha band holds only $0.10\,\mu V^2$ of noise, a twelfth of the eyes-open alpha power of about $1.2\,\mu V^2$, so the system records brain activity, not its own noise. An honest report shows all five blocks, including the weak fourth one, and the spectra.

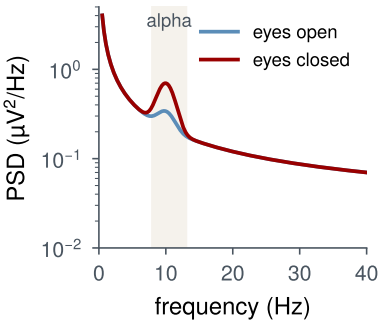


Figure 14.6: The eyes-closed alpha test (illustrative spectra, not a recording). Closing the eyes raises the alpha peak near 10 Hz above a background that falls with frequency.

Managing the data

Name every file so that it sorts and explains itself: date, participant code, condition, and device version (Fig. 14.7). Store raw data read-only and never edit it; put processed data in a separate folder, made by scripts in version control, so that any figure in the report can be regenerated. Write a small metadata file beside each recording with the sampling rate, gain, channel map, electrode type, firmware version, and notes. Use participant codes (S01, S02), keep the list that links codes to names apart from the data, and delete the data when the consent says so. Store samples in binary: eight channels of 3-byte samples at 250 samples per second fill 10.8 MB in a 30-minute session, while the same data written as text in a CSV file take about four times as much space and load more slowly.

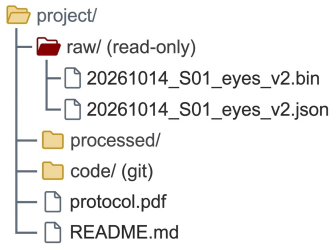


Figure 14.7: A data folder. Raw files are never edited; each has a metadata file; processed data can always be regenerated from raw data and code.

Classifiers and the demonstration

A classifier turns a recording into a decision: relaxed or not, which object was grasped, a normal or an irregular beat. Keep it in proportion. Start with features you can explain (band powers for EEG, such as the alpha/beta ratio of Chapter 8; beat-to-beat intervals for ECG; the pressure map for the glove) and a simple classifier, and add a neural network only if it beats that baseline. Split the data into training and test sets *by participant* or by session, never by random samples from the same recording: neighboring samples are almost identical, so a random split tests the model on data it has already seen and reports an accuracy it will not have on a new person. Report a *confusion matrix* (a table of true against predicted classes), not only an accuracy, and say how many participants and trials it rests on. A small network fits on the wearable itself: a dense network for the glove with about 23,000 weights needs 23 kB at 8 bits, a few percent of the 1 MB of flash on a typical BLE microcontroller board. The limit is training data, not memory.

The demonstration is part of the result. Rehearse it on the hardware you will use, keep a recorded file as a backup, and decide in advance what you will show if the live system fails. A short, working demonstration with a known input (eyes closed, a fist, a labeled object) convinces more than a long, uncertain one.

The report and the presentation

Write the final report as a short journal paper, and write it as you go: the introduction and literature review early, each design section when its block is finished, the results last. A structure that works: an *abstract* with the goal, the system, and the main measured result with

a number; an *introduction* with the need and the project statement; the *state of the art* from the literature review; the *system design*, from the specification table and block diagram through the sensor, front end, firmware, and processing; *testing and results*, bench tests first and then tests on people, each against its specification row; a *discussion* of what met the specification, what did not, and why; and a statement of how AI tools were used, the contributions of each member, and checked references.

Every figure has axes with units, every number has its uncertainty or range, and every claim of success points to a measurement. When something did not work, say so and show the evidence that tells the reader why. "The board read the generator's signal correctly but no ECG was recorded from the printed electrodes; their impedance in saline was 4 kΩ, and the most likely cause is . . . " is a useful result. "The device is comfortable and user-friendly" with no test behind it is not. In the presentation, let every member present the part they owned, show the block diagram rather than every schematic, and put each measured number next to its target.

14.7 *Lessons from Past Projects*

THREE PROJECTS FROM PAST SEMESTERS show what teams build in one term and the problems they share.

A wearable ECG. A team printed six silver electrodes on thermoplastic polyurethane (TPU) and connected them through an FFC to a board with a three-channel, 24-bit front end (ADS1293) and a small Bluetooth microcontroller board. After cutting a trace that reset the front end whenever the microcontroller printed data, the board read a supply and a generator correctly, but on the body the printed electrodes gave only noise, even with gel. The lessons: measure the printed electrodes' impedance before using them on a person (Lab 2), chloride silver to Ag/AgCl for a stable interface (Chapter 6), and keep a known-good path (an evaluation board with commercial electrodes) to tell electrode problems from board problems.

An around-the-ear EEG. Another team built a version of the system of Fig. 14.2, with ten printed silver electrodes joined by serpentine interconnects on TPU and an ADS1299 on a 40×25 mm board. The registers read and wrote reliably on the first day, but an offset and a scaling error, traced to the reference and channel configuration, used up the time left, and the second board revision had lost its ground pours. The lessons: break every channel out to a test point, test at

the real gain with the signal of Worked Example 14.3, and review the layout before each new board.

A capacitive glove. A knitted glove with several hundred pressure sensors and a neural network has learned to identify objects from the way they are grasped.[2] A team built a smaller version, a 4×4 array of 1 cm plates printed on a flexible film over a porous dielectric, read either by one FDC2214 with a multiplexer switching the columns or by two FDC2214 chips. One chip and a multiplexer is cheaper but slower and suffers crosstalk from the floating columns; two chips are faster and simpler. Lab 2 builds a version of this array.

[2] S. Sundaram, P. Kellnhofer, Y. Li, *et al.*, *Nature* (2019).

Across all three, the hardware usually worked before the sensors did, and the time ran out before the data were collected. Plan backward from the end of the term, and test the riskiest part, the electrodes, first.

14.8 What Is a Medical Device?

Suppose a team's printed ECG patch works on the bench: clean P waves, a heart rate that matches a commercial monitor, a battery that lasts a week. Before a cardiologist can prescribe it, new questions arise. Is it a medical device at all? Which rules apply, and who decides? What evidence shows that it is safe and that it works? Could its adhesive irritate the skin, could it shock someone through a charging cable, could someone hack its radio? The rest of this chapter answers them in the order a product meets them.

In the United States, devices are regulated by the Food and Drug Administration (FDA), mainly through its Center for Devices and Radiological Health (CDRH). FDA rules apply to devices sold or distributed in the United States, wherever they are made, and the US is the largest single device market, so most device startups plan for FDA first. In the European Union, devices fall under the Medical Device Regulation (MDR): private *notified bodies* assess most devices, and the maker then affixes the CE mark. Japan, China, Canada, and other countries have their own regulators. All of them sort devices by risk and ask for more evidence as the risk grows, so the logic of this chapter carries over even where the names differ. We follow FDA.

The definition

The Federal Food, Drug, and Cosmetic Act (section 201(h)) defines a *device* in three parts. It is an instrument, apparatus, machine, implant, *in vitro* reagent, or a component or accessory of one. It is intended for

the diagnosis, cure, mitigation, treatment, or prevention of disease, or to affect the structure or a function of the body. And it does not achieve its primary purpose through chemical action in or on the body or by being metabolized; those products are drugs.

The key word is *intended*. FDA reads intended use from the claims a company makes in its labeling, advertising, and website, so the same hardware can be a device or not depending on what is said about it. A sensor patch sold to "track your hydration during training" is, legally, a different product from the same patch sold to "screen for dehydration in elderly patients" (Fig. 14.8).

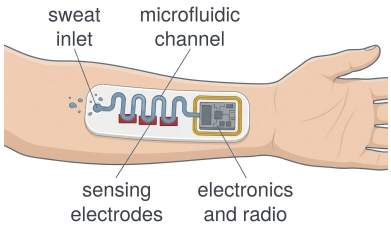


Figure 14.8: A wearable sweat sensor on the forearm. Hydration advice for athletes is a wellness claim; monitoring electrolytes in patients is a device claim. The words on the box, not the hardware, decide which.

Wellness products and software

For wearables the most important edge of the definition is FDA's policy on *general wellness* products: low-risk products intended only to maintain or encourage a healthy lifestyle, which FDA does not regulate as devices. Its 2026 revision of the policy allows a non-invasive wearable to show estimates such as heart rate variability, oxygen saturation, or blood pressure for sleep, fitness, or recovery, provided the values are validated and the product makes no claim about a disease. It may suggest seeing a professional when readings fall outside wellness ranges, but it may not name a condition, call a value abnormal, or provide disease monitoring and alerts. An invasive or implanted product is never a wellness product.

Software is a device when it meets the definition; FDA calls software that is a device on its own *software as a medical device*. Some *clinical decision support* software is excluded: software that displays medical information such as clinical guidelines to a health care professional, who can review the basis of its recommendation independently. But software that analyzes a physiological signal (an ECG, EEG, or PPG) or a medical image is a device however it presents its output, and time-critical alarms, for stroke or sepsis for example, remain devices.

14.9 *Device Classes and Premarket Pathways*

FDA SORTS DEVICES INTO THREE CLASSES by risk, and the class sets the *controls* that give "reasonable assurance" of safety and effectiveness (Fig. 14.9).

Class I devices have the lowest potential for harm: elastic bandages, examination gloves, hand-held surgical instruments. They are subject to the *general controls*: registration of the manufacturer and listing of the device with FDA, rules against adulteration and misbranding, labeling, the quality system (some class I devices are exempt), and re-

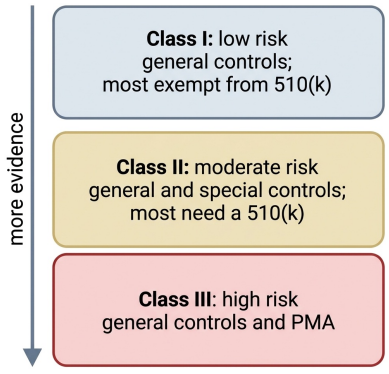


Figure 14.9: FDA device classes. Each class adds controls to the one above it.

porting of adverse events. Most are exempt from premarket review. *Class II* devices have a moderate potential for harm, such as infusion pumps, powered wheelchairs, and ECG electrodes. They also carry *special controls*: performance standards, device-specific testing and labeling, or postmarket surveillance, written into the device's classification regulation. Most need a 510(k). *Class III* devices support or sustain life, are of substantial importance in preventing impairment of health, or present a potential unreasonable risk: replacement heart valves, implanted stimulators. They need premarket approval (PMA). Most active implants are class III.

A device's class is not its maker's opinion. It comes from the *classification regulation* that matches the device's intended use and technology, found in Title 21 of the Code of Federal Regulations (21 CFR parts 862 to 892, one part per medical specialty, such as part 870 for cardiovascular and part 882 for neurological devices) and through three-letter *product codes* in FDA's online classification database. Technology matters as well as use: a spot-reading electronic thermometer is class II and exempt from 510(k), while one that measures continuously needs a 510(k).

The routes to market

The class sets the route to market (Fig. 14.10), and each route ends in a decision with its own word: a 510(k) device is *cleared*, a De Novo request is *granted*, and a PMA device is *approved*. Only class III devices are "FDA approved"; a company that calls its 510(k) device "FDA approved" misbrands it. (And it is "510(k)", after the section of the Act that created it, not "501(k)".)

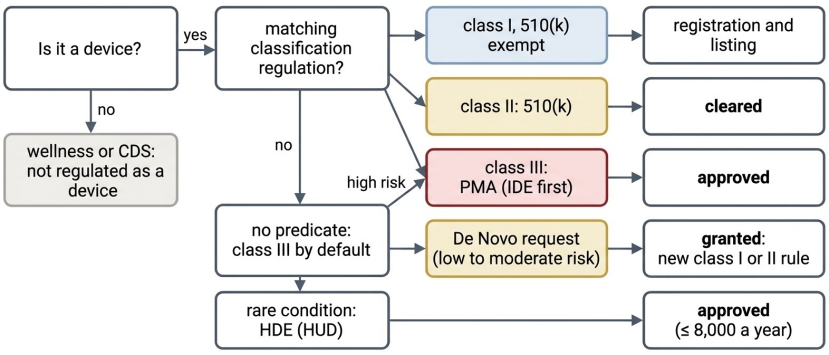


Figure 14.10: Paths to market for a medical device in the United States. The first question, whether the product is a device at all, is the subject of Section 14.8. Most devices follow the class of an existing regulation. A device type with no predicate is class III by default; a De Novo request asks FDA to classify it as class I or II, otherwise it needs a PMA. The Humanitarian Device Exemption serves rare conditions.

510(k) exempt. For most class I and some class II devices, the maker registers, lists the device, follows the general controls, and sells it.

510(k) premarket notification. A *510(k)* shows that a device is *substantially equivalent* (SE) to a legally marketed device, the *predicate*, that does not itself need a PMA. A device is substantially equivalent if it has the same intended use as the predicate and either the same technological characteristics, or different ones that do not raise different questions of safety and effectiveness, with performance data (bench, animal, or clinical) showing that it is at least as safe and effective (Fig. 14.11). A new hip implant that differs from its predicate only in a smaller ceramic head raises the same questions of wear and fracture strength, answered by bench tests, so it can be cleared; a new bearing polymer would raise new questions of wear debris and biocompatibility and could need clinical data or another route. If FDA finds the device not substantially equivalent (NSE), it cannot be sold until a new 510(k), a De Novo, or a PMA succeeds.

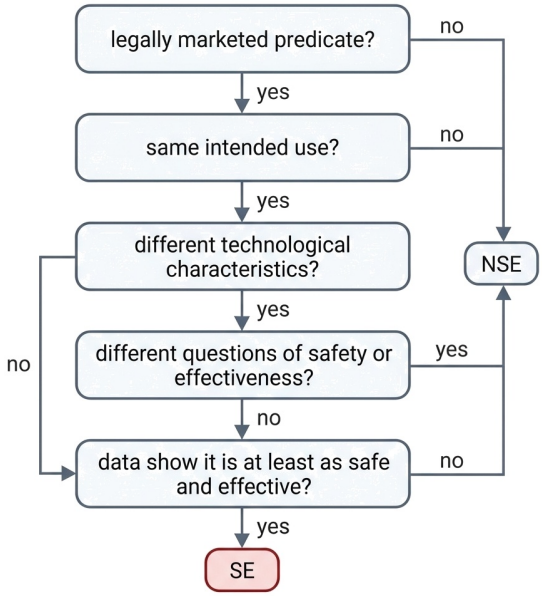


Figure 14.11: The substantial-equivalence decision of a 510(k), simplified from FDA's 510(k) decision flowchart. A "no" to the predicate or the intended use, new questions of safety or effectiveness, or data that fall short all lead to "not substantially equivalent".

De Novo. When a new device has no predicate, the law makes it class III automatically, whatever its actual risk. A *De Novo* request asks FDA to classify such a device, if its risk is low to moderate, into class I or II instead. FDA grants it if general controls, or general and special controls, together with valid scientific evidence (usually including clinical data) give reasonable assurance of safety and effectiveness. A granted De Novo creates a new classification regulation with its special controls, and the device becomes a predicate, so competitors can follow by 510(k). De Novo is not an approval as class III; it is the way out of class III for a novel device whose risk does not warrant a PMA. The landmark case for this book: continuous glucose monitors (Chapter 6)

were long approved as class III devices, until in 2018 FDA granted a De Novo request for an *integrated* CGM, one designed to work with insulin pumps, and created a class II regulation (21 CFR 862.1355) with special controls on accuracy, data transmission, and labeling. Later CGMs of that type have been cleared by 510(k).

Premarket approval. A *PMA* is FDA's full review of the safety and effectiveness of a class III device: a detailed device description, extensive bench and animal data, clinical data from studies run under an investigational device exemption (Section 14.10), and manufacturing information. An advisory panel of outside experts reviews many PMAs, and FDA inspects the manufacturer before approval. Changes to an approved device need PMA supplements.

Humanitarian Device Exemption. A *Humanitarian Use Device* (HUD) treats or diagnoses a condition that affects no more than 8,000 people in the United States per year. A *Humanitarian Device Exemption* (HDE) approves it without proof of effectiveness: the applicant shows that it poses no unreasonable risk, that its probable benefit outweighs its risks, and that no comparable device is available. In return the device generally may not be sold for profit, and each facility's IRB must approve its use there.

FDA's *Breakthrough Devices Program* gives priority review and more contact with FDA to devices that treat or diagnose life-threatening or irreversibly debilitating conditions better than existing options; it is not a separate pathway. Whatever the route, talk to FDA early: a *Pre-Submission* gets written feedback and a meeting on the classification, the predicate, the test plan, and the clinical protocol before most of the money is spent.

Worked Example 14.4 *Six devices from this book*

Give the likely class and premarket pathway for: (a) a printed ECG patch that records for a week and flags arrhythmias; (b) a continuous glucose monitor that works with insulin pumps; (c) a deep brain stimulation (DBS) system for Parkinson's disease; (d) a retinal implant for blindness from retinitis pigmentosa; (e) an EEG headband sold to help people meditate; (f) an ingestible capsule camera for the small intestine.

Solution. (a) Class II: ECG electrodes (21 CFR 870.2360) and arrhythmia detection software (870.1025, with special controls) are both class II. A 510(k) against a cleared ambulatory ECG patch, with bench tests to the ECG standards, biocompatibility, electrical safety, and clinical validation of the algorithm. (b) Class II under

the iCGM regulation created by the 2018 De Novo; a new one uses a 510(k) and must meet its special controls. (c) Class III: DBS systems (Chapter 9) are approved by PMA, and changes go through PMA supplements. (d) Class III; the Argus II epiretinal implant (Chapter 10) reached the US market in 2013 through an HDE, because few people a year have retinitis pigmentosa severe enough to qualify. A new retinal implant would need an HDE or a PMA, with an IDE study first. (e) A wellness product if it claims only relaxation or focus; a diagnostic electroencephalograph is class II, so "detects seizures" would move it into device regulation. (f) Class II with special controls (21 CFR 876.1300), cleared by 510(k). ■

14.10 *Clinical Evidence*

Testing an unapproved device on people is itself regulated. The *investigational device exemption* (IDE, 21 CFR part 812) allows a device that is not yet cleared or approved to be used in a clinical study to collect safety and effectiveness data, while protecting the people who test it.

Every study falls into one of three groups (Fig. 14.12). Some are *exempt*, for example studies of a device used according to its cleared labeling. The rest are significant risk or nonsignificant risk. A *significant risk* (SR) study involves an implant, a device that supports or sustains life, a device of substantial importance in diagnosing or treating disease, or another device with a potential for serious risk to the participants. It needs an IDE approved by FDA and approval by the IRB of each institution (21 CFR part 56).

A *nonsignificant risk* (NSR) study, such as testing a wrist PPG sensor on healthy volunteers with no clinical decisions made from it, needs only IRB approval, though the sponsor must still obtain informed consent (21 CFR part 50), monitor the study, keep records, and report adverse events. The sponsor proposes a risk level, but the IRB decides whether a study is NSR, and FDA has the final word if asked. A cortical microelectrode array implanted in people with paralysis is SR; a sweat patch that only records, beside the blood draws a patient's care already requires, is usually NSR.

Device studies grow their evidence in steps. An *early feasibility study* tests a device early in development, often before the design is final, in a few participants, to learn whether the concept works and how to improve it. A *feasibility study* gathers preliminary safety and effectiveness data on a nearly final design to plan the main study. A *pivotal study* collects the definitive evidence of safety and effectiveness for a stated

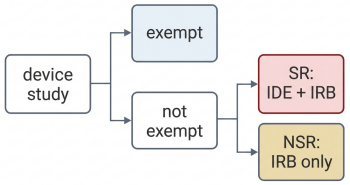


Figure 14.12: Study types under the IDE rules. Significant risk studies need FDA approval of an IDE and IRB approval; nonsignificant risk studies need IRB approval and follow abbreviated requirements.

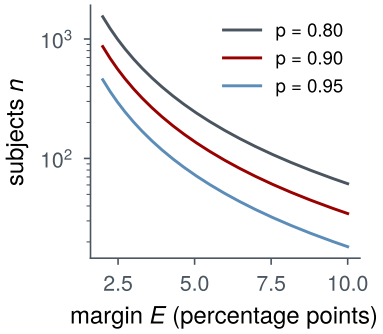


Figure 14.13: Participants needed, Eq. (14.4), versus the margin of the 95 % confidence interval for three expected proportions.

intended use, with a sample size justified by statistics.

How many participants?

A pivotal study must show its result with a stated confidence. For an accuracy claim estimated as a proportion p, such as the *sensitivity* of an arrhythmia detector (the fraction of true cases it detects), the number of participants that gives a 95 % confidence interval of half-width E is about

$$n = \frac{z^2\, p(1-p)}{E^2}, \qquad z = 1.96, \tag{14.4}$$

which grows rapidly as the margin shrinks (Fig. 14.13): halving the margin quadruples the study. Real studies use endpoints agreed with FDA, hypothesis tests, and allowances for dropouts, but this estimate sets the scale.

KEY EQUATION

$n = z^2 p(1-p)/E^2$

Participants needed to estimate a proportion within $\pm E$ at 95 % confidence ($z = 1.96$).

WORKED EXAMPLE 14.5 *Sizing a pivotal study*

An ECG patch is expected to detect atrial fibrillation with a sensitivity of 90 %. (a) How many patients with atrial fibrillation are needed to estimate the sensitivity within ±5 %? Within ±3 %? (b) If 20 % of the people enrolled turn out to have atrial fibrillation and 10 % drop out, how many must be enrolled for the ±5 % study?

SOLUTION. (a) $n = (1.96)^2(0.90)(0.10)/(0.05)^2 = 138.3$, so 139 patients with atrial fibrillation; for ±3 %, $n = 0.3457/0.0009 = 384.1$, so 385. (b) $139/(0.20 \times 0.90) = 772.2$: enroll about 773 people. The sensitivity claim is limited by the number of positive cases, not by the total enrollment. ■

14.11 *Quality Systems and Standards*

FDA REGULATES HOW A DEVICE IS MADE, not only what it is. Class II and III devices, and many class I devices, must be designed and manufactured under a *quality system*, and its records feed every submission. Since February 2026 FDA's rule for it, the Quality Management System Regulation, incorporates the international standard ISO 13485, so one quality system can satisfy FDA and most other regulators.

Design controls

The heart of the quality system for an engineer is *design control* (Fig. 14.14), and it is the design project of this chapter done formally. User needs become written *design inputs*, requirements that can be tested, like the rows of Table 14.1. The design produces *design outputs*: drawings, specifications, software, labeling. *Verification* proves that the

outputs meet the inputs ("did we build the device right?"): the bench tests, analyses, and inspections of Sections 14.3 and 14.4. *Validation* proves that the finished device meets the user needs in actual or simulated use ("did we build the right device?"): usability testing and clinical evaluation, the clinical version of the alpha test. Formal design reviews happen at planned stages, changes are controlled, and everything is recorded in the design file. Devices made for a clinical study under an IDE are not exempt from design controls, so a startup starts its quality system with its design inputs, not after its prototype works.

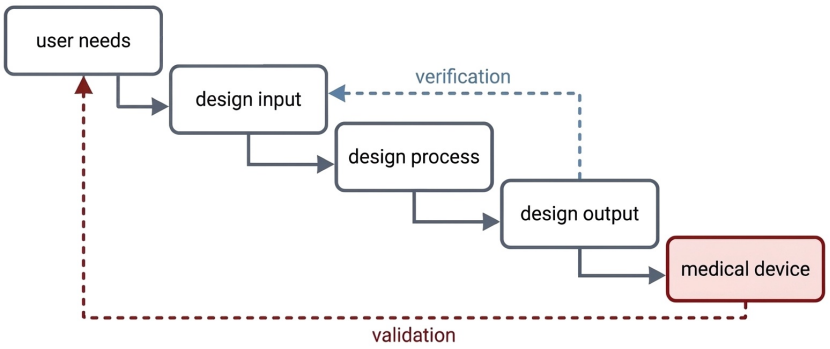


Figure 14.14: Design controls as a waterfall. Verification checks each design output against its design input; validation checks the finished device against the user needs. Reviews take place between stages, and in practice the stages repeat.

Risk management

Risk management follows ISO 14971. The maker identifies *hazards*, potential sources of harm, the situations in which people are exposed to them, and the resulting *harms*. It estimates each risk from the severity of the harm and the probability that it occurs, judges it against criteria set in advance (Fig. 14.15), and controls unacceptable risks in a fixed order: inherent safety by design first, then protective measures, then information for safety (warnings and training).

For a rechargeable ECG patch with Bluetooth, three hazards stand out. *Skin irritation from the adhesive and gel* over a week of wear: choose materials with a record of biocompatibility and test them, limit the wear time, then warn in the labeling. *Electric shock through the charging cable*: design the patch so that it cannot charge while worn, or isolate the charger to meet the leakage limits, then warn. *A missed arrhythmia* because of a dropped link: store data on the patch and resend it, alert the user when the link is lost, and validate the detection algorithm clinically. Each risk is then re-estimated, and a residual risk is accepted only if it cannot practicably be reduced further and the benefit outweighs it.

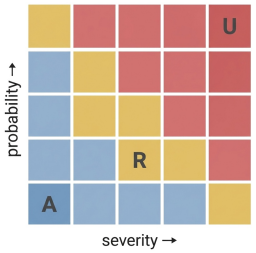


Figure 14.15: An example risk acceptability matrix: A, acceptable; R, reduce as far as possible and justify; U, unacceptable. ISO 14971 asks the manufacturer to define its own criteria; this one is illustrative.

The standards behind the tests

FDA recognizes consensus standards, and a declaration of conformity to one can replace much of the test data in a submission. Four matter for almost every device in this book.

- **Biocompatibility, ISO 10993-1**: a risk-based biological evaluation of everything that touches the body, set by the type of contact (skin, mucosa, blood or tissue, implant) and its duration (up to 24 h, up to 30 days, or longer). A skin patch usually needs cytotoxicity, sensitization, and irritation tests; an implant needs much more (Chapter 11).
- **Electrical safety, IEC 60601-1**, with companion standards for electromagnetic compatibility and home use; it sets the leakage limits of Section 14.5.
- **Software, IEC 62304**: software life-cycle processes, scaled by whether a software failure could cause no injury, a non-serious injury, or death or serious injury.
- **Cybersecurity**: the submission of a device whose software can connect to a network must include a plan to monitor and fix vulnerabilities, with updates and patches, and a *software bill of materials* listing every software component.

After the market, and getting there

Clearance or approval starts a second set of duties. The maker reports deaths, serious injuries, and malfunctions that could cause them; FDA publishes these reports in its MAUDE database, a good place to learn how similar devices fail before you design your own. A correction or removal made to reduce a risk to health, a recall, must also be reported. Every device label carries a unique device identifier, and a change that could significantly affect safety or effectiveness, or a new intended use, needs a new 510(k) or a PMA supplement.

Put together, the path from a working prototype to a cleared device runs in this order: the intended use and the classification first, because they set everything else; a quality system with design controls from the design inputs on; a Pre-Submission; development with verification to the standards above; clinical studies where needed; the submission; and then manufacturing, launch, postmarket surveillance, and the separate question of whether insurers will pay for the device. It takes years, not months, and for a class III device much longer. Regulation is a barrier to entry that cuts both ways: it makes a device company slow and expensive to start, and it protects one that gets through. So talk early to the people who will use, prescribe, and pay for the device, and put the regulatory strategy into the first business plan.

14.12 *Ethics of Human Studies and Devices*

Regulation sets minimum rules; ethics asks what we owe the people who test and wear our devices. Bioelectronics raises the old questions of human research and some new ones, because its devices read the body continuously and some of them read, or change, the brain.

The Belmont principles and informed consent

The *Belmont Report* (1979), written after abuses such as the Tuskegee syphilis study, states three principles behind the US rules for research on people (the Common Rule, and FDA's parts 50 and 56). *Respect for persons*: treat people as autonomous agents and protect those with diminished autonomy, which leads to informed consent. *Beneficence*: do no harm, and maximize benefits while minimizing risks, which leads to a systematic weighing of risks and benefits by the investigator and the IRB. *Justice*: distribute the burdens and benefits of research fairly, which leads to fair selection of participants, neither exploiting vulnerable groups nor excluding groups that will use the device.

Informed consent is a process, not a signature. The participant must understand the purpose, the procedures, the foreseeable risks and discomforts, the possible benefits, the alternatives, how their data will be kept confidential, whom to contact, and that taking part is voluntary and can be stopped at any time without penalty. Payment large enough to be hard to refuse, or a request from someone with power over the participant, such as an instructor, undermines it.

Privacy

In the United States the HIPAA Privacy Rule protects health information held by health plans, providers, and their business associates. A hospital's remote monitoring service is covered; a consumer fitness band and its app, sold directly to users, usually are not, although the Federal Trade Commission requires many health apps and connected devices outside HIPAA to tell users about data breaches. Data from wearables are detailed enough to reveal pregnancy, illness, location, and daily routines, so good practice collects only what is needed, keeps it on the device when possible, encrypts it, and states plainly who can see it.

Neural devices

Brain-computer interfaces and neurostimulators (Chapter 9) raise questions that wearables only hint at. *Mental privacy*: neural data can reveal intentions, emotions, and health conditions the person has not chosen to share; Colorado and California added neural data to the sensitive data protected by their privacy laws in 2024, and in 2025 UNESCO adopted a global Recommendation on the Ethics of Neurotechnology. *Agency and identity*: when a stimulator changes mood or behavior, or a decoder turns attempted speech into text, who is the author of the result? *Device abandonment*: an implant depends on its maker for programming, repairs, parts, and software. The maker of the Argus II retinal prosthesis stopped producing it in 2019 and then cut back its support, leaving implanted users facing the loss of the vision it gave them when parts failed. A first-in-human study of a neural implant should therefore plan, and fund in advance, long-term support or explantation if the company stops, and its consent form should say so, along with what neural data are recorded, who owns them, and what happens to the device at the end of the study.

Equity and bias

A device that works less well for some groups is unsafe for them. Pulse oximeters, which can read falsely high in people with darker skin and hide low oxygen levels (Chapter 7), are the documented case, and FDA now asks for validation across skin tones. The lesson generalizes: calibrate and validate on the population that will use the device, report performance by subgroup, and check algorithms for bias in the data they were trained on. A project that tests only on its own team has learned how the device works on a few young adults, and its report should say so.

Summary

- Start from a measurable goal. Every row of the specification (signal, bandwidth, noise, channels, resolution, power, battery life, size, safety, validation) has a target, a unit, and a reason.
- Give each block and its interfaces one owner, test the riskiest assumption (usually the electrodes) first, build a thin end-to-end path early, and know the critical path. Check every reference and every AI-produced number.
- Budgets on paper: input range $\pm V_{\text{ref}}/G$ and LSB $2V_{\text{ref}}/(G2^N)$; noise adds as a root sum of squares; imbalance hum $V_{\text{cm}}\Delta Z/Z_{\text{cm}}$; BLE throughput $(\text{MTU}-3)\times n/\text{interval}$; battery life $0.8\,Q/I_{\text{avg}}$.

- Bring up in stages with known inputs, from supplies to a person. On a person: consent, battery power only, no mains-connected instruments, limited bias current, IRB review before any data you might publish. A mains laptop can drive tens of microamperes through the chest (limits 100 µA BF, 10 µA CF).
- Write a protocol, record a reference, keep raw data read-only with metadata, split data by participant, and report measured numbers against the specification, including what failed.
- A device is defined by its intended use (claims); wellness products are outside, and software that analyzes a physiological signal is a device. Class I: general controls, mostly 510(k) exempt. Class II: special controls, mostly a 510(k) (cleared, by substantial equivalence to a predicate). Class III: PMA (approved). With no predicate, a De Novo grant creates a new class I or II regulation. HDE: up to 8,000 patients a year.
- Clinical evidence: SR studies need an FDA-approved IDE and the IRB, NSR studies the IRB; early feasibility, feasibility, and pivotal studies; $n = z^2 p(1-p)/E^2$.
- Quality system: design controls (verification and validation), ISO 14971 risk management, ISO 10993, IEC 60601-1, IEC 62304, cybersecurity. Ethics: respect for persons, beneficence, and justice; informed consent; privacy; device abandonment; validation across the people who will use the device.

This chapter closes the book. The ideas of its fourteen chapters, from the membrane of a neuron to a regulated product, meet in the design project, and the four labs of Part V build its hardware: Lab 1 lays out the sensor breakout board, Lab 2 prints the electrodes and a capacitive array, Lab 3 designs the analog front-end board, and Lab 4 streams its data over SPI and Bluetooth.

14.13 *Problems*

14.1 • Write a specification table (at least ten rows, each with a target, a unit, and a reason) for a single-lead ECG chest patch meant to record heart rate and rhythm for 24 hours.

14.2 • An ECG patch averages 1.2 mA while recording and storing to memory, and 4.0 mA for 2 min each hour while it uploads by BLE.

(a) Find the average current.

(b) What battery capacity gives seven days with $\eta_{\text{use}} = 0.8$?

14.3 • A laptop charger has 4.7 nF of isolation capacitance and the line is 120 V, 60 Hz (the output floats at half the line voltage). Find the leakage current through a person connected as in Fig. 14.5. Repeat for a 230 V, 50 Hz line.

14.4 ◦ Is each product a medical device? Explain.

(a) A smartphone app that gives a nightly "sleep score" from the phone's microphone.

(b) A thermometer patch for infants labeled "alerts you to fever".

(c) A smart ring that reports heart rate variability and a "readiness" score for training.

(d) Software that reads a 12-lead ECG and reports "possible myocardial infarction" to the emergency physician.

14.5 •• A wrist device is to be validated for detecting atrial fibrillation, with an expected sensitivity of 92 % and specificity of 96 % (the fraction of people without atrial fibrillation it correctly clears).

(a) How many participants with atrial fibrillation give a 95 % interval of ± 4 % on the sensitivity?

(b) How many without atrial fibrillation give ± 2 % on the specificity?

(c) If 15 % of enrolled participants have atrial fibrillation, which requirement sets the enrollment, and how many must be enrolled (no dropouts)?

Part V

Laboratory

Lab 1
PCB Layout and Gerber Files

Lab 1 at a glance

What you build. A two-layer *sensor breakout* board, 30 mm by 20 mm, that brings the ten contacts of a printed sensor out to pins you can plug into. The sensor arrives on a flat flexible cable (FFC), a thin plastic ribbon with ten printed conductors 1.0 mm apart (the spacing of neighboring contacts is called the *pitch*). The board carries a 10-position FFC connector, routed to a 10-pin header with the common 2.54 mm pitch of breadboards; a decoupling capacitor (a small capacitor across the supply that keeps its voltage steady); a power LED and its resistor; and two test points, bare pads where a meter probe can touch. You take it from schematic to manufacturing files in KiCad, a free circuit-board design program. The board is for bench work with printed sensors that make no electrical contact with the body, such as the capacitive array of the glove project of Chapter 14, whose four rows and four columns fit the eight signal pins. Its pin 1 carries 3.3 V and pin 10 ground, so an electrode tail (ECG or EEG) must never plug into it: it would put 3.3 V DC between two electrodes on the skin. Electrode tails plug only into the FFC connector of the front-end board of Lab 3.

Time. About 2.5 hours.

Before the lab. Install KiCad (version 8 or later) and open it once. Read Section 1.11 on the anatomy of a bioelectronic system and the overview of the design project in Chapter 14; if resistors, capacitors, and Ohm's law are new to you, read the circuit essentials of Chapter 1 first. Download the datasheet of the FFC connector you will use, and find the capability table of the board manufacturer you plan to order from.

You hand in. A zip file with the Gerber and drill files; the BOM as a CSV file; a PDF of the schematic; a screenshot of

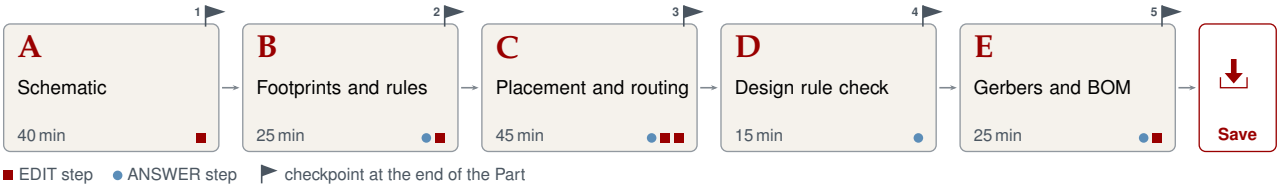
the DRC report with zero errors; a screenshot of the board in a Gerber viewer; and your answers to the ANSWER steps and the analysis questions. The Background explains each of these terms.

What you need

Hardware. A computer (Windows, macOS, or Linux) with a mouse that has a scroll wheel. Optional: the FFC connector and a cable, calipers, and a 1:1 paper print of your board to check footprints against real parts.

Software. KiCad 8 or later (free, from `kicad.org`), which includes the schematic editor, the PCB editor, the footprint assigner, and the GerbView Gerber viewer. A manufacturer's online Gerber viewer is a useful second check. Menu names below are from KiCad 8; later versions move a few items.

Files. Your connector's datasheet and your manufacturer's capability table. Older designs you inherit are often Autodesk EAGLE projects (`.sch` and `.brd`). Autodesk ended EAGLE on June 7, 2026, and its commercial successor is the electronics workspace of Autodesk Fusion (Fusion Electronics). KiCad opens the old files directly: in the KiCad project manager choose *File → Import Non-KiCad Project* and select the EAGLE project; then check every footprint and the design rules, which do not always translate exactly.



Objectives

- Translate a set of front-end decisions into a schematic with symbols, power nets, and net labels, and check it with the electrical rules checker.
- Choose and verify footprints against a datasheet.
- Set a board stack-up and design rules from a manufacturer's capability table, and size trace widths for current with the IPC-2221 rule.

- Place and route a two-layer board with a ground plane, readable silkscreen, and no design rule violations.
- Generate Gerber, drill, and BOM files and inspect them in a Gerber viewer before ordering.

Background

What a PCB is. A printed circuit board is an assembly that uses copper conductors on an insulating board to connect electronic components. It also holds the components mechanically, so that the device can be mounted in an enclosure. Copper foil is laminated to a glass-fiber and epoxy board (FR-4), etched into a pattern, coated with a protective layer, drilled, and plated. Two drawings describe a board. The *schematic* shows which component pins are connected; each set of pins joined together is a *net*. The *layout* shows where each part sits and where the copper runs. Before computer-aided design (CAD), layouts were drawn by hand on large sheets over a light table and the sheets were used in manufacturing; today a layout program writes standard manufacturing files that the machines read directly.

Traces. A *trace* (or track) is a copper path on the board that connects two pads. Its thickness is set by the copper weight, the mass of copper per square foot of board (1 oz is 35 µm thick; 2 oz is 70 µm), and its width by you. The width must be at least the manufacturer's minimum, and wide enough to carry the current without heating too much. The IPC-2221 design standard gives a rule of thumb, fitted to measurements, for the current I (in amperes) that a trace of cross-section A (in square mils, 1 mil = 0.0254 mm) carries with a temperature rise ΔT (in °C):

$$I = k\,\Delta T^{0.44} A^{0.725}, \tag{L1.1}$$

with $k = 0.048$ for outer layers and $k = 0.024$ for inner layers. Inner layers get half the current because they cannot shed heat to the air. The width is A divided by the copper thickness (1.378 mil for 1 oz). For 1 A and $\Delta T = 10$ °C on an outer layer of 1 oz copper the rule asks for about 0.30 mm; a 0.25 mm trace carries about 0.9 A. Sensor signals carry microamperes, so for them the manufacturer's minimum, not current, sets the width (Fig. L1.1).

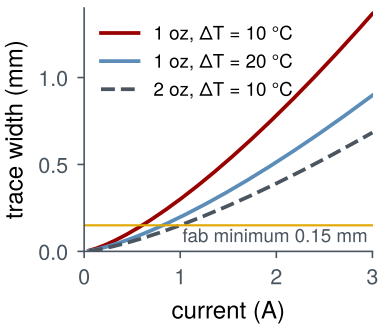


Figure L1.1: IPC-2221 trace width versus current on an outer layer. Below about 0.6 A the manufacturing minimum of 0.15 mm (gold line) is wider than the current requires.

KEY EQUATION

$I = k\,\Delta T^{0.44} A^{0.725}$

IPC-2221 trace current; A in mil^2, k = 0.048 outside, 0.024 inside.

Pads. *Pads* are exposed metal areas where component leads are soldered. *Through-hole* pads surround a plated hole for a component with wire leads, such as a pin header. *Surface-mount* pads are flat copper areas for surface-mount devices (SMD), parts that sit on the surface,

such as the 0805 resistor, capacitor, and LED of this board (0805 means 0.08 in × 0.05 in, or 2.0 mm by 1.25 mm).

Layers and stack-up. Modern boards use several copper layers so that traces need not crowd onto one. The *stack-up* lists them with the insulating layers between them: a rigid *core* and sheets of *prepreg*, glass cloth with partly cured resin that bonds the layers in a press (Fig. L1.2). Some layers are given to one net; an inner layer of solid ground, for example, gives every signal a short return path (the path its current takes back to the source). Our board has two layers on a 1.6 mm core.

Vias. A *via* is a plated hole that carries a connection from one layer to another. A *through* via goes through the whole board; a *blind* via joins an outer layer to an inner one; a *buried* via joins two inner layers and is invisible from outside. Blind and buried vias save space in dense boards and cost more. A via is described by its drill diameter and its pad diameter; the ring of copper between them is the *annular ring*.

Solder mask, paste, and silkscreen. The *solder mask* is the thin polymer coating (usually green) that covers the copper except at the pads, so that solder goes only where it should. The *paste* layer defines the openings of the stencil through which solder paste is applied to SMD pads. The *silkscreen* is the printed ink layer with component outlines, reference designators (R1, C1), pin-1 marks, part values, and other text that helps people assemble, test, and identify the board. It must never cover pads.

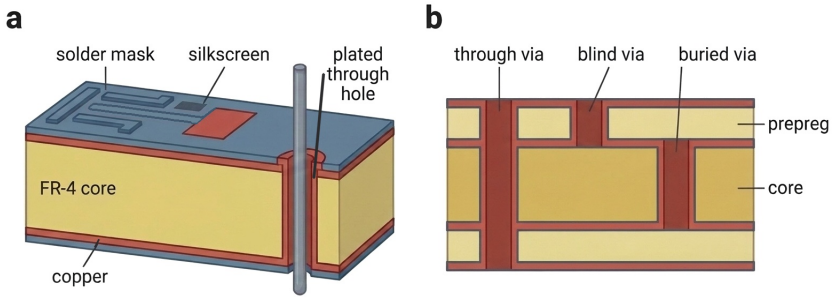


Figure L1.2: Board cross sections (not to scale). (a) The two-layer board of this lab: copper (cardinal) on both faces of an FR-4 core, solder mask (steel) everywhere except pads, silkscreen ink on top, and a plated through hole. (b) A four-layer stack-up of core and prepreg; a through via crosses all layers, a blind via starts at an outer layer, a buried via stays inside.

Design rules and the capability table. A manufacturer can only make features down to certain sizes. Its *capability table* (Table L1.1) lists the minimum trace width and spacing, the smallest drill and annular ring, the board thicknesses and copper weights it offers, and the tolerances. You enter these numbers as *design rules*, and the design rule checker (DRC) flags anything that breaks them. Designing to the minimum

costs more and fails more often, so use the recommended values when the board has room, as ours does.

Item	Capability	This lab uses
Number of layers	1 to 10	2
Material	FR-4, aluminum	FR-4
Board thickness	0.2 to 2.4 mm	1.6 mm
Thickness tolerance	$\pm 10\,\%$ ($t \geq 1$ mm)	
Outline tolerance	± 0.2 mm (CNC routing)	
Minimum trace width	0.1 mm (4 mil); 0.15 mm suggested	0.15 mm minimum, 0.25 mm signal, 0.5 mm power
Minimum spacing	0.1 mm (4 mil); 0.15 mm suggested	0.15 mm minimum, 0.2 mm default
Outer copper	1, 2, or 3 oz (35, 70, 105 µm)	1 oz
Minimum drill	typically 0.3 mm	0.3 mm (vias)
Minimum via pad	typically 0.6 mm	0.6 mm
Annular ring	typically 0.15 mm	0.15 mm
Copper to board edge	typically 0.3 mm	0.3 mm
Silkscreen line, text	typically 0.15 mm, 1 mm high	0.15 mm, 1 mm

Table L1.1: A manufacturer capability table and the rules used in this lab. The first group of rows follows a typical low-cost prototype service; the rows marked "typically" are common values. Always use your own manufacturer's table.

Gerber, drill, and BOM files. The standard manufacturing files are *Gerber* files (format RS-274X, or the newer X2 with layer attributes), one for each layer: front and back copper, mask, paste, and silkscreen, and the board outline. The holes go in a separate *drill* file in Excellon format. A *bill of materials* (BOM) lists every part with its reference, value, footprint, and manufacturer part number, and a *position* file gives the location and rotation of each part for a pick-and-place machine. A bare-board order needs only the Gerbers and the drill file; an assembled order also needs the BOM and the position file.

The design flow. Figure L1.3 shows the steps. The *front-end* decisions fix what the board does: the parts it connects to, its voltages, chips, layers, and size. The *schematic* connects component symbols with wires and net labels, and each symbol gets a *footprint*, its pattern of pads. *Placement* groups parts by function in standard orientations, keeps noisy and quiet sections apart and parts away from heat, puts connectors at the edges, and leaves room for traces and a solid ground plane. *Routing* replaces the *ratsnest* lines (which show the connections still to make) with copper. Fast digital signals, and pairs of signals that travel together (differential pairs), need traces of a controlled width and spacing; an *autorouter*, a program that draws traces by itself, can do the noncritical rest. The *design review* runs the DRC for overlaps, spacing, edge distance, via size, and annular ring. Last come the manufacturing files.

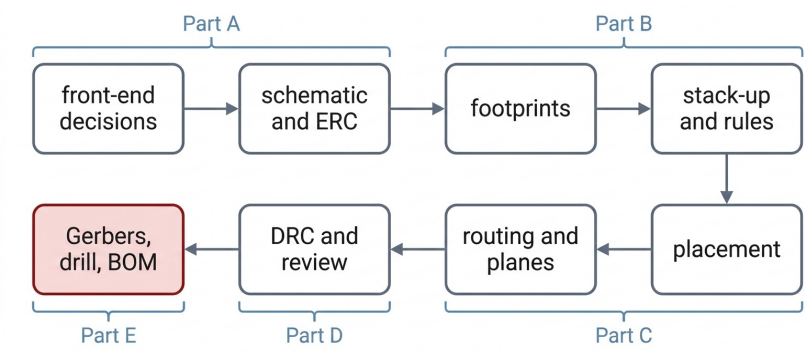


Figure L1.3: The PCB design flow and the Parts of this lab. Errors found late are expensive: a wrong footprint found after the boards arrive costs a new order.

The board. The front-end decisions for this lab are already made:

- Function: connect a printed sensor on a flat flexible cable to the 2.54 mm headers of a breadboard, an evaluation kit, or a microcontroller.
- Connector J1: 10 positions, 1.0 mm pitch, FFC, at the top edge. Pin 1 is 3.3 V, pins 2 to 9 are signals S1 to S8, pin 10 is ground. Because pins 1 and 10 carry the supply, an electrode tail must never plug into J1.
- Connector J2: a 1×10 pin header at 2.54 mm pitch at the bottom edge, with the same pin order.
- C1: 100 nF decoupling capacitor between 3.3 V and ground, next to J1. R1 and D1: a power LED with its series resistor. TP1 and TP2: test points on 3.3 V and ground.
- Two layers, 1.6 mm FR-4, 1 oz copper, 30 mm $\times$ 20 mm; ground pour on both layers; supply current at most 0.5 A.

Figure L1.4 is the schematic and Fig. L1.5 a finished layout.

Procedure

Work through the Parts in order and save after every step (*File* $\rightarrow$ *Save*, or Ctrl+S). Each Part ends with a checkpoint; do not go on until it passes. Keep a lab notebook (a text file is fine) for the ANSWER steps.

Part A: Front-End Decisions and the Schematic

GOAL Draw the schematic of Fig. L1.4 and pass the electrical rules check. about 40 min

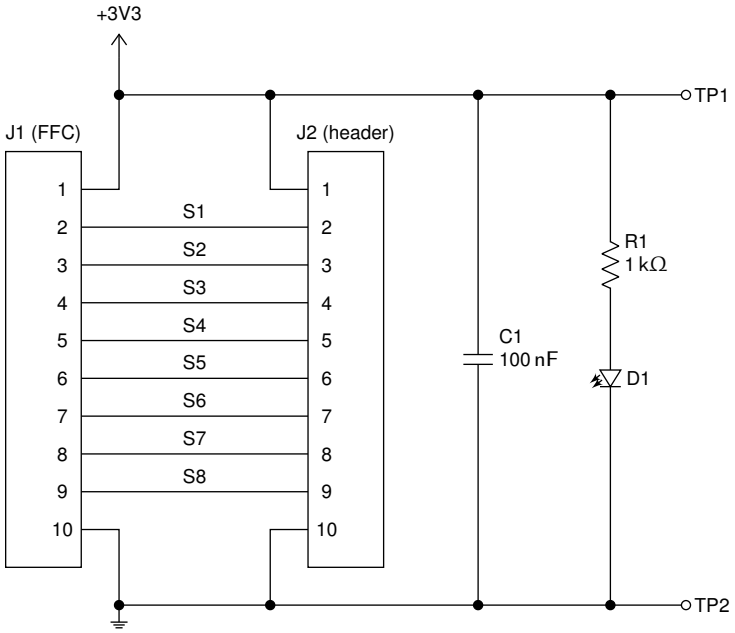


Figure L1.4: Schematic of the sensor breakout. Pins 2 to 9 of J1 and J2 share the net labels S1 to S8; pin 1 is 3.3 V (+3V3) and pin 10 is ground on both. C1 decouples the supply, R1 and D1 show that power is present, and TP1 and TP2 give a probe a place to land. The board has 11 nets.

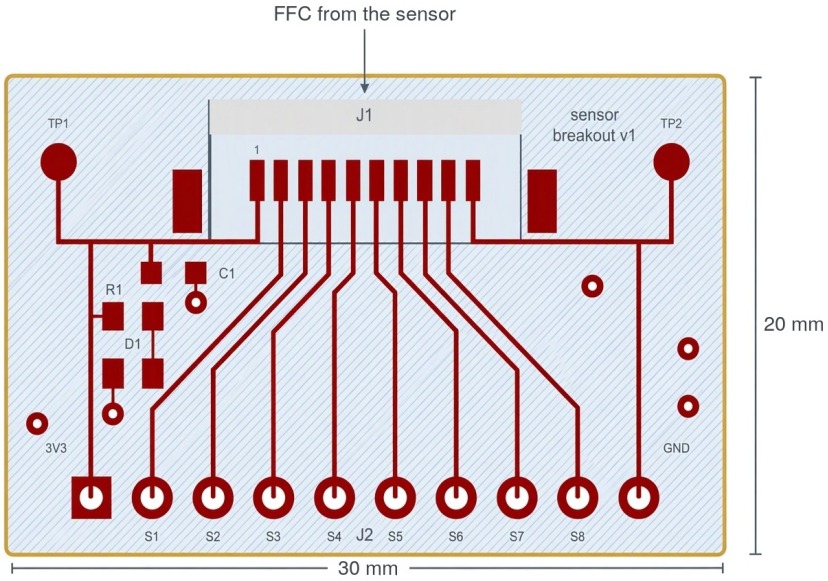


Figure L1.5: A finished layout of the 30 mm × 20 mm board, top view. Front copper is cardinal; the hatched area is the ground pour on the back layer (the front pour of C4 is not drawn); gold is the board outline; slate is silkscreen. The eight signal traces fan out at 45° from the 1.0 mm FFC pitch to the 2.54 mm header pitch without crossing; small circles are vias to the ground pour. Route the supply traces wider than drawn here, at 0.5 mm (C3).

A1 DO Create the project

Start KiCad and choose *File* → *New Project*. Name it `sensor_breakout` in a folder of its own. KiCad creates the project file (`.kicad_pro`), an empty schematic (`.kicad_sch`), and an empty board (`.kicad_pcb`). Open the schematic editor from the project manager. In *File* → *Page Settings*, enter a title, your name, and a revision (`v1`).

A2 EDIT Choose the LED resistor

The power LED is a red 0805 LED with a forward voltage (the nearly constant voltage across a lit LED) of about 2.0 V. A resistor in series sets its current. It only needs to be visible, so about 1 mA to 2 mA is plenty.

A2 · EDIT

WHAT TO DO Choose R1 from the E12 series (the standard values 1.0, 1.2, 1.5, 1.8, 2.2, 2.7, 3.3, 3.9, 4.7, 5.6, 6.8, and 8.2 times a power of ten) so that the LED current from 3.3 V is between 1 and 2 mA. Write the value into your notebook with the calculation.

HINT $I = (V_S - V_F)/R$. Check the resistor's power, I^2R, against the 0.125 W rating of an 0805 part.

DONE WHEN your value gives a current in range and you can state the resistor power in milliwatts.

A2 · ANSWER in your lab notebook

RECORD the LED current and the resistor power for your R1.

UNIT mA and mW.

WHERE TO FIND IT your calculation in A2.

A3 DO Place the symbols

A *symbol* is the drawing of a part in the schematic. Press A (*Place* → *Add Symbol*) and place these symbols from the standard libraries:

J1, J2: two `Connector_Generic:Conn_01x10`

C1: `Device:C`; R1: `Device:R`; D1: `Device:LED`

TP1, TP2: two `Connector:TestPoint`

Press P (*Place* → *Add Power*) for `power:+3V3` and `power:GND`, and place two `power:PWR_FLAG` symbols. Double-click each symbol and set its value: `100nF`, your R1 value, `RED`, and `FFC_10P_1.0mm` and `HDR_1x10` for the connectors. Rotate (R) and mirror (X or Y) symbols so that the drawing reads left to right as in Fig. L1.4.

A4 ◆DO Wire the circuit

Press W to draw wires and L to add net labels (a net label gives a net its name). Connect pin 1 of both connectors to +3V3 and pin 10 to GND. Instead of eight long wires, attach a label S1 to pin 2 of J1 and of J2, S2 to pin 3, and so on to S8 on pin 9: pins with the same label are connected. Connect C1, the series pair R1 and D1 (LED anode toward R1, cathode to GND), and TP1 and TP2 between the rails. Attach one PWR_FLAG to the +3V3 net and one to GND; they tell the checker that power enters the board through the connector.

A5 ◆DO Annotate

Choose *Tools* → *Annotate Schematic* and accept the defaults, so that every symbol gets a unique reference (no "R?" left).

A6 ✓CHECK Electrical rules check

Choose *Inspect* → *Electrical Rules Checker* and run it. Typical first errors: "Input Power pin not driven by any Output Power pins" (add the missing PWR_FLAG), "pin not connected" (a wire ends near but not on a pin; zoom in and redraw), or a label typed differently on the two connectors (S1 and s1 are different nets). Fix each one and rerun.

YOU SHOULD SEE After the fixes the checker reports 0 errors and 0 warnings. Hovering over a wire highlights its net, and the netlist (*File* → *Export* → *Netlist*) lists 11 nets: +3V3, GND, S1 to S8, and the unnamed net between R1 and D1.

✓CHECK **Checkpoint 1** ERC shows no errors, every reference is annotated, and the schematic matches Fig. L1.4. Plot it to PDF (*File* → *Plot*) for your report.

Part B: Footprints, Stack-Up, and Design Rules

GOAL Give every symbol a verified footprint and enter the board's stack-up and design rules. about 25 min

B1 ◆DO Assign footprints

Choose *Tools* → *Assign Footprints* and assign these footprints from the standard libraries:

J2: `Connector_PinHeader_2.54mm:PinHeader_1x10_P2.54mm_Vertical`

C1: `Capacitor_SMD:C_0805_2012Metric`

R1: `Resistor_SMD:R_0805_2012Metric`

D1: `LED_SMD:LED_0805_2012Metric`

TP1, TP2: `TestPoint:TestPoint_Pad_D1.5mm`

For J1, search the `Connector_FFC-FPC` library for 10-position footprints with `P1.0mm` in the name and choose the one whose name matches your connector's manufacturer part number.

B2 ✓CHECK Check the FFC footprint against the datasheet

Open the footprint in the footprint viewer and compare it with the land pattern in the connector datasheet: pitch (1.0 mm), pad width and length, the mounting tabs, the position of pin 1, and the side the cable enters. Also note whether the connector's contacts are on its top or bottom. A *same-side* cable has its exposed contacts on the same face at both ends; an *opposite-side* cable has them on opposite faces. Which type carries pin 1 to pin 1 depends on the contact side and orientation of both the sensor tail and J1, so sketch both from above, mark pin 1 and the contact faces, and choose the cable type so that the sensor's pin 1 reaches J1 pin 1.

B3 ◆DO Send the design to the board

Open the PCB editor and choose *Tools* → *Update PCB from Schematic* (F8). The footprints appear in a cluster, joined by thin ratsnest lines.

B4 ✎EDIT Stack-up and design rules

Choose *File* → *Board Setup*.

B4 · ✎EDIT

WHAT TO DO Under *Board Stackup* → *Physical Stackup*, set 2 copper layers and a board thickness of 1.6 mm with 35 µm copper. Under *Design Rules* → *Constraints*, enter the rules of Table L1.1: minimum clearance 0.15 mm, minimum track width 0.15 mm, minimum annular width 0.15 mm, minimum via diameter 0.6 mm, minimum through hole 0.3 mm, and copper to edge clearance 0.3 mm. Under *Net Classes*, set the Default class to a 0.25 mm track, 0.2 mm clearance, and a 0.6 mm/0.3 mm via; add a class `Power` with a 0.5 mm track and assign the nets +3V3 and GND to it.

HINT if your manufacturer's table differs from Table L1.1, use its numbers; the recommended values, not the absolute minimums.

DONE WHEN *Board Setup* shows your values and the net class assignment lists +3V3 and GND under `Power`.

B5 ▣ANSWER Check the trace widths with IPC-2221

The short script below evaluates Eq. (L1.1). Run it (or do the arithmetic by hand) for the power net at 0.5 A and a signal at 10 mA, both with $\Delta T = 10\,^{\circ}\mathrm{C}$ on 1 oz outer copper.

B5 · code

```
def ipc2221_width_mm(I, dT=10.0, oz=1.0, outer=True):
    """Trace width (mm) for current I (A) by IPC-2221."""
    k = 0.048 if outer else 0.024
```

```
    area_mil2 = (I / (k * dT**0.44)) ** (1 / 0.725)
    return area_mil2 / (1.378 * oz) * 0.0254

for I in (0.5, 0.01):
    print(I, "A ->", round(ipc2221_width_mm(I), 4), "mm")
```

B5 · ANSWER in your lab notebook

RECORD the two widths, and which rule (current or manufacturing) sets the width of each class.

UNIT mm.

WHERE TO FIND IT the script output and Table L1.1.

CHECK **Checkpoint 2** Every symbol has a footprint (the Assign Footprints window shows none missing), the FFC footprint matches the datasheet, the board shows all footprints with ratsnest lines, and the rules and net classes are entered.

Part C: Placement and Routing

GOAL Place the parts, route every connection, pour ground, and label the board. about 45 min

C1 DO Draw the board outline

Select the `Edge.Cuts` layer and draw a rectangle 30 mm by 20 mm (*Place* → *Draw Rectangle*); double-click it to type exact coordinates. Rounded corners of about 1 mm are kinder to hands and cables. Set the grid to 0.25 mm (or 0.1 mm near the FFC, a tenth of its 1.0 mm pitch) from the toolbar.

C2 DO Place the components

Move (M) and rotate (R) footprints into the arrangement of Fig. L1.5: J1 centered on the top edge with its cable opening facing out over the edge; J2 along the bottom edge with pin 1 at the left; C1 within a few millimeters of J1's supply pins; R1 and D1 together where the LED can be seen; TP1 and TP2 near the corners where a probe can reach them. Keep all copper at least 0.3 mm from the outline. Press Alt+3 (*View* → *3D Viewer*) to see whether the connectors overhang correctly.

C3 EDIT Route the signals and the power

Press X (*Route* → *Single Track*) and route each ratsnest line on the front copper.

C3 · EDIT

WHAT TO DO Route S1 to S8 as 0.25 mm traces that fan out from the FFC pitch to the header pitch, and route +3V3 and GND as 0.5 mm traces (the

net class sets the width automatically). Use 45° bends, not right angles. Press V while routing to drop a via and change layers if two traces must cross.

HINT route the two outermost signals (S1 and S8) first and work inward; with the pattern of Fig. L1.5 no crossings are needed. Keep the supply traces outside the signal fan.

DONE WHEN the status bar or the board statistics report 0 unrouted connections.

C4 EDIT Pour ground

C4 · EDIT

WHAT TO DO A *pour* (or filled zone) is a large copper area joined to one net. Draw a filled zone (*Place → Add Filled Zone*) covering the whole board on B.Cu, assigned to GND, with a 0.3 mm clearance; add a second GND zone on F.Cu. Add a few vias from GND traces and pads to the back pour, including one next to C1's ground pad. Fill the zones with *Edit → Fill All Zones* (B).

HINT a zone fills around other nets automatically; if it does not appear, check that it is assigned to GND and refill.

DONE WHEN both pours are visible and every GND pad connects to them (no GND ratsnest lines remain).

C5 ANSWER Measure the layout

C5 · ANSWER in your lab notebook

RECORD the length of the longest signal trace, the number of vias, and the distance from C1 to J1's supply pin.

UNIT mm, count, mm.

WHERE TO FIND IT *Inspect → Measure* (Ctrl+Shift+M), and *Inspect → Board Statistics*.

C6 DO Clean up the silkscreen

Move every reference designator next to its part, readable from one or two directions, and off all pads. Add text on `F.Silkscreen`: the board name and revision, "3V3" and "GND" at the header ends, S1 to S8 along the header, and a pin-1 mark at each connector. Text should be at least 1 mm tall with 0.15 mm lines.

> CHECK **Checkpoint 3** No ratsnest lines remain, both ground zones are filled, signal and power traces have their net-class widths, and the silkscreen is readable and clear of pads.

Part D: Design Rule Check and Review

GOAL Find and fix every rule violation, and compare the board with the schematic. about 15 min

D1 RUN Run the design rule checker

Choose *Inspect* → *Design Rules Checker*, tick *Refill all zones before performing DRC* and *Test for parity between PCB and schematic*, and click *Run DRC*. Click each violation to jump to it.

D2 ANSWER Record and fix the violations

> **D2** · ANSWER in your lab notebook
>
> RECORD the number of errors and warnings in the first run, the three most common types, and the count after your fixes (it must be zero errors).
>
> UNIT counts.
>
> WHERE TO FIND IT the DRC dialog's Violations and Unconnected Items tabs.

D3 DO Review in 3D

Open the 3D viewer again and check: J1's cable opening faces off the board, J2's pins point the right way, the LED's polarity mark matches the schematic, and no part overhangs another. Print the board at 1:1 (*File* → *Print*, scale 1:1) and lay the real connector on the paper if you have one.

> CHECK **Checkpoint 4** DRC reports 0 errors and 0 unconnected items, the parity test passes, and the 3D view shows the connectors and the LED correctly oriented.

Part E: Gerbers, Drill File, and BOM

GOAL Generate the manufacturing files and inspect them as the factory will see them.

about 25 min

E1 DO Plot the Gerber files

Choose *File* → *Fabrication Outputs* → *Gerbers (.gbr)*. Select the layers F.Cu, B.Cu, F.Paste, B.Paste, F.Silkscreen, B.Silkscreen, F.Mask, B.Mask, and Edge.Cuts; set the output folder to `gerbers/`; keep *Check zone fills before plotting* ticked; and click *Plot*. If your manufacturer asks for Protel file extensions (`.GTL`, `.GBL`, ...) tick that option.

E2 DO Generate the drill file

In the same dialog click *Generate Drill Files*. Choose Excellon, units of millimeters, the absolute origin (the same origin as the Gerbers), and either separate or merged plated (PTH) and non-plated (NPTH) files as your manufacturer prefers. Write the files to `gerbers/`.

E3 EDIT Make the bill of materials

E3 · EDIT

WHAT TO DO In the schematic editor choose *Tools* → *Generate Bill of Materials*, which opens the symbol fields table. Add a field `MPN` and fill in a manufacturer part number for J1, J2, C1, R1, and D1 (test points need none). Group by value and footprint, and export a CSV with the columns Reference, Value, Footprint, Quantity, and MPN.

HINT the capacitor's MPN must match 100 nF, an 0805 size, and a voltage rating above 3.3 V (16 V or more is common); choose an X7R dielectric.

DONE WHEN the CSV has one row per distinct part and the quantities add up to 7 parts.

E4 DO Inspect the files in a Gerber viewer

Open GerbView from the KiCad project manager and load every `.gbr` file and the drill file. Toggle the layers one at a time: the outline closes; copper on both layers matches the PCB editor; mask openings sit over every pad; no silkscreen falls on a pad; the drill hits land in the centers of the through-hole pads and vias. Then upload the zip to your manufacturer's online viewer as a second check.

E5 ANSWER Summarize the fabrication package

E5 · ANSWER in your lab notebook

RECORD the number of files in the package, the board area, the narrowest trace and the smallest clearance in the design, and the smallest drill.

UNIT count, cm^2, mm.

WHERE TO FIND IT the `gerbers/` folder, *Inspect* → *Board Statistics*, and the drill report.

E6 CHECK Package the files

Zip the contents of `gerbers/` (the Gerbers, the job file, and the drill files) without the KiCad project files. Open the zip once more in a viewer to confirm that nothing is missing.

CHECK **Checkpoint 5** The zip opens in a Gerber viewer with all layers aligned and the drill hits centered; the BOM lists all 7 parts, with part numbers for all but the test points; the narrowest feature is at least 0.15 mm.

Hand in

FINISH THE LAB

You designed a two-layer sensor breakout board and produced a fabrication package that a manufacturer can build.

- ☐ **A2** EDIT LED resistor value, current, and power
- ☐ CHECK Checkpoint 1: schematic PDF with a clean ERC
- ☐ **B4** EDIT Stack-up, constraints, and net classes entered
- ☐ **B5** ANSWER IPC-2221 widths for the power and signal nets
- ☐ **C3** EDIT All signals and power routed
- ☐ **C4** EDIT Ground pours on both layers
- ☐ **C5** ANSWER Longest trace, via count, C1 distance
- ☐ **D2** ANSWER DRC counts before and after; final screenshot with 0 errors
- ☐ **E3** EDIT BOM CSV with part numbers
- ☐ **E5** ANSWER Fabrication package summary
- ☐ CHECK Checkpoint 5: Gerber viewer screenshot
- ☐ Analysis questions 1 to 8
- ☐ **Save** SAVE Zip of the Gerber and drill files, the BOM, and the schematic PDF

Typical results

With R1 $= 1\,\mathrm{k}\Omega$ and $V_F \approx 2.0\,\mathrm{V}$ the LED draws about 1.3 mA and R1 dissipates about 1.7 mW, far below its rating. The IPC-2221 widths at $\Delta T = 10\,^{\circ}\mathrm{C}$ are about 0.12 mm for 0.5 A and under 0.001 mm for 10 mA, so the 0.5 mm power and 0.25 mm signal widths are set by the manufacturing rules and by margin, not by heating. The finished board has 11 nets, 7 parts, about 4 to 8 vias, and its longest signal trace is about 15 mm to 17 mm. A first DRC run commonly reports 3 to 10 violations: silkscreen over pads, a reference outside the outline, a zone not refilled, or a trace too close to an FFC pad; all are fixed in a few minutes. A KiCad 8 Gerber package for this board contains the nine layer files named `sensor_breakout-F_Cu.gbr`, `-B_Cu.gbr`, `-F_Mask.gbr`, `-B_Mask.gbr`, `-F_Paste.gbr`, `-B_Paste.gbr`, `-F_Silkscreen.gbr`, `-B_Silkscreen.gbr`, and `-Edge_Cuts.gbr`, a job file (`.gbrjob`), and either two drill files (`-PTH.drl` and `-NPTH.drl`) or one merged `.drl` file; the board area is $6.0\,\mathrm{cm}^2$.

Troubleshooting

- *ERC: "Input Power pin not driven."* The power net has no source on the board. Add a PWR_FLAG to each supply net that enters through a connector.
- *A symbol has no footprint, or "footprint not found" on update.* Assign it in *Tools* → *Assign Footprints*; check that the library name before the colon exists in *Preferences* → *Manage Footprint Libraries*.
- *Ratsnest lines remain after routing.* A trace ends near a pad but not on it, or a label was misspelled. Use the DRC's Unconnected Items tab to find it.
- *Clearance errors at the FFC pads.* The pads are 1.0 mm apart; a 0.25 mm trace leaving a pad between neighbors needs care. Leave each pad straight out before bending, or reduce the width only at the pad (a neck down to 0.2 mm).
- *The ground zone does not appear, or appears in the Gerbers with holes.* Refill with B before plotting; check the zone's net and its priority.
- *Silkscreen over pads.* Move or shrink the text; the manufacturer clips silkscreen at mask openings, and the clipped text becomes unreadable.
- *The drill file is offset from the copper in the viewer.* The Gerbers and the drill file used different origins. Regenerate both with the same origin (absolute, or the same auxiliary origin).

- *The outline is open or doubled.* The manufacturer cannot cut it. Make Edge.Cuts one closed shape with no overlapping segments.
- *An imported EAGLE board shows odd pads or rules.* Check each imported footprint against its datasheet and re-enter the design rules; EAGLE's DRU settings do not carry over one for one.

Analysis questions

1. Use Eq. (L1.1) to find the width for 1 A at $\Delta T = 10\,^{\circ}\mathrm{C}$ on an outer and on an inner layer of 1 oz copper. Explain the difference physically.
2. Why should C1 sit next to J1 rather than across the board? A trace has an inductance of roughly 1 nH/mm; estimate the impedance at 10 MHz of 30 mm of trace between the capacitor and the connector (an inductance L has an impedance of magnitude $2\pi f L$).
3. Your sensor flex has its pin 1 on the left with contacts on top, and J1 has bottom contacts. Which cable type (same-side or opposite-side) makes the sensor's pin 1 reach J1 pin 1? What would go wrong with the other type?
4. In a capacitive glove sensor (one of the design projects of Chapter 14), the stray capacitance of the traces and cable adds in parallel to each sensor and shrinks the fraction by which a press changes the total. Name two layout choices on this board that reduce stray capacitance or coupling between neighboring signal lines, and one that increases it.
5. What resistor gives an LED current of 0.5 mA if $V_F = 1.8\,\mathrm{V}$? Why might a wearable use such a small current?
6. Why does this board need only through vias? When would you pay for blind or buried vias?
7. Which files does a bare-board manufacturer need, and which additional files does an assembly service need?
8. You import an older EAGLE board into KiCad. List three things you would check before ordering it.

Lab 2
Printed Electrodes and a Capacitive Sensor Array

Lab 2 at a glance

What you build: a 4×4 row-column capacitive pressure array from two stencil-printed silver layers with a compressible dielectric (an insulating layer) between them. You read it with an LCR meter, a bench instrument that measures capacitance, inductance, and resistance, and with a TI FDC2214, a chip that turns capacitance into a number, on an Arduino Nano 33 BLE Sense microcontroller board. You also print a pair of Ag/AgCl ECG electrodes (silver coated with silver chloride, the usual material of skin electrodes) and measure their impedance in saline, and you measure the sheet resistance of your cured silver.
Time: two sessions of about three hours: Parts A to D (design, stencil, print, resistance), then Parts E to G (assembly, capacitance, electrodes).
Before the lab: read Sections 12.6 (stencil printing) and 12.11 (printed electrodes) of Chapter 12 and Section 5.5 on capacitive sensing (Chapter 5), and compute (a) the capacitance of a 1 cm^2 parallel-plate capacitor with a 1 mm dielectric of $\varepsilon_r = 1.5$ and of 3 (1.33 and 2.66 pF) and (b) the resistance of a 0.5 mm$\times$50 mm line at 30 mΩ per square (3.0 Ω).
You hand in: a short report with your artwork, photographs, the tables and plots of Parts D, F, and G, and the analysis answers.

What you need

Hardware.

- Stencil: a craft vinyl cutter and mat; adhesive vinyl (about 70 to 100 µm) and transfer tape.
- Printing: PET, polyimide, or TPU film (50 to 125 µm); low-temperature silver flake paste and Ag/AgCl paste; a squeegee

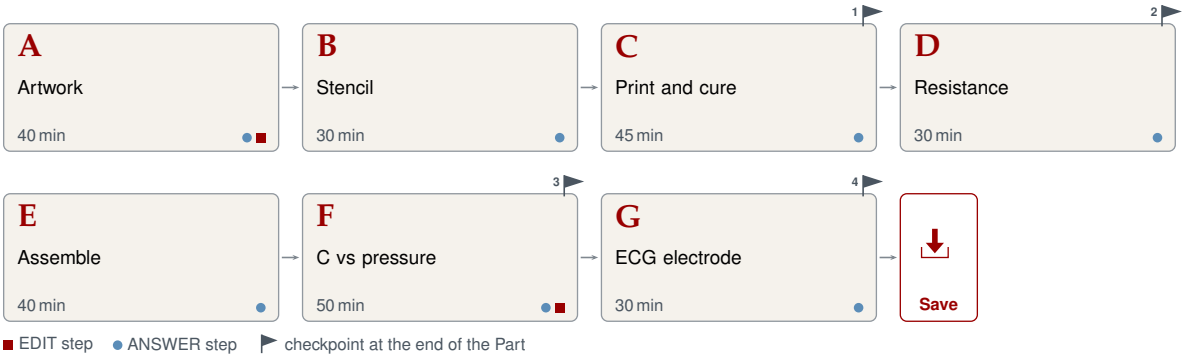
or stiff card; a glass plate, polyimide tape, and isopropanol; an oven or hot plate.

- Measuring: a micrometer; a multimeter with four-wire mode; an LCR meter with Kelvin (four-wire) clips.
- Assembly: 1 mm polyurethane foam or porous PDMS; copper tape and conductive epoxy.
- Readout: an FDC2214 evaluation module; an Arduino Nano 33 BLE Sense; a 74HC4051 multiplexer (an electronic switch) on a small breakout board.
- Testing: masses of 20 to 500 g and a 10 mm square presser foot; 0.9 % NaCl saline.

Software. A PCB layout tool (KiCad, as in Lab 1), the cutter's software (Silhouette Studio), Arduino IDE, Python.

Files. The bitmaps you export in Part A (for example `top.bmp`, one layer of a 4×4 stencil: four columns of pads, contact fingers, corner fiducials, which are alignment marks), and a note of the sizes you enter in the cutter software.

SAFETY Pastes contain glycol ethers and esters: work with ventilation, nitrile gloves, and safety glasses, no food or drink, and put paste and wipes in labeled waste, never down the drain (silver is toxic to aquatic life). The oven reaches 130 °C: use tongs, set a timer, and stay with it; overheated PET and TPU warp and fume. Cut away from your body and cap blades when you put them down. Never mix bleach with acids or ammonia. The electrode test is done in saline, not on skin: the LCR meter is mains-powered, and no mains-powered instrument should ever be connected to a person (Chapter 14).



Objectives

Turn a sensor drawn in PCB software into a vinyl stencil with a 500 µm minimum line; print and cure silver and measure its sheet resistance; assemble a row-column capacitive array and measure each pixel and its response to pressure; read the array with an LC-tank converter over I^2C; and compare printed Ag/AgCl and plain silver electrodes in saline.

Background

A printed film of resistivity ρ and thickness t has a *sheet resistance* $R_s = \rho/t$: the resistance from one edge to the opposite edge of any square of it, whatever the size of the square. A trace of length L and width W counts L/W *squares* in a row, and

$$R = R_s \frac{L}{W}, \qquad R_s = \frac{\rho}{t}. \tag{L2.1}$$

Measure it with four wires, two carrying current and two inside them sensing voltage, so lead and contact resistance drop out. A pixel is a parallel-plate capacitor where a row pad faces a column pad across a dielectric of thickness d and relative permittivity ε_r (the factor by which the material raises the capacitance over that of vacuum):

$$C = \frac{\varepsilon_0 \varepsilon_r A}{d} = 0.885\,\text{pF} \times \varepsilon_r \quad (A = 1\,\text{cm}^2,\ d = 1\,\text{mm}). \tag{L2.2}$$

Pressing thins the dielectric and squeezes air out of a porous one, raising ε_r; both raise C (Fig. L2.3). 1 kPa on a 1 cm^2 foot is 0.1 N, the weight of 10.2 g. Rows and columns make 16 pixels from 8 wires, as in the glove sensor a team in a past semester proposed: 1 cm pads 1 cm apart, a sponge or porous-PDMS dielectric, and an FDC2214 on a Nano 33 BLE with a 74HC4051 to switch columns.

The FDC2214 puts the sensor in parallel with the capacitor of an *LC tank*, an inductor and a capacitor in parallel that oscillate at one frequency (on the evaluation module $L = 18\,\mu\text{H}$ and $C_{\text{tank}} = 33\,\text{pF}$). The chip keeps the tank oscillating and counts its frequency against a reference clock; more capacitance gives a lower frequency:

$$f_{\text{sensor}} = \frac{1}{2\pi\sqrt{L\,(C_{\text{tank}} + C_p + C)}}, \qquad C + C_p = \frac{1}{L\,(2\pi f_{\text{sensor}})^2} - C_{\text{tank}}, \tag{L2.3}$$

with C_p the parasitic capacitance, the unwanted capacitance of traces, cable, and multiplexer. Each channel reports a 28-bit code D, with $f_{\text{sensor}} = F_{\text{IN}} f_{\text{REF}} D / 2^{28}$, where $F_{\text{IN}} = 2$ in the single-ended setting used here and $f_{\text{REF}} = 40\,\text{MHz}$ on the evaluation module. The chip talks to

the Nano over I^2C, a two-wire digital bus with a data line (SDA) and a clock line (SCL), at address `0x2A` (ADDR pin low), on the Nano's pins A4 (SDA) and A5 (SCL) at 3.3 V. Between two electrodes in saline the impedance is the solution resistance $d/(\sigma A)$, with $\sigma \approx 1.5\,\mathrm{S/m}$, in series with two electrode interfaces (Chapter 6).

Part A: Design the artwork

GOAL Draw both electrode layers, test lines, and an ECG electrode with a 500 µm minimum line. about 40 min

A1 ◆DO Draw the array and export bitmaps

On the top layer of a two-layer board draw four *columns* of four 10 mm square pads on a 20 mm pitch, joined by 1.0 mm traces and ending in 1.5 mm×15 mm fingers on a 2.54 mm pitch; on the bottom layer four *rows* the same way, so row and column traces meet only inside the pixels (Fig. L2.1). Add fiducial crosses in the corners, and on a third sheet test lines 0.5 mm and 1.0 mm wide and 50 mm long with 5 mm end pads, and two pairs of ECG electrodes (a 15 mm disk, a 2 mm×60 mm trace, a 10 mm×15 mm tab). Export each layer as a 300 dpi black-and-white bitmap (a pixel is 84.7 µm). A past team's file for a layout like this one was 1046 × 1236 pixels, or 88.6 × 104.6 mm, with pads about 49 pixels (4.2 mm) wide; in your file a 10 mm pad should be about 118 pixels wide.

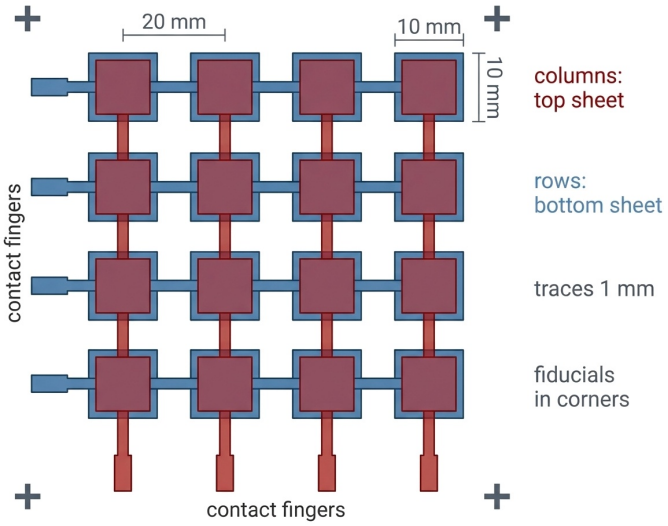


Figure L2.1: The two electrode layers of the 4 × 4 array, overlaid as they face each other in the assembled sensor. Column pads on the top sheet lie over row pads on the bottom sheet; the traces cross only inside the pixels. The top (column) layer is mirrored in its artwork because it is printed face up and turned over, ink down, during assembly. The contact fingers are drawn at the end of each row and column; in the artwork the traces bring them together on a 2.54 mm pitch.

A2 EDIT Design rules and the mirrored layer

A2 · EDIT

WHAT TO DO set a minimum trace of 0.50 mm and a minimum clearance of 1.0 mm, and pass the design-rule check. The two sheets face each other ink to foam, so the column (top-layer) sheet is printed face up and then flipped ink down: mirror its artwork.

HINT paste creeps 50 to 200 µm under a stencil, so lines closer than about 0.5 mm may bridge.

DONE WHEN the check passes and the mirrored top layer, flipped back over the bottom layer, shows 16 coincident pads.

A3 ANSWER Expected values

A3 · ANSWER in your lab notebook

RECORD squares and expected resistance of each test line at 30 mΩ per square; expected pixel capacitance for $\varepsilon_r = 1.5$.

UNIT squares, Ω, pF

WHERE TO FIND IT Eqs. (L2.1) and (L2.2).

Part B: Cut the stencil

GOAL Turn each bitmap into a vinyl stencil whose openings are the electrodes. about 30 min

B1 DO Size, cut, and weed

Trace the bitmap in the cutter software and set its width and height exactly, with the aspect ratio locked; the software ignores the bitmap's resolution. Type the size in millimeters, at 0.0847 mm per pixel; for the past team's file of **A1**, 1046 × 1236 pixels, that is 88.6 mm by 104.6 mm, an aspect ratio of $1046/1236 = 0.846$. Cut a test of one sensor pixel first and check its pad with calipers. Then cut the vinyl, peel out (*weed*) the features so the openings are the electrodes, and cover the stencil with transfer tape. A tree of pads and traces weeds cleanly because it has no islands, loose pieces of vinyl inside a closed opening; a ring would leave its center as an island.

B2 ANSWER Stencil thickness and release

B2 · ANSWER in your lab notebook

RECORD the vinyl thickness, the narrowest opening, and its area ratio $W/(2t)$ (release needs above about 0.66).

UNIT µm, dimensionless

WHERE TO FIND IT micrometer and calipers; Section 12.6.

Part C: Print and cure

GOAL Print silver for the array and test lines, and Ag/AgCl on the ECG pads, and cure them. about 45 min

C1 DO Prepare and print

Wipe the film with isopropanol. A water drop that beads up on a scrap, with a contact angle well above 90°, means the paste will not wet (spread); treat the film briefly with plasma or corona (Section 12.2). Tape the film to the glass, apply the stencil, remove the transfer tape, and press every edge down. Lay a bead of stirred paste above the pattern and pull it across in one firm stroke at 45 to 60°; peel the stencil at once, slowly and at a low angle. For the ECG electrodes print silver first, cure it, then print Ag/AgCl over the disks of one pair through a second stencil aligned on the fiducials.

C2 DO Cure

Follow the paste datasheet, typically 120 to 130 °C for 10 to 30 min, with the film on the glass. Keep TPU at the lowest temperature the datasheet allows, because it softens.

C3 ANSWER Print fidelity

> **C3** · ANSWER in your lab notebook
>
> RECORD the printed width of each test line at five places (mean and standard deviation) and any bridges or gaps.
>
> UNIT µm
>
> WHERE TO FIND IT microscope or calipers; compare with the width you drew.

> CHECK **Checkpoint 1** Every row and column conducts end to end (a few ohms) and neighboring traces are open (above 20 MΩ).

Part D: Thickness and sheet resistance

GOAL Measure the film thickness, sheet resistance, and resistivity. about 30 min

D1 DO Thickness and four-wire resistance

Measure film plus silver and bare film at three places with a micrometer; the difference is t. Clip the current leads to the outer ends of the test-line end pads and the sense leads to their inner ends, and record the four-wire resistance; then a two-wire reading with the same clips.

D2 ANSWER Sheet resistance and resistivity

D2 · ANSWER in your lab notebook

RECORD $R_s = RW/L$ for both lines (measured widths), $\rho = R_s t$, its ratio to bulk silver (1.59 µΩ cm), and the two-wire error.

UNIT mΩ per square, µΩ cm, Ω

WHERE TO FIND IT Eq. (L2.1).

✓CHECK **Checkpoint 2** The two lines agree within 20 %, between about 10 and 100 mΩ per square.

Part E: Assemble the array

GOAL Stack rows, dielectric, and columns into 16 pixels with leads.

about 40 min

E1 ◆ DO Stack the layers

Cut the dielectric to cover the 16 pads but not the fingers, and measure its thickness under a glass slide. Tape the row sheet ink up, lay the dielectric on it, lower the column sheet ink down, align the fiducials through the clear film, and tape the edges. Attach leads to the fingers with copper tape and a dot of conductive epoxy, and insulate them with polyimide tape (Fig. L2.2).

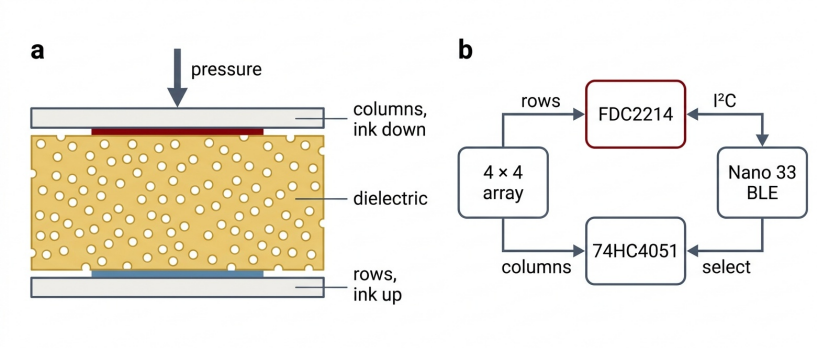


Figure L2.2: (a) Cross section of one pixel: the printed pads face each other across the porous dielectric. (b) The readout: each row drives one channel of the FDC2214, and the multiplexer connects one column at a time to ground, so the four channels read the four pixels of that column.

E2 ANSWER Isolation and a first capacitance

E2 · ANSWER in your lab notebook

RECORD the DC resistance from each row to each column (open, above 20 MΩ) and the capacitance of pixel (1,1) at 100 kHz after open and short compensation of the leads (the meter measures its leads open and shorted and subtracts them).

UNIT MΩ, pF

WHERE TO FIND IT LCR meter in its parallel-capacitance mode, test level 1 V or less.

Part F: Capacitance versus pressure

GOAL Map the 16 pixels, sweep one with pressure, and read the array with the FDC2214. about 50 min

F1 ◆DO LCR map, pressure sweep, and wiring

Measure all 16 pixels at 100 kHz into a 4×4 table. Load pixel (2,2) through the presser foot with 0, 20, 50, 100, 200, and 500 g (0 to 49 kPa), reading after 10 s, and unload in reverse to see the hysteresis (a different reading on the way down than on the way up). Then wire the readout of Fig. L2.2b: each row to the sensor input of one FDC2214 channel (pins IN0A to IN3A), the columns to Y0 to Y3 of the 74HC4051 with its common pin Z and enable to ground and its supply pin to 3.3 V, S0 to S2 to Nano pins D2 to D4, and the FDC2214's 3.3 V, ground, SDA, and SCL to the Nano, with ADDR and SD (shutdown) low. Keep leads short and fixed: moving wires change C_p.

F2 ✎EDIT The readout sketch

A *sketch* is an Arduino program. The FDC2214 is set up by writing 16-bit numbers into its *registers*, numbered memory cells inside the chip that hold its settings and its results. The sketch configures four channels for automatic scanning and steps the multiplexer through the columns. Two lines are yours.

F2 · code

```
#include <Wire.h>
const uint8_t FDC = 0x2A;              // ADDR pin low
const float FREF = 40.0e6;             // EVM reference clock (Hz)
const float L_TANK = 18e-6, C_TANK = 33e-12;
const int SEL[3] = {2, 3, 4};          // 74HC4051 S0..S2

void writeReg(uint8_t reg, uint16_t v) {
  Wire.beginTransmission(FDC);
  Wire.write(reg); Wire.write(v >> 8); Wire.write(v & 0xFF);
  Wire.endTransmission();
}
uint16_t readReg(uint8_t reg) {
  Wire.beginTransmission(FDC); Wire.write(reg);
  Wire.endTransmission(false);           // repeated start
  Wire.requestFrom(FDC, (uint8_t)2);
  uint16_t v = Wire.read() << 8;  v |= Wire.read();
  return v;
}
uint32_t readData(uint8_t ch) {          // MSB register first
  uint16_t msb = readReg(2 * ch);        // bits 13:12 are error flags
  uint16_t lsb = readReg(2 * ch + 1);
```

```
  return ((uint32_t)(msb & 0x0FFF) << 16) | lsb;
}
void setup() {
  Serial.begin(115200);  Wire.begin();
  for (int i = 0; i < 3; i++) pinMode(SEL[i], OUTPUT);
  Serial.println(readReg(0x7F), HEX);    // DEVICE_ID, expect 3055
  for (uint8_t ch = 0; ch < 4; ch++) {
    writeReg(0x08 + ch, 0x2000);         // RCOUNT
    writeReg(0x10 + ch, 0x0100);         // SETTLECOUNT
    writeReg(0x14 + ch, 0x2001);         // CLOCK_DIVIDERS: FIN_SEL
      2, FREF/1
    writeReg(0x1E + ch, 0x7800);         // DRIVE_CURRENT
  }
  writeReg(0x1B, 0xC20D);                // MUX_CONFIG: scan ch
     0-3, 10 MHz
  writeReg(0x1A, 0x1E01);                // CONFIG: active,
     external clock
}
void loop() {
  for (int col = 0; col < 4; col++) {
    for (int i = 0; i < 3; i++) digitalWrite(SEL[i], (col >> i) &
     1);
    delay(30);                           // two fresh scans
    for (uint8_t row = 0; row < 4; row++) {
      uint32_t d = readData(row);
      float f = 0;   // EDIT: sensor frequency from d
      float c = 0;   // EDIT: capacitance added to the tank
      Serial.print(c * 1e12, 3);
      Serial.print((row == 3 && col == 3) ? '\n' : ',');
    }
  }
}
```

F2 · EDIT

WHAT TO DO the two `EDIT` lines: $f = F_{IN} f_{REF} D / 2^{28}$ with $F_{IN} = 2$, and the capacitance of Eq. (L2.3).

HINT write 2^{28} as `268435456.0`. Check each bit of the registers CONFIG and MUX_CONFIG against the tables in the datasheet; on a board without the 40 MHz clock, use the internal-oscillator setting and its own `FREF`.

DONE WHEN the ID prints `3055`, each line has 16 values, and an untouched pixel reads a steady $C + C_p$ of a few to a few tens of picofarads.

F3 ANSWER Codes, sensitivity, and noise

F3 · ANSWER in your lab notebook

RECORD for pixel (2,2): D and f_{sensor} unloaded, the change in D for 100 g, and $\Delta C/\Delta P$ from 0 to 10 kPa from both the LCR and the FDC data; the standard deviation of the unloaded pixel over a minute, and the smallest pressure change it resolves (noise over sensitivity).

UNIT counts, MHz, fF/kPa, fF, kPa

WHERE TO FIND IT the serial output and Eq. (L2.3).

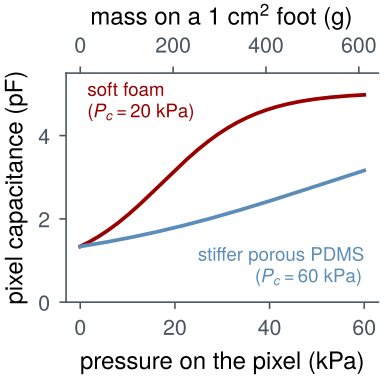


Figure L2.3: Model of one 1 cm² pixel with a 1 mm porous dielectric, 30 % solid with a solid permittivity of 2.7 (illustrative, not a measurement). A softer dielectric is more sensitive at low pressure and saturates sooner.

✓CHECK **Checkpoint 3** Each pixel responds to a fingertip on that pixel (neighbors move by a small fraction), and the sweep of pixel (2,2) rises monotonically.

Part G: A printed ECG electrode

GOAL Compare printed Ag/AgCl and plain silver electrodes in saline.

about 30 min

G1 ◆DO Measure two pairs

Fix the two Ag/AgCl electrodes face to face 10 mm apart in 0.9 % NaCl with only the disks wet, set the LCR meter to its lowest test level, and record $|Z|$ and phase at its lowest frequency (often 20 Hz), 100 Hz, 1 kHz, and 10 kHz. Repeat with a pair of plain silver disks. If time allows, chloride a silver pair (Example 12.5, or ten minutes in bleach and a rinse) and measure it again.

G2 ⊡ANSWER Solution resistance and the two electrodes

G2 · ⊡ANSWER in your lab notebook

RECORD $R_{sol} = d/(\sigma A)$ and the measured $|Z|$ at 10 kHz, where the interfaces are nearly short circuits; $|Z|$ and phase of both pairs at the lowest frequency, and their ratio.

UNIT Ω, degrees

WHERE TO FIND IT Background and **G1**.

✓CHECK **Checkpoint 4** The Ag/AgCl pair has the lower low-frequency impedance and a phase nearer zero; both approach R_{sol} at 10 kHz.

Hand in

FINISH THE LAB

You printed and measured a capacitive pressure array and a pair of ECG electrodes.

- ☐ **A2** ✎EDIT Artwork passing the design-rule check, top layer mirrored
- ☐ **A3** ⊡ANSWER Expected squares, resistances, and capacitance
- ☐ **B2** ⊡ANSWER Stencil thickness and area ratio
- ☐ **C3** ⊡ANSWER Printed widths and defects
- ☐ ✓CHECK Checkpoint 1: continuity and no shorts
- ☐ **D2** ⊡ANSWER Sheet resistance, resistivity, two-wire error
- ☐ ✓CHECK Checkpoint 2: consistent sheet resistance
- ☐ **E2** ⊡ANSWER Isolation and the first capacitance

☐ **F2**	EDIT	Working readout sketch
☐ **F3**	ANSWER	Codes, sensitivity, noise, and pressure resolution
☐	CHECK	Checkpoint 3: all 16 pixels respond
☐ **G2**	ANSWER	Solution resistance, Ag versus Ag/AgCl
☐	CHECK	Checkpoint 4: electrode impedance
☐ **Save**	SAVE	Artwork, photographs, data files, and plots in the team folder

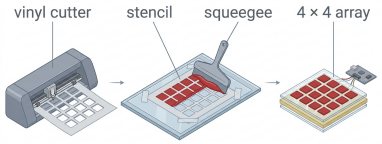


Figure L2.4: Lab 2 on the bench: a craft cutter cuts the vinyl stencil, a squeegee pulls silver paste across it on film taped to glass, and the finished array is wired to a small readout board.

Typical results

Figure L2.4 sums up the bench work. These ranges follow from the chapter and typical datasheets; yours depend on paste, film, and foam. An 80 µm vinyl stencil leaves roughly 10 to 30 µm of cured silver, and lines print 50 to 200 µm wider than drawn. Cured flake paste gives about 10 to 60 mΩ per square, 10 to 40 times the resistivity of bulk silver; the 0.5 mm test line reads a few ohms, and the two-wire reading is higher by 0.2 Ω or more. An unloaded pixel reads 1.5 to 4 pF, above the parallel-plate value because of fringing and traces; a firm press raises it by tens of percent to more than double, with some hysteresis because foams are viscoelastic. The FDC2214 runs near 5 to 6 MHz; a picofarad moves it by tens of kilohertz and the code by hundreds of thousands of counts. In saline the electrode pairs approach the solution resistance, tens to a hundred ohms, at high frequency, and plain silver has the higher, more capacitive impedance at low frequency.

The sketch of **F2** reads 7 to 8 frames per second. A conversion takes $(16\,\mathrm{RCOUNT} + 4)/f_{\mathrm{REF}} = 3.28$ ms, a four-channel scan with settling 13.5 ms, and the sketch waits two scans per column, so with the I²C reads a 16-pixel frame takes about 0.13 s. For the 20 or so that the glove project of Chapter 14 aims at, set RCOUNT to `0x0800` (3.7 ms per scan), use `delay(8)`, and add `Wire.setClock(400000)` after `Wire.begin()`: about 25 frames per second, with two bits less resolution (TI's rule of thumb is RCOUNT $= 2^{N-4}$ for N bits), so compare the noise of **F3** at both settings.

Troubleshooting

Paste bleeds or traces bridge: burnish the stencil, use less paste and pressure, peel at once, widen clearances. *Missing areas:* the paste did not wet (treat the film) or the opening's area ratio is too small. *High resistance:* under-cured, or cracked by bending while hot. *Row shorted to column:* torn dielectric or paste on its edge. *No answer from the FDC2214* (ID reads `FFFF` or 0): check 3.3 V, ground, SDA and SCL, ADDR and SD low, and pull-up resistors. *Frozen codes or error bits:* the tank is not

oscillating; try another DRIVE_CURRENT and read STATUS (`0x18`). *A whole column moves when one pixel is pressed:* unselected columns float and couple to the rows (Analysis question 3). *Slow drift:* temperature and humidity; warm up, keep hands off the leads, subtract a running baseline.

Analysis questions

1. From the unloaded capacitance and the measured dielectric thickness, find the effective ε_r. Why does Eq. (L2.2) underestimate the measured value?
2. Explain why the top (column) layer was mirrored, and what an unmirrored array would look like.
3. When column 2 is selected, the other columns float. Sketch the capacitances that couple a row to ground through a floating column, explain the crosstalk, and propose a circuit change that removes it.
4. Your sensitivity falls at high pressure. Explain with Fig. L2.3, and choose a dielectric for grasps up to 50 kPa and one for touches below 5 kPa.
5. For $L = 18\,\mu\text{H}$, $C_{\text{tank}} = 33\,\text{pF}$, $C_p = 10\,\text{pF}$, and a 1.3 pF pixel, find f_{sensor}, D, and the change in D for 0.1 pF. What limits the real resolution far above one code?
6. Explain the low-frequency difference between Ag/AgCl and plain silver with the interface of Chapter 6, and why it matters for an ECG amplifier (Chapter 8).
7. Measure the fiducial offset between the two layers and compare it with the registration numbers of Section 12.10. What else contributes?

Lab 3
Designing an Analog Front-End Daughter Board

Lab 3 at a glance

What you build: the schematic, layout, and manufacturing files of a small *daughter board*, a board that plugs onto another one. It carries your team's analog front end (AFE), the chip that amplifies, filters, and digitizes the electrode signals (the ADS1293 for the wearable ECG, the ADS1299 for the wearable EEG), with its power, clock, and input network; a 10-pin, 1 mm pitch FFC connector for the printed sensor; and headers onto which an Arduino Nicla Sense ME, a thumb-sized microcontroller board with a Bluetooth radio, plugs.
Time: two lab sessions of about three hours, plus time outside the lab.
Before the lab: read Sections 8.4 to 8.8 of Chapter 8; work through Lab 1, which teaches the KiCad tools; download the documents listed in the box below.
You hand in: the KiCad project (zipped), the schematic as a PDF, the DRC report, the BOM as a CSV file, the Gerber and drill files, the component-position file, and the review checklist of Checkpoint 1, signed by the reviewing team.

What you need

Hardware (optional, for checking): an Arduino Nicla Sense ME, a caliper, your printed sensor with its FFC tail, and, if the lab has one, the TI evaluation kit of your AFE (ADS1293EVM or ADS1299EEGFE-PDK).

Software: KiCad 8 or later, as in Lab 1. Autodesk retired the EAGLE layout program on June 7, 2026, so this lab uses KiCad, which can import old EAGLE projects such as the board of a past team.

Files: the AFE datasheet; the evaluation kit's user guide, with its schematic, bill of materials, and layout; the Nicla Sense ME datasheet and full pinout; the datasheets of the regulator, crystal or oscillator, and FFC connector you choose; the drawing of your printed sensor's tail (10 pads, 1 mm pitch).



Objectives

- Read an evaluation-kit schematic and keep only what your board needs.
- Choose regulators, decoupling, and a clock that meet the AFE's datasheet, and check that every interface pin matches the Nicla's logic level and pinout.
- Draw a schematic that passes the electrical rules check (ERC) and a design review by another team.
- Place and route a mixed-signal board that keeps digital noise out of the analog inputs, and pass the design rules check (DRC).
- Export a bill of materials and the files a board house needs to build and assemble the board.

Background

The design projects split the electronics into two boards (Fig. L3.1). The Nicla Sense ME, 22.86 mm square, brings the processor, the Bluetooth radio, the battery charger, and the USB port. Your daughter board brings the analog front end and everything it needs, and joins the printed sensor through a flat flexible cable. The electrode tail plugs into this board's FFC connector and nowhere else: never into the sensor breakout of Lab 1, whose pins 1 and 10 carry 3.3 V and ground and would put that voltage between two electrodes on the skin. Companies that make an AFE also sell an evaluation kit with a tested schematic, so you do not start from a blank page: you start from that schematic and remove what your board does not need (the USB interface, the kit's

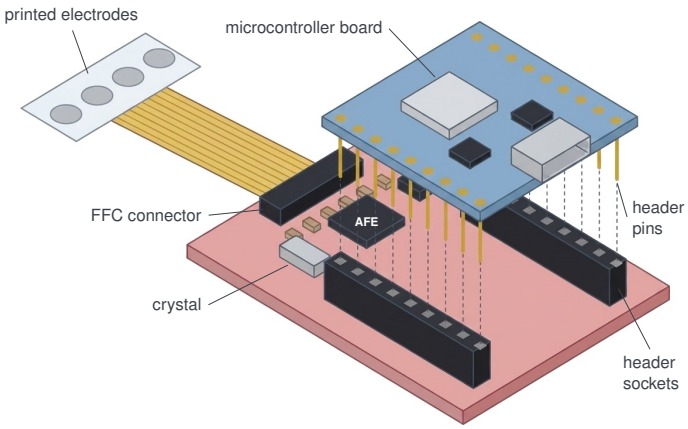


Figure L3.1: The daughter board, laid out like the floor plan of Fig. L3.2. The AFE, its crystal, regulator, and capacitors sit on the analog side; the Nicla Sense ME (drawn lifted off) plugs onto the digital side. Its two rows of header pins run parallel to the two header sockets and sit directly above them, so the board drops straight down. A flat flexible cable (FFC) connects the printed electrodes.

own microcontroller, current sensing, jumpers, EEPROM, and most test points).

Five groups of parts survive. *Power:* find every supply pin of the AFE, its allowed range, and its current. The ADS1293 needs an analog supply of 2.7 to 5.5 V and an I/O supply of 1.65 to 3.6 V; the ADS1299 needs 5 V across its analog supply (unipolar 5 V or $\pm$2.5 V) and 1.8 to 3.6 V for its digital supply. A low-dropout regulator (LDO) makes a clean, steady supply voltage (a *rail*) from the battery; its *dropout* is the smallest difference between input and output at which it still regulates. You may substitute an LDO for the one on the kit if its parameters match: output voltage, maximum current, dropout at that current, input range, power-supply rejection at low frequency (how well it keeps ripple on its input from reaching its output), output noise, quiescent current (the current it draws itself), the output capacitor it needs to be stable, and an enable pin if you want one. *Decoupling:* a 0.1 µF ceramic capacitor at every supply pin, a bulk capacitor of 1 to 10 µF on each rail, and the reference capacitors the datasheet asks for (on the ADS1293 kit, 1 µF on CVREF and 0.1 µF on RLDREF). *Clock:* the converter inside the AFE is timed by a clock. The ADS1293 needs a 4.096 MHz crystal (a small quartz resonator that sets an exact frequency) between XTAL1 and XTAL2 with a 20 pF capacitor on each pin (the kit fits 22 pF); the ADS1299 has no crystal pins and runs from its internal 2.048 MHz oscillator or from a 2.048 MHz oscillator module on its CLK pin. *Input network and drive:* the series resistors and filter capacitors of the inputs, and the right-leg or bias components (Section 8.5). *Connectors:* the FFC connector and the Nicla headers.

The interface to the Nicla is SPI, the four-wire serial link of Section 8.8, plus two or three control lines (Table L3.1). Every Nicla header pin passes through a level translator, a chip that shifts logic signals between two voltages, powered by VDDIO_EXT, which software sets between 1.8 V and 3.3 V, so the AFE's digital supply must equal VD-

DIO_EXT. The translators can drive very little current: do not power the AFE from a signal pin.

Nicla pin	ADS1293 pin	ADS1299 pin
D6 (CS)	CSB	$\overline{\text{CS}}$
D9 (SCLK)	SCLK	SCLK
D8 (COPI)	SDI	DIN
D7 (CIPO)	SDO	DOUT
D0	DRDYB	$\overline{\text{DRDY}}$
D5 (spare)	RSTB (with pull-up)	START or $\overline{\text{RESET}}$
VDDIO_EXT	VDDIO	DVDD
GND	DGND	DGND

Table L3.1: A suggested connection between the Nicla Sense ME and the AFE. The SPI pins are fixed by the Nicla (D6 to D9); D0, which a past ECG team used for DRDYB, and D5 are general-purpose pins. Check every pin against the full pinout of your board revision.

Layout rules. The datasheet's layout example and the planning documents of past teams agree on the rules. Keep the analog section (FFC connector, input network, AFE inputs, crystal, references) on one side of the board and the digital section (Nicla headers, SPI) on the other. Give the board a solid ground plane; if you split analog and digital ground, join them at one point near the AFE, as the kit does, and never route a trace across the gap. Keep return paths short, put each decoupling capacitor next to its pin with a via straight to the plane, place bypass capacitors (ceramic) across differential inputs, keep input traces short and of equal length, and put small series resistors on long digital lines near the driver.

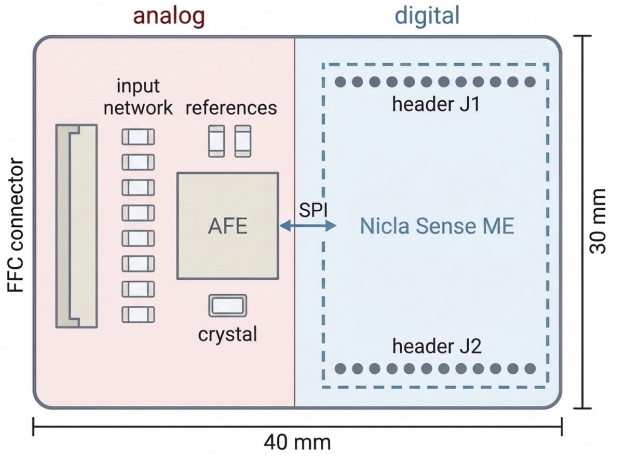


Figure L3.2: A floor plan for a 40 mm × 30 mm daughter board. The FFC connector, input network, AFE, crystal, and references form the analog side; the Nicla's headers, which sit under the Nicla's outline (dashed), and the SPI lines form the digital side. The outline and pin positions shown are schematic: the Nicla is a 22.86 mm square, so take the real ones from its mechanical drawing.

Lessons from past teams. Student ECG and EEG teams have built boards like this one, and their reports are a checklist of what to avoid. An ECG team's parts list called for a 12 MHz crystal (its part number, ECS-120-20-5PX, codes the frequency) although the ADS1293 expects 4.096 MHz. The same team's first board reset the AFE every time the sketch printed to the serial port, because the AFE's reset pin RSTB was

wired to a Nicla header pin; they cut the trace and let the pull-up hold RSTB high. The team also ran the AFE's analog supply at 5.5 V, the very top of its range, where a 3.0 V or 3.3 V rail is the safer choice. An EEG team built a 40 mm by 25 mm four-layer board with the Nicla on two rows of headers; its second revision lost the top and bottom ground pours by mistake, and its SPI lines ran under the ADS1299, both sources of noise. That team's report describes AVDD and AVSS of +2.5 V and −2.5 V made by LDOs from the Nicla's 3.3 V. An LDO can only lower a voltage: a negative rail from a positive supply needs an inverting charge pump or converter in front of the negative LDO, and +2.5 V from 3.3 V leaves only 0.8 V of headroom for the positive one. Check every rail against the physics as well as against the block diagram.

Procedure

Work in your project team. Steps tagged EDIT change your design files; steps tagged ANSWER go in your lab notebook; the two checkpoints are reviews that an instructor or another team signs.

Part A: Read the reference design

GOAL Know every rail, clock, and reference your AFE needs before you draw anything. about 40 min

A1 DO Collect the documents

Download the AFE datasheet and the evaluation kit's user guide, and open its schematic. Open the Nicla Sense ME datasheet at its pinout and its mechanical drawing. In the AFE datasheet, find the pin table, the recommended operating conditions, the power-supply and layout recommendations, and the application example closest to your project (TI's 3-lead and 5-lead ECG examples for the ADS1293).

A2 ANSWER The power table

A2 · ANSWER in your lab notebook

RECORD a table with one row per supply pin of the AFE: pin name, allowed range, the voltage you choose, typical current, and the regulator that makes it on the kit (part number, input, output, maximum current, dropout).

UNIT volts and milliamperes.

WHERE TO FIND IT the datasheet's pin table and recommended operating conditions; the power page of the kit schematic.

A3 ANSWER Clock and references

A3 · ANSWER in your lab notebook

RECORD the clock source (crystal or oscillator), its frequency and load capacitance, and the capacitors on XTAL1 and XTAL2 or the CLK and CLKSEL connections; each reference pin and its capacitor.

UNIT megahertz, picofarads, microfarads.

WHERE TO FIND IT the datasheet's clock and reference sections and the kit schematic around the AFE.

A4 DO Mark up the kit schematic

On a printout or a PDF copy of the kit schematic, mark in one color what you keep (AFE, regulators, decoupling, clock, references, input network, RLD or bias network, pull-ups on reset and sync pins) and in another what you drop (USB, microcontroller, current sense, jumpers, EEPROM, test points). Note every jumper whose position you must hard-wire, and how.

YOU SHOULD SEE about 20 to 30 parts left for an ADS1293 board: the AFE, one or two LDOs, six or so capacitors around the AFE, two crystal capacitors and the crystal, R_i, R_f, C_f for the RLD, input resistors, two pull-ups, the FFC connector, and the Nicla headers. A past ECG team's parts list had 24 parts. An ADS1299 board has roughly twice as many, mostly capacitors and the input network.

Part B: Schematic

GOAL Draw a schematic that passes ERC and a review by another team. about 80 min

B1 DO Create the project

Create a KiCad project named after your team (Lab 1 shows how) with three schematic sheets: `power`, `afe`, and `connectors`. Use global net labels for the nets that cross sheets: `AVDD`, `VDDIO` (or `DVDD`), `AGND`, `DGND`, `CS`, `SCLK`, `COPI`, `CIPO`, `DRDY`, and the electrode nets.

B2 EDIT Draw the power sheet

B2 · EDIT

WHAT TO DO the power input (battery or the Nicla's VIN, through a switch if you want one), the LDO for the analog supply with its input and output capacitors, the digital supply (VDDIO from the Nicla's VDDIO_EXT, or a second LDO), and the ground symbols. For the ADS1299 in bipolar mode, draw the positive regulator and the inverting charge pump followed by the negative regulator.

HINT a 3.7 V LiPo cell falls to about 3.2 V near empty, so an LDO for a 3.3 V rail runs out of headroom early; a 3.0 V rail with a low-dropout part runs the ADS1293 (minimum 2.7 V) to the end of the cell. Put the regulator's capacitors exactly as its datasheet asks: some LDOs need a minimum capacitance or a minimum ESR (equivalent series resistance, the small resistance inside a real capacitor) to be stable.

DONE WHEN every rail has a regulator, an input capacitor, an output capacitor, and a net label, and the rail voltages match your table from A2.

B3 EDIT Draw the AFE sheet

B3 · EDIT

WHAT TO DO the AFE symbol with a 0.1 µF capacitor at every supply pin, the reference capacitors, the clock (crystal and its two capacitors, or the oscillator module), the RLD or bias network (R_i, R_f, C_f for the ADS1293), pull-ups on RSTB and SYNCB (the kit uses 1 MΩ to VDDIO), the input series resistors and filter capacitors, and the exposed pad of the ADS1293 tied to ground as the kit ties it.

HINT copy values from the kit, then check each against the datasheet text. Read the crystal's part number: the frequency is coded in it. Leave the ALARMB output connected to a test pad if you have no pin for it.

DONE WHEN each AFE pin is connected, tied off, or marked no-connect, and each value on the sheet appears either in the datasheet or in the kit schematic.

B4 EDIT Draw the connector sheet

B4 · EDIT

WHAT TO DO the 10-pin, 1 mm pitch FFC connector, with its pins named after your electrode map (for the ECG: RA, LA, LL, RL, V1, and spares; for the EEG: eight channels, the reference to SRB1, and the bias electrode), and the two Nicla headers J1 and J2 with the nets of Table L3.1.

HINT decide whether the cable's exposed contacts face up or down when it enters the connector, and choose a top-contact or bottom-contact part to match your sensor's tail; a mirrored pin map is the most common FFC mistake. A past ECG team used a 1.00 mm pitch, horizontal, zero-insertion-force (ZIF) connector.

DONE WHEN the pin-1 end of the FFC is marked on the sheet, and the electrode order matches the drawing of your sensor's tail pad by pad.

B5 ANSWER The interface table

B5 · ANSWER in your lab notebook

RECORD for each AFE digital pin: the net, the Nicla pin, the direction, and the logic level on each side.

UNIT volts for the levels.

WHERE TO FIND IT your schematic and the Nicla pinout.

B6 ▸ RUN Electrical rules check

Run the ERC in the schematic editor. Add power-flag symbols where KiCad cannot see that a rail is driven, and fix every error.

✓CHECK **Checkpoint 1** Schematic review. Another team checks your schematic against this list and signs it: every supply pin connected and decoupled; rail voltages inside the datasheet ranges, with LDO headroom at the lowest battery voltage; the clock frequency in the part number equals the datasheet's; reset and sync pins pulled up and driven only by a pin you control; the digital supply of the AFE equals VDDIO_EXT; the RLD or bias network present; the FFC pin map matches the sensor; ERC clean. You are now ready to commit to copper.

Part C: Footprints and bill of materials

GOAL Give every symbol a real footprint and a real part.

about 30 min

C1 ✎EDIT Assign footprints

C1 · ✎EDIT

WHAT TO DO a footprint for every symbol: 0402 or 0603 passives (resistors and capacitors); the ADS1293 in its 28-pin 5 mm WQFN package, with the exposed pad (a metal pad under the chip) and the thermal vias of the datasheet's land pattern, or the ADS1299 in its 64-pin TQFP package; the FFC connector from its manufacturer's drawing; and a custom footprint for the Nicla headers drawn from the Nicla's mechanical drawing.

HINT measure the Nicla's pin pitch and the spacing between its two rows from the drawing, and check them with a caliper on a real board; print the footprint at 1:1 and set a Nicla on the paper.

DONE WHEN the footprint assignment tool shows no unassigned symbol, and the printed Nicla footprint lines up with a real board.

C2 ◆DO Fill in the part fields

For every part, fill in the value, the package, the manufacturer's part number (MPN), and the distributor's part number, and mark parts you do not fit as DNP (do not populate). If a part is out of stock, choose a substitute whose datasheet matches the parameters you need, and note why it is equivalent.

C3 ANSWER Export the BOM

C3 · ANSWER in your lab notebook

RECORD export the bill of materials from the schematic editor as a CSV file with one line per part number (reference designators, quantity, value, footprint, MPN); record the number of lines, the total number of parts, and any substitution.

UNIT parts.

WHERE TO FIND IT the BOM export of the schematic editor; the assembler's instructions for the column format.

Part D: Layout

GOAL Place and route a quiet, DRC-clean board that a Nicla plugs onto. about 100 min

D1 DO Outline, stack-up, and fixed parts

Draw the board outline (about 40 mm by 30 mm is enough; smaller is better on the body) and choose four layers if you can: signals, ground plane, power, signals. Place the Nicla headers first, exactly at the positions of your footprint from **C1**, then the FFC connector at the board edge facing the sensor, then mounting holes or pads for strain relief.

D2 EDIT Place the parts

D2 · EDIT

WHAT TO DO place the parts by the floor plan of Fig. L3.2: the AFE close to the FFC connector, the input network between them, the crystal within a few millimeters of its pins, each decoupling capacitor at its pin, the reference capacitors at theirs, the LDO near the power entry, and the digital parts toward the headers.

HINT place, then look at the ratsnest: crossing lines between the analog and digital sides mean a part is in the wrong place. Keep the space under the Nicla free of tall parts.

DONE WHEN no analog net crosses the digital side, and each decoupling capacitor is within about 2 mm of its pin.

D3 EDIT Route

D3 · EDIT

WHAT TO DO route the input traces short and side by side in pairs; route SPI from the AFE straight to the headers, never under the AFE's analog pins or the crystal; pour ground on the outer layers and stitch the pours to the plane with vias; give each decoupling capacitor its own via to the plane.

HINT route the clock first and keep it short, then the power, then the

inputs, then SPI. A 22 Ω to 51 Ω series resistor near the driver on SCLK and on the AFE's data output tames ringing on long lines (the ADS1293 kit uses 51 Ω on two of them).

DONE WHEN every net is routed, and the ground pours are present on both outer layers after you refill the zones.

D4 ▸ RUN Design rules check

Set the design rules to your board house's capabilities (minimum trace width, spacing, drill, and annular ring) and run the DRC. Fix every error and every unconnected item; justify any warning you leave.

✓CHECK **Checkpoint 2** DRC-clean layout. The DRC reports no errors and no unconnected items; the 3D view shows the Nicla seated on its headers with no collision; a 1:1 printout of the copper and silkscreen lines up with a real Nicla and with the sensor's FFC tail; an instructor has looked at the placement of the clock, the decoupling, and the SPI routing.

Part E: Fabrication files

GOAL Produce the files a board house needs to build and assemble your board. about 20 min

E1 ◆ DO Plot Gerbers and drill files

Plot the copper, solder mask, silkscreen, paste, and edge-cut layers as Gerbers and generate the drill files. Open them in KiCad's Gerber viewer and check every layer, especially the paste openings of the exposed pad and the FFC connector.

E2 ◆ DO Export the assembly files

Export the component-position (pick-and-place) file and save the BOM of C3 in the assembler's format. Zip the Gerbers and drill files for the board house.

E3 ⊡ ANSWER Design summary

E3 · ⊡ ANSWER in your lab notebook

RECORD board size, layer count, number of parts and of unique part numbers, smallest trace and drill, and the quote and lead time from the board house.

UNIT millimeters, parts, dollars, days.

WHERE TO FIND IT the board setup, the BOM, and the board house's quote page.

Hand in

Finish the lab

You designed a mixed-signal board: an AFE with its power, clock, and protection, a sensor connector, and a microcontroller interface, ready to be built.

- ☐ **A2** ANSWER the power table
- ☐ **A3** ANSWER clock and references
- ☐ **B2** EDIT power sheet
- ☐ **B3** EDIT AFE sheet
- ☐ **B4** EDIT connector sheet with the FFC pin map
- ☐ **B5** ANSWER the interface table
- ☐ CHECK Checkpoint 1: schematic review signed by another team
- ☐ **C1** EDIT footprints, including the Nicla headers checked at 1:1
- ☐ **C3** ANSWER BOM as CSV
- ☐ **D2** EDIT placement
- ☐ **D3** EDIT routing with ground pours
- ☐ CHECK Checkpoint 2: DRC-clean layout
- ☐ **E3** ANSWER design summary with the quote
- ☐ Analysis questions 1 to 7 answered in your notebook
- ☐ **Save** SAVE the zipped project, schematic PDF, DRC report, BOM, Gerbers, drill and position files

Typical results

Quantity	Typical value
Board size	25 to 40 mm on a side; a past EEG board was 40 mm × 25 mm
Layers	two for a simple ADS1293 board; four for the ADS1299
Parts	about 25 (ADS1293) to 50 (ADS1299)
Supplies	ADS1293: one 3.0 V to 3.3 V rail; ADS1299: 5 V or ±2.5 V plus DVDD
Clock	4.096 MHz crystal (ADS1293); internal or 2.048 MHz oscillator (ADS1299)
Revisions	past teams needed two board revisions

Table L3.2: Typical outcomes of Lab 3, from the boards of past teams and the evaluation kits.

Most teams finish the schematic in the first session and the layout in the second (Table L3.2). The review of Checkpoint 1 is where most mistakes are found: a crystal of the wrong frequency, a reset pin on a busy header pin, an FFC pin map mirrored, or a rail that cannot be made from the battery. Boards come back from fabrication and

assembly in one to three weeks, so a second revision costs a large part of a project schedule; it pays to review slowly.

Troubleshooting

- *ERC errors on power pins:* add a power-flag symbol to each rail that comes from a connector or a regulator KiCad cannot trace.
- *The Nicla does not fit:* the header footprint does not match the drawing; measure again from the drawing and with a caliper, and reprint at 1:1.
- *The FFC cable makes no contact, or the electrodes are in reverse order:* the connector's contact side or pin 1 does not match the tail; check against the sensor drawing.
- *The AFE resets or stops when the firmware does something unrelated:* a reset or sync pin is on a header pin that the firmware also uses, or is floating; keep the pull-up and drive it from a dedicated pin.
- *No conversions on the first board (Lab 4 reports no data):* check the crystal frequency, its capacitors, and the solder joints of the exposed pad.
- *Noisy data:* look for SPI traces under the AFE, missing ground pours, a split ground crossed by a trace, or decoupling capacitors far from their pins.

Analysis questions

1. List what you removed from the evaluation kit and why each part was safe to remove.
2. Which LDO did you choose? Compare its dropout, noise, power-supply rejection, and quiescent current with the kit's part, and find the lowest battery voltage at which your analog rail stays in regulation.
3. How long will your board run from the Nicla's battery? Use the method of Example 13.1 with the currents from your power table.
4. Why must the digital supply of the AFE equal VDDIO_EXT? What would happen if the AFE ran its inputs and outputs (I/O) at 3.3 V and the Nicla at 1.8 V?
5. For the ADS1299 in bipolar mode, draw the chain that makes -2.5 V from a single-cell battery, and explain why an LDO alone cannot.

6. Explain, with Eq. (8.11), why the input traces should be short, close together, and of equal length.
7. Which design-review finding of your team (or of the past teams) would have cost the most to discover after fabrication, and why?

Lab 4
Streaming Biopotentials over SPI and Bluetooth

Lab 4 at a glance

What you build: *firmware*, the program that runs on the microcontroller, for the Arduino Nicla Sense ME. It configures your AFE over SPI, reads every sample when the AFE signals that data are ready, and streams the samples over Bluetooth Low Energy (BLE), the low-power radio link of most wearables, to a Python program on your laptop. The program filters them and computes a heart rate (ECG teams) or EEG band powers, the strength of a brain rhythm in a band of frequencies (EEG teams).
Time: about three hours.
Before the lab: read Sections 8.7 to 8.9 of Chapter 8 and Section 8.10 on electrical safety; install the software below; charge the Nicla's battery.
You hand in: your sketches and Python script, a 60 s recording received over BLE (CSV), the plots of Parts C and E, and your answers.

What you need

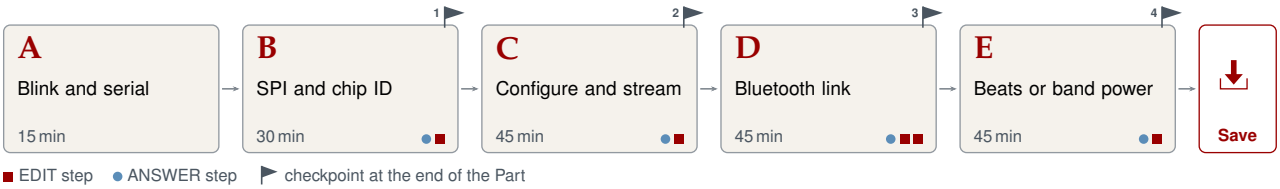
Hardware: an Arduino Nicla Sense ME on your Lab 3 daughter board, or, until that board arrives, wired to an AFE breakout board (such as the ProtoCentral ADS1293 breakout) or to the TI evaluation kit; a 3.7 V single-cell LiPo battery for the Nicla; a laptop that can run on its own battery; three to five gel Ag/AgCl electrodes and leads (or your printed sensor, whose tail plugs only into the daughter board's FFC connector; see Lab 3); a multimeter.

Software: Arduino IDE 2 with the board package *Arduino Mbed OS Nicla Boards*; the libraries ArduinoBLE and Nicla_System

(installed with the board package); for reference, the ProtoCentral ADS1293 ECG Library (version 2.0 or later, from the Library Manager) and the OpenBCI ADS1299 Arduino code; Python 3.10 or later with `bleak`, `numpy`, `scipy`, and `matplotlib`.

Files: the listings of this lab; your Lab 3 interface table.

SAFETY Electrodes go on a person only when nothing in the system connects to the mains (Section 8.10). Before you attach electrodes, unplug the laptop's charger, disconnect any oscilloscope, function generator, or desktop computer from the board, and run the Nicla from its battery. Never connect a function generator and a person to the board at the same time. Remove the electrodes if the skin itches or reddens. This is a teaching system, not a medical device: an odd reading says more about the setup than about your heart.



Objectives

- Program the Nicla Sense ME and read its serial output.
- Read and write AFE registers over SPI in the right mode, and verify the wiring with the chip's ID register.
- Configure the ADS1293 (or ADS1299) register by register, explain each write, and read samples on the data-ready signal.
- Stream samples over BLE notifications with a sequence number and receive them in Python without losses.
- Filter the ECG and detect R peaks (or compute EEG band power), and check the result against a measurement you make by hand.

Background

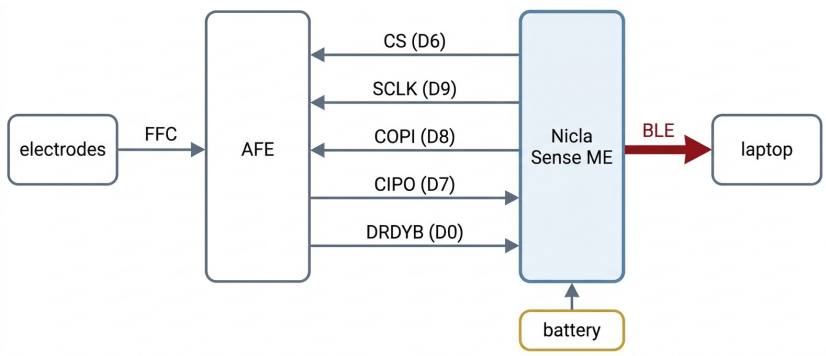


Figure L4.1: The system of Lab 4. The Nicla is the SPI controller: it selects the AFE with CS (Nicla pin D6, to the ADS1293's CSB), clocks it on SCLK (D9), sends commands on COPI (D8, to SDI), and reads data on CIPO (D7, from SDO); the AFE pulls DRDYB (to D0) low when a sample is ready. Samples leave over BLE, and the whole system runs from the battery.

Figure L4.1 shows the system. The AFE is controlled through its *registers*, numbered one-byte memory cells inside the chip that hold its settings and its results. The firmware does four things. It sets up SPI in the AFE's mode, which fixes on which edge of the clock the data change and are read (mode 0 for the ADS1293, mode 1 for the ADS1299; Section 8.8). It writes the configuration registers and reads them back. It waits for the data-ready signal and reads each sample in one SPI *transaction* (one period with CS held low). And it packs samples into 20-byte BLE *notifications*, short messages that the Nicla pushes to the laptop without being asked. Two open libraries make good references: ProtoCentral's ADS1293 library and an ADS1299 library written by Conor Russomanno for OpenBCI. You will write the register access yourself, in a few lines, so that you know what every byte does; the libraries are references. Two warnings. The ProtoCentral library was rewritten as version 2.0 in 2026 with a new interface, and its changelog records that version 1.x mislabeled its data rates and decoded the ADS1293's offset-binary output as two's complement, two different ways of writing a signed number in bits (**C2** explains them). The original ADS1299 library writes the SPI registers of AVR processors (`SPCR` and `SPDR`) directly, so it compiles for an Arduino Uno but not for the Nicla; on the Nicla use the Arduino `SPI` library with `SPISettings` instead.

Procedure

ECG teams follow the ADS1293 listings; EEG teams use the ADS1299 variants where a step gives one. Each listing is short; the lines marked `EDIT` are yours to write.

Part A: Blink and serial

GOAL Program the Nicla and see it print. about 15 min

A1 ◆ DO Set up the IDE

In Arduino IDE 2, install the board package *Arduino Mbed OS Nicla Boards* from the Boards Manager and the ArduinoBLE library from the Library Manager. Plug the Nicla in with a USB data cable, choose *Arduino Nicla Sense ME* and its port.

A2 ▸ RUN Blink the LED and print

The Nicla's RGB LED hangs on an I^2C driver, so it is switched through the Nicla_System library rather than with `digitalWrite`.

A2 · code

```
#include "Nicla_System.h"   // RGB LED and power chip of the Nicla

void setup() {
  nicla::begin();
  nicla::leds.begin();
  Serial.begin(115200);
}

void loop() {
  nicla::leds.setColor(green);  delay(250);
  nicla::leds.setColor(off);    delay(750);
  Serial.print("alive at ms = ");
  Serial.println(millis());
}
```

YOU SHOULD SEE a green flash once per second and, in the serial monitor at 115200 baud, a line per second whose number grows by about 1000.

Part B: SPI and the chip ID

GOAL Prove the SPI wiring and mode by reading the chip's ID. about 30 min

B1 ◆ DO Wire and power

On the daughter board the wiring is done; on a breakout, connect the pins of Table L3.1 (Lab 3) with short wires and power the AFE's digital supply from the same voltage as the Nicla's VDDIO_EXT. Check with the multimeter that the AFE's analog supply is at its design voltage before you load any code. VDDIO_EXT, and with it the AFE's digital supply, is set by the sketch below (`enable3V3LDO()` gives 3.3 V, `enable1V8LDO()` 1.8 V); measure it once the sketch runs. On a breakout whose reset and sync lines pass through a level translator (the

ProtoCentral board routes RSTB and SYNCB through a TXB0108), do not leave them floating: an undriven line can hold the chip in reset or keep resetting its filters. Wire RSTB to D5 and drive it high at the start of `setup()` (`pinMode(5, OUTPUT);` `digitalWrite(5, HIGH);`), and tie SYNCB to the digital supply.

B2 EDIT Register access

B2 · code

```
#include <SPI.h>
#include "Nicla_System.h"  // power chip: sets VDDIO_EXT
const int CS_PIN = 6;      // Nicla D6 -> ADS1293 CSB
const int DRDY_PIN = 0;    // Nicla D0 <- ADS1293 DRDYB
SPISettings afeSPI(1000000, MSBFIRST, SPI_MODE0);  // 1 MHz

uint8_t readReg(uint8_t addr) {
  SPI.beginTransaction(afeSPI);
  digitalWrite(CS_PIN, LOW);
  SPI.transfer(addr | 0x80);         // bit 7 = 1: read
  uint8_t v = SPI.transfer(0x00);
  digitalWrite(CS_PIN, HIGH);
  SPI.endTransaction();
  return v;
}
void writeReg(uint8_t addr, uint8_t v) {
  // EDIT B2: the same frame with bit 7 = 0, then send v
}
void setup() {
  nicla::begin();
  nicla::enable3V3LDO();     // VDDIO_EXT = 3.3 V; see B1
  Serial.begin(115200);
  pinMode(CS_PIN, OUTPUT);
  digitalWrite(CS_PIN, HIGH);
  pinMode(DRDY_PIN, INPUT);
  SPI.begin();
  delay(100);
  Serial.print("REVID 0x");
  Serial.println(readReg(0x40), HEX);
  writeReg(0x0A, 0x07);              // CMDET_EN, used in Part C
  Serial.print("CMDET 0x");
  Serial.println(readReg(0x0A), HEX);
}
void loop() {}
```

B2 · EDIT

WHAT TO DO the body of `writeReg`: begin the transaction, pull CS low, send the address with bit 7 cleared (`addr & 0x7F`), send `v`, raise CS, and end the transaction.

HINT compare with `readReg`; only the first byte and the meaning of the second change.

DONE WHEN the serial monitor prints `REVID 0x1` and `CMDET 0x7` every time you press reset.

For the ADS1299, change the mode to `SPI_MODE1` and use its commands:

B2-EEG · code

```
SPISettings eegSPI(1000000, MSBFIRST, SPI_MODE1);  // mode 1
void command(uint8_t op) {           // RESET 0x06, SDATAC 0x11
  SPI.beginTransaction(eegSPI);
  digitalWrite(CS_PIN, LOW);
  SPI.transfer(op);
  digitalWrite(CS_PIN, HIGH);
  SPI.endTransaction();
}
uint8_t readRegEEG(uint8_t addr) {
  SPI.beginTransaction(eegSPI);
  digitalWrite(CS_PIN, LOW);
  SPI.transfer(0x20 | addr);         // RREG + start address
  SPI.transfer(0x00);                // number of registers - 1
  uint8_t v = SPI.transfer(0x00);
  digitalWrite(CS_PIN, HIGH);
  SPI.endTransaction();
  return v;
}
// setup(): with PWDN and RESET high, delay(150) first
//          (power-on reset: 2^18 clocks = 128 ms), then
//          command(0x06); delay(1); command(0x11);
//          then print readRegEEG(0x00)
```

A register write is the same frame with `0x40` in place of `0x20` and the value as a third byte. The ADS1299 starts in RDATAC mode and ignores register commands until it receives SDATAC.

B3 ANSWER The ID, ten times

B3 · ANSWER in your lab notebook

RECORD the ID you read (ADS1293: REVID at `0x40`; ADS1299: ID at `0x00`) in ten resets, and what you read when you deliberately set the wrong SPI mode.

UNIT hexadecimal bytes.

WHERE TO FIND IT the serial monitor.

YOU SHOULD SEE ADS1293: `0x01` every time. ADS1299: a byte whose low five bits are `1 1110`, usually `0x3E`. In the wrong mode the bytes come back shifted by one bit or as `0x00` or `0xFF`.

✓CHECK **Checkpoint 1** The ID reads correctly ten times in ten resets, and a register you write reads back the same value. The Nicla and the AFE are talking.

Part C: Configure the AFE and stream to the serial plotter

GOAL Configure the AFE register by register and see a live ECG (or EEG). about 45 min

C1 ▸ RUN The ADS1293 configuration, line by line

A past ECG team configured the ADS1293 with the fifteen steps of the 5-lead example in TI's datasheet. Their register values were right, but several of their comments did not match the values written, a step was written twice, and one had no comment. Table L4.1 gives the sequence with the meaning of each value; the electrodes are RA on IN1, LA on IN2, LL on IN3, RL on IN4, and V1 on IN5. The TI example converts at 853 samples per second ($R_2 = 5$, $R_3 = 6$; R_2 and R_3 are the decimation ratios of the AFE's digital filters, which set the sample rate). A serial link at 115200 baud carries at most 11520 characters per second, and three numbers with separators take about 25 characters, so 853 samples per second would need more than 21000. This lab therefore changes only the rate registers, to $R_2 = 4$ and $R_3 = 32$: 200 samples per second with a 40 Hz bandwidth, enough for heart rate and for BLE (Example 8.5). The listing first writes `0x00` to CONFIG: the AFE keeps converting when the Nicla resets, and while conversions run it ignores writes to the clock, rate, and DRDYB registers. It then starts the crystal oscillator (OSC_CN = `0x00`) and waits 20 ms, longer than the oscillator's 15 ms start-up, before the table sets STRTCLK (`0x04`) to feed the clock to the digital core, the order the datasheet asks for.

Address	Value	Register	Effect
`0x01`	`0x11`	FLEX_CH1_CN	CH1: positive IN2 (LA), negative IN1 (RA): lead I
`0x02`	`0x19`	FLEX_CH2_CN	CH2: positive IN3 (LL), negative IN1 (RA): lead II
`0x03`	`0x2E`	FLEX_CH3_CN	CH3: positive IN5 (V1), negative IN6 (WCT): lead V1
`0x0A`	`0x07`	CMDET_EN	common-mode detector on IN1, IN2, IN3
`0x0C`	`0x04`	RLD_CN	RLD output to IN4, the right-leg electrode (no comment in the team's code)
`0x0D` to `0x0F`	`0x01`, `0x02`, `0x03`	WILSON_EN1 to 3	Wilson buffers on IN1, IN2, IN3
`0x10`	`0x01`	WILSON_CN	Wilson terminal to IN6
`0x12`	`0x04`	OSC_CN	external crystal; clock to the digital core (written twice in the team's code)
`0x21`	`0x02`	R2_RATE	$R_2 = 5$ (team's comment said 8, code `0x08`); this lab: `0x01`, $R_2 = 4$
`0x22` to `0x24`	`0x02`	R3_RATE_CH1 to 3	$R_3 = 6$ (team's comment said 16, code `0x10`); this lab: `0x20`, $R_3 = 32$
`0x27`	`0x08`	DRDYB_SRC	DRDYB follows CH1
`0x2F`	`0x70`	CH_CNFG	loop read-back of CH1, CH2, CH3 ECG data
`0x00`	`0x01`	CONFIG	start conversions

Table L4.1: The ADS1293 5-lead configuration of TI's datasheet, as a past team wrote it, with corrected comments. In FLEX_CHn_CN, bits 5 to 3 select the positive input and bits 2 to 0 the negative input (001 = IN1, ..., 110 = IN6); for example `0x11` is 00 010 001, IN2 and IN1.

C1 · code

```
const uint8_t cfg[][2] = {            // {address, value}
  {0x01,0x11}, {0x02,0x19}, {0x03,0x2E}, {0x0A,0x07},
  {0x0C,0x04}, {0x0D,0x01}, {0x0E,0x02}, {0x0F,0x03},
  {0x10,0x01}, {0x12,0x04},
```

```
  {0x21,0x01},                        // R2 = 4
  {0x22,0x20}, {0x23,0x20}, {0x24,0x20},   // R3 = 32
  {0x27,0x08}, {0x2F,0x70}};
void setupECG() {                     // 102400/(4*4*32) = 200 sps
  writeReg(0x00, 0x00);               // stop conversions first
  writeReg(0x12, 0x00);               // oscillator on, clock off
  delay(20);                          // 15 ms start-up; the table
                                      // then sets STRTCLK (0x04)
  for (auto &r : cfg) {
    writeReg(r[0], r[1]);
    if (readReg(r[0]) != r[1]) {
      Serial.print("verify failed 0x");
      Serial.println(r[0], HEX);
    }
  }
  writeReg(0x00, 0x01);               // start conversions last
}
```

C2 EDIT Read a sample on DRDYB

The AFE pulls DRDYB low when a sample is ready. An *interrupt* makes the processor run a short function the moment that pin falls; here the function only sets a flag, and `loop()` does the reading. The ADS1293 writes each sample as an unsigned 24-bit code in *offset binary*: zero volts reads as half the full-scale code (ADCMAX/2), so subtracting that half gives a signed count. Two's complement, the usual signed format of computers, instead marks negative numbers by a set top bit.

C2 · code

```
volatile bool ready = false;
void onDrdy() { ready = true; }       // DRDYB fell: new sample
const long HALF = 0x400000;           // ADCMAX/2 (R2 = 4, R3 = 32)
const float UV_PER_COUNT =            // 0.1635 uV per count
    2 * 2.4e6 / (3.5 * 0x800000);

void readSample(long c[3]) {          // one loop read, CH1..CH3
  SPI.beginTransaction(afeSPI);
  digitalWrite(CS_PIN, LOW);
  SPI.transfer(0x50 | 0x80);          // DATA_LOOP, read
  // EDIT C2: 3 bytes per channel, MSB first, minus HALF
  digitalWrite(CS_PIN, HIGH);
  SPI.endTransaction();
}
void loop() {
  if (!ready) return;
  ready = false;
  long c[3];
  readSample(c);
  Serial.print(c[0] * UV_PER_COUNT);    // lead I, uV
  Serial.print(' ');
  Serial.println(c[1] * UV_PER_COUNT);  // lead II, uV
}
// setup(): add setupECG(); and
//   attachInterrupt(digitalPinToInterrupt(DRDY_PIN),
//                   onDrdy, FALLING);
```

C2 · EDIT

WHAT TO DO the loop over the three channels: shift in three bytes with `SPI.transfer(0)` into an unsigned 32-bit value, then subtract `HALF` to get a signed count.

HINT the output is offset binary (Eq. (8.25)): zero volts reads ADC-MAX/2. Do not sign-extend as for two's complement. Never call SPI from inside the interrupt routine; it only sets the flag.

DONE WHEN with the inputs shorted together and to the right-leg pin the plotter shows two flat lines within a few microvolts of zero.

For the ADS1299, the start-up and read sequence are: RESET (`0x06`), SDATAC (`0x11`); CONFIG3 (`0x03`) = `0xE0` to turn on the internal reference buffer and a wait for the reference to settle; CONFIG1 (`0x01`) = `0x96`, 250 samples per second, its reset value; CH1SET to CH8SET (`0x05` to `0x0C`) = `0x60` for gain 24 and a normal electrode input, or `0x50` for gain 12 with printed electrodes (Example 8.5); MISC1 (`0x15`) = `0x20` to connect SRB1, the reference electrode, to every channel's negative input; BIAS_SENSP (`0x0D`) = `0xFF` and CONFIG3 = `0xEC` to derive the bias from all channels and turn on the bias buffer; then START (`0x08`) and RDATAC (`0x10`). Each time DRDY falls, read 27 bytes (3 status bytes, then 3 per channel) and convert each channel from two's complement. These values are a common starting point; check every bit against the register tables of your datasheet revision before you rely on them.

C3 ANSWER Rate and noise

C3 · ANSWER in your lab notebook

RECORD the sample rate, from the number of samples printed in 10 s (add a counter and `millis()`), and the rms noise of one channel with its inputs shorted together and to the right-leg pin.

UNIT samples per second; microvolts rms.

WHERE TO FIND IT the serial monitor; compute the rms in Python from a pasted block of values.

C4 DO Your own ECG (or EEG)

Follow the safety box: battery only, charger unplugged, USB only to a laptop running on its own battery, or skip to Part D and watch over BLE. Place gel electrodes on the right and left shoulders (RA, LA), the left lower abdomen (LL), and the right lower abdomen (RL), or on the chest positions of your printed sensor. Sit still and open the serial plotter. EEG teams place the array or gel electrodes behind the ear (or at O1 with the reference on the earlobe) and close their eyes for 20 s.

YOU SHOULD SEE lead II with an R wave of about 0.5 to 2 mV (500 to 2000 µV) once per heartbeat, a smaller lead I, and a baseline that wanders when you breathe or move. For EEG: a 10 Hz ripple that appears when the eyes close, a few microvolts behind the ear (about 2 µV rms) and tens of microvolts at O1.

✓CHECK **Checkpoint 2** A live ECG with recognizable P, QRS, and T waves (or a visible alpha rhythm with the eyes closed), at the sample rate you designed. You have built the analog front end of an electrocardiograph.

Part D: The Bluetooth link

GOAL Stream samples over BLE to Python with no lost packets.

about 45 min

D1 ✎EDIT The BLE peripheral

A BLE device offers *services*, each holding *characteristics*, data fields that a laptop can read or subscribe to; each is named by a 128-bit UUID (universally unique identifier). The Nicla is the *peripheral*, the device that advertises and is connected to, and the laptop is the central.

D1 · code

```
#include <ArduinoBLE.h>
#define SVC_UUID "0b7e0001-5c0e-4d4a-9a59-0e1d2c3b4a59"
#define CHR_UUID "0b7e0002-5c0e-4d4a-9a59-0e1d2c3b4a59"
BLEService ecgService(SVC_UUID);
BLECharacteristic ecgChar(CHR_UUID, BLERead | BLENotify,
                          20, true);  // 20 bytes, fixed
uint8_t pkt[20];
int nInPkt = 0;
uint16_t seq = 0;

void bleSetup() {
  if (!BLE.begin()) { Serial.println("no BLE"); while (1); }
  BLE.setLocalName("ECG-team1");
  BLE.setAdvertisedService(ecgService);
  ecgService.addCharacteristic(ecgChar);
  BLE.addService(ecgService);
  BLE.setConnectionInterval(6, 12);  // 7.5 to 15 ms
  BLE.advertise();
}
void addSample(long a, long b) {     // seq, then 3 x (a, b)
  if (nInPkt == 0) { pkt[0] = seq & 0xFF; pkt[1] = seq >> 8; }
  uint8_t *p = pkt + 2 + 6 * nInPkt;
  // EDIT D1: a, then b, as 3 bytes each, MSB first
  if (++nInPkt == 3) {
    if (ecgChar.subscribed()) ecgChar.writeValue(pkt, 20);
    nInPkt = 0;
    seq++;
  }
```

```
}
// loop(): BLE.poll(); then addSample(c[0], c[1]) per sample
```

D1 · EDIT

WHAT TO DO six lines that write `a` and `b` into `p[0]` to `p[5]`: `(a » 16) & 0xFF`, `(a » 8) & 0xFF`, `a & 0xFF`, and the same for `b`. Then replace the serial printing of **C2** by `addSample`, and call `BLE.poll()` at the top of `loop()`.

HINT the signed counts of **C2** fit in 24 bits, so their low three bytes are their two's-complement form. Keep the payload at 20 bytes: the default ATT MTU, the largest packet the link carries, is 23 bytes, of which 20 are data, and ArduinoBLE truncates longer values without warning.

DONE WHEN a BLE scanner app on a phone lists `ECG-team1` with your service, and its characteristic changes when you subscribe.

EEG teams need more than this packet carries: eight channels of 3 bytes at 250 samples per second are 6000 bytes/s, 300 default notifications per second (Example 8.5). Test the link first with two channels in this packet, then stream what the project needs in one of the ways of that example and of Example 14.2: 16-bit samples at 125 samples per second, band powers computed on the Nicla and sent once a second, or a larger MTU negotiated with the laptop (check the value bleak reports in **D3**).

D2 EDIT The Python receiver

D2 · code

```
import asyncio
import numpy as np
from bleak import BleakScanner, BleakClient
UUID = "0b7e0002-5c0e-4d4a-9a59-0e1d2c3b4a59"
rows, seqs = [], []

def s24(b):                     # 3 bytes, MSB first -> signed
    v = (b[0] << 16) | (b[1] << 8) | b[2]
    return v - (1 << 24) if v & 0x800000 else v

def on_packet(sender, data: bytearray):
    seqs.append(data[0] | (data[1] << 8))
    # EDIT D2: three (lead I, lead II) pairs from data[2:20]

async def main(seconds=60):
    dev = await BleakScanner.find_device_by_name(
        "ECG-team1", timeout=10.0)
    async with BleakClient(dev) as client:
        print("MTU", client.mtu_size)
        await client.start_notify(UUID, on_packet)
        await asyncio.sleep(seconds)
        await client.stop_notify(UUID)
    np.savetxt("ecg_ble.csv", np.array(rows) * 0.1635,
               delimiter=",", comments="",
               header="leadI_uV,leadII_uV")
```

```
26
27 asyncio.run(main())
```

D2 · EDIT

WHAT TO DO a loop over `k = 0, 1, 2` that takes the six bytes starting at `2 + 6*k` and appends `(s24(...), s24(...))` to `rows`. Then, after the recording, find the gaps with `np.diff(seqs) % 65536 != 1` (a step of k means $k-1$ lost packets) and compute the received rate as `len(rows) / seconds`.

HINT `data[i:i+3]` slices three bytes. Run the script with the laptop on battery.

DONE WHEN the script saves `ecg_ble.csv` and prints zero lost packets for a 60 s recording.

D3 ANSWER The link budget, measured

D3 · ANSWER in your lab notebook

RECORD the MTU bleak reports, the notifications per second, the received sample rate, and the number of lost packets in 60 s, at 1 m and at 5 m from the laptop; compare with Eq. (8.26).

UNIT bytes, packets per second, samples per second.

WHERE TO FIND IT the output of your Python script.

CHECK **Checkpoint 3** A 60 s recording received over BLE with a continuous sequence counter and a sample rate within 1 % of the design. Your wearable is now wireless.

Part E: Heart rate (ECG) or band power (EEG)

GOAL Turn the stream into a number and check it by hand.

about 45 min

E1 EDIT Filter and find the beats

E1 · code

```
import numpy as np
from scipy.signal import (butter, sosfiltfilt, iirnotch,
                          filtfilt, find_peaks)
fs = 200.0
x = np.loadtxt("ecg_ble.csv", delimiter=",", skiprows=1)
x = x[:, 1] / 1000                      # lead II in mV
bp = butter(2, [0.5, 40], btype="bandpass", fs=fs,
            output="sos")
y = sosfiltfilt(bp, x)
b, a = iirnotch(60.0, 30.0, fs)
y = filtfilt(b, a, y)
qrs = sosfiltfilt(butter(2, [5, 15], btype="bandpass",
```

```
                          fs=fs, output="sos"), y)
w = np.ones(30) / 30                    # 150 ms window
e = np.convolve(np.gradient(qrs) ** 2, w, mode="same")
# EDIT E1: peaks of e above a threshold, >= 0.25 s apart;
#          ibi = np.diff(peaks) / fs; hr = 60 / ibi.mean()
```

E1 · EDIT

WHAT TO DO one call of `find_peaks` on `e` with `height` about 0.3 times the 99th percentile of `e` and `distance=int(0.25*fs)`, then the inter-beat intervals, the heart rate, and RMSSD (Eq. (8.5)), the root mean square of the differences between successive inter-beat intervals, a common measure of heart-rate variability. Plot `y` with the peaks marked. The smoothed energy `e` peaks near the R wave, not on it, so move each mark to the largest `abs(y)` within 15 samples (75 ms).

HINT the window of 30 samples is 150 ms at 200 samples per second; recompute it if you change the rate. In NumPy 1.x, `np.trapezoid` (used in E1-EEG) is called `np.trapz`.

DONE WHEN every R wave in the plot carries one mark and nothing else does.

For EEG teams, the band powers replace the beat finder. The function `welch` estimates the power spectrum, the power of the signal per unit of frequency; the area under the spectrum across a band is the band power:

E1-EEG · code

```
from scipy.signal import welch
# x_uV: one EEG channel in uV, loaded from your CSV as in E1
f, P = welch(x_uV, fs=250, nperseg=500)   # 0.5 Hz bins
def bandpower(lo, hi):
    m = (f >= lo) & (f < hi)
    return np.trapezoid(P[m], f[m])        # uV^2
# EDIT E1: alpha (8-13 Hz) and beta (13-30 Hz) power,
#          eyes open and eyes closed
```

E2 ANSWER Check against your own count

E2 · ANSWER in your lab notebook

RECORD ECG: count your pulse at the wrist for 30 s while the recording runs, twice, and compare with the detector's heart rate over the same 30 s. EEG: alpha power and the alpha/beta ratio for 20 s with the eyes open and 20 s closed, three times each.

UNIT beats per minute; μV^2 and a ratio.

WHERE TO FIND IT your notebook and the output of E1.

CHECK **Checkpoint 4** The detector's heart rate agrees with your hand count within 3 bpm (or the alpha power is at least twice as large with the eyes closed in all three trials). You have measured a vital sign with a

system you designed from the electrode to the screen.

Hand in

FINISH THE LAB

You wrote the firmware of a wireless biopotential recorder: SPI register access, configuration, interrupt-driven sampling, BLE streaming, and the processing that turns samples into a heart rate or a band power.

- ☐ **B2** EDIT `writeReg` (or the ADS1299 commands)
- ☐ **B3** ANSWER the ID in ten resets and in the wrong mode
- ☐ CHECK Checkpoint 1: the chip answers
- ☐ **C2** EDIT `readSample` with the offset-binary conversion
- ☐ **C3** ANSWER sample rate and noise
- ☐ CHECK Checkpoint 2: a live ECG or alpha rhythm
- ☐ **D1** EDIT the BLE packing
- ☐ **D2** EDIT the Python receiver with loss count
- ☐ **D3** ANSWER the measured link budget
- ☐ CHECK Checkpoint 3: 60 s over BLE, no losses
- ☐ **E1** EDIT the beat finder (or band powers)
- ☐ **E2** ANSWER comparison with your hand count (or eyes open and closed)
- ☐ CHECK Checkpoint 4: agreement within 3 bpm (or alpha doubles)
- ☐ Analysis questions 1 to 8 answered
- ☐ **Save** SAVE sketches, Python script, `ecg_ble.csv`, plots

Typical results

Table L4.2 and Fig. L4.2 show what a working system gives. Past teams show what a half-working one gives. One past ECG team saw clean steps and polarity changes when it applied a voltage from a generator to its board, but only noise from printed silver electrodes on the chest, even with gel, and spikes that did not look like an ECG from conventional electrodes. Their firmware polled the data registers without waiting for DRDYB and printed every sample at 853 samples per second through a 115200 baud serial port, and their parts list carried the crystal error of Lab 3; any one of these can spoil a recording. An EEG team read and wrote every ADS1299 register correctly and saw a 60 to 70 mV test voltage at gain 1 follow a potentiometer, and all channels responded when a fingertip touched the connector, but its data showed a constant offset and a scale error that the team suspected came from

Quantity	Typical result
ID register	ADS1293 REVID `0x01`; ADS1299 ID usually `0x3E`
Sample rate	200 samples per second (ADS1293, $R_2 = 4$, $R_3 = 32$); 250 (ADS1299)
Noise, inputs shorted	ADS1293: about 1.5 µV rms, or 10 µV p-p (datasheet: 200 samples per second, 40 Hz band, low-power mode, the default used here); 1.3 µV rms, 7 µV p-p, in high-resolution mode (Table 8.2); more with long wires
R wave, lead II	0.5 to 2 mV with gel electrodes; smaller and noisier with printed silver
BLE	67 notifications per second (3 samples each); MTU 23, the ArduinoBLE default; no losses within a few meters
Heart rate	within 3 bpm of a 30 s hand count
EEG alpha	two to ten times more alpha power with the eyes closed

Table L4.2: Expected results of Lab 4 with a working daughter board and gel electrodes.

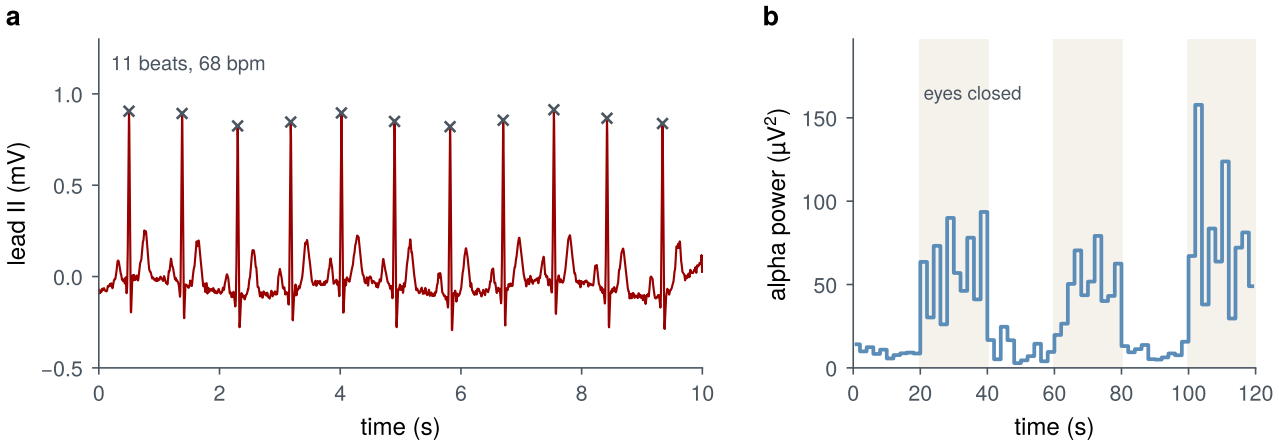


Figure L4.2: What Parts D and E produce when everything works (illustrative data). (a) Lead II at 200 samples per second after the band-pass and the notch, with the R peaks found by the detector of **E1**. (b) Alpha power of one scalp EEG channel at O1 in 2 s windows while the subject alternates 20 s with eyes open and 20 s with eyes closed (shaded). Behind the ear, eyes-closed alpha power is only about 4 µV² (Example 14.1).

the reference configuration; the internal reference buffer, which is off at power-up, is the first thing to check. Start from a known input, a few millivolts from a battery through a resistor divider, before you put electrodes on anyone.

Troubleshooting

- *The ID reads `0x00` or `0xFF`:* check CS, SCLK, COPI, CIPO, the AFE's supplies, and VDDIO against VDDIO_EXT; try the other SPI mode to see whether the bytes shift.
- *Registers read back correctly but DRDYB never falls:* the clock is not running (wrong crystal, missing capacitors, OSC_CN not set), conversions were not started, or DRDYB_SRC is zero. The ADS1293 also masks DRDYB for its first few samples after a start.
- *Data jump between huge positive and negative values:* the conversion treats offset binary as two's complement or the reverse.
- *Samples repeat or go missing:* the code reads without waiting for DRDYB, reads the bytes in separate transactions, or spends too long in `Serial.print`; use the loop read and the interrupt flag.
- *The AFE resets during printing:* see Lab 3; a reset pin is on a busy header pin.
- *A flat line at full scale (ADS1299):* the electrode offset exceeds the full scale; lower the gain to 12 or 8.
- *Large* 60 Hz *hum:* the right-leg or bias electrode is off, the leads form a big loop, or the laptop is on its charger (which is also a safety violation).
- *BLE packets lost:* reduce the distance, keep the payload at 20 bytes, call `BLE.poll()` often, and check the sequence counter for where the gaps occur.

Analysis questions

1. What happened when you used the wrong SPI mode (**B3**)? Explain it with Fig. 8.18.
2. Explain each line of Table L4.1 for your configuration. What would you change for a 3-lead ECG, and what does TI's 3-lead example do with the unused third channel?
3. Why is one loop read better than nine single-register reads? Give the SPI time of each at your clock rate.

4. Compare the noise you measured in **C3** with the datasheet. What adds noise in your setup?
5. Compute the BLE budget of your link with Eq. (8.26) and compare it with **D3**. How much margin is left, and how would you add two more channels?
6. How well did your detector agree with your hand count, and what limits the agreement: the detector, the counting, or the window?
7. EEG teams: by what factor did alpha power rise with the eyes closed? Is the alpha/beta ratio a good stress index for one person over a day?
8. Which steps of this lab made the system safe to connect to a person, and what single fault would each one guard against?

A
Answers to Chapter Problems

Final answers to every end-of-chapter problem are collected here. Check your result only after you have a complete solution of your own; when your answer differs, the worked examples in the chapter are the best place to find the step that went wrong. Numerical answers are rounded, usually to two or three significant figures. Design problems have many correct answers; the one given is a reasonable choice, with the reasoning that leads to it.

Chapter 1: Introduction to Bioelectronics

1.1 (a) $R_{\min} = \sqrt{D/(2\gamma)} = 1.21\,\text{mm}$ (with $D = 2.94 \times 10^{-8}\,\text{N m}$). (b) 3.84 mm. (c) A continuous sheet can bend around one axis but must stretch in its plane to cover a doubly curved surface; narrow mesh ribbons bend and twist independently, carry little bending stiffness per unit area, and let fluid drain so capillary forces pull the film onto the tissue.

1.2 (a) $0.02\,\mu\text{m/h} \times 336\,\text{h} = 6.7\,\mu\text{m}$. (b) The coil lasts 0.4 h once exposed, so the encapsulation itself must last the full 14 days. (c) The silicon lasts $50/5 = 10$ days after exposure, so it is gone about 10 days after the Mg.

1.3 (a) $R = 1/(2\pi f_c C) = 22.7\,\text{k}\Omega$ (the standard 22 kΩ gives 0.72 Hz); $\tau = RC = 0.23\,\text{s}$. (b) -0.73 dB at 0.3 Hz (gain 0.92) and -9.6 dB at 2 Hz (gain 0.33). (c) $0.92/0.33 = 2.8$, or 8.9 dB. A second-order filter falls 40 dB per decade above its corner instead of 20, so it cuts the motion at 2 Hz by about twice as many decibels while breathing is barely touched; an accelerometer that measures the walking motion so that it can be subtracted is another cure.

1.4 (a) 36.6 µV. (b) No; with a 0.6 V range it would need 20 bits, and with the full 2.4 V range 22 bits. (c) A gain of at least 37 (use 40); the input range is then ± 30 mV around the midpoint.

1.5 One reasonable design: two chest electrodes for the ECG (heart rate), which can also sense breathing through the impedance change of the chest, or a chest accelerometer for breathing, and a thermistor for skin temperature. Sampling: ECG 250 Hz with a 24-bit biopotential front end (or 12 bits after AC coupling and gain); breathing 25 Hz (three accelerometer axes at 16 bits); temperature once per second, 12 to 16 bits. Raw rate about $6000 + 1200 + 16 \approx 7.2\,\text{kbit/s}$, or 78 MB per day. Compute heart rate every 5 s, breathing rate every 30 s, and temperature every minute on the patch (about 23 kB per day), and store or send short ECG snippets only when a rhythm alarm fires.

Chapter 2: Cells and Biomolecules

2.1 (a) 278 mOsm/L, close to isotonic; once the cells use the glucose, the water is left behind and acts as hypotonic free water. (b) $36 + 54 \times 290/600 = 62.1$ fL. (c) $(40\,\text{mol/m}^3)RT = 1.03\,\text{bar}$.

2.2 (a) 0.182 µA, 1.00 µA, 1.82 µA. (b) $c = 9K_m = 90\,\text{mM}$. (c) $c \leq K_m/9 = 1.11\,\text{mM}$.

2.3 (a) $L = nFAD/S = 48\,\mu\text{m}$; $t \approx L^2/2D = 116\,\text{s}$. (b) $S = nFAD/L$ fixes L/D, so $t = L^2/2D = (nFA/S)^2 D/2 \propto D$. With $D = 2.5 \times 10^{-12}\,\text{m}^2/\text{s}$, $L = 12.1\,\mu\text{m}$ and $t = 29\,\text{s}$. (c) Pinholes and uneven thickness in thin films, damage during insertion, and protein and cell layers on the surface, which add their own diffusion resistance.

2.4 (a) MCV 78.9 fL and MCH 25.0 pg, both low; MCHC 31.7 g/dL, slightly low: small, pale cells, typical of iron deficiency. (b) 12.8 mL/dL ($1.34 \times 9.5 \times 0.98 + 0.3$), about 65 % of normal. (c) Pump $19.8/12.8 = 1.55$ times more blood (about 7.7 L/min instead of 5 L/min), or extract more oxygen per pass.

2.5 (a) $(650 - 140) \times 0.33 = 168\,\Omega\,\text{cm}^2$. (b) A cell layer is many small paths in parallel; a larger area has more of them and less resistance, so $R \times A$ is the tissue property, independent of the insert. (c) $168/1.12 + 140 = 290\,\Omega$.

Chapter 3: Neurons and the Action Potential

3.1 (a) 59.2 mV per decade at 25 °C, 61.5 mV at 37 °C, and 30.8 mV for Ca^{2+} (half, since $z = 2$). (b) E_K rises by 61.5 mV, from −89.1 mV to −27.5 mV.

3.2 (a) −64.6 mV, about 5 mV depolarized. (b) −56.9 mV, near threshold. (c) A depolarized resting potential first brings heart cells closer to threshold, then leaves many sodium channels inactivated, so conduction slows and fails; the result can be arrhythmia and cardiac arrest.

3.3 (a) $r_a = 4.8 \times 10^9\,\Omega/\text{cm}$, $r_m = 4.8 \times 10^7\,\Omega\,\text{cm}$, $\lambda = 1.0\,\text{mm}$, $\tau = 30\,\text{ms}$. (b) $e^{-0.5} = 0.61$.

3.4 $20 + 0.7 + 20 + 0.5 = 41\,\text{ms}$, about 2.7 times the normal 15 ms of conduction and synaptic delay.

3.5 (a) $e^{-0.1/\tau} = 1 - 0.40/1.00 = 0.6$, so $\tau = 0.1/\ln(1/0.6) = 0.196\,\text{ms}$; chronaxie $\tau \ln 2 = 0.136\,\text{ms}$. (b) 1.77 mA and 89 nC; 0.625 mA and 125 nC; 0.402 mA and 402 nC. (c) $I_{\text{rh}}\tau = 78\,\text{nC}$, approached with very short, very strong pulses. (d) I^2T is 0.157, 0.078, and $0.162\,\text{mA}^2\,\text{ms}$: the 0.2 ms pulse. The energy is lowest near $T = 1.26\tau \approx 0.25\,\text{ms}$, so a width near the chronaxie, 0.1 to 0.3 ms, balances charge against battery energy.

Chapter 4: Sensors and Signal Conditioning

4.1 (a) Each side has 1 C, 2 N, 3 O, 9 H and a charge of +1. (b) Potentiometric: an ammonium or pH electrode under urease. Field effect: an ISFET under urease (local pH). Conductometric: urease on interdigitated electrodes (one neutral molecule becomes three ions). Optical: a pH indicator dye in the enzyme gel. Thermal: urease on a thermistor (heat of reaction). (c) Amperometric: no product is easily oxidized or reduced at an electrode.

4.2 (a) $v_n = \sqrt{4kTRB} = 2.07\,\mu\text{V}$ rms; SNR = 96.6, or 39.7 dB. (b) $q_n = 0.806\,\text{mV}/\sqrt{12} = 0.233\,\text{mV}$ at the ADC, 2.33 µV at the input; total $\sqrt{2.07^2 + 2.33^2} = 3.12\,\mu\text{V}$, SNR 36.2 dB. (c) The ADC, slightly. Raise the gain to 1000 (the ADC then adds only 0.23 µV at the input) or use more bits.

4.3 (a) Mean 7.050, $s = 0.013$, CV 0.19 %, bias +0.05. (b) Calibration: the bias is four times the spread. (c) $a = 3/2.96 = 1.0135$, $b = 7.00 - 1.0135 \times 7.05 = -0.145$.

4.4 (a) 57.18 mV/pH. (b) Gain −17.49: $R_f = 174.9\,\text{k}\Omega$, $R_2 = R_f \times 2.50/7.00 = 62.5\,\text{k}\Omega$. (c) 100 fA: 20 µV at the input, 0.35 mV at the output (0.0003 pH). 50 pA: 10 mV at the input, 175 mV at the output (0.17 pH). (d) $200/(5 \times 10^4 + 200) = 0.40\,\%$ low in slope; at pH 10, $3 \times 0.0040 = 0.012$ pH.

4.5 (a) $V_{\text{out}} = V_b + IR_f$: put $V_b = 0.10\,\text{V}$ and $R_f = 3.0\,\text{V}/200\,\text{nA} = 15\,\text{M}\Omega$. (b) $1/(2\pi \times 15\,\text{M}\Omega \times 2\,\text{Hz}) = 5.3\,\text{nF}$ (5.6 nF gives 1.9 Hz). (c) At $V_b = 0.10\,\text{V}$.

Chapter 5: Physical and Field-Effect Transducers

5.1 (a) 37 °C: 5.99 kΩ, −4.11 %/K; 42 °C: 4.89 kΩ, −3.98 %/K. (b) 114.2 Ω, 0.337 %/K; 116.2 Ω, 0.331 %/K.

5.2 (a) The fractional change is $\delta = 2.0 \times 0.0008 = 0.0016$, so $V_s\delta = 5.28\,\mathrm{mV}$. (b) $2.0/0.00528 = 379$. (c) Full bridge: no change, because all four arms rise together and the ratios stay equal. Quarter bridge: $3.3 \times 0.002/4.004 = 1.65\,\mathrm{mV}$, which reads as 1000 microstrain of false strain, more than the real signal.

5.3 (a) $x = 2 \times 9.81/(2\pi \times 3000)^2 = 55.2\,\mathrm{nm}$. (b) $x/d = 0.0276$; $C_{w1}/C_0 - 1 = 0.0284$, which exceeds the linear x/d by 7.8×10^{-4}, or 2.8 % of the change. (c) $2.5 \times 0.0276 = 69.0\,\mathrm{mV}$. Each divider ratio is a ratio of gaps, $(d \mp x)/(2d)$, which is linear in x.

5.4 (a) $R_e = 52\,\Omega$; $1/R_i = 1/38 - 1/52$, so $R_i = 141\,\Omega$. (b) The tissue voltage is $300\,\mu\mathrm{A} \times 52\,\Omega = 15.6\,\mathrm{mV}$; the amplifier receives $15.6 \times 500/504 = 15.48\,\mathrm{mV}$, 0.8 % low; the source must deliver $300\,\mu\mathrm{A} \times 4.052\,\mathrm{k}\Omega = 1.22\,\mathrm{V}$. (c) Two electrodes would read $4052\,\Omega$: the contacts are 77 times the tissue, and they change with sweat, pressure, and motion by far more than $1\,\Omega$ (a 1 % change of the contacts is $40\,\Omega$), so the tissue change is lost.

5.5 (a) $\beta_{\mathrm{int}} = 0.2\,\mathrm{nm}^{-2}$: $\alpha = 0.777$, $47.8\,\mathrm{mV/pH}$. $\beta_{\mathrm{int}} = 5\,\mathrm{nm}^{-2}$: $\alpha = 0.989$, $60.8\,\mathrm{mV/pH}$. (b) 2.48 nm and 0.79 nm. (c) $0.4 \times 48 = 19.2\,\mathrm{mV}$, or $19.2/60.8 = 0.32$ pH.

Chapter 6: Electrochemical Biosensors

6.1 (a) Anode (oxidation): $\mathrm{Zn} \longrightarrow \mathrm{Zn^{2+}} + 2\,\mathrm{e^-}$; cathode (reduction): $\mathrm{Cu^{2+}} + 2\,\mathrm{e^-} \longrightarrow \mathrm{Cu}$. $E_{\mathrm{cell}} = 0.342 - (-0.762) = 1.104\,\mathrm{V}$; $\Delta G = -2FE = -213\,\mathrm{kJ/mol}$. (b) Oxidation at the electrode, so an anodic current; two electrons per molecule.

6.2 (a) A: $R_s \approx 40\,\Omega$, $R_{ct} \approx 2.5\,\mathrm{k}\Omega$. B: $R_s \approx 12\,\Omega$, $R_{ct} \approx 14\,\Omega$. (b) A is controlled by slow charge transfer (kinetics); the Warburg tail appears only at the lowest frequencies. B has fast kinetics: its small semicircle ends near 4 Hz, and below that the long 45° line of diffusion (mass transport) takes up most of the plot. (c) $C_{dl} = 1/(2\pi f R_{ct})$: 20 µF for A and 41 µF for B.

6.3 (a) $E_{\mathrm{ISE}} = K + s\log a_{\mathrm{Na^+}}$ and $E_{\mathrm{ref}} = E^\circ - s\log a_{\mathrm{Cl^-}}$; their difference is $K - E^\circ + s\log(a_{\mathrm{Na^+}} a_{\mathrm{Cl^-}})$. (b) $2s\log_{10} 3 = 58.0\,\mathrm{mV}$ ($s = 60.7\,\mathrm{mV}$); assuming slope s one would infer a ninefold rise in sodium instead of threefold. (c) A solid-state reference with a KCl-loaded membrane (or a liquid-junction reference), or a separate $\mathrm{Cl^-}$ sensor to correct the reading.

6.4 (a) $\Delta R_{ct}/R_{ct,0} = 0.45$, $\theta = 0.45/1.2 = 0.375$, $C = K_D\theta/(1-\theta) = 30\,\mathrm{nM}$. (b) About $0.1K_D$ to $10K_D$, 5 to 500 nM. (c) Electrodes differ in area and coverage; the relative change cancels much of that variation.

6.5 (a) The channel holds 2.7 µL and the flow is 0.96 µL/min, so it fills in 2.8 min. (b) $v = 0.36\,\mathrm{mm/s}$, $D_h = 200\,\mu\mathrm{m}$, $\mathrm{Re} = 0.071$.

Chapter 7: Optical Biosensing

7.1 (a) 405 nm: 3.06 eV; 850 nm: 1.46 eV; 1550 nm: 0.800 eV. (b) 405 and 850 nm (above 1.12 eV, or below 1107 nm); 1550 nm passes through silicon undetected. (c) 8.56×10^{15} photons/s.

7.2 Solving $1.2 \times 10^4 x + 2.0 \times 10^3 y = 0.360$ and $3.0 \times 10^3 x + 1.8 \times 10^4 y = 0.615$: $c_X = 25.0\,\mu\text{M}$, $c_Y = 30.0\,\mu\text{M}$.

7.3 (a) $k_r = 6.0 \times 10^7\,\text{s}^{-1}$, $k_{nr} = 4.0 \times 10^7\,\text{s}^{-1}$. (b) $\Phi = k_r\tau = 0.30$; $k_q[Q] = 1/5\,\text{ns} - 1/10\,\text{ns} = 1.0 \times 10^8\,\text{s}^{-1}$, so $k_q = 1.0 \times 10^{10}\,\text{M}^{-1}\,\text{s}^{-1}$ (the diffusion limit) and $K_{SV} = k_q\tau_0 = 100\,\text{M}^{-1}$.

7.4 (a) $R' = R(0.975) = 0.327$ from Eq. (7.14) with equal paths. (b) $L_r/L_{ir} = 0.50/0.327 = 1.53$. (c) $R = 1.3 \times 0.327 = 0.425$, and the line reports $110 - 25(0.425) = 99.4\,\%$: about two points high for this person, from path length alone.

7.5 (a) $P_d = 0.80\,\mu\text{W}$, $I_{DC} = 0.24\,\mu\text{A}$, AC 2.4 nA peak to peak. (b) The total current can reach 5.24 µA, so $R_f \le 477\,\text{k}\Omega$; use 470 kΩ (2.46 V in daylight, 0.11 V indoors in the dark). (c) The AC rms is 0.85 nA. Without ambient light the shot noise is 0.88 pA and SNR = 968 (60 dB); with 5 µA of ambient light it is 4.1 pA and SNR = 207 (46 dB): ambient light costs 13.4 dB and forces a smaller R_f, even after it is subtracted.

Chapter 8: Biopotential Recording

8.1 (a) Lead II is 1.00 mV. Then aVR $= -(\text{I} + \text{II})/2 = -0.90\,\text{mV}$, aVL $= (\text{I} - \text{III})/2 = 0.30\,\text{mV}$, and aVF $= (\text{II} + \text{III})/2 = 0.60\,\text{mV}$. (b) Lead I is $M\cos\theta = 0.80$. Lead III is $M\cos(\theta - 120^\circ) = -0.5M\cos\theta + (\sqrt{3}/2)M\sin\theta$, so $-0.40 + 0.866\,M\sin\theta = 0.20$ and $M\sin\theta = 0.693$. Then $\theta = \arctan(0.693/0.80) = 40.9^\circ$ and $M = 1.06\,\text{mV}$. The axis is normal (between -30° and 90°).

8.2 (a) $1/\omega C = 159\,\text{M}\Omega$ in parallel with 1 GΩ gives $|Z_{\text{cm}}| = 157\,\text{M}\Omega$, so $v_d \approx 5\,\text{mV} \times 30\,\text{k}\Omega/157\,\text{M}\Omega = 0.95\,\mu\text{V}$. (b) It needs $|Z_{\text{cm}}| \ge 1.5\,\text{G}\Omega$. No capacitance achieves that while the 1 GΩ resistance is in parallel; the resistance must also exceed 1.5 GΩ and the capacitance fall below 2.1 pF. It is easier to lower v_{cm} with a right-leg drive.

8.3 (a) The input stage amplifies the difference by G_1 but passes the common mode with gain 1, so the difference stage sees a ratio of differential to common-mode input G_1 times larger; with the difference-stage result $(1 + G_d)/(4\epsilon)$ of Chapter 4, CMRR $\approx G_1(1 + G_d)/(4\epsilon)$. (b) (10, 1): $10 \times 2/0.004 = 5000$, 74.0 dB; (5, 2): 3750, 71.5 dB; (2, 5): 3000, 69.5 dB. (c) (10, 1): the offset gives $10 \times 0.3\,\text{V} = 3\,\text{V}$ at the first stage, inside ±4.5 V, and that split has the highest CMRR. On a 3.3 V supply the offset would force G_1 down to about 5.

8.4 (a) $\pm 187.5\,\text{mV}$; LSB $= 2(4.5)/(24 \times 2^{24}) = 22.4\,\text{nV}$. (b) OSR $= 1.024\,\text{MHz}/250 = 4096$. (c) $-3000 \times 22.35\,\text{nV} = -67.1\,\mu\text{V}$. (d) The ADS1293's range of $\pm 400\,\text{mV}$ holds a 300 mV offset plus the ECG; the ADS1299 at gain 24 spans only $\pm 187.5\,\text{mV}$, so with such electrodes it needs $4.5/G > 0.3$, a gain of 12 or less.

8.5 (a) Type BF: $100\,\mu\text{A}$ in normal condition and $500\,\mu\text{A}$ with a single fault (AC; lower for DC). (b) The charging laptop ties the circuit's ground to the mains through the charger, so a fault in the charger or the building wiring can drive current through the gel electrodes into the chest. Unplug the charger and run the laptop on its battery, or run the Nicla from its battery and use only the BLE link (a USB isolator is a third option). (c) $R \geq 3.3\,\text{V}/10\,\mu\text{A} = 330\,\text{k}\Omega$.

Chapter 9: Neural Recording and Stimulation

9.1 (a) EEG: sleep stages are defined by scalp EEG rhythms, and a home study must be noninvasive. (b) ECoG: it covers a large area of cortex with a resolution of millimeters to a centimeter, set by the contact spacing, and captures the high-frequency onset of seizures. (c) Single units from a penetrating array: the direction tuning of individual motor-cortex neurons carries the most information. (d) LFP from the DBS contacts: the beta rhythm is a population signal, and the large contacts cannot resolve single neurons anyway.

9.2 (a) 46 dB is a gain of 200, so $C_{\text{in}} = 20\,\text{pF}$. (b) $R_p = 1/(2\pi f_L C_f) = 3.2 \times 10^{12}\,\Omega$. (c) $|Z_{\text{in}}| = 8.0\,\text{M}\Omega$; the input receives $8.0/(8.0+1) = 0.89$ of the signal, so about 11 % is lost.

9.3 $A = 1.96 \times 10^{-3}\,\text{cm}^2$. (a) $C = 58.9\,\text{nF}$, $Q = 35\,\text{nC}$. (b) Platinum $0.20\,\mu\text{C}$, iridium oxide $3.9\,\mu\text{C}$. (c) Shannon at $k = 1.85$ allows $0.37\,\mu\text{C}$. For platinum the CIC ($0.20\,\mu\text{C}$) is the lower limit; for iridium oxide the Shannon limit ($0.37\,\mu\text{C}$) is.

9.4 (a) $V = 200\,\mu\text{A} \times (5\,\text{k}\Omega + 10\,\text{k}\Omega) = 3.0\,\text{V}$ (1.0 V step plus 2.0 V ramp); the anodic phase ends at $+1.0\,\text{V}$. (b) $V_{\text{DDL}} \geq 3.3\,\text{V}$ and $V_{\text{DDH}} \geq 1.3\,\text{V}$; symmetric rails of $\pm 3.3\,\text{V}$. (c) $P_{\text{stim}} = 12\,\mu\text{W}$ (upper estimate with the peak voltage); $\eta = 12/20 = 0.60$ at $\pm 5\,\text{V}$ and $12/13.2 = 0.91$ at $\pm 3.3\,\text{V}$.

9.5 (a) Halving the noise from 10 to $5\,\mu\text{V}$ costs four times the power: $1.8\,\mu\text{W}$ per amplifier, 0.46 mW for 256. (b) $256 \times 50 \times 2 = 25.6\,\text{kb/s}$, so $51\,\mu\text{W}$. (c) Total 1.01 mW. The surface is $0.56\,\text{cm}^2$, so the budget is 22.4 mW: about 22 times the need. Each added channel costs about $2.0\,\mu\text{W}$ (amplifier plus radio), so about 11 000 channels fit in all, some 10 700 more; chip area and the ADC would limit you first.

Chapter 10: Sensory Prostheses

10.1 (a) 8 mm from the base ($x = 27/35$): 6.7 kHz; 18 mm ($x = 17/35$): 1.6 kHz. (b) 500 Hz is 25.1 mm from the base; 2 kHz is 16.4 mm. (c) 4 kHz is at 11.7 mm, so the octave from 2 to 4 kHz spans 4.8 mm.

10.2 (a) $16 \times 1200 = 19\,200$ pulses/s, a slot of 52.1 µs. (b) $(52.1 - 10)/2 = 21$ µs. (c) A pulse needs $2 \times 50 + 10 = 110$ µs, so at most 9091 pulses/s in total, or about 570 per channel.

10.3 (a) $20 \log_{10} 5 = 14$ dB. (b) 160, 280, and 400 µA. (c) $Q = 12$ nC, $Q/A = 6\,\mu\text{C/cm}^2$, $k = \log 6 + \log 0.012 = -1.1$, far below 1.85.

10.4 (a) $1024 \times 6 + 16 = 6160$ bits per frame, so 185 kb/s. (b) The camera produces 221 Mb/s, about 1200 times more.

10.5 (a) $p = 288 \times 160/1200 = 38.4$ µm; the 6° field is 1.73 mm, so $45 \times 45 = 2025$ pixels. (b) $A = (38.4\,\mu\text{m})^2 = 1.47 \times 10^{-5}\,\text{cm}^2$, so $Q/A = 1.36\,\text{mC/cm}^2$, below $3\,\text{mC/cm}^2$ for iridium oxide; $k = \log 1356 + \log 0.02 = 1.43 < 1.85$, although Shannon's criterion is not established for electrodes this small. (c) The active electrode is smaller than the pixel, which raises the charge density; thresholds do not fall in proportion to electrode size; current spreads between neighbors; and each pixel must harvest enough light to deliver its charge.

Chapter 11: Materials and Mechanics of Soft Electronics

11.1 $F = E\varepsilon A$ with $A = 4 \times 10^{-6}\,\text{m}^2$: 0.056 N (Ecoflex), 1.2 N (PDMS), 16 N (TPU), and 3200 N (PET). The skin hardly notices Ecoflex, about the weight of a few grams.

11.2 (a) $y_n = [(2.5)(12.5)(6.25) + (130)(1)(13)]/[(2.5)(12.5) + 130] = 11.7$ µm, below the silicon. The silicon top sees $(13.5 - 11.69)/1000 = 0.18\,\%$ and its bottom 0.08 %, both tensile. (b) The estimate $13.5/2000 = 0.68\,\%$ is almost four times too high, because the silicon's Et (130) is four times the polyimide's (31) and pulls the neutral plane close to itself. (c) 12.5 µm: a symmetric stack puts the neutral plane in the middle of the silicon.

11.3 $R_0 = 0.588\,\Omega$; at 30 %, $1.69R_0 = 0.994\,\Omega$. Gauge factor $(1.69 - 1)/0.30 = 2.3$; at small strain the gauge factor is 2, close to a metal foil gauge but over a range a hundred times larger.

11.4 $\rho = 3.98 \times 10^{-6}\,\Omega\,\text{cm}$, $R_s = 79.5\,\text{m}\Omega$ per square, 333 squares, $R = 26.5\,\Omega$. For $1\,\Omega$ the line would need to be 13 µm thick. One inkjet pass deposits well under a micrometer after drying; a thick wet line would bulge and spread, so thickness is built in several dried passes (or a thicker printing method is used).

11.5 $Wd/L = 2 \times 10^{-5}$ cm, $g_m = (2 \times 10^{-5})(1.5)(39)(0.3) = 0.35$ mS. Doubling the thickness (or W/L) doubles g_m; a thicker channel re-

sponds more slowly because ions must fill a larger volume, and a wider channel adds capacitance and area.

Chapter 12: Printed Bioelectronics

12.1 (a) $\eta(0.05\,\mathrm{s}^{-1}) = 80 \times 0.05^{-0.65} = 561\,\mathrm{Pa\,s}$; in the mesh $\dot{\gamma} \approx 0.15\,\mathrm{m/s}/40\,\mu\mathrm{m} = 3750\,\mathrm{s}^{-1}$ and $\eta = 0.380\,\mathrm{Pa\,s}$, about 1500 times thinner. (b) $\tau = \eta\dot{\gamma} = 1.43\,\mathrm{kPa}$. (c) Too slow: the print slumps and spreads, edges blur, and close lines merge. Too fast: the paste stiffens before the mesh marks level, leaving a textured film or pinholes.

12.2 (a) $Z = 2.77$ (jets well), 0.52 (too viscous to eject), and 11.5 (jets with satellites; add a more viscous cosolvent). (b) $Z = 1.26$: just inside the window, jettable but near the viscous limit.

12.3 (a) $b = 22.7\,\mu\mathrm{m}$, $p_{\max} = 13.6\,\mu\mathrm{m}$. (b) $w = 26.5\,\mu\mathrm{m}$ at $10\,\mu\mathrm{m}$, which is 2540 dpi. (c) 2000 drops at 5 kHz: 0.40 s. (d) Print a row of test lines at spacings from about 4 to $20\,\mu\mathrm{m}$ in $1\,\mu\mathrm{m}$ steps, at the speed, drop delay, and substrate temperature of production; image them and measure the width along each. $p_{\min}$ is the smallest spacing without periodic bulges; the onset of scalloping checks $p_{\max}$.

12.4 (a) Area ratio $500/180 = 2.78$, aspect ratio 5.6: releases easily. (b) Area ratio 0.67 and aspect ratio 1.33: marginal; use a thinner foil or wider openings. (c) The vinyl between the disk and the ring is an island, completely surrounded by the ring opening, and falls out when cut. Add narrow bridges across the ring opening (and connect the ring segments with a trace), hold the island in place with transfer tape, or use a mesh screen.

12.5 (a) Screen printing, sheet-fed or rotary: thick, low-cost Ag/AgCl in volume. (b) Inkjet (or aerosol jet): digital and fine enough for $30\,\mu\mathrm{m}$ fingers. (c) Aerosol jet or a dispenser: non-contact, millimeter standoff, follows the curved surface. (d) Slot-die coating (or gravure) roll to roll: pre-metered, uniform. (e) A dispenser with liquid metal or a stretchable silver paste on TPU, or a stencil for the paste.

Chapter 13: Powering Bioelectronics

13.1 (a) $40 + 15 + 25 = 80\,\mu\mathrm{A}$ from the active blocks plus $3\,\mu\mathrm{A} \times 0.939 = 2.8\,\mu\mathrm{A}$ of sleep current: $82.8\,\mu\mathrm{A}$. (b) 1208 h, about 50 days. (c) The microcontroller ($40\,\mu\mathrm{A}$), then the sensor ($25\,\mu\mathrm{A}$).

13.2 (a) $\dot{Q} = 14/800 = 17.5\,\mathrm{mW}$, $\Delta T_{\mathrm{TEG}} = 2.63\,\mathrm{K}$, $V_{\mathrm{oc}} = 120\,\mathrm{mV}$, $P = V^2/(4R) = 59.7\,\mu\mathrm{W}$. (b) $\dot{Q} = 22.4\,\mathrm{mW}$, $\Delta T_{\mathrm{TEG}} = 3.36\,\mathrm{K}$, $V_{\mathrm{oc}} = 153\,\mathrm{mV}$, $P = 97.8\,\mu\mathrm{W}$. (c) Yes: 120 mV open circuit, 60 mV at the matched load, both above 20 mV.

13.3 (a) 0.555 Wh, 2.0 kJ. (b) 300 mA for about 30 minutes. (c) 500 cycles at one every two days: about 2.7 years.

13.4 (a) $k = 40/\sqrt{2000 \times 800} = 0.0316$ and $k^2 Q_1 Q_2 = 2.0$, so $\eta_{\max} = 2.0/(1+\sqrt{3})^2$, which is 26.8 %. (b) $R_2 = \omega L_2 / Q_2 = 1.36\,\Omega$; $R_L = R_2\sqrt{1 + k^2 Q_1 Q_2} = 2.36\,\Omega$. (c) $\eta = 0.5$ needs $k^2 Q_1 Q_2 = 8$, so $Q_2 = 100$.

13.5 (a) $\lambda = 0.77$ mm; $4 \times 2 \times 0.5 = 4$ dB, leaving 40 %. (b) $A = 1\,\text{mW}/(720\,\text{mW/cm}^2 \times 0.30) = 4.6 \times 10^{-3}\,\text{cm}^2 = 0.46\,\text{mm}^2$, for example 0.7×0.7 mm.

Chapter 14: From Prototype to Patient, and the Design Project

14.1 Open-ended. One answer: one lead plus a reference electrode; signal 0.5 to 4 mV; bandwidth 0.5 to 40 Hz for rhythm monitoring (0.05 to 150 Hz if the waveform must be diagnostic); DC offset ±300 mV; 250 samples/s; LSB below 5 µV at the input; input noise below about 5 µV rms; electrode impedance below 10 kΩ with gel; battery life at least 24 h; data rate 6 kbit/s (store locally, upload in bursts); size about 60 × 30 × 8 mm and under 20 g; adhesive rated for 24-hour wear; battery only, leakage below 10 µA; validation: heart rate within ±3 bpm of a chest-strap or clinical reference in 9 of 10 one-minute segments, at rest and walking. Each row needs its reason.

14.2 (a) $(58 \times 1.2 + 2 \times 4.0)/60 = 1.29$ mA. (b) $1.29 \times 168\,\text{h}/0.8 = 272$ mAh; choose a 300 mAh cell.

14.3 60 V across 4.7 nF at 60 Hz: 106 µA. 115 V at 50 Hz: 170 µA. Both exceed even the 100 µA normal-condition limit for a body-floating part.

14.4 (a) Not a device if it claims only general sleep wellness; "screens for sleep apnea" would make it one. (b) A device: detecting fever is a claim about a disease or condition. (c) A wellness product: training readiness and heart rate variability for fitness, with no disease claims. (d) A device: it analyzes a physiological signal and suggests a diagnosis, and a time-critical alert stays a device in any case.

14.5 (a) $n = 1.96^2(0.92)(0.08)/0.04^2 = 176.7$, so 177 with atrial fibrillation. (b) $1.96^2(0.96)(0.04)/0.02^2 = 368.8$, so 369 without. (c) Enrolling 1,180 people gives 177 with atrial fibrillation and 1,003 without, more than enough for the specificity: the sensitivity requirement sets the enrollment (the specificity alone would need 435).

About the Author



Yasser Khan is the Andrew and Erna Viterbi Early Career Chair and an Assistant Professor in the Ming Hsieh Department of Electrical and Computer Engineering and Biomedical Engineering at the University of Southern California, which he joined in 2022. There he teaches Bioelectronics, the graduate course this book was written for, and Introduction to Electrical Engineering.

He earned his B.S. in Electrical Engineering from the University of Texas at Dallas and his M.S. in Electrical Engineering from KAUST. He completed his Ph.D. in Electrical Engineering and Computer Sciences at the University of California, Berkeley, in 2018, and then conducted postdoctoral research in the Department of Chemical Engineering at Stanford University.

His research uses additive manufacturing and hardware-enabled AI to develop next-generation wearables, implantables, and ingestibles for precision health and psychiatry. His honors include the 2026 DARPA Young Faculty Award, the 2025 Air Force Office of Scientific Research Young Investigator Award, the 2025 IEEE Sensors Council Early Career Technical Achievement Award, the 2025 USC Viterbi Junior Faculty Research Award, the 2024 Packard Fellowship, and the 2023 Google Research Scholar Award. Earlier in his career he received the EECS Departmental Fellowship at UC Berkeley, the Discovery Scholarship and Graduate Fellowship at KAUST, and the Academic Excellence Scholarship at UT Dallas. His work has been published in journals of the *Nature*, *Science*, and *Cell* families, *PNAS*, the *Advanced Materials* family of journals, and other leading venues, and has been featured by BBC News, *The Wall Street Journal*, *Newsweek*, and NSF News.