BIO-INSPIRED AND NANOSCALE INTEGRATED COMPUTING

EDITED BY

Mary Mehrnoosh Eshaghian-Wilner

WILEY

A JOHN WILEY & SONS, INC., PUBLICATION

BIO-INSPIRED AND NANOSCALE INTEGRATED COMPUTING

BIO-INSPIRED AND NANOSCALE INTEGRATED COMPUTING

EDITED BY

Mary Mehrnoosh Eshaghian-Wilner

WILEY

A JOHN WILEY & SONS, INC., PUBLICATION

Published by John Wiley & Sons, Inc., Hoboken, New Jersey
Published simultaneously in Canada

Library of Congress Cataloging-in-Publication Data:

Bio-inspired and nanoscale integrated computing / edited by Mary Mehrnoosh Eshaghian-Wilner.
p. cm. – (Nature-inspired and nanoscale integrated computing)
Includes bibliographical references and index.
ISBN 978-0-470-11659-3 (cloth)
1. Molecular computers. 2. Biologically-inspired computing. 3. Nanotechnology. 4. Nanomedicine. I. Eshaghian-Wilner, Mary Mehrnoosh.
QA76.887.B56 2009
004.1–dc22
2009013331

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

CONTENTS

FOREWORD

Bio-inspired and Nanoscale Integrated Computing (edited by Professor Mary Mehrnoosh Eshaghian-Wilner) is the first book in a new and exciting Wiley book series, the *Nature-Inspired Computing Series*. The proliferation of computing devices in every aspect of our lives increases the demand for better understanding of emerging computing paradigms. The new book series seeks to provide an opportunity for researchers to explore the new computational paradigms and their impact on computing in the new millennium.

The series is quite timely since the field of computing as a whole is undergoing many changes. A vast literature exists today on such new paradigms and their implications on a wide range of applications—a number of studies have reported on the success of such techniques in solving difficult problems in all key areas of computing. The list of topics covered by the series is by no means exhaustive but it serves as a guide to the diversity of the topics covered here. It is also hoped that the topics covered will get the readers to think of the implications of such new ideas on the developments in their own fields. These fields are: Quantum Computing, DNA Computing, Genetic Algorithms, Evolutionary Paradigms, Cellular Automata, Neural Networks, Swarm Algorithms, Fuzzy Logic, Computational Synthesis, Machine Learning Techniques, Computational Methods for Biological Systems. The series will also publish books related to technologies that are relevant to nature-inspired computing: Silicon Neuron Processing, Molecular Scale Computers and NanoTech, Optics, Evolvable Hardware, Quantum Hardware, Reconfigurable Hardware, and others.

Bio-inspired and Nanoscale Integrated Computing is a solid inaugural contribution that overviews some of the recent developments in the area of nanocomputing. This is an important emergent discipline which is facilitated by scientific advances in the field of nanotechnology. This book, through a number of chapters covering a wide range of topics, increases the awareness of the potential of nanocomputing as a new paradigm that holds great promise. It is well-positioned to generate more interest and concerted effort in studying the different facets of nanocomputing and applying it to a wider range of problems in science and engineering.

Bio-Inspired and Nanoscale Integrated Computing can be used by scientists, engineers, graduate students, senior undergraduate students, researchers, instructors, and practitioners, who want to learn more about nanocomputing. It is definitely an excellent reference on this topic.

Albert Y. Zomaya

The University of Sydney
Sydney, New South Wales, Australia

PREFACE

The word “bionics” is often described as a compound of “biology” and “electronics.” Some say it was coined in the 1950s and that it originates from the Greek word “βίον,” meaning “unit of life,” and the suffix “-ic,” meaning “like” or “in the manner of”–hence “like life.” But just like many people of a certain age, I first came to know the term when I was a child (in my case growing up in Tehran) watching the then-famous duo of American TV series, “The Six Million Dollar Man” and “The Bionic Woman.” Back then, I had no idea how much of those programs was just fantasy. Not only was I too young to know better, but I was also under the impression that America was so much more advanced than where I lived, and therefore everything must have already been possible in the United States.

As I grew older and eventually moved to the United States during my early teen years, I began to realize more and more just how far removed the bionic fantasy was. The absence of those advancements I had expected to see was certainly a big disappointment, yet it motivated me to begin my undergraduate studies in biomedical and electrical engineering at the University of Southern California (USC).

As part of my studies, it became apparent that computer science and engineering were key disciplines that would make the bionic fantasy a reality. I therefore majored in Computer Engineering in the USC graduate school and received a Master s degree and a Ph.D. My ultimate dream was to make an actual bionic eye. To move one step closer to that dream, I concentrated on designing ultra-dense multiprocessor VLSI chips that had the capability of massive electro-optical intercommunications. I also designed a large number of algorithms that could be mapped onto those chips for image processing and computer vision. Before long, my career took off and my work received recognition in the area of parallel processing. I published, gave talks, taught courses, and organized conferences in various areas within parallel processing. Over time, my bionic dream was diminished to a small side issue.

About two decades later, while employed as a professor and department head at the Rochester Institute of Technology, I learned about the intriguing advancements that were rapidly taking place in the new fields of nanotechnology and bioinformatics. I applied for and received a fellowship at the National Science Foundation’s Institute of Pure and Applied Mathematics at the University of California Los Angeles (UCLA). There, I interacted with an interdisciplinary group of scientists who were all interested in the applications of nanotechnology to their fields of expertise. Among them were biologists, chemists, physicists, mathematicians, electrical engineers, mechanical engineers, computer scientists, and so forth. I knew that if I could somehow get all of them to understand each other, we could then also sit and design some of the world's most challenging devices, e.g., those that were part of my "bionic dream." Towards that goal, I began working with a small group of interdisciplinary scientists, and together we wrote a joint article about various ways computing could be done at a nanoscale level. The collaboration also led to the production of some very interesting breakthrough results that were widely disseminated and publicized through various news media. Shortly after that, I was contacted by an editor for the publisher, John Wiley & Sons, who told me Wiley would be interested in a book on the type of collaborative multidisciplinary work in which I was engaged.

"Bio-inspired and Nanoscale Integrated Computing (BioNIC)," the name I had given to my research team at UCLA, was the title that I proposed to Wiley. I proposed a collection of the latest findings of leading scientists who were working in the proposed topic area, with the intent to get one step closer to the creation of actual bionic parts.

Paul Petralia, the Wiley Editor, liked my proposal, then sent it for external review for which it received very positive and high peer ratings. I began working on the book immediately; and after about three years, it is now complete. The quality of the work presented in this book is all to the credit of the outstanding group of scientists with whom I had the distinct privilege and honor to collaborate. I must mention here that this book was initially planned to contain significantly fewer chapters than what you see before you. Only through the remarkable dedication and contributions of the co-authors did this book evolve to its current size. I am immensely thankful and indebted to all of them.

I invite all of you to begin reading with a special emphasis on Chapter 1. That chapter was produced primarily by a group of amazing interdisciplinary UCLA students who took a seminar course on nanocomputing that I taught. Each student studied the application of nanocomputing to their prospective fields and, based on those, they all got together and wrote the chapter. The chapter not only contains a brief but comprehensive background, but it also serves as a guide to various chapters within the book where additional information can be found. Special credit goes to Shawn Singh, who took a superb leading role among all the students that were involved. I also thank Professor Alireza Nojeh for his final review of their work.

There are many additional people to whom I am grateful. First and foremost, I thank Paul Petralia of Wiley, who had the foresight, leadership, and vision to help me produce this book from a simple idea to what it is now. He is also responsible for creating the Nature-Inspired Computing Series that I am co-editing with my dear colleague, Professor Albert Zomaya. This book will serve as the first in the new series and it will be followed by several other interesting titles, some of which are currently in progress. I am extremely grateful and pleased for the opportunity to be working with the renowned Professor Zomaya on this important series.

My undying gratitude also extends to the numerous other wonderful people at Wiley and UCLA who have worked with me nonstop on this book. At Wiley, a special shout-out goes to Lisa Morano Van Horn for her outstanding production work, to George Telecki for seamlessly transitioning into Paul Petralia's role, and to Michael Christian for his routine help with all diverse tasks. Over at UCLA, I would like to thank my administrative assistant, Rose Weaver, and all of my research assistants, including Dr. Shiva Navab, David Shen, Jon Lau, Mike Yip, and Stephen Chu, and Eric Mlinar.

Finally, I don't think this book would have been possible without the various contributions of my family. My parents, Mehdi and Molly, my brothers, Michael and Mark, and my sister, Maggie, have always supported me throughout the various stages of my life and career. I am forever indebted to them. My husband, Arthur, who is an excellent writer himself, has been a great adviser to me. I thank him with all my heart for all his valuable feedback. But most of all, I owe the most credit to my beautiful daughter, Ariana, who has been so patient and understanding about Mommy's crazy schedule. I love her more than words can describe, and I dedicate this book to her and our family.

Thank you for reading this book. Enjoy!

MARY MEHRNOOSH ESHAGHIAN-WILNER

Los Angeles, California
March 2009

CONTRIBUTORS

Sumit Ahuja, Research Assistant, Fermat Lab, Virginia Polytechnic and State University, Blacksburg, Virginia

Lisong Ai, Department of Biomedical Engineering and Division of Cardiovascular Medicine, University of Southern California, Los Angeles, California

Debayan Bhaduri, Research Assistant, Fermat Lab, Virginia Polytechnic and State University, Blacksburg, Virginia

Varun Bhojwani, Analyst, Wachovia Securities, Los Angeles, California

Elaine Ann Ebreo Cara, Department of Electrical Engineering, University of California, Los Angeles, California

R. Chomko, Assistant Research Professor, Department of Electrical Engineering, University of California, Riverside, California

Stephen Chu, Department of Electrical Engineering, University of California, Los Angeles, California

Dr. I. Dumer, Professor, Department of Electrical Engineering, University of California, Riverside, California

Dr. Mary Mehrnoosh Eshaghian-Wilner, Adjunct Professor, Department of Electrical Engineering, University of California, Los Angeles, California

Dr. Robert A. Freitas Jr., Senior Research Fellow, Institute for Molecular Manufacturing, Palo Alto, California

Aaron K. Friesz, Department of Electrical Engineering, University of California, Los Angeles, California

Tzung K. Hsiai, Department of Biomedical Engineering and Cardiovascular Medicine, University of Southern California, Los Angeles, California

Dr. Graham A. Jullien, Professor, Department of Computer and Electrical Engineering, University of Calgary, Calgary, Alberta, Canada

Alex Khitun, Research Engineer, Department of Electrical Engineering, University of California, Los Angeles, California

Dr. Sakhrat Khizroev, Professor, Department of Electrical Engineering, University of California, Riverside, California

Dr. Thomas H. LaBean, Professor, Department of Chemistry and Department of Computer Science, Duke University, Durham, North Carolina

Ling Lau, Hardware Engineer, Cisco Systems, San Jose, California

Dr. Dimitri Litvinov, Director, Center for Nanomagnetic Systems, Associate Professor, Electrical and Computer Engineering and Chemical and Biomolecular Engineering Departments, University of Houston, Houston, Texas

Dr. Tulin Mangir, Professor, Electrical Engineering Department, College of Engineering, Director, Nano Technology Lab, and Director, Wireless and Mobile Security Lab, California State University, Long Beach, California

Dr. Shiva Navab, Staff Scientist, Broadcom Corporation, Irvine, California

Dr. Alireza Nojeh, Assistant Professor, Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, British Columbia, Canada

Dr. Alice C. Parker, Professor, Department of Electrical Engineering, University of California, Los Angeles, California

Dr. John H. Reif, A. Hollis Edens Distinguished Professor of Computer Science, Duke University, Durham, North Carolina

Fady Rofail, Department of Electrical Engineering, University of California, Los Angeles, California

Mahsa Rouhanizadeh, Department of Biomedical Engineering, University of Southern California, Los Angeles, California

Dr. Mario Ruben, Institut fur Nanotechnologie, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

Michael M. Safaee, David Geffen School of Medicine, University of California, Los Angeles, California

David D. Shen, School of Law, University of California, Davis, California

Sandeep Shukla, Associate Professor, Department of Electrical and Computer Engineering, Virginia Polytechnic and State University, Blacksburg, Virginia

Gaurav Singh, Fermat Lab, Virginia Polytechnic and State University, Blacksburg, Virginia

Shawn Singh, Department of Electrical Engineering, University of California, Los Angeles, California

Dr. James M. Tour, Chao Professor of Chemistry, Professor of Computer Science, and Professor of Mechanical Engineering and Materials Science, Rice University, Houston, Texas

Ko-Chung Tseng, Department of Electrical Engineering, University of California, Los Angeles, California

Dr. Konrad Walus, Associate Professor, Electrical and Computer Engineering, University of British Columbia, Vancouver, British Columbia, Canada

Dr. Kang L.Wang, Raytheon Chair Professor of Physical Electronics, Director, Marco Focus Center on Functional Engineered Nano Architectonics (FENA), and Director, Western Institute of Nanoelectronics (WIN), University of California, Los Angeles, California

Alexander D. Wissner-Gross, Environmental Fellow, Harvard University, Cambridge, Massachusetts

Daniel Wu, Department of Electrical Engineering, University of California, Los Angeles, California

Chun Wing Mike Yip, Department of Electrical Engineering, University of California, Los Angeles, California

Dr. Hongyu Yu, Assistant Professor, School of Earth and Space Exploration, Department of Electrical Engineering, Arizona State University, Tempe, Arizona

Dr. Lin Zhong, Professor, Department of Electrical and Computer Engineering, Rice University, Houston, Texas

1

AN INTRODUCTION TO NANOCOMPUTING

Elaine Ann Ebreo Cara, Stephen Chu, Mary Mehrnoosh Eshaghian-Wilner, Eric Mlinar, Alireza Nojeh, Fady Rofail, Michael M. Safaee, Shawn Singh, Daniel Wu, and Chun Wing Yip

The continuous shrinking of transistors has made it possible to do amazing feats with computers, and a major theme of the microelectronics era has been "smaller is better." Today, technology has already shrunk to the nanometer scale, causing many practical challenges and motivating the search for new nanoscale materials and designs. In this chapter, we present a brief introduction to the concept of nanocomputing and provide a high level overview of nanocomputing devices and paradigms. We also discuss some applications of nanocomputing such as biomedical engineering and neuroscience.

1.1. INTRODUCTION

In 1959, Nobel laureate Richard Feynman posed this question to his fellow physicists: *"Why cannot we write the entire 24 volumes of the Encyclopedia Britannica on the head of a pin?"* In that lecture, aptly named "There's Plenty of Room at the Bottom," Feynman challenged scientists and engineers to imagine what could be possible using nanoscale structures. He used computing as a prime example, suggesting that wires and circuits could be shrunk to only hundreds of angstroms in size [1], making it possible to combine billions of devices to perform amazing tasks.

The transistor, today's most prevalent modern computing device, can already be manufactured as small as hundreds of atoms across. This could mark the

Bio-Inspired and Nanoscale Integrated Computing. Edited by Mary Mehrnoosh Eshaghian-Wilner

end of the microcomputing era and the beginning of the *nanocomputing era*. Nanocomputing could be fundamentally different from microcomputing: individual particles could play a significant role; quantum effects enter the game much more directly. Even though Feynman's vision has been partially realized with such tiny transistors, researchers are only beginning to explore the fundamental questions: How small can we make computers? How would such computers work? How can we embrace quantum mechanics for computation? What applications are possible with nanocomputing that were not possible with microcomputing? We will explore these questions throughout the book.

In this introductory chapter, we will give a brief overview of devices, paradigms, and applications of nanocomputing. Of course, we cannot include all existing ideas about nanocomputing in the introduction, or even in the entire book. Instead, we aim to inspire the reader with a variety of ideas that appear commonly in nanocomputing research. We will first define computing and nanocomputing and provide some historical context of the microcomputing era. The limitations of today's microcomputers motivate a discussion of nanoscale devices and paradigms, many of which are detailed in later chapters. We then consider two major fields that can greatly benefit from nanocomputing: biology and neurology.

1.2. WHAT IS NANOCOMPUTING?

Computing is the *representation* and *manipulation* of information. Computer games, surfing the Internet, solving complex math equations, and even verbal communication are examples of computing. While we certainly compute using our own thinking power, it is often more useful to create a machine that computes on its own so we can use it to enhance our daily lives. Indeed, our world has become dependent on machines that automatically compute for us. These "computers" are used for entertainment, education, safety, and a vast number of other applications, all of which require manipulation of abstract information.

To build a computing machine, abstract data must eventually be represented by something that physically exists. For example, digital states (such as 1's and 0's) can be represented as high and low voltages on a wire. Similarly, manipulations of abstract data, such as adding two numbers, must eventually be performed by a physical phenomenon that affects the data. Therefore, to explore the world of computing, we must ask the most fundamental question: *How can we use physics to represent and manipulate abstract information?*

Directly or indirectly, researchers from all over the world are exploring this fundamental question. This endeavor spans across several disciplines, including mathematics, computer science, physics, chemistry, and biology. Throughout this book, there are many examples of how physics can be used for computation; some of these ideas may eventually lead to more powerful computers.

Nanocomputing can be interpreted literally as *computing at the nanometer scale*. It is generally agreed that the terms *nanotechnology*, *nanocomputing*, and *nanoscale* are used when considering devices that are at least one dimension smaller than 100 nanometers (nm). Today, devices such as transistors have channels that are well below 100 nanometers in length. Eventually the size of entire computers may also be measured in nanometers.

Just how small is a nanometer? A nanometer is one-billionth of a meter (10^{-9} m). Figure 1.1 shows the approximate size of various physical entities at the nanometer scale. If it were possible to arrange 10 hydrogen atoms side by side, their combined width would measure approximately 1 nanometer. A typical processor found in a modern desktop computer is roughly 10 millimeters wide and 10 millimeters long—in terms of nanometers, this is nearly 10 million nanometers in width and length! Other computers are commonly much smaller; embedded processors used today in cell phones, cars, and many other devices are as small as a fraction of a millimeter squared. Nanocomputers will be even smaller; electrons, atoms, DNA, and proteins are all nanometer scale or smaller and offer a huge variety of ways to represent and manipulate data.

At first this description may seem to be simply a matter of size. For more than 50 years, computers have continued to get smaller and faster, and so it may not be obvious that nanocomputing is drastically different than microcomputing. Looking deeper, however, the nanometer size opens a new world of possibilities. Nanocomputers will be able to fit anywhere, even inside our own bodies. They may be nearly undetectable, certainly invisible to the naked eye. Millions of such computers could work together and intelligently collect data about the world. Quantum physics makes it difficult to continue using old microcomputing techniques, but it also gives us infinitely more possibilities.

With all this in mind, **nanocomputing** can be defined as *the study of devices, paradigms, and applications that surpass the domain of traditional microcomputers by using physical phenomena and objects measuring 100 nm or less*. This definition clarifies that nanocomputing encompasses a large variety of challenges, ranging from effectively fabricating nanoscale devices to creating revolutionary applications for nanocomputers.

In the nanoscale world, it is unlikely that computers will work the same way they work today. Figure 1.2 lists some of the novel paradigms that can be realized with nanoscale physics, roughly estimating how they may be compared to each other given today's understanding of nanocomputing. As the figure shows, there is a vast amount of untapped potential computational power beyond CMOS logic, today's dominant technology.

Another important aspect of nanocomputing is the new set of applications that will be possible with such tiny, powerful computers. In turn, the plethora of applications raises ethical, social, and economic questions that are also of great interest.

The applications and impact of nanocomputing will be discussed later in this chapter, but let us first turn our attention to the fundamental question stated above, that is, exploring how physics can be used for computation. Over the next several sections we will discuss this idea, providing a historical context and generic taxonomy of nanocomputing topics that are detailed in the rest of this book.

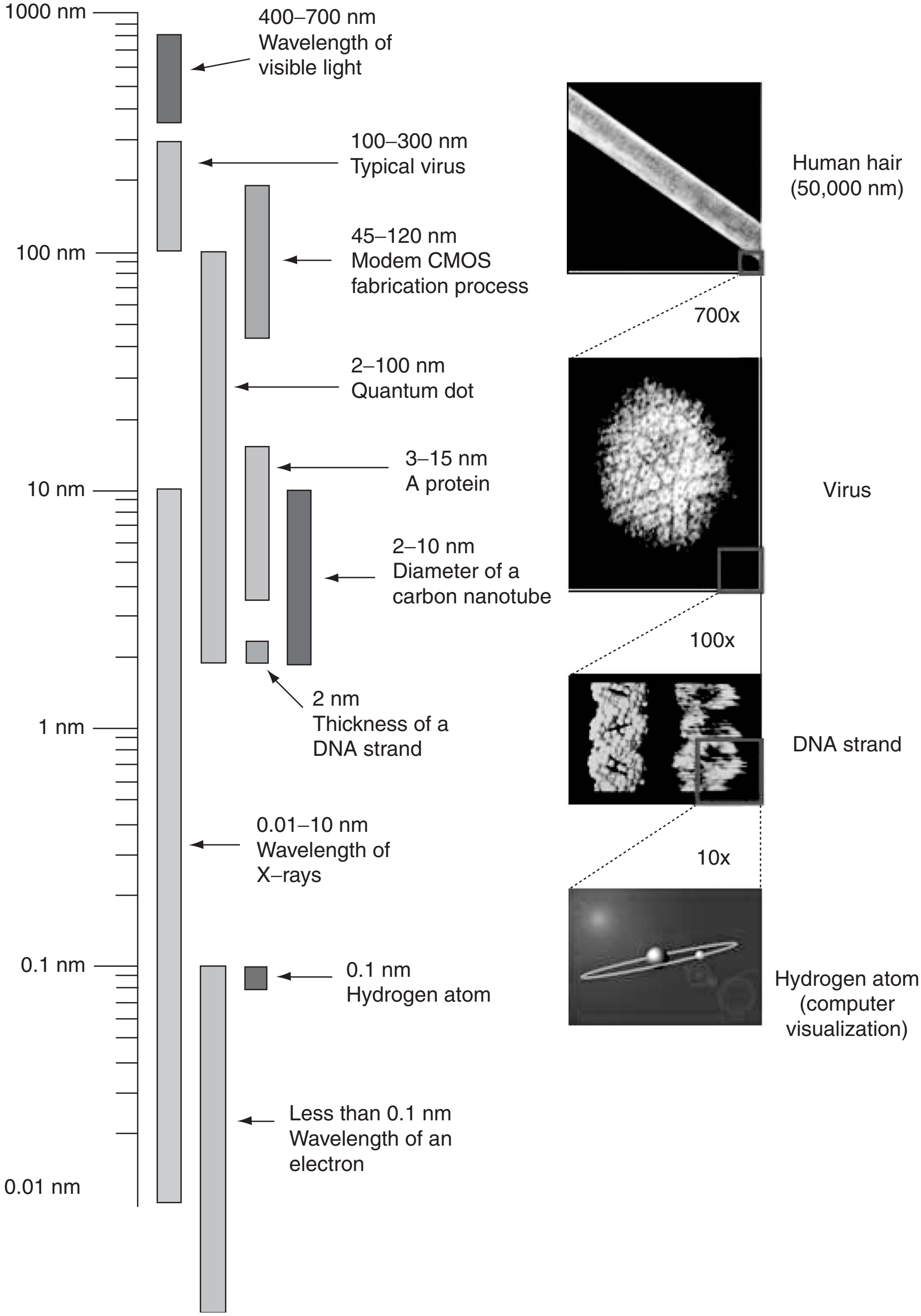

Figure 1.1. Left: Size of various objects, measured on a logarithmic scale. Right: Visual depiction of some of these objects, to compare relative size. Permissions obtained for the DNA strand from CalTech. The human hair and the virus pictures were taken from the Wikipedia public domain.

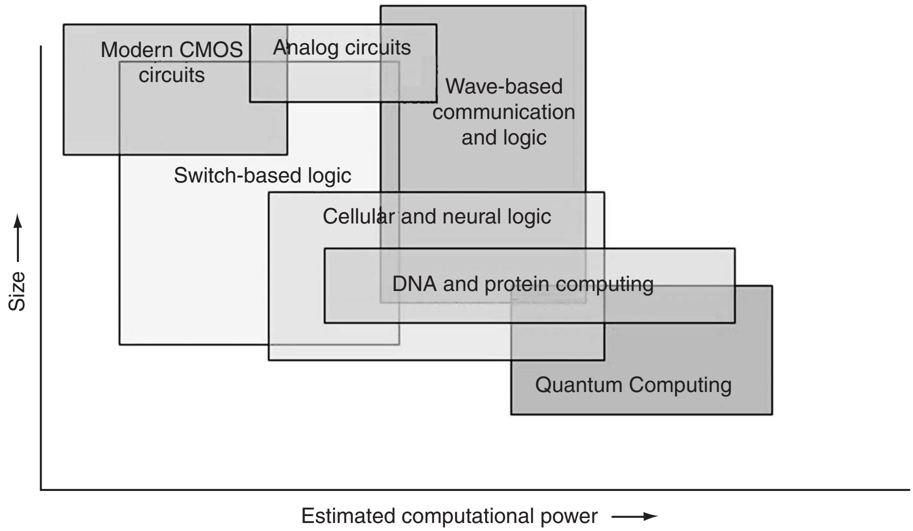

Figure 1.2. Visualization of how future paradigms in nanocomputing may compare to today's CMOS technology.

1.3. THE MICROCOMPUTING ERA: THE TRANSISTOR AS A SWITCH

Traditionally, the most common way to use physics for computation is to cleverly control electricity. Figure 1.3 shows a simplified *transistor*. We can add or remove electrons from the gate. When there is no charge in the gate, the wire can easily transmit its own electrons. If there are electrons in the gate, an electric field of negative charge is created, and this repelling force makes it difficult for electrons to flow through the wire. In a sense, we can control how much current flows through the wire by controlling how much charge we put into the gate.

With this physical device, an abstract 0 or 1 is represented as a low or high current on the wire. This is known as the *digital abstraction*. The transistor's

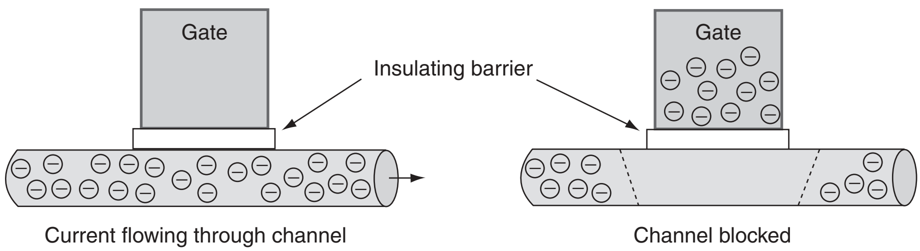

Figure 1.3. A simplified field-effect transistor. Ideally, the gate can "switch" current on or off.

behavior represents a simple *switch*: the gate can allow or prevent current from flowing through the wire. This is the *switch abstraction*. In practice, there are many more abstractions placed on top of these two (for example, representing integers in binary form with a series of 0's and 1's). However, the digital and switch abstractions are particularly significant because they bridge between a physical phenomenon, moving electrons and electric fields, to an entirely abstract world, manipulating 0's and 1's with switches. This use of transistors is the cornerstone of modern computing.

One particularly interesting achievement occurred in 1959, when both Robert Noyce and Jack Kilby independently developed the *integrated circuit*. With integrated circuits, one fabrication process simultaneously creates many transistors, all of them integrated on a single crystalline structure such as silicon. As fabrication techniques began to improve, it became possible to pack more transistors together. By 1965 Gordon Moore, the co-founder of Intel, predicted that the number of transistors that fit into a given area would double every 18 months due to continued improvements in the fabrication process. Following this prediction known as Moore's Law, transistor size, speed, and power consumption have exponentially improved for almost 50 years. Today it is possible to construct hundreds of millions, even billions, of tiny transistors on a small piece of silicon the size of a thumbnail (Fig. 1.4). In turn, it has become practical to create abstract computers that use millions or billions of switches.

Because the fabrication process produces all transistors simultaneously, the cost of fabricating these computers is largely independent of the number of transistors. There is typically a large initial cost, and this initial cost can be amortized over thousands or millions of processors, which can be produced cheaply. The economics of this situation is staggering—with a smaller transistor, performance improves, power consumption decreases, more abstract computation fits onto a single processor, and all this happens as the price of each transistor decreases! With this persistent exponential improvement, it is very easy to manipulate large amounts of abstract information, and computers are used for a prolific number of applications today. All of this has hinged on the fact that transistors continue to get smaller, and this has led to the general trend that "smaller is better."

1.3.1. Difficulties with Transistors at the Nanometer Scale

Transistor sizes are already at the nanometer scale, and this causes many practical difficulties. At the time of publication of this book, many consumer products are using a 45 nm fabrication process, and 32 nm technology has already been demonstrated. At these small sizes, fundamental limitations have to be considered. Entire books have been written on the subject, and here we describe only a few such challenges.

One primary example of these difficulties is a quantum phenomenon known as *tunneling*, visualized in Figure 1.5. Due to the wave nature of particles, electrons can "jump," or tunnel, through barriers with some nonzero probability. This probability increases exponentially as the size of the barrier decreases. The size of

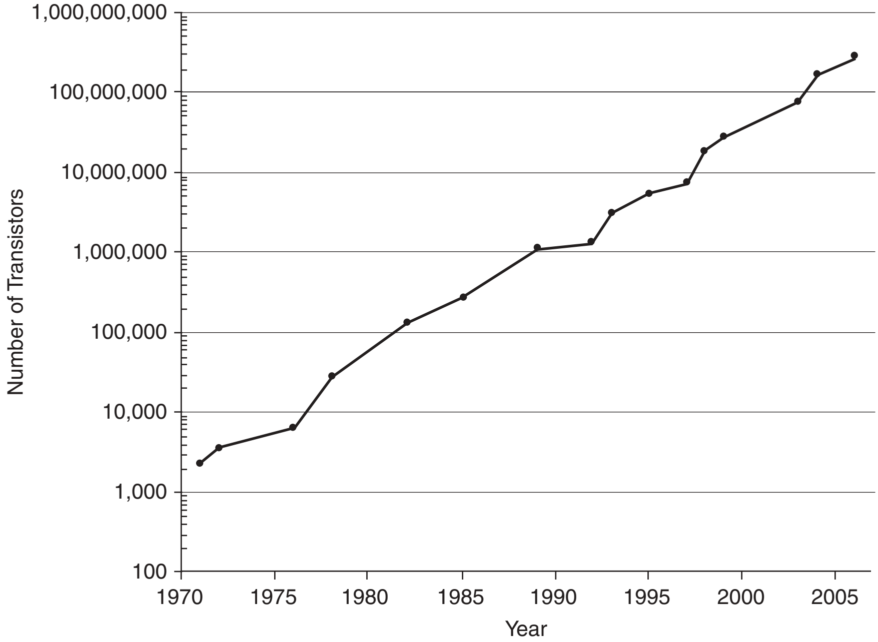

Figure 1.4. The number of transistors found in commercial processors. Note that the y-axis is logarithmic. This exponential trend is mostly due to the decreasing size of transistors. It is expected that very soon transistors will become as small as physically possible, which motivates the exploration of other devices that may be smaller and more powerful. Data acquired from [2].

transistors has decreased so much that in today's tiny transistors, electrons regularly tunnel between the gate and the wire. (Fig. 1.6). Since electrons and charge cannot be controlled as easily at the nanometer scale, the transistor behaves less and less like an ideal switch.

Tunneling has become part of a larger tradeoff between performance and power consumption. The size of transistors has reached the point where traditional models of transistors cannot be applied without a detailed understanding of nonideal characteristics [3]. There are many reasons that electrons can unintentionally leak across the wire, even when the gate tries to block current. Furthermore, the smaller the wires become, the more difficult it becomes for electrons to move through wires; that is, thinner wires have greater resistance. Because of this, even more power is required to push electrons through the wires quickly. Most processors today are limited to about 4 GHz, largely because power requirements beyond this speed are too costly and generate too much heat for a processor to function properly.

Many creative solutions have kept transistors useful despite these limitations. For example, by placing the appropriate stress or strain on the crystalline

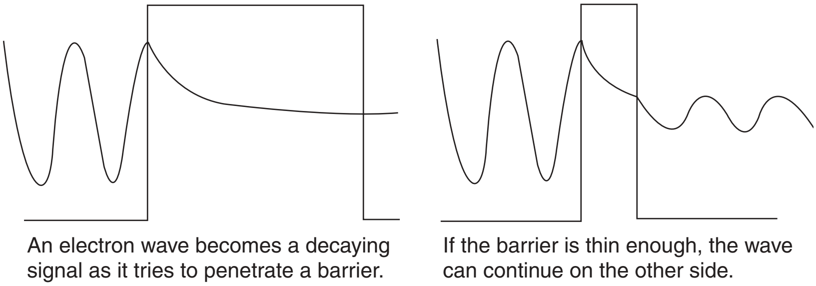

Figure 1.5. Visualization of the tunneling phenomenon. Mathematically, the wave changes into an exponential decay when it enters a region of "high potential"—the barrier—and resumes as a wave after it exits. The probability that an electron will tunnel across the barrier is related to how much amplitude the wave has left once it exits the barrier.

structure of silicon, electrons can move through a transistor more easily. This can be done by adding materials on top of transistors that naturally want to bend, thus pulling or pushing on the silicon. The so-called *strained silicon* [4] has quickly become a standard technique to improve the performance of transistors at 90 nm or less. Another example is the development of better insulating materials, known as *high-K dielectrics* [5]. The right combination of conducting and insulating materials can reduce the amount of undesirable tunneling between the gate and channel, even when the barrier is only a few layers of molecules thick. This advancement has been the key towards 45-nm technology. In the future, it may be necessary to use multiple gates to reliably control the current along a wire. *FinFETs* [6] or *trigate transistors* [7] are two multigate variations of transistors that may take us beyond 45 nm.

There are several more limitations when using tiny transistors that motivate the nanocomputing ideas presented in this book. First, the wiring that interconnects transistors is becoming a very signficant limitation for performance, power, and size of devices. There are even theoretical limitations about how much

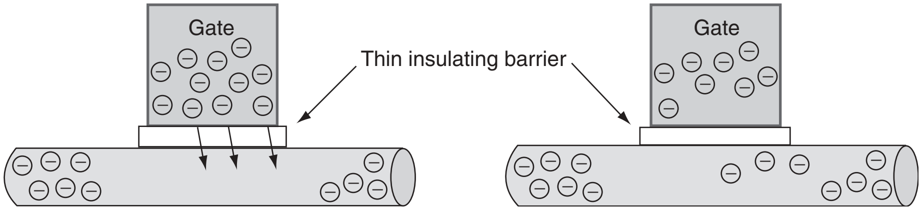

Figure 1.6. One of many nonideal effects in a transistor is that electrons in the gate may tunnel into the wire. This occurs more often as the thickness of the insulating barrier decreases.

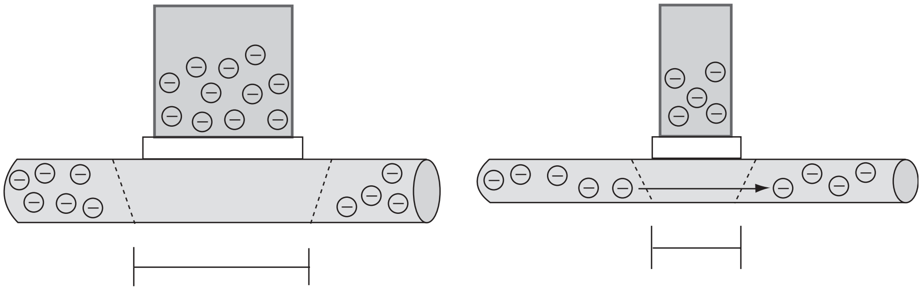

Figure 1.7. When transistors become very tiny, electrons can tunnel across the channel when the gate tries to block current flow. If the channel length is small enough, electrons will regularly tunnel in this way, and the gate would no longer effectively control current flowing through the wire.

area interconnections require as long as we connect transistors with traditional wires [8]. As we will see in the next few sections, there are many ideas in this book that reduce the limitations of wiring.

Second, variations during fabrication are now becoming a very significant problem. Relative to such tiny transistors, variations in geometry or chemical concentrations can easily change or break the behavior of the transistor. This variability decreases the yield and reliability of devices. *Fault-tolerant methodology* (Chapter 10) is desired for computing under unreliable conditions, and new fabrication methods, such as *self-assembly* (Chapter 12), may be better for reliable fabrication at the nanometer scale. Furthermore, *reconfigurability* (Chapter 5) offers a way to keep a device useful by updating or fixing its functionality.

Finally, when transistors become very small (below 5 to 15 nm approximately), electrons will be able to tunnel in a different, much more challenging way: electrons would be able to tunnel through the channel itself, even when the gate tries to block current, defeating the purpose of a gate entirely (Fig. 1.7). It is currently not clear how to overcome this upcoming problem, except to find a better nanoscale device that can behave like a switch [9].

1.4. BEYOND THE TRANSISTOR: NANOSCALE DEVICES

In practice, the use of transistors has been so successful that so far it remains unchallenged as the “best way” to use physics for computation. However, as mentioned above, it is not clear that the transistor will continue to be the best device to use as a switch at the nanometer scale. One major facet of nanocomputing research is finding new devices that exhibit switching or other behaviors that are useful for computing. Unlike the classical transistor, these devices very directly embrace the properties of quantum physics to serve their function. In this section, we briefly describe various nanoscale device technologies, referring to the specific chapters where topics are discussed in more detail.

It should be noted that this introductory material is not intended to be a comprehensive list of nanoscale devices; such information can be found in later chapters. In fact, this section only describes a mere fraction of the devices that are being explored at the nanometer scale. Instead, the purpose of this section is to give an intuitive understanding for several common aspects of nanoscale devices.

1.4.1. Molecular Devices

In general, there are a huge variety of molecules and structures that can be explored (for example DNA, proteins, rotaxanes, nanotubes, and more) [10]. In some sense, atoms and molecules are just highly complicated toy blocks: there are an infinite number of ways to assemble molecules into something useful, limited only by the creativity of future research.

Molecular structures can be used to create very tiny switches, ranging from 1 to 10 nm in size. One possible approach is to control how easily electrons can flow through the molecule, very much like a transistor, but with different underlying physics (e.g., [11]). Another possible approach is to control how light is absorbed or scattered by the molecules (e.g., [12]). These interactions with molecules can be controlled in many different ways, for example, by applying a nearby voltage or by changing the structure of molecules. Molecular switches and molecular computing are discussed further in Chapter 11.

A big challenge with molecular switches—and many nanoscale devices—is to effectively fabricate and interconnect them to perform complex logic functions. In an attempt to circumvent these problems, one proposed molecular device is the *NanoCell* [13]. The NanoCell tolerates defects and variability that occur during self-assembly fabrication. To provide reliability, the NanoCell depends on post-fabrication "training" to create the desired logic function. This approach is interesting for two reasons. First, the logic function of the NanoCell can (ideally) be reconfigured instead of permanently fixed; second, it allows the use of larger and fewer wires to connect between different cells. The function implemented by a single cell would be equivalent to using many transistors, thus simplifying the arrangement of large-scale computations.

1.4.2. Nanotubes

One interesting class of molecular devices is nanotubes, particularly *carbon nanotubes*. Recall that pure carbon has two common crystalline forms: diamond, where carbon atoms form a three-dimensional structure; and graphite, where carbon atoms form flat sheets that can easily slide and peel from each other. A single-wall carbon nanotube (Fig. 1.8) can be visualized as a single sheet of graphite rolled into a tube (though it is not created in this way), with a diameter of only a few nanometers.

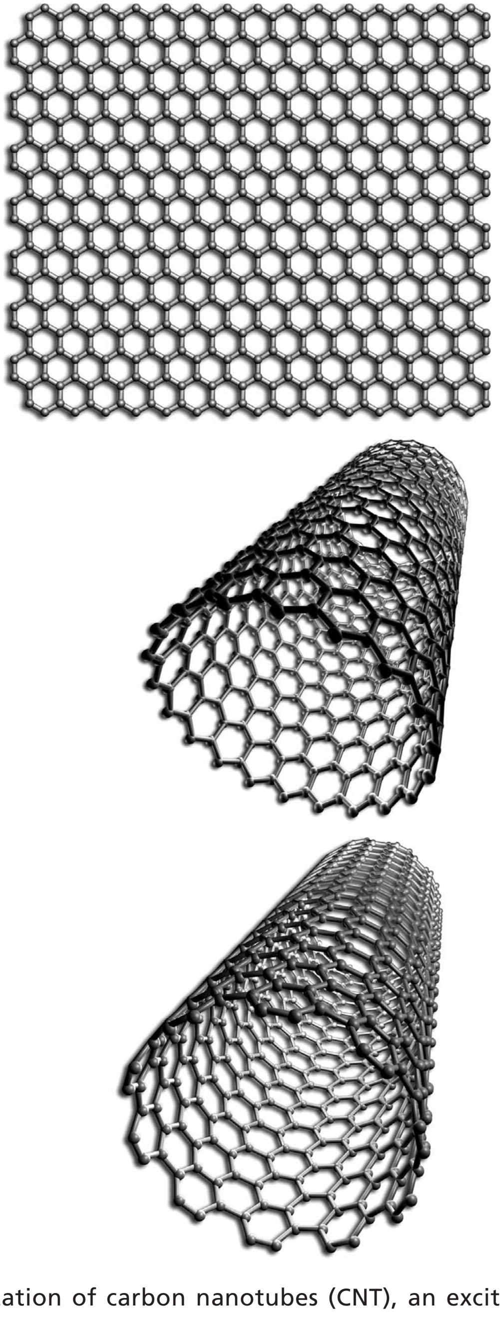

Figure 1.8. Visualization of carbon nanotubes (CNT), an exciting class of molecular devices. CNTs can be visualized as a sheet of graphene (top) rolled into a tube, though they are not created this way. Note that these two nanotubes have slightly different configurations. Various configurations of nanotubes can result in drastically different physical properties. Images created by Gabriele Nateneli.

Carbon nanotubes have very interesting mechanical and electrical properties. Mechanically, they are extremely strong and versatile. The tensile strength and stiffness of carbon nanotubes are extremely high, especially relative to their weight. Also, two tubes arranged with one inside the other can have a very low friction interface, similar to the way two flat sheets of graphite can very easily slide over each other. Therefore, two nanotubes can have very efficient telescoping and rotation motions, useful for nanomachines [14].

Electrically, a nanotube can potentially behave as a *ballistic conductor*. This means that an electron travels through the tube with small, quantized levels of resistance. The typical levels of resistance are much lower than traditional conductors. Nanotubes can exhibit properties of a metal or a semiconductor, depending on how the sheet of graphite is rolled into a tube. It has even been suggested that nanotubes can behave like a *waveguide*, guiding the wave-like properties of an electron similar to the way electromagnetic (optical) waves are guided through a fiber-optic cable [15]. All of these properties are being investigated for future switches and wires. In fact, switches, wires, and support structures have all been demonstrated with carbon nanotubes, but, as with many nanoscale devices, the ability to fabricate a practical nanoscale device with nanotubes and nanowires is still an open challenge. Carbon nanotubes are discussed further in several chapters in this book. See Chapter 2, Chapter 12, and Chapter 18.

1.4.3. Quantum Dots and Tunneling Devices

Many quantum phenomena occur when confining electrons to a very small space, such as the nanoscale range. For example, an electron confined to a small area can only have a select few discrete levels of energy, similar to the discrete levels of energy that an electron may have as part of an atom. When a group of electrons is confined in all axes of movement (i.e., in three dimensions), a *quantum dot* is formed. Similarly, a *quantum wire* is a group of electrons confined along a 1-dimensional line, and a *quantum well* restricts electrons to a 2-dimensional plane. These structures can exhibit properties similar to electrons in atoms or molecules, even if there is no nucleus of protons and neutrons. Their properties can be fine-tuned with more freedom than atoms or molecules, making them very interesting structures to use for computing.

Often, the phenomenon of tunneling, described previously, is combined with quantum dots, wires, and wells to create useful devices. This is in contrast to traditional transistors, where tunneling is very undesirable. Three such nanoscale devices are the *resonant tunneling diode* (RTD), the *single electron transistor* (SET), and *quantum-dot cellular automata* (QCA). An RTD is a device that has a quantum well where electrons can be confined; therefore, electrons in this region can assume only a few possible discrete energy states. When the energy of an incoming electron is close to one of these "resonant" discrete energy states, the electron can tunnel through with high probability. This device can emit extremely high frequencies, in the hundreds of GHz, making it interesting for