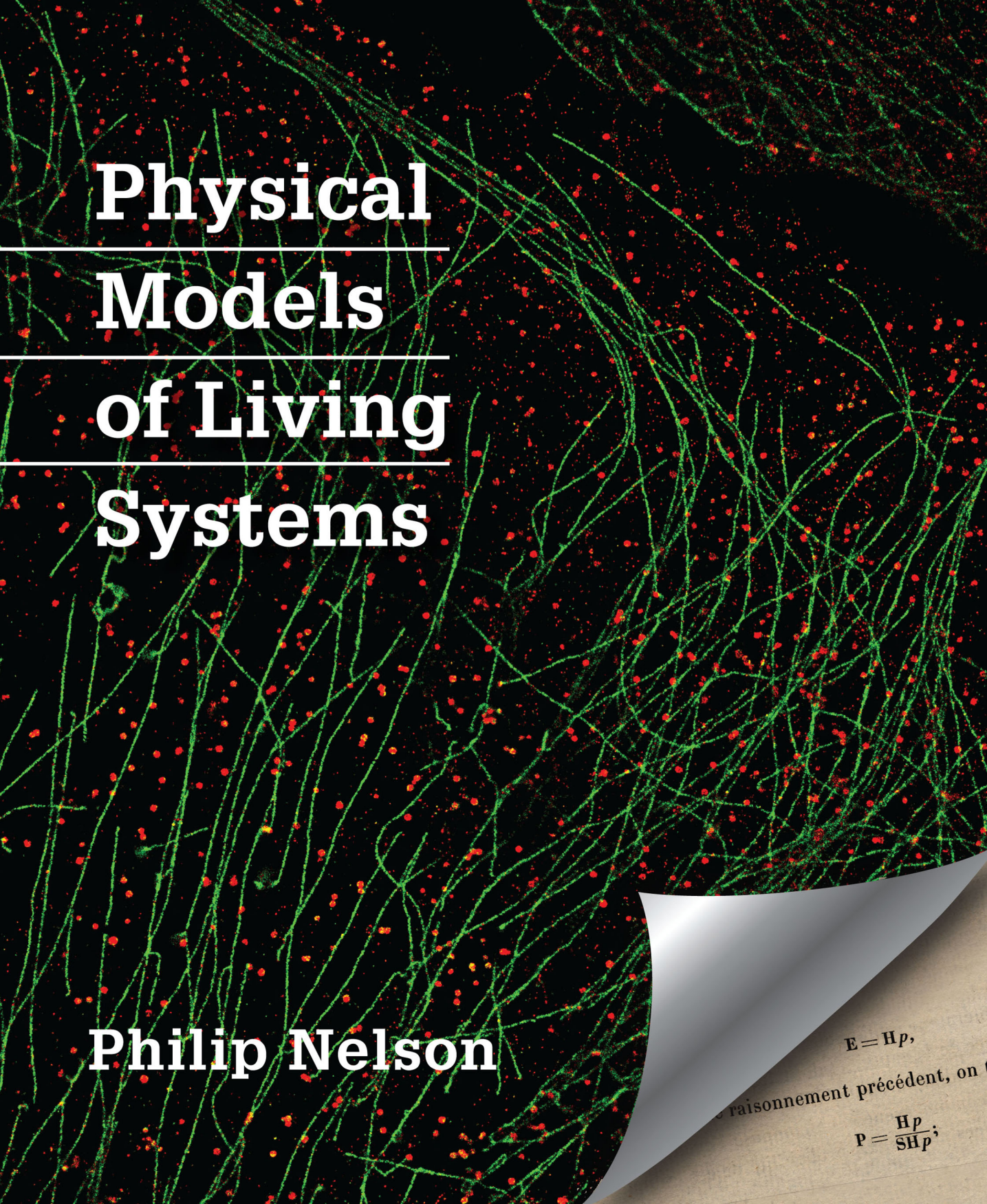
Physical Models of Living Systems

Philip Nelson

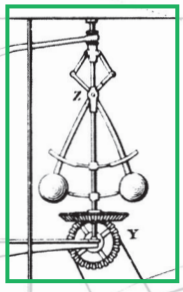


Physical Models of Living Systems

Philip Nelson

University of Pennsylvania

with the assistance of Sarina Bromberg, Ann Hermundstad, and Jason Prentice



W. H. Freeman & Company

A Macmillan Education Imprint

Publisher: Kate Parker
Acquisitions Editor: Alicia Brady
Senior Development Editor: Blythe Robbins
Assistant Editor: Courtney Lyons
Editorial Assistant: Nandini Ahuja
Marketing Manager: Taryn Burns
Senior Media and Supplements Editor: Amy Thorne
Director of Editing, Design, and Media Production: Tracey Kuehn
Managing Editor: Lisa Kinne
Project Editor: Kerry O'Shaughnessy
Production Manager: Susan Wein
Design Manager and Cover Designer: Vicki Tomaselli
Illustration Coordinator: Matt McAdams
Photo Editors: Christine Buese, Richard Fox
Composition: codeMantra
Printing and Binding: RR Donnelley

Cover: [Two-color, superresolution optical micrograph.] Two specific structures in a mammalian cell have been tagged with fluorescent molecules via immunostaining: microtubules (false-colored *green*) and clathrin-coated pits, cellular structures used for receptor-mediated endocytosis (false-colored *red*). See also Figure 6.5 (page 138). The magnification is such that the height of the letter "o" in the title corresponds to about 1.4 μm. [Image courtesy Mark Bates, Dept. of NanoBiophotonics, Max Planck Institute for Biophysical Chemistry, published in Bates et al., 2007. Reprinted with permission from AAAS.] *Inset:* The equation known today as the "Bayes formula" first appeared in recognizable form around 1812, in the work of Pierre Simon de Laplace. In our notation, the formula appears as Equation 3.17 (page 52) with Equation 3.18. (The letter "S" in Laplace's original formulation is an obsolete notation for sum, now written as $\sum$.) This formula forms the basis of statistical inference, including that used in superresolution microscopy.
Title page: Illustration from James Watt's patent application. The green box encloses a centrifugal governor. [From *A treatise on the steam engine: Historical, practical, and descriptive* (1827) by John Farey.]

Library of Congress Preassigned Control Number: 2014949574
ISBN-13: 978-1-4641-4029-7
ISBN-10: 1-4641-4029-4

Printed in the United States of America

First printing

W. H. Freeman and Company, 41 Madison Avenue, New York, NY 10010
Houndmills, Basingstoke RG21 6XS, England
www.whfreeman.com

For my classmates Janice Enagonio, Feng Shechao, and Andrew Lange.

Whose dwelling is the light of setting suns,
And the round ocean and the living air,
And the blue sky, and in the mind of man:
A motion and a spirit, that impels
All thinking things, all objects of all thought,
And rolls through all things.

– William Wordsworth

Brief Contents

PART III Control in Cells

Detailed Contents

Web Resources

The book's Web site (`http://www.macmillanhighered.com/physicalmodels1e`) contains links to the following resources:

- The *Student's Guide* contains an introduction to some computer math systems, and some guided computer laboratory exercises.
- *Datasets* contains datasets that are used in the problems. In the text, these are cited like this: Dataset 1, with numbers keyed to the list on the Web site.
- *Media* gives links to external media (graphics, audio, and video). In the text, these are cited like this: Media 2, with numbers keyed to the list on the Web site.
- Finally, *Errata* is self-explanatory.

To the Student

Learn from science that you must doubt the experts.
—Richard Feynman

This is a book about physical models of living systems. As you work through it, you'll gain some skills needed to create such models for yourself. You'll also become better able to assess scientific claims without having to trust the experts.

The *living systems* we'll study range in scale from single macromolecules all the way up to complete organisms. At every level of organization, the degree of inherent complexity may at first seem overwhelming, if you are more accustomed to studying physics. For example, the dance of molecules needed for even a single cell to make a decision makes Isaac Newton's equation for the Moon's orbit look like child's play. And yet, the Moon's motion, too, is complex when we look in detail—there are tidal interactions, mode locking, precession, and so on. To study any complex system, we must first make it manageable by adopting a *physical model*, a set of idealizations that focus our attention on the most important features.

Physical models also generally exploit analogies to other systems, which may already be better understood than the one under study. It's amazing how a handful of basic concepts can be used to understand myriad problems at all levels, in both life science and physical science.

Physical modeling seeks to account for experimental data quantitatively. The point is not just to summarize the data succinctly, but also to shed light on underlying mechanisms by testing the different predictions made by various competing models. The reason for insisting on quantitative prediction is that often we can think up a cartoon, either as an actual sketch or in words, that sounds reasonable but fails quantitatively. If, on the contrary, a model's numerical predictions are found to be confirmed in detail, then this is unlikely to be a fluke. Sometimes the predictions have a definite character, stating what should happen every time; such models can be tested in a single experimental trial. More commonly, however, the output of a model is probabilistic in character. This book will develop some of the key ideas of probability, to enable us to make precise statements about the predictions of models and how well they are obeyed by real data.

Perhaps most crucially in practice, a good model not only guides our interpretation of the data we've got, but also suggests what *new* data to go out and *get next*. For example, it may suggest what quantitative, physical intervention to apply when taking those data, in order to probe the model for weaknesses. If weaknesses are found, a physical model may suggest how to improve it by accounting for more aspects of the system, or treating them more realistically. A model that survives enough attempts at falsification eventually earns the label "promising." It may even one day be "accepted."

This book will show you some examples of the modeling process at work. In some cases, physical modeling of quantitative data has allowed scientists to deduce mechanisms whose key molecular actors were at the time unsuspected. These case studies are worth studying, so that you'll be ready to operate in this mode when it's time to make your own discoveries.

Skills

Science is not just a pile of facts for you to memorize. Certainly you need to know many facts, and this book will supply some as background to the case studies. But you also need skills. Skills cannot be gained just by reading through this (or any) book. Instead you'll need to work through at least some of the exercises, both those at the ends of chapters and others sprinkled throughout the text.

Specifically, this book emphasizes

- *Model construction skills:* It's important to find an appropriate level of description and then write formulas that make sense at that level. (Is randomness likely to be an essential feature of this system? Does the proposed model check out at the level of dimensional analysis?) When reading others' work, too, it's important to be able to grasp what assumptions their model embodies, what approximations are being made, and so on.
- *Interconnection skills:* Physical models can bridge topics that are not normally discussed together, by uncovering a hidden similarity. Many big advances in science came about when someone found an analogy of this sort.
- *Critical skills:* Sometimes a beloved physical model turns out to be . . . wrong. Aristotle taught that the main function of the brain was to cool the blood. To evaluate more modern hypotheses, you generally need to understand how raw *data* can give us *information,* and then *understanding.*
- *Computer skills:* Especially when studying biological systems, it's usually necessary to run many trials, each of which will give slightly different results. The experimental data very quickly outstrip our abilities to handle them by using the analytical tools taught in math classes. Not very long ago, a book like this one would have to content itself with telling you things that faraway people had done; you couldn't do the actual analysis yourself, because it was too difficult to make computers do anything. Today you can do industrial-strength analysis on any personal computer.
- *Communication skills:* The biggest discovery is of little use until it makes it all the way into another person's brain. For this to happen reliably, you need to sharpen some communication skills. So when writing up your answers to the problems in this book, imagine that you are preparing a report for peer review by a skeptical reader. Can you take another few minutes to make it easier to figure out what you did and why? Can you label graph axes better, add comments to your code for readability, or justify a step? Can you anticipate objections?

You'll need skills like these for reading primary research literature, for interpreting your own data when you do experiments, and even for evaluating the many statistical and pseudostatistical claims you read in the newspapers.

One more skill deserves separate mention. Some of the book's problems may sound suspiciously vague, for example, "Comment on . . . ." They are intentionally written to make you ask, "What is interesting and worthy of comment here?" There are multiple "right" answers, because there may be more than one interesting thing to say. In your own scientific research, *nobody will tell you the questions.* So it's good to get the habit of asking yourself such things.

Acquiring these skills can be empowering. For instance, some of the most interesting graphs in this book do not actually appear anywhere. You will create them yourself, starting from data on the companion Web site.

What computers can do for you

A model begins in your mind as a proposed mechanism to account for some observations. You may represent those ideas by sketching a diagram on paper. Such diagrams can help you to think clearly about your model, explain it to others, and begin making testable experimental predictions.

Despite the usefulness of such traditional representations, generally you must also carry out some calculational steps before you get predictions that are detailed enough to test the model. Sometimes these steps are easy enough to do with pencil, paper, and a calculator. More often, however, at some point you will need an extremely fast and accurate assistant. Your computer can play this role.

You may need a computer because your model makes a statistical prediction, and a large amount of experimental data is needed to test it. Or perhaps there are a large number of entities participating in your mechanism, leading to long calculations. Sometimes testing the model involves *simulating* the system, including any random elements it contains; sometimes the simulation must be run many times, each time with different values of some unknown parameters, in order to find the values that best describe the observed behavior. Computers can do all these things very rapidly.

To compute responsibly, you also need some insight into what's going on under the hood. Sometimes the key is to write your own simple analysis code from scratch. Many of the exercises in this book ask you to practice this skill.

Finally, you will need to understand your results, and communicate them to others. *Data visualization* is the craft of representing quantitative information in ways that are meaningful, and honest. From the simplest *xy* graph to the fanciest interactive 3D image, computers have transformed data visualization, making it faster and easier than ever before.

This book does not include any chapters explicitly about computer programming or data visualization. The *Student's Guide* contains a brief introduction; your instructor can help you find other resources appropriate for the platform you'll be using.

What computers *can't* do for you

Computers are *not* skilled at formulating imaginative models in the first place. They do not have intuitions, based on analogies to past experience, that help them to identify the important players and their interactions. They don't know what sorts of predictions can be readily measured in the lab. They cannot help you choose which mode of visualization will communicate your results best.

Above all, *a computer doesn't know whether it's appropriate to use a computer* for any phase of a calculation, or whether on the contrary you would be better off with pencil and paper. Nor can it tell you that certain styles of visualization are misleading or cluttered with irrelevant information. Those high-level insights are your job.

Structure and features

- Every chapter contains "Your Turn" questions. Generally these are short and easy (though not always). Beyond these explicit questions, however, most of the formulas are consequences of something said previously, which you should derive yourself. Doing so will greatly improve your understanding of the material—and your fluency when it's time to write an exam.
- Most chapters end with a "Track 2" section. These are generally for advanced students; some of them assume more background knowledge than the main, "Track 1," material. (Others just go into greater detail.) Similarly, there are Track 2 footnotes and homework problems, marked with the glyph $\boxed{T_2}$.
- Appendix A summarizes mathematical notation and key symbols that are used consistently throughout the book. Appendix B discusses some useful tools for solving problems. Appendix C gathers a few constants of Nature for reference.
- Many equations and key ideas are set off and numbered for reference. The notations "Equation x.y" and "Idea x.y" both refer to the same numbered series.
- When a distant figure gets cited, you may or may not need to flip back to see it. To help you decide, many figure references are accompanied by an iconified version of the cited figure in the margin.

Other books

The goal of this book is to help you to teach yourself some of the skills and frameworks you will need in order to become a scientist, in the context of physical models of living systems. A companion book introduces a different slice through the subject (Nelson, 2014), including mechanics and fluid mechanics, entropy and entropic forces, bioelectricity and neural impulses, and mechanochemical energy transduction.

Many other books instead attempt a more complete coverage of the field of biophysics, and would make excellent complements to this one. A few recent examples include
General: Ahlborn, 2004; Franklin et al., 2010; Nordlund, 2011.
Cell biology/biochemistry background: Alberts et al., 2014; Berg et al., 2012; Karp, 2013; Lodish et al., 2012.
Medicine/physiology: Amador Kane, 2009; Dillon, 2012; Herman, 2007; Hobbie & Roth, 2007; McCall, 2010.
Networks: Alon, 2006; Cosentino & Bates, 2012; Vecchio & Murray, 2014; Voit, 2013.
Mathematical background: Otto & Day, 2007; Shankar, 1995.
Probability in biology and physics: Denny & Gaines, 2000; Linden et al., 2014.
Cell and molecular biophysics: Boal, 2012; Phillips et al., 2012; Schiessel, 2013.
Biophysical chemistry: Atkins & de Paula, 2011; Dill & Bromberg, 2010.
Experimental methods: Leake, 2013; Nadeau, 2012.
Computer methods: Computation: DeVries & Hasbun, 2011; Newman, 2013. Other computer skills: Haddock & Dunn, 2011.

Finally, no book can be as up-to-date as the resources available online. Generic sources such as Wikipedia contain many helpful articles, but you may also want to consult `http://bionumbers.hms.harvard.edu/` for specific numerical values, so often needed when constructing physical models of living systems.

To the Instructor

Physicist: "I want to study the brain. Tell me something helpful."
Biologist: "Well, first of all, the brain has two sides . . . ."
Physicist: "Stop! You've told me too much!"
—V. Adrian Parsegian

This book is the text for a course that I have taught for several years to undergraduates at the University of Pennsylvania. The class mainly consists of second- and third-year science and engineering students who have taken at least one year of introductory physics and the associated math courses. Many have heard the buzz about synthetic biology, superresolution microscopy, or something else, and they want a piece of the action.

Many recent articles stress that future breakthroughs in medicine and life science will come from researchers with strong quantitative backgrounds, and with experience at systems-level analysis. Answering this call, many textbooks on "Mathematical Biology," "Systems Biology," "Bioinformatics," and so on have appeared. Few of these, however, seem to stress the importance of physical models. And yet there is something remarkably—unreasonably—effective about physical models. This book attempts to show this using a few case studies.

The book also embodies a few convictions, including[1]

- The study of living organisms is an inspiring context in which to learn many fundamental physical ideas—even for physical-science students who don't (or don't yet) intend to study biophysics further.
- The study of fundamental physical ideas sheds light on the design and functioning of living organisms, and the instruments used to study them. It's important even for life-science students who don't (or don't yet) intend to study biophysics further.

[1] See also "To the Student."

In short, this is a book about how *physical science and life science illuminate each other.*

I've also come to believe that

- Whenever possible, we should try to relate our concepts to familiar experience.
- All science students need some intuitions about probability and inference, in order to make sense of methods now in use in many fields. These include likelihood maximization and Bayesian modeling. Other universal topics, often neglected in undergraduate syllabi, include the notion of convolution, long-tail distributions, feedback control, and the Poisson process (and other Markov processes).
- Algorithmic thinking is different from pencil-and-paper analysis. Many students have not yet encountered it by this stage of their careers, yet it's crucial to the daily practice of almost every branch of science. Recent reports have commented on this disconnect and recommended changes in curricula (e.g., Pevzner & Shamir, 2009; National Research Council, 2003). The earlier students come to grips with this mode of thought, the better.
- Students need explicit discussions about Where Theories Come From, in the context of concrete case studies.

This book is certainly not intended as a comprehensive survey of the enormous and protean field of Biophysics. Instead, it's intended to develop the *skills and frameworks* that students need in many fields of science, engineering, and applied math, in the context of understanding how living organisms manage a few of their remarkable abilities. I have tried to tell a limited number of stories with sufficient detail to bring students to the point where they can do research-level analysis for themselves. I have selected stories that seem to fit a single narrative, and that seem to open the most doors to current work. I also tried to stick with stories for which the student can actually do all the calculations, instead of resorting to "Smith has shown . . . ."

Students in the course come from a wide range of majors, with a correspondingly wide range of backgrounds. This can lead to some tricky, yet valuable, cross-cultural moments, like the one in the epigraph to this section. I have found that a little bit of social engineering, to bring together students with different strengths, can start the process of interdisciplinary contact at the moment when it is most likely to become a habit.

Ways to use this book

Most chapters end with "Track 2" sections. Some of these contain material appropriate for students with more advanced backgrounds. Others discuss topics that are at the undergraduate level, but will not be needed later in the book. They can be discussed a la carte, based on your and the students' interests. The main, "Track 1," sections do not rely on any of this material. Also, the *Instructor's Guide* contains many additional bibliographic references, some of which could be helpful for starting projects based on primary literature.

This book could serve as the basis of a course on the science underpinning contemporary biological physics. Or it can be used as a supplement in more specialized courses on physics, biophysics, or several kinds of engineering or applied math. Although Track 1 is meant as an undergraduate course, it contains a lot of material not generally included in undergraduate physics curricula. Thus, it could easily form the basis of a graduate course, if you add all or part of Track 2, and perhaps some reading from your own specialty (or work cited in the *Instructor's Guide*).

This book is not a sequel to my earlier one (Nelson, 2014). Indeed there is very little overlap between these books, which partly explains why certain topics are not covered here. Still other topics will appear in a forthcoming book on light, imaging, and vision. A few of the many other recent books with overlapping goals are listed in "To the Student"; others appear at the ends of chapters.

There are many ways to organize the material: by organism type, by length scale, and so on. I have tried to arrange topics in a way that gradually builds up the framework needed to understand an important and emblematic system in Chapter 11.

Computer-based assignments

The difference between a text without problems and a text with problems is like the difference between learning to read a language and learning to speak it.
—Freeman Dyson

All of the problems set in this book have been tested on real students. Many ask the student to use a computer. One can learn some of the material without doing this, but I think it's important for students to learn how to write their own short codes, from scratch. It's best to do this not in the vacuum of a course dedicated to programming, but in the context of some problems of independent scientific interest—for example, biophysics. The book's companion Web site features a collection of real experimental datasets to accompany the homework problems. Many reports stress the importance of students working with such data (for example, see National Research Council, 2003).

To do research, students need skills relevant for data visualization, simulation of random variables, and handling of datasets, all of which are covered in this book's problems. Several general-purpose programming environments would work well for this, depending on your own preference, for example, *Mathematica*®, MATLAB®, Octave, Python, R, or Sage. Some of these are free and open source. It's hugely motivating when that beautiful fit to data emerges, and important for students to have this experience early and often.

In my own course, many students arrive with no programming experience. A separate *Student's Guide* gives them some computer laboratory exercises and other suggestions for how to get started. The *Instructor's Guide* gives solutions to these exercises, and to the Problems and Your Turn questions in this book. Keep in mind that programming is very time consuming for beginners; you can probably only assign a few of the longer problems in a semester, and your students may need lots of support.

Classroom demonstrations

One kind of experiential learning is almost unique to physical science classes: We bring a piece of apparatus into the class and show the students some surprising *real* phenomenon—not a simulation, not a metaphor. The *Instructor's Guide* offers some suggestions for where to give demonstrations.

New directions in education

Will life-science students really need this much background in physical science? Although this is not a book about medicine per se, nevertheless many of its goals mesh with recent

guidelines for the preparation of premedical students, and specifically for the revised MCAT exam (American Association of Medical Colleges, 2014):[2]

1. "Achieving economies of time spent on science instruction would be facilitated by breaking down barriers among departments and fostering interdisciplinary approaches to science education. Indeed, the need for increased scientific rigor and its relevance to human biology is most likely to be met by more interdisciplinary courses."
2. Premedical students should enter medical school able to
   - "Apply quantitative reasoning and appropriate mathematics to describe or explain phenomena in the natural world."
   - "Demonstrate understanding of the process of scientific inquiry, and explain how scientific knowledge is discovered and validated," as well as "knowledge of basic physical and chemical principles and their applications to the understanding of living systems."
   - "Demonstrate knowledge of how biomolecules contribute to the structure and function of cells."
   - "Apply understanding of principles of how molecular and cell assemblies, organs, and organisms develop structure and carry out function."
   - "Explain how organisms sense and control their internal environment and how they respond to external change."
3. At the next level, students *in* medical school need another set of core competencies, including an understanding of technologies used in medicine.
4. Finally, practicing physicians need to explain to patients the role of complexity and variability, and must be able to communicate approaches to quantitative evidence.

This book may be regarded as showing one model for how physical science and engineering departments can address these goals in their course offerings.

Standard disclaimers

This is a textbook, not a monograph. Many fine points have been intentionally banished to Track 2, to the *Instructor's guide*, or even farther out into deep space. The experiments described here were chosen simply because they illustrated points I needed to make. The citation of original works is haphazard. No claim is made that anything in this book is original. No attempt at historical completeness is implied.

[2] See also American Association of Medical Colleges / Howard Hughes Medical Institute, 2009. Similar competencies are listed in the context of biology education in another recent report (American Association for the Advancement of Science, 2011), for example, "apply concepts from other sciences to interpret biological phenomena," "apply physical laws to biological dynamics," and "apply imaging technologies."

Prolog: A Breakthrough on HIV

Los Alamos, 1994

Alan Perelson was frustrated. For some years, he, and many other researchers, had been staring at an enigmatic graph (Figure 0.1). Like any graph, it consisted of dry, unemotional squiggles. But like any graph, it also told a story.

The enigmatic feature of the graph was precisely what made HIV so dangerous: After a brief spike, the concentration of virus particles in the blood fell to a low, steady level. Thus, after a short, flu-like episode, the typical patient had no serious symptoms, but remained

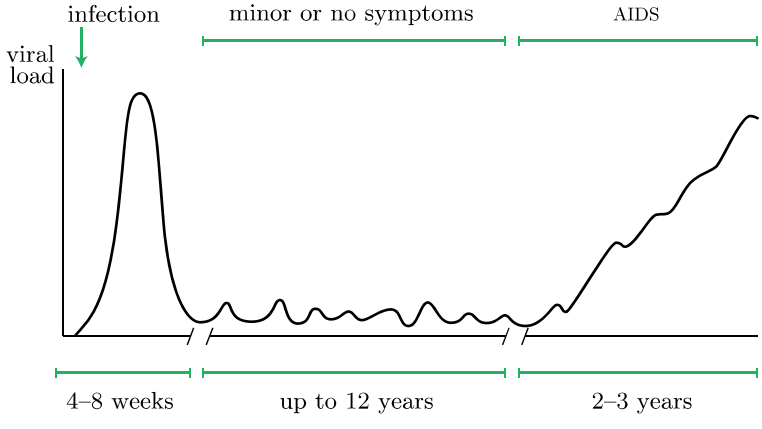


Figure 0.1 [Sketch graph.] **The time course of HIV infection,** representing the progression of the disease as it was understood in the early 1990s. After a brief, sharp peak, the concentration of virus particles in the blood ("viral load") settled down to a low, nearly steady level for up to ten years. During this period, the patient showed no symptoms. Ultimately, however, the viral load increased and the symptoms of full AIDS appeared. [After Weiss, 1993.]

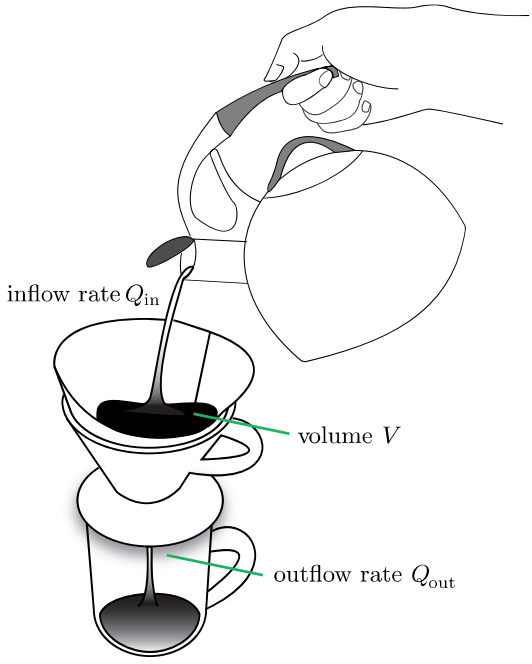


Figure 0.2 [Metaphor.] **Steady state in a leaky container.** Inflow at a rate Q_{in} replenishes the container, compensating outflow at a rate Q_{out}. If we observe that the volume V of liquid in the container is steady, we can conclude that Q_{out} matches Q_{in}, but we can't determine the actual value of either quantity without more information. In the analogy to viral dynamics, Q_{in} corresponds to the body's production of virus particles and Q_{out} to the immune system's rate of virus clearance (see Chapter 1).

contagious, for up to ten years. Inevitably, however, the virus level eventually rose again, and the patient died.

In the early 1990s, many researchers believed that these facts implied that HIV was a slow virus, which remained in the body, nearly dormant, for years before rising sharply in number. But how could such a long latency period be possible? What was happening during those ten years? How could the patient's immune system fight the virus effectively at first, and then ultimately succumb?

Perelson and others had suspected for some time that maybe HIV was not slow or dormant at all during the apparent latent period. He made an analogy to a physical system: If we see a leaky container that nevertheless retains water at some constant level, we can conclude that there must be water flowing into it (Figure 0.2). But we can't determine *how fast* water is flowing in. All we can say is that the rate of inflow equals the rate of outflow. Both of those rates could be small—or both could be large. Applying this idea to HIV, Perelson realized that, during the long period of low blood concentration, the virus might actually be multiplying rapidly, but after the brief initial episode, it could be eliminated by the body just as rapidly.

A real leaky container has another simple property reminiscent of the HIV data: Because the outflow rate $Q_{\text{out}}(V)$ increases as the volume of the water (and hence its pressure at the exit point) goes up, the system can *self-adjust* to a steady state, no matter what inflow rate Q_{in} we select. Similarly, different HIV-infected patients have quite different steady levels of virus concentration, but all maintain that steady level for long periods.

Perelson was head of the Theoretical Biology and Biophysics Group at Los Alamos National Laboratory. By 1994, he had already developed a number of elaborate mathematical models in an attempt to see if they could describe clinical reality. But his models were full of unknown parameters. The available data (Figure 0.1) didn't help very much. How could he make progress without some better knowledge of the underlying cellular events giving rise to the aggregate behavior?

New York City, 1994

David Ho was puzzled. As the head of the Aaron Diamond AIDS Research Center, he had the resources to conduct clinical trials. He also had access to the latest anti-HIV drugs and had begun tests with ritonavir, a "protease inhibitor" designed to stop the replication of the HIV virus.

Something strange was beginning to emerge from these trials: The effect of treatment with ritonavir seemed to be a very *sudden* drop in the patient's total number of virus particles. This was a paradoxical result, because it was known that ritonavir by itself didn't destroy existing virus particles, but simply stopped the creation of new ones. If HIV were really a slow virus, as many believed, wouldn't it also *stay around* for a long time, even once its replication was stopped? What was going on?

Also, it had been known for some time that patients treated with antiviral drugs got much better, but only temporarily. After a few months, ritonavir and other such drugs always lost their effectiveness. Some radically new viewpoint was needed.

Hilton Head Island, 1994

Perelson didn't know about the new drugs; he just knew he needed quantitative data. At a conference on HIV, he heard a talk by one of Ho's colleagues, R. Koup, on a different topic. Intrigued, he later phoned to discuss Koup's work. The conversation turned to the surprising results just starting to emerge with ritonavir. Koup said that the group was looking for a collaborator to help make sense of the strange data they had been getting. Was Perelson interested? He was.

Ho and his colleagues suspected that simply measuring viral populations before and after a month of treatment (the usual practice at the time) was not showing enough detail. The crucial measurement would be one that examined an asymptomatic patient, not one with full AIDS, and that monitored the blood virus concentration *every day* after administering the drug.

More clinical trials followed. Measurements from patient after patient told the same story (Figure 0.3): *Shutting down the replication of virus particles brought a hundredfold drop in their population in 2–3 weeks.*

Perelson and Ho were stunned. The rapid drop implied that the body was constantly clearing the virus at a tremendous rate; in the language of Figure 0.2, Q_{out} was huge. That could only mean that, without the drug, the production rate Q_{in} was also huge. Similar results were soon obtained with several other types of antiviral drugs. The virus wasn't dormant at all; it was replicating like mad. Analysis of the data yielded a numerical value for Q_{out}, as we'll see in Chapter 1. Using this measurement, the researchers estimated that the typical asymptomatic patient's body was actually making at least *a billion* new virus particles each day.[3]

As often happens, elsewhere another research group, led by George Shaw, independently pursued a similar program. This group, too, contained an "outsider" to AIDS

[3]Later, more refined estimates showed that the average production rate was actually even larger than this initial lower bound.

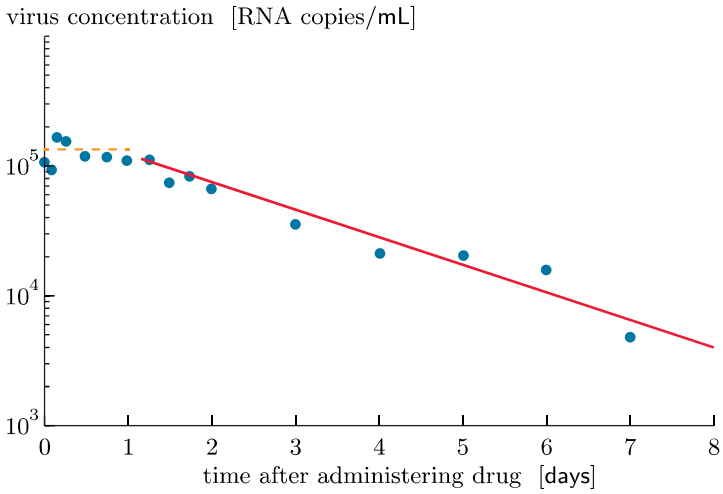


Figure 0.3 [Experimental data with preliminary fit.] **Virus concentration in a patient's blood ("viral load") after treatment with a protease inhibitor**, showing the rapid decline after treatment. In this semilog plot, the *solid line* shows the time course corresponding to elimination of half the total viral population every 1.4 days. The *dashed line* highlights a deviation from this behavior at early times (the "initial plateau"); see Chapter 1. [Data from Perelson, 2002; see Dataset 1.]

research, a mathematician named Martin Nowak. Both groups published their findings simultaneously in *Nature*. The implications of this work were profound. Because the virus is replicating so rapidly, it can easily mutate to find a form resistant to any given drug.[4] Indeed, as we'll see later, the virus mutates often enough to generate every possible single-base mutation every few hours. Hence, every infected patient *already* has some resistant mutant viruses before the drug is even administered; in a couple of weeks, this strain takes over and the patient is sick again. The same observation also goes to the heart of HIV's ability to evade total destruction by the body: It is constantly, furiously, playing cat-and-mouse with the patient's immune system.

But what if we simultaneously administer *two* antiviral drugs? It's not so easy for a virus to sample every possible *pair* of mutations, and harder still to get three or more. And in fact, subsequent work showed that "cocktails" of three different drugs can halt the progression of HIV infection, apparently indefinitely. The patients taking these drugs have not been cured; they still carry low levels of the virus. But they are alive, thanks to the treatment.

The message

This book is about basic science. It's not about AIDS, nor indeed is it directly about medicine at all. But the story just recounted has some important lessons.

The two research groups mentioned above made significant progress against a terrible disease. They did this by following some general steps:

1. Assemble (or join) an interdisciplinary team to look at the problem with different sets of tools;
2. Apply simple physical metaphors (the leaky container of water) and the corresponding disciplines (dynamical systems theory, an area of physics) to make a hypothesis; and

[4] Actually the *fact* of mutation had already been established a few years earlier. Prior to the experiments described here, however, it was difficult to understand how mutation could lead to fast evolution.

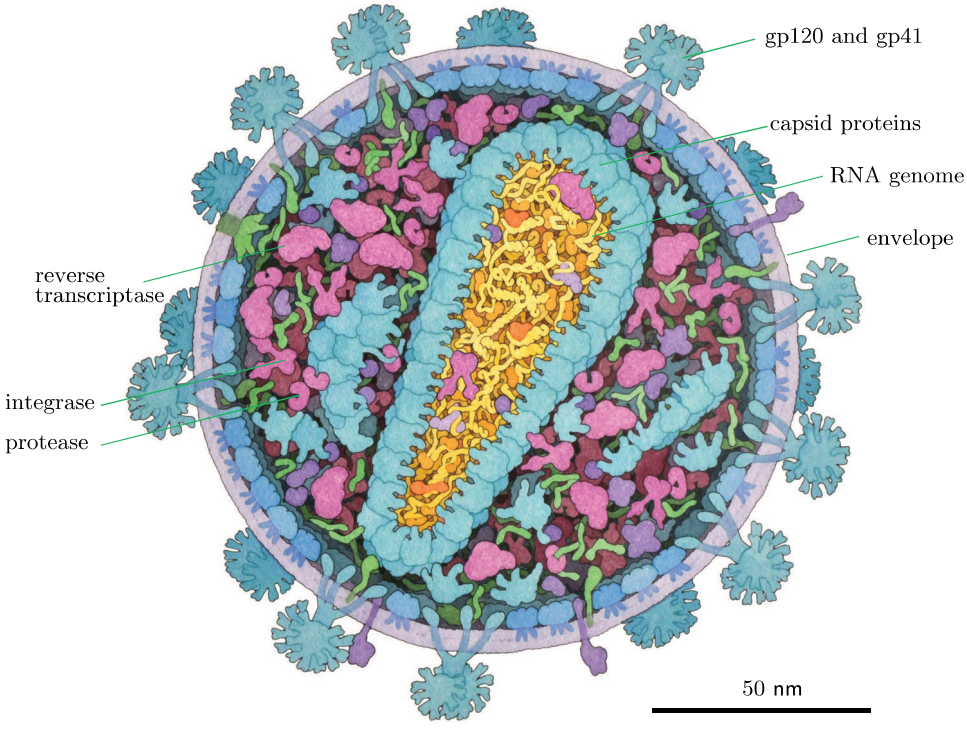
3. Perform experiments specifically designed to give new, quantitative data to support or refute the hypothesis.

This strategy will continue to yield important results in the future.

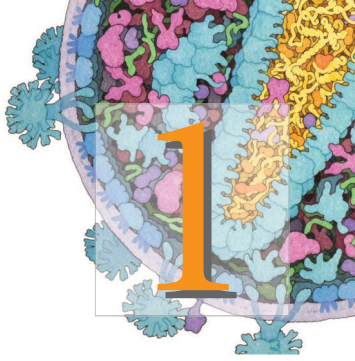
The rest of the book will get a bit dry in places. There will be many abstract ideas. But abstract ideas do matter when you understand them well enough to find their concrete applications. In fact, sometimes their abstractness just reflects the fact that they are so widely applicable: Good ideas can jump like wildfires from one discipline to another. Let's get started.

PART I

First Steps



[Artist's reconstructions based on structural data.] **A human immunodeficiency virus particle** (virion), surrounded by its lipid membrane envelope. The envelope is studded with gp120, the protein that recognizes human T cells. The envelope encloses several enzymes (proteins that act as molecular machines), including HIV protease, reverse transcriptase, and integrase. Two RNA strands carrying the genome of HIV are packaged in a cone-shaped protein shell called the capsid. See also Media 1. [Courtesy David S Goodsell.]



1 Virus Dynamics

We all know that Art is not truth. Art is a lie that makes us realize the truth.
—Pablo Picasso

1.1 First Signpost

The Prolog suggested a three-step procedure to make headway on a scientific problem (see page 4). Unfortunately, the experiment that can be performed usually does not directly yield the information we desire, and hence does not directly confirm or disprove our original hypothesis. For example, this chapter will argue that testing the viral mutation hypothesis in the Prolog actually requires information not directly visible in the data that were available in 1995.

Thus, a fourth step is almost always needed:

4. Embody the physical metaphor (or **physical model**) in mathematical form, and attempt to fit it to the experimental data.

In this statement, **fit** means "adjust one or more numbers appearing in the model." For each set of these **fit parameter** values that we choose, the model makes a prediction for some experimentally measurable quantity, which we compare with actual observations. If a successful fit can be found, then we may call the model "promising" and begin to draw tentative conclusions from the parameter values that yield the best fit. This chapter will take a closer look at the system discussed in the Prolog, illustrating how to construct a physical model, express it in mathematical form, fit it to data, evaluate the adequacy of the fit, and draw conclusions. The chapter will also get you started with some of the basic computer skills needed to carry out these steps.