This page intentionally left blank
Advances in
COMPUTERS

Nanotechnology

EDITED BY
MARVIN V. ZELKOWITZ

Department of Computer Science
and Institute for Advanced Computer Studies
University of Maryland
College Park, Maryland

VOLUME 71
# Contents

<table>
<thead>
<tr>
<th>Contributors</th>
<th>ix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preface</td>
<td>xiii</td>
</tr>
</tbody>
</table>

## Programming Nanotechnology: Learning from Nature

**Boonserm Kaewkamnerdpong, Peter J. Bentley and Navneet Bhalla**

1. Introduction ............................................................................................................ 2
2. Development in Nanotechnology ................................................................. 4
3. Benefits of Computer Science for Nanotechnology ........................................ 10
4. Swarm Intelligence ............................................................................................. 14
5. Perceptive Particle Swarm Optimisation ....................................................... 19
6. Perceptive Particle Swarm Optimisation for Nanotechnology ....................... 23
7. Self-Assembling Nanotechnology ..................................................................... 26
8. Conclusions ......................................................................................................... 33

References .................................................................................................................. 34

## Nanobiotechnology: An Engineer's Foray into Biology

**Yi Zhao and Xin Zhang**

1. Introduction ............................................................................................................ 40
2. Nanofabrication ..................................................................................................... 42
3. Nanobiotechnologies for Sensing and Actuating ............................................. 74
4. Nanobiotechnology for Drug Delivery and Therapeutics .................................. 91
5. Concluding Remarks ............................................................................................. 92

References .................................................................................................................. 92
Toward Nanometer-Scale Sensing Systems: Natural and Artificial Noses as Models for Ultra-Small, Ultra-Dense Sensing Systems

Brigitte M. Rolfe

1. Introduction .......................................................... 104
2. The Physiology of the Sense of Smell ............................ 108
3. Electronic Noses: Chemical Sensing Systems ................. 126
4. Nanosensors ......................................................... 143
5. Designing a Nanometer-Scale Nose-Like Sensing System ... 151
6. Summary .................................................................. 157
   Acronyms ................................................................. 158
   Acknowledgements .................................................. 159
   References ............................................................... 159

Simulation of Nanoscale Electronic Systems

Umberto Ravaioli

1. Introduction .......................................................... 168
2. Simulation Hierarchy for Semiconductor Devices ............ 173
3. Simulation Issues in Nanoscale Silicon Devices .............. 188
4. Organic Molecular Devices ........................................... 197
5. Simulation of Molecular Conduction ........................... 200
6. Carbon Nanotubes .................................................... 208
7. Ionic Channels ........................................................ 220
8. Conclusions ............................................................ 241
   Acknowledgements .................................................. 242
   References ............................................................... 242

Identifying Nanotechnology in Society

Charles Tahan

1. Introduction .......................................................... 252
2. Definitions ad infinitum (and More Politics) ................... 258
3. Perspectives from Science .......................................... 263
4. Conclusions ............................................................ 267
   Acknowledgements .................................................. 268
The Convergence of Nanotechnology, Policy, and Ethics

Erik Fisher

1. Introduction ........................................................................................................ 274
2. Converging Paths ................................................................................................ 277
3. From Convergence to Collaboration .................................................................. 282
4. Conclusion ........................................................................................................... 290
   References ........................................................................................................... 291

AUTHOR INDEX .................................................................................................... 297
SUBJECT INDEX .................................................................................................... 315
CONTENTS OF VOLUMES IN THIS SERIES ...................................................... 327
This page intentionally left blank
Contributors

Dr. Peter J. Bentley is a Senior Research Fellow at the Department of Computer Science, University College London (UCL), Collaborating Professor at the Korean Advanced Institute for Science and Technology (KAIST) and Visiting Research Fellow of the University of Kent. Peter runs the Digital Biology Interest Group at UCL and is known for his prolific research covering all aspects of Evolutionary Computation and Digital Biology. His research investigates evolutionary algorithms, computational development, artificial immune systems, swarming systems and other complex systems, applied to diverse applications including design, control, novel robotics, nanotechnology, fraud detection, security, art, and music composition. Peter was nominated for the $100,000 Edge of Computation Prize in 2005. He regularly gives plenary speeches at international conferences and is a consultant, convenor, chair and reviewer for workshops, conferences, journals and books in the field of evolutionary computation and complex systems. He has published over 130 scientific papers and is editor of the books “Evolutionary Design by Computers,” “Creative Evolutionary Systems” and “On Growth, Form and Computers,” and author of “The PhD Application Handbook” and “The Book of Numbers.”

Navneet Bhalla achieved a B.Sc. in Computer Science at the University of Ottawa, Canada and an M.Sc. in Intelligent Systems at University College London, winning first prize in a student paper competition. He has worked in the areas of mobile web authoring, new market development, quantum genetic algorithms, and physical self-assembly, with a patent application for the later. Navneet is currently performing a Ph.D. on evolutionary self-assembling systems at the University of Calgary under the supervision of Professor Christian Jacob and Dr. Peter Bentley.

Dr. Erik Fisher is a postdoctoral research associate at the Center for Nanotechnology in Society and the Consortium for Science, Policy, and Outcomes at Arizona State University. He has held faculty appointments at the Center for Science and Technology Policy Research and the Lockheed Martin Engineering Management Program at the University of Colorado at Boulder, where he was an advisor to the
College of Engineering and Applied Science, and where he earned a Ph.D. in Environmental Studies, a certificate in Science and Technology Policy, and an M.A. in Classics. He holds a B.A. from St. John’s College, Annapolis, MD.

Boonserm Kaewkamnerdpong achieved a first class B. Eng in Computer Engineering, King Mongkut’s University of Technology Thonburi (KMUTT), Thailand and an M. Eng in Mechatronics at the Asian Institute of Technology (AIT), Thailand. During this time she won three awards for best academic performance and won first prize in the Software Development Competition organized by National Electrical and Computer Technology Centre (NECTEC), Thailand. Boonserm has published in the areas of Machine Learning, Image and Speech Processing, Wavelet Transform, Multi-agent Systems, Artificial Life and Nanotechnology. She is currently completing her Ph.D. at University College London on the use of particle swarm optimization for Nanotechnology, under the supervision of Dr. Peter Bentley.

Prof. Umberto Ravaioli obtained the Laurea Dr. in Electronics Engineering in 1980 and the Laurea Dr. in Physics in 1982, both from the University of Bologna, Italy. He then received the Ph.D. in Electrical Engineering from Arizona State University in 1986. He then joined the University of Illinois at Urbana-Champaign, where he is now a Professor of Electrical and Computer Engineering and Assistant Dean of Academic Affairs. His current research interests are in the numerical simulation of charge transport in solid-state devices as well as biological and bio-inspired systems, nanotechnology, supercomputation, advanced visualization, and interactive computer approaches for engineering education. He is a Fellow of the IEEE and of the Institute of Physics.

Ms. Brigette Rolfe has a B.A. and M.A. in mathematics from the University of California at Los Angeles, as well as an M.B.A. from Johns Hopkins University. Over the years, she has been engaged in software systems engineering, strategic planning and portfolio management, and served as technical advisor to numerous government program managers. Since February, 2002, she has been Group Leader of the MITRE Nanosystems Group, where she also leads the Nanosensing Task. Ms. Rolfe has won numerous commendations and awards for her work. She is an Invited Speaker for the 2007 IEEE International Conference in Nanotechnology, and she is Guest Editor for an upcoming IEEE Sensors Journal Special Issue on Nanosensor Technology.

Dr. Charles Tahan is a National Science Foundation Distinguished International Postdoctoral Research Fellow located at the Cavendish Laboratory of the University of Cambridge (UK) with extended visits to the Centre for Quantum Computer Technology (Australia) and the University of Tokyo (Japan). His research focuses on the
intersection of quantum physics, nanotechnology, and information theory, including quantum electronics and optics, quantum many-body physics in mesoscopic systems, and quantum information processing devices. While in the physics department of the University of Wisconsin-Madison, he worked with professors in sociology, public affairs, history of science, and engineering to co-develop and teach a course on nanotechnology and its societal implications.

**Prof. Xin Zhang** received the Ph.D. degree in Mechanical Engineering from Hong Kong University of Science and Technology, Hong Kong, in 1998. From 1998 to 2001, she was a Postdoctoral Researcher and then a Research Scientist with Massachusetts Institute of Technology (MIT), Cambridge, MA, where she had worked at the MIT MicroEngine project. She joined the Department of Manufacturing Engineering and the Department of Aerospace and Mechanical Engineering, Boston University, Boston, MA, in November 2001, where she is currently an Associate Professor. She leads the Laboratory for Microsystems Technology (LMST), Boston University, to perform research on MEMS and NEMS. Specifically, she is interested in applying materials science, micro/nanomechanics, and micro/nanomanufacturing technologies to solve various engineering problems that are motivated by practical applications in MEMS/NEMS and emerging bionanotechnologies. Dr. Zhang is the recipient of the Boston University SPRInG Award in 2002, the NSF Faculty CAREER Award in 2003, and the Boston University Technology Development Award in 2004. Her current work has been sponsored by NSF, AFOSR, AFRL, ARL, and a host of industries. She serves in the Technical Program Committee (TPC) for both IEEE International Conference on MEMS and IEEE Solid State Sensors and Actuators Workshop; she has more than 100 refereed technical publications; she is also a member of IEEE and ASME societies.

**Prof. Yi Zhao** received the B.Sc. and M.Sc. degrees from Tsinghua University, China, and Ph.D. degree from Boston University. Presently, he is an assistant professor in the Department of Biomedical Engineering at the Ohio State University. His main research interests include biomedical microelectromechanical systems (BioMEMS) and nanobiotechnology, in particular the development and application of new fabrication technologies for biosensing and bioactuating micro/nanosystems. He has about 30 publications in the field of MEMS. Dr. Zhao is a member of IEEE and ASME.
This volume of *Advances in Computers* is number 71 in the series that began back in 1960. This series presents the ever changing landscape in the continuing evolution of the development of the computer within the field of information processing. Each year three volumes are produced presenting approximately 20 chapters that describe the latest technology in the use of computers today. Volume 71 presents the exciting new field of nanotechnology. Rapidly moving from the realm of science fiction to commercial products, nanotechnology involves the creation of microscopic devices that can provide support at a smaller scale previously unavailable. (Think of the 1966 movie “Fantastic Voyage,” where a miniaturized submarine is injected into a patient to clear a blood clot, come to life.) Miniature computing devices may someday provide help, especially in the health and medical field, which is currently unavailable.

In the first chapter, “Programming Nanotechnology: Learning from Nature” by Boonserm Kaewkamnerdpong, Navneet Bhalla and Peter J. Bentley, the authors focus on defining the nanotechnology landscape. What are the primitive building blocks used to create nanotechnological devices? What are nanomanipulators and what do we mean by nanofabrication? Additionally, since this is a volume concerning computer advances, what do we really mean by a nanocomputer? These and other related concepts is the focus of this chapter.

Yi Zhao and Xin Zhang in Chapter 2, “Nanobiotechnology: An Engineer’s Foray into Biology,” focus on applications of nanotechnology in the biological/medical field. Like the technology presented in Chapter 1, Chapter 2 describes nanobiosensors and nanobiomotors made out of nanomaterials. The use of DNA as drivers of nanoengines is discussed.

“Toward Nanometer-Scale Sensing Systems: Natural and Artificial Noses as Models for Ultra-Small, Ultra Dense Sensing Systems” by Brigitte Rolfe in Chapter 3 describes one particular nanotechnology application. In this chapter, Dr. Rolfe looks at the application of nanotechnology in the development of an artificial nose, “a highly sensitive, selective, reliable, and compact sensing system to detect toxic chemicals and biological agents.” The chapter discusses the biological basis for smell as well as the technological developments for implementing these functions in hardware.

The first 3 chapters discuss nanotechnology and its application in the biological and medical fields. However, how do we design and build such devices? In Chapter 4,
Umberto Ravaioli discusses “Simulation of Nanoscale Electronic Systems.” In this chapter, Prof. Ravaioli discusses the fabrication of nanodevices out of metal-oxide-semiconductor field-effect transistors (MOSFET) and then how one uses simulation in order to test these designs before construction.

In Chapter 5, “Identifying Nanotechnology in Society,” Charles Tahan looks at what we mean by nanotechnology. It has become a catchall term to collect many different technologies under one banner. It binds together computer technology, material science, physics, electronics, manufacturing, biology and medicine under the general rubric of “nanotechnology.” In this chapter, the various aspects of these technologies are explored.

The final chapter, “The Convergence of Nanotechnology, Policy, and Ethics” by Erik Fisher, presents a different point of view in building nanotechnological devices. What are the ethical and societal issues associated with nanotechnology? Will the public accept nanodevices unleashed into the environment? What are the dangers of such devices causing harm? Will medical advances be available to all or will costs be prohibitive and therefore available to only a few? This chapter discusses these and many other ethical issues.

I hope that you find these chapters to be of value to you. I am always looking for new material to present. If you would like to contribute to the Advances in Computers, I can be reached at mvz@cs.umd.edu. Draft chapters are generally due June 30 of each year, with final versions submitted during October, for publication the following spring.

Marvin Zelkowitz
University of Maryland and Fraunhofer Center, Maryland
College Park, Maryland
Programming Nanotechnology: Learning from Nature

BOONSERM KAEWKAMNERDPONG AND
PETER J. BENTLEY

Department of Computer Science
University College London
UK

NAVNEET BHALLA

Department of Computer Science
University of Calgary
Canada

Abstract
For many decades, nanotechnology has been developed with cooperation from researchers in several fields of studies including physics, chemistry, biology, material science, engineering, and computer science. In this chapter, we explore the nanotechnology development community and identify the needs and opportunities of computer science research in nanotechnology. In particular we look at methods for programming future nanotechnology, examining the capabilities offered by simulations and intelligent systems. This chapter is intended to benefit computer scientists who are keen to contribute their works to the field of nanotechnology and also nanotechnologists from other fields by making them aware of the opportunities from computer science. It is hoped that this may lead to the realisation of our visions.

1. Introduction ........................................ 2
2. Development in Nanotechnology ............... 4
   2.1. Nanomanipulators .............................. 4
   2.2. Nanofabrication ................................. 6
   2.3. Nanocomputers ................................ 7
   2.4. Nanorobots .................................... 9
   2.5. Nanomedicine .................................. 9
1. **Introduction**

In 1959, Richard Feynman, a future Nobel Laureate, gave a visionary talk entitled “There’s Plenty of Room at the Bottom”\(^1\) on miniaturisation to nanometre-scales. Later, the work of Drexler\(^2\) also gave futuristic visions of nanotechnology. Feynman and Drexler’s visions inspired many researchers in physics, materials science, chemistry, biology and engineering to become nanotechnologists. Their visions were fundamental: since our ancestors made flint axes, we have been improving our technology to bring convenience into our everyday life. Today a computer can be carried with one hand—40 years ago a computer (hundreds of times slower) was the size of a room. Miniaturisation of microprocessors is currently in process at nanometre-scales\(^3\). Yet, the style of our modern technology is still the same as ancient technology that constructed a refined product from bulk materials. This style is referred to as *bulk* or *top–down technology*\(^1\). As conventional methods to miniaturise the size of transistors in silicon microprocessor chips will soon reach its limit\(^2\) and the modification of today’s top–down technology to produce nanoscale structures is difficult and expensive\(^3\), a new generation of computer components will be required. Feynman and Drexler proposed a new style of technology, which assembles individual atoms or molecules into a refined product\(^1\). This Drexler terms *molecular technology* or *bottom–up technology*\(^1\). This bottom–up technology could be the answer for the computer industry. Though top–down technology currently remains the choice for constructing mass-produced devices, nanotechnologists are having increasing success in developing bottom–up technology\(^3\).

---

\(^1\) For more information, see http://www.zyvex.com/nanotech/feynman.html.

There are some concerns regarding emergent bottom–up technology. First, the laws of physics do not always apply at nanometre-scales [4]. The properties of matter at nanometre-scales are governed by a complex combination of classical physics and quantum mechanics [4]. Nevertheless, bottom–up fabrication methods have been successfully used to make nanotubes and quantum dots [3]. These methods are not yet suitable for building complex electronic devices such as computer processors, not to mention nanoassemblers that can make copies of themselves and work together at a task. Furthermore, and significantly, once knowledge of nanotechnology is advanced and real-world nanoassemblers are realised, they must be properly controllable to prevent any threats to our world.

More recently computer science has become involved in nanotechnology. Such research is wide ranging and includes: software engineering, networking, Internet security, image processing, virtual reality, human–machine interface, artificial intelligence, and intelligent systems. Most work focuses on the development of research tools. For example, computer graphics and image processing have been used in nanomanipulators that provide researchers an interactive system interface to scanning-probe microscopes, which allow us to investigate and manipulate the surface at atomic scales [5,6]. In addition, genetic algorithms have been used as a method in automatic system design for molecular nanotechnology [7].

Computer science offers more opportunities for nanotechnology. Soft Computing techniques such as swarm intelligence, genetic algorithms and cellular automata can enable systems with desirable emergent properties, for example growth, self-repair, and complex networks. Many researchers have successfully applied such techniques to real-world problems including complex control systems in manufacturing plants and air traffic control. With some modifications towards nanotechnology characteristics, these techniques can be applied to control a swarm of a trillion nanoassemblers or nanorobots (once realised). It is anticipated that soft computing methods such as these will overcome concerns about implications of nanotechnology, and prevent the notorious scenario of self-replicating nanorobots multiplying uncontrollably.

This chapter reviews nanotechnology from different points of view in different research areas. We discuss the development of the field at the present time, and examine some concerns regarding the field. We then focus on the needs and benefits of computer science for nanotechnology, as well as existing and future computer science research for nanotechnology. The second half of this chapter introduces the area of swarm intelligence and then summarises investigations into how nanotechnology and self-assembling devices may be controlled by such techniques.

---

3 For more information, see http://www.cs.unc.edu/Research/nano/cismm/nm/index.html.
2. Development in Nanotechnology

To describe Feynman’s grand visions that have inspired many researchers in several fields of study, Drexler\(^5\) introduced the term “Nanotechnology” and “Molecular Engineering” in his book, “Engines of Creation” \([1]\). He explored and characterised an extensive view of Feynman’s visions in many aspects including potential benefits and possible dangers to humanity. According to the vision, building products with atomic precision by bottom–up technology could offer a dramatic widespread of potential and a decrease in environmental impact which would improve our way of life. A simple example of potential benefits from nanotechnology is that information stored on devices could be packed into much smaller spaces so that less pollution from discarding those devices would be produced. The aspect that would be directly beneficial to humankind is nanomedicine, which involves medical research at nanoscale \([1,8]\). For example, a group of programmable nanorobots that could flow along our bloodstreams without harm to our bodies could be injected to treat our bodies from within.

Nanotechnology has indeed promised a great future for humanity. However, the down side of the technology should not be neglected. Drexler suggested the potential threats to life on Earth of uncontrollably replicating assemblers \([1]\). In order to prevent any threat to society, it is crucial that nanotechnology is developed under acceptable standards with regard to ethical and social considerations. Recently, the Foresight Institute, which is a non-profit organisation for nanotechnology, gave version 4.0 of its guidelines as a self-assessment list for research and development in the field of nanotechnology.\(^6\) The Science Media Centre has also produced a document describing nanotechnology for use by the media.\(^7\) Today nanotechnology is gaining public attention. Many companies have been doing research and development in nanotechnology for commercial purposes. The governments of several countries have begun funding for research in this area. This recent development of nanotechnology is described further in the following sections.

2.1 Nanomanipulators

One important concept of nanotechnology is building products using bottom–up technology. Instead of sculpting bulk materials into desired products, bottom–up technology suggests a new method that assembles individual atoms into products. The first step to bottom–up technology is to acquire the ability to manipulate individual atoms at the scale of nanometres as desired. Therefore, the development of a


\(^6\) For more information, see http://www.foresight.org/guidelines/current.html#Principles.

\(^7\) For more information, see http://www.sciencemediacentre.org/nanotechnology.htm.
nanomanipulator, which is a tool for manipulating nanoscopic materials, is seen by some as being crucial to the progress of nanotechnology.

The first imaging in nanoscale was from the electron microscope developed by M. Knoll and E. Ruska in 1931 [9]. Later in 1981, G. Binnig and H. Rohrer invented the scanning tunnelling microscope (STM)\(^8\) that can image individual atoms, and earned the Nobel Prize [9]. The success of the scanning tunnelling microscope leads to the development of other scanning probe microscopes (SPM) including the atomic force microscope (AFM). Instead of using lenses like traditional microscopes, all these scanning probe microscopes use a probe to scan atoms over the surface, measure a local property and result the image. Different types of scanning probe devices are designed for different tasks. For example, the STM is only appropriate when the material conducts electricity, while the AFM can work with non-conducting materials.

Apart from resembling a surface at atomic scale into a high-resolution image, scanning probe microscopes can be used to manipulate individual atoms. In 1990, D.M. Eigler of IBM used an STM to precisely place xenon atoms on a nickel plate into the name “IBM”\(^9\) [10,2]. In 1993, W. Robinett and R.S. Williams developed a virtual reality system that allowed user to see and touch atoms via the scanning tunnelling microscope [9,11]. This was the beginning of a nanomanipulator.\(^10\) At the University of North Carolina, another nanomanipulator has been developed in a multi-disciplinary project involving in the collaboration of several departments including computer science, physics and chemistry.\(^11\) This nanomanipulator is a virtual-reality interface to scanning probe devices. Using technology in computer graphics, the features that are faint in the image can be enhanced. The system allows scientists to investigate and manipulate the surface of materials at the atomic scale. As a result, it has led to new discoveries in biology, material science and engineering.\(^11\) For example, scientists have used the nanomanipulator system to examine the mechanical and electrical properties of carbon nanotubes [6]. Nanomanipulators are now commercially available. However, the ability to manipulate individual atoms alone could not yet enable us to build reliable nanomachines, unless the physical principles at nanoscales are comprehended.

\(^8\) The STM can show surface topography imaging in high resolution by scanning its electrically conducting tip over a surface at the distance of a few atoms. The STM measures the electrical current called the tunnelling current between the tip and each point of the surface. As the distance between the tip and the surface changes, the tunnelling current is altered. The STM adjusts the vertical position of the tip to maintain a constant distance to the surface. These adjustments are recorded as a grid of values and finally transformed into image. For more information, see http://www.almaden.ibm.com/vis/stm/.


\(^11\) For more information, see http://www.cs.unc.edu/Research/nano/cismm/nm/.
2.2 Nanofabrication

After scientists have gained the ability to manipulate individual atoms directly, the next step is to manufacture structures at nanometre scale, i.e., structures smaller than 100 nanometres across. In this section, we discuss nanofabrication methods, which can be divided into two categories: top–down methods and bottom–up methods [3]. Akin to the concept of technology styles discussed previously, the top–down methods involve carving out or adding a small number of molecules to a surface, while the bottom–up methods assemble atoms or molecules into nanostructures.

A top–down method that has been used in the electronics industry is photolithography. Photolithography is the process that transfers the geometric shape on a mask to the surface of a silicon wafer by exposure to UV light through lenses. The computer industry uses this technology to fabricate microprocessor chips [3]. However, the use of photolithography to fabricate nanostructures is limited by the wavelength of the UV light. One modification can be made by using electron-beam lithography, which is a technique for creating fine patterns on a thin polymer film with a beam of electrons [3,12]. Because electron-beam lithography is very expensive and slow, the development of soft lithography, which is a process that creates elastic stamp in order to transfer structures to a surface, allows researchers to reproduce patterns inexpensively in a wide range of materials. Nevertheless, this technique is not yet ideal for manufacturing complex multi-layered structure electronic devices. The need for methods to fabricate complex nanostructures that are simpler and less expensive has stimulated researchers to explore unconventional approaches.

Another top–down method involves using the scanning probe microscopes that were used to manipulate individual atoms to spell IBM. Researchers can manipulate atoms with an STM in three modes: pushing, pulling and sliding. Apart from mechanical manipulation, the STM can be used to assist in fabrication by chemistry catalysing. In 1995, W.T. Muller et al. proposed a method to use scanning probe microscopes in nanofabrication [13]. They used a platinum-coated AFM tip to scan over the surface coated with a monolayer of azide (–N$_3$) compounds. As a result, amino groups are formed by catalytic conversions of azide and can be used to generate more complex structures. Another nanofabrication method using scanning probe devices was introduced by E.S. Snow and P.M. Campbell [14]. Their technique was to add a bias current to the AFM tip and monitor the electrical resistance of the structure during the fabrication process. When the target resistance is reached, the bias is switched off. This innovative feedback mechanism has been modified and used in later research. Recently, F. Rosei et al. proposed a novel nanofabrication method for metal structures [15]. This method uses organic molecules as templates for the re-arrangement of copper atoms on a surface. At low temperature where the copper atoms are static, the template molecules can be moved away without damaging the copper surface by precisely controlling the STM tip. For more information on using
the scanning probe devices in fabrication, a review by S.W. Hla and K.H. Rieder is recommended\textsuperscript{12} [16].

In contrast, bottom–up methods are truly representing a new style of technology. Although the advancement of the bottom–up methods may not yet be suitable for the production of electronic devices or allow us to replace conventional top-up methods in fabrication, researchers can inexpensively assemble atoms and molecules into nanostructures with dimensions between 2 and 10 nanometres by self-assembly chemical reactions. One innovation created with a bottom–up method is a carbon nanotube discovered by S. Iijima of NEC in 1991 [17,9]. A carbon nanotube is a tube-shaped carbon material that is measured in nanometre scales. It became the fifth type of solid-state carbon\textsuperscript{13} after diamond structures, graphite structures, non-crystalline structures and fullerene molecules or buckyballs, which were discovered by R.F. Curl et al. in 1985 [9]. Since then, researchers have been studying the properties and characteristics of carbon nanotubes. Different structures of carbon nanotubes varying in length, thickness, type of spiral and number of layers have been developed for various purposes. Recently, NEC announced the world’s first compact fuel cell for mobile devices that uses spiral-shaped carbon nanotubes or nanohorns for the electrodes.\textsuperscript{14} Carbon nanotubes are expected to be a key material in the future.

Another new material, quantum dots, is made by bottom–up methods. Quantum dots are crystals that emit only one wavelength of light when their electrons are excited. Because the electrical, magnetic and optical properties of the dot are regulated by the size of the dot, the production of quantum dots must maintain their size and composition [3]. The size of the dots can be selected by varying the amount of time for the chemical reaction. The emission of light by quantum dots could be used in medicine as a biological marker [3]. Alternatively, quantum dots could be used as quantum bits and to form the basis of computers.\textsuperscript{15}

2.3 Nanocomputers

In 1965, G. Moore, the co-founder of Intel, predicted a trend that the number of transistors contained in a microprocessor chip would double approximately every 18 months. This became known as Moore’s law. As exemplified in Intel’s chips,\textsuperscript{16} the prediction appears surprisingly correct. However, without the development of nanotechnology researchers will struggle to meet the prediction of Moore’s law.

\textsuperscript{12} From http://www.imm.org/Reports/Rep040.html.
\textsuperscript{13} For more information, see http://www.labs.nec.co.jp/Eng/innovative/E1/02.html.
\textsuperscript{14} For more information, see http://www.smalltimes.com/document_display.cfm?document_id=7563.
\textsuperscript{15} For more information, see http://news.uns.purdue.edu/html4ever/010917.Chang.quantum.html.
\textsuperscript{16} For more information, see http://www.intel.com/research/silicon/mooreslaw.htm.
One of the first achievements in nanocomputer research was perhaps the development of single-electron tunnelling (SET) transistors by D. Averin and K. Likharev in 1985.\(^\text{17}\) Later in 1987, T.A. Fulton and G.J. Dolan at Bell Laboratories fabricated single-electron transistors and made an observation on the quantum properties and effects of electrons when transistors are in operation\(^\text{18}\). As techniques in nanofabrication advances, researchers have successfully created electronic components including transistors, diodes, relays and logic gates from carbon nanotubes\(^\text{19,20}\). The next step is providing the interconnection between components. Researchers have been working on a different type of nanoscale wire called a semiconductor nanowire and studied how to interconnect and integrate the components\(^\text{19}\). The final step to build a computer processor is to fabricate the designed circuit. Recently, the semiconductor industry has successfully built 70-megabit memory chips containing over half billion transistors.\(^\text{18}\) As the advancement in nanofabrication progresses, the silicon-based nanocomputer becomes closer to reality.

Another approach to nanocomputers is DNA computing. Deoxyribonucleic acid (DNA) is a nucleic acid that carries genetic information for the biological development of life. In 1994, L. Adleman introduced the idea of solving a well-known complex mathematical problem, called the travelling salesman problem, by using DNA\(^\text{21}\). His DNA computer showed that DNA could indeed be used to calculate complex mathematics; however, it is not yet comparable to conventional computer in terms of speed and ease of use. Nevertheless, his work has encouraged the development in DNA computing. In 1997, researchers at the University of Rochester built DNA logic gates, another step towards a DNA computer. The fact that a DNA molecule can store more information than any conventional memory chip and that DNA can be used to perform parallel computations make the area very appealing.\(^\text{19}\) Regardless of the success of DNA computers, the development of silicon-based nanocomputers could use the advantages of DNA computing.

Apart from silicon-based nanocomputers and DNA computers, researchers believe that quantum computers may be another promising approach that overcomes the limits of conventional computers\(^\text{22}\). Feynman began one of the first research groups to explore computational devices based on quantum mechanics. In 1982, he demonstrated how computations could be done by quantum systems according to the principles of quantum physics\(^\text{23}\). In quantum computers, the binary data in conventional computers are represented by quantum bits, or qubits, which can be in a state of 0, 1 and superposition (simultaneously both 0 and 1). As a quantum computer can hold multiple states simultaneously, it is argued that it has the potential to

\(^{17}\) For more information, see http://physicsweb.org/articles/world/11/9/7/1.
\(^{19}\) From http://www.news.wisc.edu/view.html?id=3542.
perform a million computations at the same time.\textsuperscript{20} However, quantum computers are based on quantum-mechanical phenomena, which are vulnerable to the effects of noise. A scheme for quantum error correction is required.\textsuperscript{21} Researchers have been working to overcome this obstacle. To date, quantum computing is still in the very early stages.

\section*{2.4 Nanorobots}

One vision of a nanoassembler or nanorobot is a device with robotic arms, motors, sensors and computer to control the behaviour, all at the scale of nanometres. In 1992, the book called “Nanosystem” by Drexler gives an analysis of the feasibility of machine components for such nanorobots \cite{Drexler1992}. However, even to build a molecular motor, researchers have to consider laws of thermodynamics when motors are actually in operation \cite{ThermodynamicsMotor}. Just building a miniature version of an ordinary motor is not adequate. Recently, a controversy arose surrounding Feynman’s vision of nanorobots. In 2003, an open debate through letters between K.E. Drexler and R.E. Smalley (who was awarded a Nobel Prize for the discovery of fullerenes) was presented to public.\textsuperscript{22} Smalley was not convinced that such molecular assemblers envisioned by Drexler are physically possible, while Drexler insists on his previous findings. Certainly, the study of similarly-sized biological machines—organic cells—suggests there may be more effective alternatives to Drexler’s nanorobots. Even if nanorobots can be realised, they will not be available in the near future \cite{Freitas2003}.

\section*{2.5 Nanomedicine}

Nanotechnology promises a great future for medical research including improved medical sensors for diagnostics, augmentation of the immune system with medical nanomachines, rebuilding tissues, and tackling aging. Proponents claim that the application of nanotechnology to medicine, so-called nanomedicine, offers ultimate benefits for human life and society by eliminating all common diseases and all medical suffering.\textsuperscript{23} Eventually, it is argued that nanomedicine would allow the extension of human capabilities.\textsuperscript{23} In 2003, R.A. Freitas Jr. commented that nanometre-scale structures and devices held great promises for the advancement of medicine includ-

\textsuperscript{20} For more information, see http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume4/hogg96a-html/node6.html.

\textsuperscript{21} For more information, see http://www.theory.caltech.edu/~quic/errors.html.

\textsuperscript{22} The details of those letters can be found at http://pubs.acs.org/cen/coverstory/8148/8148counterpoint.html.

\textsuperscript{23} From http://www.foresight.org/Nanomedicine/NanoMedFAQ.html#FAQ19.
In this initial stage of nanomedicine, nanostructured materials are being tested in various potential areas; for example, tagging nanoparticles using quantum dot nanocrystals as biological markers and smart drugs that become active only in specific circumstances. In addition, researchers have found a method to control the size of densely packed DNA structures, one of nature’s efficient ways for transporting gene information. This could improve the efficiency of gene therapy for medical treatment and disease prevention. It is hoped by many that the next stage of nanomedicine, where nanorobots or nanocomputers are fully available, would expand enormously the effectiveness, comfort and speed of future medicine treatments with fewer risks and costs.

3. Benefits of Computer Science for Nanotechnology

Recently, M.C. Roco of the National Nanotechnology Initiative (NNI), an organization officially founded in 2001 to initiate the coordination among agencies of nanometre-scale science and technology in the USA, gave a timeline for nanotechnology to reach commercialisation. For the next twenty years, the NNI has divided the development of nanotechnology into four generations. The first generation, which just ended in 2004, involved the development of passive nanostructures such as coatings, nanoparticles, nanostructured metals, polymers and ceramics. At the time of writing, we begin the second generation, during which we should manufacture active nanostructures including transistors, amplifiers, targeted drugs, actuators and adaptive structures. Later, from the year 2010, nanotechnology should enter the third generation. It is estimated that systems of nanosystems, for example: guided molecular assembling systems, 3D networking and new system architectures for nanosystems, robotics and supramolecular devices, would be developed. Finally, from the year 2020, the fourth generation of nanotechnology should be the generation of molecular nanosystems, which would integrate evolutionary systems to design molecules as devices or components at atomic levels.

To date, nanotechnology has been developed mostly from the basis in physics, chemistry, material science and biology. As nanotechnology is a truly multidisciplinary field, the cooperation between researchers in all related areas is crucial to
the success of nanotechnology. Until now, computer science has taken a role mostly in research tools, for example: a virtual-reality system coupled to scanning probe devices in nanomanipulator project. However, according to M.C. Roco, the third and fourth generation of nanotechnology would rely heavily on research in computer science.

Perhaps reflecting the extensive use of computers in the modern world, computer science is today a broad field, with many aspects that may affect nanotechnology. Earlier sections have outlined the use of graphics and imaging with nanomanipulators. Other current uses of computer science for nanotechnology include developing software systems for design and simulation. A research group at NASA has been developing a software system, called Nanodesign, for investigating fullerene nanotechnology and designing molecular machines. The software architecture of Nanodesign is designed to support and enable their group to develop complex simulated molecular machines.

However, here we focus on intelligent systems. Research in intelligent systems involves the understanding and development of intelligent computing techniques as well as the application of these techniques for real-world tasks, often including problems in other research areas. The techniques in intelligent systems comprise methods or algorithms in artificial intelligence (AI) including knowledge representation/reasoning, machine learning and natural computing or soft computing.

An exciting new development at the time of writing is a project called PACE (programmable artificial cell evolution). This large interdisciplinary project aims to create a “nanoscale artificial protocell able to self-replicate and evolve under controlled conditions.” The protocells in this work are intended to be the “simplest technically feasible elementary living units (artificial cells much simpler than current cells).” These are intended to act as nanorobots, comprising an outer membrane, a metabolism, and peptide-DNA to encode information. Evolutionary modelling is being used extensively in PACE, to analyse real and simulated protocell dynamics, their possible evolution, and the evolution of (potentially noisy) protocellular networks. Evolution is also being used within microfluidic FPGA chips to produce stable self-replicating cell-membranes, with a genetic algorithm using physical populations on the chip and evaluated by a computer vision system. In addition to this work, computer modelling of embryogenesis and developmental systems is becoming increasingly popular in computer science [31]. Should artificial cells become a

29 From http://complex.upf.es/~ricard/PACEsite.html.
30 From http://134.147.93.66/bmcmyp/Data/BIOMIP/Public/bmcmyp/Data/PACE/Public/paceprosheet.html.
reality, such models will provide a method for their genes to be programmed in order
to enable the growth of larger, multicellular forms.

Apart from genetic algorithms and other evolutionary algorithms that have promising potential for a variety of problems (including automatic system design for molecular nanotechnology [7]), another emerging technique is swarm intelligence, which is inspired by the collective intelligence in social animals such as birds, ants, fish and termites. These social animals require no leader. Their collective behaviours emerge from interactions among individuals, in a process known as self-organisation. This collective intelligence in social animals often cannot emerge from direct interaction among individuals. Instead, indirect social interaction (stigmergy) must be employed. Each individual may not be intelligent, but together they perform complex collaborative behaviours. Typical uses of swarm intelligence are to assist the study of human social behaviour by observing other social animals and to solve various optimisation problems [27,28]. There are three main types of swarm intelligence techniques: models of bird flocking, the ant colony optimisation (ACO) algorithm, and the particle swarm optimisation (PSO) algorithm. Different techniques are suitable for different problems.

Although still a young field of computer science, swarm intelligence is becoming established as a significant method for parallel processing and simultaneous control of many simple agents or particles in order to produce a desired emergent outcome. For example, researchers at the Santa Fe Institute developed a multi-agent software platform, called Swarm31 inspired by collaborative intelligence in social insects, for simulating complex adaptive systems. Likewise, BT’s Future Technologies Group developed a software platform known as EOS, for Evolutionary Algorithms (EAs) and ecosystem simulations. The group uses EOS for research into novel EAs and ecosystem models and for rapid development of telecommunication-related applications [32]. Systems such as these will become increasingly important for modelling molecular machine systems.32 They are also being investigated as a solution to provide self-healing, adaptive and autonomous telecommunications networks. Another potential benefit of such techniques for complex adaptive systems in this area would be to control intelligently the manufacture of nanometre-scale devices, where no exact mathematical model of the system exists. Many intelligent systems’ techniques have been successfully applied in control system of various complex applications. Although at nanometre-scale the principles and properties of materials are altered, researchers have attempted to solve other dynamic problems using soft computing techniques and have been developing new techniques to cope with such problems.

31 For more information, see http://www.swarm.org/.
Also inspired by emergent collaborating behaviours of social insects, the *Autonomous Nanotechnology Swarm* (ANTS)\(^{33}\) architecture for space exploration by NASA Goddard Space Flight Center is claimed to be a revolutionary mission architecture. The ANTS architecture distributes autonomous units into swarms and organises them in hierarchy by using the concept of artificial intelligence. Researchers at the center have been developing a framework to realise the autonomous intelligent system by using an Evolvable Neural Interface (ENI). As a result, the interface allows cooperation between higher-level neural system (HLNS) for elementary purpose actions and lower-level neural system (LLNS) for problem solving as required in real-world situations. In the plan, each autonomous unit will be capable of adapting itself for its mission, and the ANTS structures will be based on carbon nanotube components.

In 1996, O. Holland and C. Melhuish investigated the abilities of single and multiple agents on a task with agents under similar circumstances as future nanorobots (minimal sensing, mobility, computation and environment) \([29]\). The task to be solved by the agents in their studies was to learn to move towards a light source by using simple rule-based algorithms. In the case of single agents, the result was efficient, but performance degraded as the amount of noise increased. In the case of multiple agents, the best result was from the algorithm that formed collective behaviours akin to genuine social insects. This investigation showed that emergent collective intelligence from social interactions among agents modelled on social insects could cope with the limited capabilities that would be inevitable in future nanoscale robots.

Recently, B. Kaewkamnerdpong and P.J. Bentley proposed a new swarm algorithm, called the *Perceptive Particle Swarm Optimisation* (PPSO) algorithm \([30]\). The PPSO algorithm is an extension of the conventional PSO algorithm for applications in the physical world. By taking into account both the social interaction among particles, and environmental interaction, the PPSO algorithm simulates the emerging collective intelligence of social insects more closely than the conventional PSO algorithm; hence, the PPSO algorithm would be more appropriate for real-world physical control problems. This is the first particle swarm algorithm to be explicitly designed with nanotechnology in mind. Because each particle in the PPSO algorithm is highly simplified (each able to detect, influence or impact local neighbours in limited ways) and the algorithm is designed for working with a large number of particles, this algorithm would be truly suitable for programming or controlling the agents of nanotechnology (whether nanorobots, nanocomputers or DNA computers), whose abilities are limited, to perform effectively their tasks as envisioned. Further details of this method are provided in the second part of this article.

This is seen as a crucial “missing link” in bottom–up nanotechnology: the control of the nanosized agents. A billion (or trillion) tiny particles, whether complex

\(^{33}\) For more information, see [http://ants.gsfc.nasa.gov/](http://ants.gsfc.nasa.gov/).
molecules or miniature machines, must all cooperate and collaborate in order to produce the desired end result. None will have, individually, sufficient computing power to enable complex programming. Like the growth of crystals, the development of embryos, or the intelligent behaviour of ants, bottom-up nanotechnology must be achieved through collective, emergent behaviours, arising through simple interactions amongst itself and its environment. Computer science, and especially fields of research such as swarm intelligence, will be critical for the future of bottom-up nanotech.

We now examine swarm intelligence more closely and provide some examples of how it may be used for nanotechnology.

4. Swarm Intelligence

Swarm intelligence is inspired by collaborative behaviours in social animals such as birds, ants, fish and termites. Collaborative behaviour among social animals exhibits a remarkable degree of intelligence. These social animals require no leader. Their collaborative behaviours emerge from interactions among individuals. Often the behaviour of flocks, swarms and insect colonies, arises through interactions among the individuals in the group and through interactions with their environment. For example, ant foraging behaviour in many ant species arises by means of attractive pheromone trail [28].

Forager ants lay pheromone trail as they move from a food source to their nest. The other foragers sense and, then, follow the trail to the food source. Pheromone—which is a chemical substance—deposited in the environment serves as an intermediate agent in indirect interactions among individuals. Similarly, termites construct their mound by depositing the pheromone and following the smell of pheromone [27]. Individual termites move towards the direction with strongest pheromone concentration and deposit a mixture of local soil and their saliva. With such simple activities, termites construct their mound to fill with chambers, passages, and ventilation system even though they have no construction plan beforehand. Termites are considered to be some of the greatest architects in the insect (and animal) world [20]. Although each individual in insect colonies may not be intelligent, together they perform complex collaborative behaviours. Swarm intelligence techniques model collective behaviours in social insects.

4.1 Stigmergy and Self-Organisation

Unlike the hierarchical organisation with centralised control in humans (i.e., our brain is in one place and controls everything), social insects have no leader to coor-
dinate other individuals to achieve their tasks; the role of so-called *queen* in insect colonies is merely a reproducer. With simple behaviours like pheromone laying and following, worker termites—which are blind—can build their sophisticated mound. Such indirect form of communication found in social insects is known as *stigmergy*. In 1959, Pierre-Paul Grasse made observations on termite building behaviour and used this term to describe task coordination and construction regulation in termites [28]. Grasse explained that the workers were guided by the construction [42]; individual worker deposits a chunk of material that stimulates the same worker or any other workers nearby to respond and deposit more material [28]. In general, stigmergy describes the indirect communication among individuals through the environment [27]; one individual modifies the environment, and the other individuals respond to the changed state of the environment leading to collaborative behaviours as seen in ants, termites, and other social insects.

In 1990, Deneubourg and his colleagues [43] showed that simulated robots with stigmergy and simple rules could achieve clustering and sorting tasks that are common activities in ants; some species of ants cluster corpses of their nestmates into cemetery and sort their larvae into piles according to size [28]. Without direct communication among robots, ant-like robots comprising with a short-term memory unit perform both tasks comparable to ants [43]. Deneubourg et al., however, note that long memory length prevents effective performance in clustering and sorting. In 1994, Beckers, Holland and Deneubourg [44] conducted similar clustering experiments on physical mobile robots with no memory [42]. The study shows that stigmergy can control and coordinate a number of robots to achieve their tasks and the number of robots is a critical factor to the performance of the system [44]. Whereas a greater number of robots reduces the time to achieve the task, increasing number of robots results in the exponentially increased number of interactions and may lead to the destruction of existing clusters [44].

Experiments on physical robots for sorting tasks were extended in [42] in 1999. Holland and Melhuish explored the effect of increasing number of agents in more details. Both direct and indirect interactions among social insects are required in the underlying mechanism to their collaborative behaviours, which is known as *self-organisation* [28]. Self-organisation is initially used to describe the mechanism of macroscopic patterns emerging from processes and interactions at microscopic level [45]. Likewise, the emergence of collective intelligence at colony level in social insects arising from the interactions among individuals with simple behaviours is from self-organisation as well. Bonabeau et al. [28] describes that self-organisation relies on four fundamental components:

1. **Positive feedback**: In foraging, when the individual comes back from the food source, it can recruit the others to this food source (either by dancing in bees
or pheromone trail in some ant species). The recruitment of the individuals and reinforcement are forms of positive feedback.

2. **Negative feedback**: To stabilise the collective pattern (i.e., suppress the positive feedback), negative feedback may be in the form of saturation, exhaustion, or competition.

3. **Amplification of fluctuations**: These fluctuations are, for instance, errors, random task-switching, and so forth. Randomness can promote the exploration and discovery of new solutions.

4. **Multiple interactions**: Self-organisation relies on multiple interactions including both direct and indirect interactions among individuals. For example, the action of pheromone-following can interact with pheromone-laying action if the density of the pheromone is sufficient. The pheromone substance can, however, evaporate over time; multiple interactions are required to maintain the pheromone density level and, hence, self-organisation.

The collective patterns and behaviours arising from self-organisation may not be completely orchestrated; termites do not know the order of activities they should do or specific location they should deposit soil to construct their mound. Nevertheless, stigmergy provides flexibility and robustness. Social insects can often collectively cope with external perturbation to their systems and exhibit the same collaborative behaviours [28]. Therefore, artificial agents adopting stigmergy and self-organisation can respond to perturbation without reprogramming [28]. Such intelligence can be transformed into a powerful tool in computer science.

### 4.2 Swarm Intelligence Techniques

The term “swarm intelligence” was first used to describe self-organised cellular robotic systems using nearest-neighbour interactions in [46,47]. Bonabeau et al. [28] later extended the definition to include: “any attempt to design algorithms or distributed problem-solving devices inspired by the collective behaviour of social insect colonies and other animal societies.” As the intelligence in social animals emphasises decentralisation, self-organisation, direct/indirect interactions among simple agents, flexibility, and robustness, swarm intelligence techniques are simulated model of such intelligence and typically used to solve optimisation problems [27]. In swarm intelligence, there are two main types of techniques: *ant colony optimisation* (ACO) algorithm and the *particle swarm optimisation* (PSO) algorithm. These techniques are described as follows.
4.2.1 Ant Colony Optimisation

Inspired by foraging behaviour of ants, ant colony optimisation (ACO) algorithms are probabilistic-based computational methods modelling such collective behaviour in ants [48,49]. When foraging, individual ants randomly travel to find food source and lay a certain amount of pheromone on the way back from food source. Ants that sense attractive pheromone follow the trail left by other ants to the food source. When more than one trail is found, the one with stronger pheromone is more preferable and foragers are recruited to that more attractive trail; entomologists have shown that ants probabilistically prefer the path holding high pheromone concentration [50]. With this simple behaviour, ants can find the shortest path from their nest to a food source. Apart from foraging, some ant species cluster their dead to clean the nest in similar way.

The pheromone substance laid by individual ants plays an important role in locally indirect communication among individual ants in the neighbourhood—within the proximity of pheromone trail to detect pheromone concentration. After some time, the most promising path to food source has the greater pheromone concentration. ACO uses this positive feedback mechanism to reinforce the system to good solutions as the increasing amount of pheromone reflects recruitment and reinforcement on the solution.

The persistent pheromone allows ACO to keep good solutions in memory and to find better solutions [28]. As pheromones evaporate over time, the pheromone evaporation serves as negative feedback to avoid premature convergence. The pheromone evaporation and probabilistic randomness in artificial ants allow the ants to explore new paths. To balance between exploitation of the current food source—solution—and exploration of new solution, the rate of pheromone evaporation must be appropriately set. If the virtual pheromone evaporates too quickly, no collective behaviour can emerge; on the other hand, if it evaporates too slowly, the system can yield premature convergence [28].

The ACO framework, or called ACO Meta-Heuristic, apply this collective behaviour to solve combinatorial optimisation problems [28]. The examples of combinatorial optimisation problems—optimising with qualitative variables—include travelling salesman problem (TSP), quadratic assignment problem, and telecommunication network routing. In 1997, Dorigo and Gambardella demonstrated the use of ACO in travelling salesman problem where the shortest path length to visit all cities is required [51]. For TSP problem of 50 cities, ACO yields comparable results to other methods in the literature including simulated annealing, neural network, genetic algorithm, and farthest insertion heuristic; ACO often produces better results than the others [51].

Even though ACO algorithms are powerful optimisation methods and can be applied to both discrete and continuous optimisation problems, they have a limitation
to be employed in physical applications. As pheromone is the essence of ACO meta-heuristic, there must be real pheromone (or other substances that would serve a similar purpose) and an appropriate environment in which pheromone is deposited. For the task of three-dimensional nanorobot coordination control, ACO is only likely to be suitable for very specialised types of application (e.g., laying down of conductive paths between electronic components).

4.2.2 Particle Swarm Optimisation

Self-organisation in bird flocking is one of intriguing phenomena in nature. A large number of birds can flock synchronously, often change direction spontaneously, sometimes scatter and, then, regroup. Bird flocking behaviour has been studied for the underlying mechanisms in their social behaviours [52]. Scientists have developed computer simulations of the movement of social animals like bird flocks and fish schools [53]. One motive in developing such simulations was to model human social behaviour [54]. In 1987, Craig Reynolds developed a model of motion in social animals such as birds and fish [55]. His flocking model applies three simple behaviours to control the movement of simulated creature, called a boid. Each boid observes its neighbours, which are boids locating within the defined distance from itself, and acts according to three behaviours: avoiding collisions with its neighbours, matching velocity with its neighbours, and staying close to its neighbours. Using the combination of simple behaviours, the model shows that group behaviours can arise from interactions among boids within the neighbourhood. This model has been used in a simulation a swarm of bats in the film "Batman Returns" in 1992[^34] and other applications in computer animation and behavioural simulation [56].

Rather than relying on manipulations for optimum distances amongst individuals as in [53], Kennedy and Eberhart simulate human social behaviours according to sociobiologist E.O. Wilson [3] that social sharing of information among individuals offers an evolutionary advantage [54]. Through simulation, they discovered and developed an optimisation method for continuous non-linear functions called particle swarm optimisation (PSO) [54]. The original PSO algorithm resembles swarm intelligence through a very simple concept whose implementation requires inexpensive computation speed and memory requirement as the algorithm uses only primitive mathematic operators [54]. PSO algorithm is a population-based technique exhibiting self-organisation through social interactions and the exchange of information which each individual experiences; a swarm of particles fly around the problem space to find a good solution (position) with the influences from their own experience and their neighbours’ experiences. Referring back to the self-organisation

[^34]: From http://www.red3d.com/cwr/boids/.
principles listed earlier, these experiences (or knowledge) from multiple social interactions according to neighbourhood topology can serve as positive feedback to influence particles to move towards a good position which may be, perhaps, an optimum. Meanwhile, these influences can serve as negative feedback as well; when a better position or optimum is found, social knowledge drives particles to leave the current optimum to pursue the better one. In any case, particles do not move directly towards a good position but rather explore around the good position as the PSO algorithm adds randomness in particle movement. Social interaction among individuals is crucial to the success of the PSO algorithm. In the conventional PSO algorithm, the exchange of information seems to arise from direct communication among particles and no stigmergy is regarded.

Since it was first introduced, the PSO algorithm has been continually modified to improve its performance to solve numerical optimisation problems. The particle swarm optimisation algorithm has been successfully employed to solve a range of optimisation problems including electric power systems [57], music [58], image classification [59], logic circuit design [60], recommender systems [61] and enhancement of other learning algorithms [62].

Using a similar representations to physical agents, the PSO algorithm seems a plausible method for application in physical applications including nanorobot coordination control. The inexpensive requirement in memory and computation suits well with nanosized autonomous agents whose capabilities may be limited by their size. Nevertheless, the conventional PSO algorithm requires complex, direct communication among particles in the neighbourhood which might not be possible in nanorobot. To apply in nanorobot control, a modification of PSO algorithm is required.

5. Perceptive Particle Swarm Optimisation

In particle swarm optimisation, all individuals in the swarm have the same behaviours and characteristics. It is assumed that the information on the position and the performance of particles can be exchanged during social interaction among particles in the neighbourhood. Importantly, conventional particle swarm optimisation relies on social interaction among particles through exchanging detailed information on position and performance. However, in the physical world, this type of complex communication is not always possible. Global communication may be impossible amongst swarm of nanorobots. Indeed, it is common for macro-sized robots to have no idea of their own performance in a given location and thus there may be little direct information that one individual can pass on to its companions.

Insects must cope with similar problems. Termites do not build their mounds by talking to each other and telling each other where to deposit material. Instead, they
perceive each other, and they perceive their environment, and their complex behaviour emerges as a result of those perceptions. There is no concept of communication, only interaction. Social interaction and environmental interaction (stigmergy) enables termites to build highly complex structures without any direct communication.

Recent work by the authors has focused on the use of swarm intelligence for physical nanotechnology applications, where these kinds of severe communication restrictions are common. In order to imitate the physical collective intelligence in social insects, we have proposed the Perceptive Particle Swarm Optimisation (PPSO) algorithm, which adds an extra dimension to the search space and enables both social interaction and environmental interaction by allowing a finite perception range for each individual [27].

The PPSO algorithm is relatively similar to the conventional particle swarm optimisation algorithm. However, instead of operating in $n$-dimensional search spaces for $n$-dimensional optimisation problems, the PPSO algorithm operates in $(n + 1)$-dimensional search space. In effect, the particles fly over a physical fitness landscape, observing its peaks and troughs from afar. Instead of directly exchanging information among particles in their neighbourhoods, each individual has a finite range of perception so that it can observe the search space, which is the environment of the swarm, and perceive the approximate positions of other individuals within its perception range as social insects observe the world and other individuals through senses. Thus, particles in the PPSO algorithm are attracted to the better positions in the search space they perceive and to the neighbours they perceive.

The added dimension represents the underlying performance of particles at their positions in $n$-dimensional space. The exact performance at a specific position in the space is unknown to the particles in the PPSO algorithm. Adding the additional dimension and the ability to observe the search space allows particles to perceive their approximate performance. Consider an $n$-dimensional function optimisation problem. In the conventional particle swarm optimisation, particles fly around the $n$-dimensional search space and search for position giving the greatest performance measured by using the function to optimise. On the other hand, in the PPSO algorithm the particles fly around $(n + 1)$-dimensional space to observe the space and find the optima of the landscape. Because particles can fly “over” discontinuities and noise, the PPSO algorithm finds a good solution to the problem regardless of non-deterministic functions or stochastic conditions. Figure 1 shows particles (red dots) in the conventional PSO algorithm and the PPSO algorithm. In Fig. 1(a) and (b), particles operate in a one-dimensional problem, while Fig. 1(c) and (d) demonstrate particles operating in a two-dimensional problem.

In more detail: particles in the PPSO algorithm observe the search space within their perception ranges by sampling a fixed number of directions to observe and sampling a finite number of points along those directions. Figure 2 shows an example of
FIG. 1. The comparison between the conventional PSO algorithm (a), (c) and the PPSO algorithm (b), (d) in one-dimensional and two-dimensional optimisation problems.
FIG. 2. An example of sampling the observation directions in two-dimensional problem.

a particle observing the landscape in six directions. The particle attempts to observe the search space for the landscape at several sampled distances from its position, in each direction. If the sampled point is within the landscape, the particle perceives the height of the landscape at that point. To be more realistic, the perception radius for observing the search space and other neighbouring particles can be separated into an inner radius and an outer radius. Within the inner perception radius, the particle has excellent perception, while its perception is less reliable in the outer perception range.

In the physical world, some social insects can perceive the presence of other individuals through other senses than those they use to observe the world. To simplify this in PPSO algorithm, particles can observe neighbouring particles in their perception range without sampling along specific directions. If there is any neighbour within the perception range, the particle perceives the approximate positions of neighbouring particles. The performance of each particle in the neighbourhood is unknown to each other. Therefore, each neighbouring particle might be in either a better or worse position than its own position. The particle chooses randomly the neighbouring particles, which will influence the particle to move towards them. The position of the chosen neighbour will be used as the local best position. If there is
more than one neighbour chosen, the \textit{lbest} position is the average position among those neighbours. The presence of the neighbouring particles influences the calculation of the new velocity for the next iteration in the same way as local social interaction, \textit{lbest}, in the conventional particle swarm optimisation \cite{28}. However, the particle will have no memory of the local best position from previous iterations. If the local best position at the current iteration does improve the performance of the particle, it will affect its personal best position in the next iteration because the \textit{pbest} position is the position with maximum fitness value that the particle has ever been.

Apart from parameters in the conventional particle swarm optimisation, the main parameters of the perceptive particle swarm optimisation are: the perception radius, the number of observing directions and the number of points to observe along each observing direction. A larger perception radius allows more social interaction and encourages particles to explore the search space. This is because when there is no neighbouring particle within the perception range, the particle moves around its personal best position. However, the larger perception radius requires more computation time to observe the search space. A greater number of observing directions and a greater number of points to observe along each observing direction require more computation time as well. However, more observing directions allow a greater chance to obtain a good solution and the greater number of points offers more accuracy in observation. Note that the observation directions can be designed so that particles observe the search space at various angles in order to increase the chance that the swarm will find a good solution with acceptable computation time.

The PPSO algorithm is designed for optimisation problems in physical applications, such as a swarm of rescue robots searching for survivors after an earthquake, or micro or nanoscale robots used to construct a desired form, where the conventional PSO algorithm cannot be applied. In \cite{30}, an experimental validation was conducted in two-dimensional function optimisation problem. Despite the limited communication and performance measurements of the particles, the experiment showed comparable results with those from the conventional PSO algorithm \cite{30}.

6. Perceptive Particle Swarm Optimisation for Nanotechnology

Using this model of particle movement and perception, computers can be used to simulate the aggregation of various desired forms. While the PPSO algorithm can be used for function optimisation as described above, a more direct simulation enables the same algorithm to model bottom–up form generation. Instead of simulated particles randomly flocking in a virtual “function optimisation space,” we can make them randomly flock in a virtual “form aggregation space.” From the computer science
FIG. 3. A two-layered membrane is built from swarming nanoparticles over time.

FIG. 4. A checkerboard pattern of two different types of particles.
perspective, when using a swarm of particles to aggregate on a surface, the optimum becomes a series of optimal points (unlike typical optimisation problems). Using different rule sets for the swarming particles it is possible to model arbitrarily complex or simple attraction and repulsion behaviours.

By the addition of simple signals emitted by particles (which in a physical system might be equivalent to chemical gradients, electromagnetic fields or different adhesive properties) particles can be programmed to form in specific groups, patterns or layers. These larger structures emerge as particles are attracted to (and adhere to) a surface in the environment, and as they selectively adhere to each other. The movement of the particles is modelled as a free-floating cloud (i.e., akin to a gas or liquid) of flocking particles. Figure 3 illustrates how swarming particles slowly accumulate on a surface resulting in a two-layer membrane. Figures 4 and 5 show the results when the attraction ruleset is altered—checkerboard or striped patterns can be made to emerge.

The patterns shown in Figs. 4 and 5 are the result of applying constraints at the connectors of particles. For example, for the chess-board pattern all four connectors are to connect to the other type of particles. With such connection constraints,
nanorobots just follow the attraction signal and then attach to the optimal ones with appropriate connector. No further consideration/computation is required.

7. Self-Assembling Nanotechnology

Swarm intelligence is not the only field that may inform future nanotechnology. The development of self-assembling robots has also taught us much.

Self-assembly (the autonomous construction of a device by itself) is a dream of robotics engineers and may be an essential requirement for future nanorobots. A payload of self-assembling components would be easier to transport to hazardous and distant locations compared to complete robots. A device that can self-assemble also has the ability to self-repair or regenerate damaged parts of itself, given replacement components. But the creation of self-assembling devices is a highly challenging problem.

The concept of self-assembling robots has been a popular theme in science fiction for many years. Only recently have robots been developed that display self-assembly characteristics. These robots are examples of netted systems [33], consisting of sensors and controllers that interact and self-assemble through data communication. These robots demonstrate the synthetic realisation of templated self-assembly [34], biological self-assembly [35], and self-reconfiguration [4,5], as examples from the disciplines of modular robotics and swarm robotics. However, such disciplines do not provide a generic methodology to creating self-assembling robots at all scales. This is largely due to scalability issues in relation to their respective methods of communication and assembly between modules or robotic-units.

Here, we refine the term self-assembly and suggest that it should be used to describe processes that can be controlled by an appropriate design of pre-existing components that interact in order to create emergent aggregate forms [33]. This view of self-assembly is used to link the principles self-assembly from nature to previous work in robotics and design.

Applying the principles of self-assembly to robotics has tremendous potential. This is especially true at the micro and nanoscale, where self-assembly is viewed as the only viable means of fabrication [33].

L.S. Penrose and R. Penrose were the first to show a mechanical analogue to natural self-assembly, specifically self-reproduction in the form of templated self-assembly [39]. They created two component shapes, labelled A and B, that connected in either an AB or BA configuration. Multiples of these A and B components were confined to a track in a random ordering, that when shaken, allowed components to move horizontally and interact with one another. By placing either an AB or a
BA seed complex on the track, it would cause neighbouring A and B or B and A components to self-assemble into AB and BA complexes respectively.

This example of templated self-assembly has recently been extended to robotics [34]. In this case, triangular-shaped programmed electromechanical components move randomly in two-dimensions on a cushion of air. When components collide, they communicate and latch and unlatch accordingly. Again, by initially placing a seed complex, free components can self-assemble and construct replicas of the seed complex [34].

In these two examples, templates are used to direct the self-assembly process of decentralised components. In contrast, swarm robotics uses swarm intelligence to direct the self-assembly process of decentralised robotic units, in a form of biological self-assembly. Of the robots produced in this discipline, Swarm-bot has shown successful results in mimicking self-assembling formations of social insects (e.g., the formation of living bridges by *Oecophylla longinoda* worker ants) [35]. Swarm-bot is the collective name to the set of cube-shaped mobile robotic units, named s-bots, which are capable of physically linking together. For example, s-bots can self-assemble into aggregate structures to move across terrain, otherwise not possible by an s-bot solely.

The discipline of modular robotics has produced self-reconfigurable robots using both centralised and decentralised modules [37]. Two of the most successful centralised modular robot implementations to date include PolyBot [36] and MTRAN [37]. These robots posses the ability to self-reconfigure a pre-existing set of modules that are physically connected together, and that move and attach/detach in terms of the degrees of freedom allowed by the components. PolyBot uses cube-shaped modules with one axis of rotation, which are capable of self-reconfiguring into various forms with movement such as in a loop, and in a snake-like and spider-like fashion [36]. MTRAN modules consist of two semi-cylindrical parts connected by a link, with each part being able to rotate 180° about its axis. These modules allow MTRAN to self-reconfigure into forms with one type of crawler and two types of quadruped movement [37].

These robots are all implementations of subsets of self-assembly, in the form of netted systems. In nature, self-assembly is primarily dictated by the design of the components within a system and the environmental conditions they are subjected to, as well as their component and environment physical and chemical properties [8,9]. The following section describes a general framework for self-assembling system, which covers the above mentioned types of self-assembly currently used in practise (templated self-assembly, biological self-assembly, and self-reconfiguration), and the potential to create self-assembling robots in the future, particularly at the nanoscale.
7.1 Framework

For the purposes of creating an artificial self-assembling system, the natural principles of self-assembly can be abstracted to four items:

- Components.
- Environment.
- Assembly protocol.
- Energy.

Components are defined by their properties. Such properties include, but are not limited to, shape, scale, material properties, and communication methods and interaction methods between components and/or their environment.

The environment in which components are subjected to can provide various functionalities, such as a boundary to which components are confined to. The physical and chemical properties of the environmental will also influence the nature in which components interact with one another, as well the way in which components self-assemble.

An assembly protocol defines the methods in which components can self-assemble (e.g., methods of attraction and repulsion). These methods are highly dependent on the scale of the system, as well as the physical and chemical properties of the components and the environment.

In order for the components to self-assemble, the components need to be mobile in their environment. This requires the components to have energy. This can either be available internally or transferred to components, for example, by the environment.

This self-assembly framework should be considered from the viewpoint of specific self-assembling systems. Physical constraints are normal in such systems, as we can observe in nature. A sand dune will only form in specific circumstances; if the wind force is not sufficient, it will not form. However, by continuing to gain a deeper understanding of self-assembly in nature, it can be leveraged for the purposes of design. This of course can be utilised by robotics, and the creation of simple self-assembling mechanical structures (e.g., pivots, joints, and levers), would be a fundamental next step.

7.2 Self-Assembly Illustration

One possible solution to creating simple structures is to utilise the relationship between component shape and an assembly protocol. Here, the relationship is investigated in the context of creating two-dimensional geometric mesoscale self-assembling structures, in a method that could be reduced to nanoscales.
Experiments were conducted to investigate whether a set of two-dimensional components (with concave and/or convex polygon shapes), could self-assemble into a desired shape. The assembly process is initiated by placing components on a tray, which is shaken in parallel to the surface of the tray. In this way energy is transferred to the components in the form of vibration, causing the components to move around and interact with one another; and magnetism is used to enable the components to attract and repel one another.

In this context, a component must have two essential properties; the first being the ability to fit together to form the desired shape and the second being the ability to join selectively to corresponding components or not to conflicting components. To achieve the first point, a set of components must include both concave and convex component shapes. By the components’ shapes being both concave and convex, components are able to create stronger joints, leading to more stable structures overall, and less likely to break apart when colliding with other components or the sidewalls of the tray, compared to if components’ shape were restricted to convex forms only. The second point in this example is achieved by placing a magnet in the interior of a non-magnetic material. The magnets allow components having opposite polarity to attract and assemble together, whereas components having similar polarity will repel each other, and therefore not assemble together. The non-magnetic material is used to determine the polygon form of the components. By not allowing the magnets in the components to join directly together, the components have a higher degree of freedom to move around in the given space and interact with one another. Figure 6 shows the principals behind the design of the components.

The components are placed on a tray, which allows a space in which the component shapes can move around and interact with one another. Movement of the components and their interaction is dictated by two-dimensional rigid body dynamics and magnetism. Figure 7 shows the three stable two-dimensional formations of magnetic discs.

**Fig. 6.** Component design. The solid black circle represents the magnetic disc. The outer circle represents the area of the magnetic field. The irregular pentagon represents the non-magnetic material that defines the shape of the component (the left and bottom of the component are the areas not affected by the force of magnetism).
FIG. 7. The three stable two-dimensional formations of magnetic discs: grid (left); chain (centre); and triangular (right).

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Number of component shapes</th>
<th>Symmetric vs. non-symmetric component shapes</th>
<th>Magnetic formations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Triangle</td>
<td>4</td>
<td>Symmetric</td>
<td>Triangular</td>
</tr>
<tr>
<td>2. Square</td>
<td>4</td>
<td>Non-symmetric</td>
<td>Grid</td>
</tr>
<tr>
<td>3. Parallelogram</td>
<td>6</td>
<td>3 sets of 2 of symmetric shapes</td>
<td>Grid and chain</td>
</tr>
<tr>
<td>4. Irregular octagon</td>
<td>7</td>
<td>1 set of 4 symmetric shapes, and 3 non-symmetric shapes</td>
<td>Chain and triangular</td>
</tr>
<tr>
<td>5. 16-Sided polygon</td>
<td>10</td>
<td>Non-symmetric</td>
<td>Grid, chain, and triangular</td>
</tr>
</tbody>
</table>

To test the validity of this design, five experiments were conducted. Components were constructed out of foam board, magnetic discs, and scotch tape. The tray was constructed out of foam board, pushpins, and general purpose adhesive. Each of the five experiments had a different number of components and different desired final forms. Symmetric and non-symmetric component shapes, along with the three stable two-dimensional magnetic disc formations, were also tested to see if they had an effect on the self-assembly process. Table I summarises the design and purposes of each of the five experiments.

7.2.1 Results

Each of the five experiments were successful in having their set of components self-assemble into their corresponding desired final form. Symmetric systems and systems with a lower number of components were able to self-assemble faster in general, in comparison to non-symmetric systems or systems with a large number of components. Table II shows the results of the five experiments.

These results demonstrate how the relationship between component shape and an assembly protocol can be used to create self-assembling entities of varying form. Although this combination of shape and an assembly protocol (magnetism), does not apply to all scales, it does however suggest that physical (as well as chemical), properties of a system can be leveraged to aide in creating netted systems.
This relationship of component shape and an assembly protocol allowed for a larger set of feasible self-assembling entities (in the context of the experiment setup and design). In particular, this combination allowed for the exploitation of an effective magnetic force (regions of a component in which the effects of magnetism were subjected to neighbouring components), to create closed self-assembled forms. These forms, in contrast to open forms, do not allow for free components to self-assemble to the entity, when it reaches a particular state. Closed self-assembled forms are of particular interest to self-assembling robotics.
Another application of these results is that they demonstrate that physical properties, in this case shape and magnetic attraction/repulsion, can be used as a physical encoding, and as a communication mechanism between components. This concept could also be extended to chemical properties. Physical and chemical properties could be used to replace, simplify, or enhance communication and interaction mechanisms between modules or robotic-units, in self-assembling robots.

Understanding and utilising the principles of self-assembly in nature, could be used for the realisation of nanorobots. At the nanoscale, self-assembly is considered as the only viable means of fabricating entities [33]. At this scale, as well as all others, numerous variables affect the process of self-assembly. Optimisation algorithms can be used to generate the specifications of the components and environment to create self-assembling systems [38].

### 7.3 Evolutionary Computation Model

With its ability to navigate through complex problem spaces, evolutionary computation has proven to be an extremely useful approach to solving optimisation problems. In addition, evolutionary computation can be used as a creative tool. This is most notably seen in its ability to generate novel designs [40]. This duality of evolutionary computation makes it a prime candidate to be incorporated into a process for designing and physically creating self-assembling systems [38].

One embodiment of the framework described earlier (and in [38]) can be described as an eight-step process. These steps include:

1. Define the properties of the desired self-assembling entity.
2. Encapsulate the component design, environment design, and/or construction process (referring to the methodology in which the components and/or environment would physically be created, e.g., using rapid prototyping techniques). This encapsulated information would be encoded into the genotype and phenotype representations of the components and environment.
3. Define the translation process in which the computer generated designs can be used to physically create the components and/or environment (e.g., translating the software representations of the components and/or environment to CAD files).
4. Create software that incorporates a computer model using evolutionary computation to virtually design and test the candidate components and/or environment, to allow for self-assembly of the components into the desired entity.
5. Execute the software to generate the designs of the components and/or environment.
6. Execute the translation process of the computer generated component/and or environment designs, to a form that can be used for physical fabrication.
7. Build the components and/or environment.
8. Place the components in their environment to allow for the components to self-assemble into the desired entity.

In this evolutionary computation model, the notion of a design space (the set of buildable designs of components and/or environment), is crucial. If this space is ill-defined, it will greatly affect the performance of the software, as well as inhibit the creation of the self-assembling system.

The encapsulation of a design space is a complicated task. However, it is of great importance, especially in using this process [38] for creating physical system, such as self-assembling robots. Using this process [38] to create simple self-assembling entities with features of simple mechanical machines (e.g., pivots, joints, and levers), would be an important next step, with benefits from the macroscale to the nanoscale. Preliminary results from an implementation of this embodiment, which used a genetic algorithm to evolve shapes and a simulator to model their interaction are described in [41]. Work on this area is ongoing by the authors.

8. Conclusions

As the development of nanotechnology progresses in several disciplines including physics, chemistry, biology and material science, computer scientists must be aware of their roles and brace themselves for the greater advancement of nanotechnology in the future. This chapter has outlined the development of nanotechnology. It is hoped that this gentle review will benefit computer scientists who are keen to contribute their works to the field of nanotechnology. We also suggested the possible methods that computer science can offer in the task of programming future nanotech, which can benefit other nanotechnologists from other fields by helping them be aware of the opportunities from computer science.

As computer scientists who are interested in the field of nanotechnology, our current work involves building systems that consist of a large number of particles automatically forming into a designed structure. By using the PPSO algorithm to control the swarm of particles, each particle performs lightweight computations and holds only a few values. It is anticipated that models such as these will lead to successful bottom–up nanotechnology systems in the future.

In addition, the principles of self-assembly in nature should be considered when creating self-assembling devices. The general framework presented here (consisting of a set of components, an environment, an assembly protocol, and energy), provides a method for understanding the requirements of a specific self-assembling system.
There can be no doubt that nanotechnology will play a major role in our future technology. In this chapter we have outlined some of the ways in which computer science is assisting this research effort.

REFERENCES

[29] Holland O.E., Melhuish C.R., “Getting the most from the least: Lessons for the nanoscale from minimal mobile agents”, in: *Artificial Life V*, 1996.


This page intentionally left blank
Nanobiotechnology: An Engineer’s Foray into Biology

YI ZHAO

Department of Biomedical Engineering
The Ohio State University
294 Bevis Hall, 1080 Carmack Road
Columbus, OH 43210
USA
zhao.178@osu.edu

XIN ZHANG

Department of Manufacturing Engineering
Boston University
15 Saint Mary’s Street
Boston, MA 02215
USA
xinz@bu.edu

Abstract
This book chapter concentrates on applications of nanotechnologies in the field of biological/medical research. The chapter is presented in two parts. In the first part, a brief overview of state-of-the-art nanofabrication technologies is provided, including bottom–up methods, top–down methods, and replication-based methods. The working principles, advantages and limitations of each technology are presented and discussed. The commonly-used nanomaterials for these nanofabrication technologies are also reviewed. Considering the increasingly rapid growth of this interdisciplinary field, several representative emerging nanofabrication technologies are demonstrated and the future direction is commented. The second part of the chapter focuses on applications of nanotechnologies in typical biomedical research fields, including biosensing, bioactuating, drug delivery and therapeutics. Since the nanodevices and nanosystems are similarly in size with many biological entities and possess a number of unique characteristics different from their micro/meso/macroscale counterparts, they can be used to address many unsolved problems in the biological/medical field. Several
The emerging nanotechnology has turned many of our dreams true by enabling construction of micro/nanodevices. Since the birth of nanotechnology, it has never been a single field technology. It is more preferably called “nanotechnologies,” as refers to a set of methods and approaches in physics and chemistry science, engineering fields, biologic and medical areas. The researchers in different fields usually have different understanding towards this technology, which sometimes causes uneven development towards nanoscale. For example, while engineers and physics scientists race to shrink the size of transistors and MEMS components through nanofabrication to create the next generation of high-performance electronic devices, biologists and life scientists have just begun to employ micropatterning and, to a more limited extent, nanopatterning techniques to build high-throughput detection systems for genomic and proteomic studies [1–3].

1. Introduction

The emerging nanotechnology has turned many of our dreams true by enabling construction of micro/nanodevices. Since the birth of nanotechnology, it has never been a single field technology. It is more preferably called “nanotechnologies,” as refers to a set of methods and approaches in physics and chemistry science, engineering fields, biologic and medical areas. The researchers in different fields usually have different understanding towards this technology, which sometimes causes uneven development towards nanoscale. For example, while engineers and physics scientists race to shrink the size of transistors and MEMS components through nanofabrication to create the next generation of high-performance electronic devices, biologists and life scientists have just begun to employ micropatterning and, to a more limited extent, nanopatterning techniques to build high-throughput detection systems for genomic and proteomic studies [1–3].
The burst of nanotechnologies is believed to be observed at the convergence of different research fields. Many challenges and new directions for research will be posed upon blurring of the boundaries. The most notable breakthroughs are expected at unification of nanotechnology and biotechnology, two promising research fields for the 21st century. Nowadays the study of biologic science has reached down to molecular and DNA level. The interaction of these basic life components is the foundation of various macroscopic behaviors of the living organisms. Since molecular and DNA components are usually nanometers or even smaller in size, current technologies at conventional scale seem insufficient for characterization and analysis. Nanotechnologies kick in right in time. The physics basis for the bio- and nanomarriage is that at molecular scale, both technologies are studying the same thing: the interaction of atoms and molecules by physical or chemical means. From this point of view, recent progress in molecular biology, more or less, takes the advantages of nanotechnologies development, although they may not be entitled “nano.” For example, the markers being widely used to label and detect tumor cells are made of nanoparticles. The DNA structures in cell nuclei are probed by using protein sized quantum dots. The nanotechnologies also borrow ideas of the living organisms to make small devices, such as blood-extraction chip (borrowed from mosquito), and nanostructured adhesive (borrowed from gecko feet pad). Meanwhile, a huge number of unsolved puzzles in biological science provide rationality for nanotechnologies, answering questions like “Do we really need devices fabricated at nanoscale?,” “How small is small?,” and more importantly, “What can be enabled by our tiny devices?” and “What we want our nanodevices to be?”

Nanobiotechnology is the word to describe the merging technology. In this chapter, we tend to interpret this technology from an engineering researcher’s point of view. We first give an outline of state-of-the-art development of nanotechnology, with emphasis on nanofabrication and machining. This section provides the answers to questions like (1) “What can be realized at the nanometer scale?” and (2) “How can we build the nanodevices?” We will also discuss the emerging technologies in this field and the future pathway.

In the second part of this chapter, we focus on biomedical applications where the nanotechnologies are currently being or may potentially be used for. Emphasis is given to biosensors, bioactuators and clinical applications (drug delivery and therapeutics). Due to the fast development and the broadness of nanobiotechnology, it is impossible for us to include every exciting aspect of nanobiotechnology in this chapter. Our expectation is to enlighten the readers’ interest to nanobiotechnology and give them a basic review of current research progress in this field, by highlighting some hot spots. We also expect that this chapter may help scientists, engineering researchers and biologists to find ways over the bridge from one side (biotechnology or nanotechnology) to the other.
2. Nanofabrication

People have been struggling thousands of years to make tools. The currently used machining technologies in the conventional scale such as lathing, milling, drilling, cutting and welding, just to name a few, all have been used for a rather long period of time. With the birth of computer about half century ago, there is need to machine smaller ever features: microscale circuits. Since the size of these features goes beyond the capacities of most machining tools at that time, the conventional tools have to be revised to meet the requirement at microscale. The revised tools and machining methods were given the name “microfabrication technology.” Although a semiconductor cleanroom looks far different from a conventional machine shop, the manufacturing methods being used in both places follow just the same path. If one look closely into the microfabrication technologies, he will find nothing really “new” than the conventional machining tools, except for the scaling effects. All microfabrication technologies can be found their counterparts in the conventional scale, which have been elaborated in the machining textbooks published one hundred years ago. This adaptation is so quite successful that people did it once again when there comes nanotechnology. Therefore, many nanofabrication technologies are comparable and can be easily understood by comparison with their conventional scale counterparts. This comparison may help us to track the path of technology development and unveil the seemly mysterious nanofabrication technologies.

2.1 Top–Down Methods

2.1.1 Counterparts in Conventional Scale

“Top–Down” method refers to a set of fabrication technologies which fabricate by removing certain parts from a bulk material substrate. The removing methods can be mechanical, chemical, electrochemical, and etc., depending on the material of the base substrate and requirement of the feature sizes. The formed structures usually share the same material with the base substrate.

There are a couple of manufacturing technologies in the conventional scale which can be categorized “Top–Down.” Milling is a representative example. In the milling process, material is selectively removed from the substrate, usually a metal sheet, forming a cavity with certain geometries. The dimensions of the cavity depend on the travel path of the mill, which can be precisely controlled with the help of computer-assisted numerical systems. The milling technique, along with similar methods such as drilling and grinding, is the most widely used technique in conventional manufacturing industry. People have attempted to extend “Top–Down” method into nanometer domain, and supplemented the mechanical removing methods with chemical and
electrochemical methods. The details of these “Top–Down” techniques are elaborated as below.

2.1.2 Electron Beam Nanolithography

Electron Beam Lithography (EBL) refers to a lithographic process that uses a focused beam of electrons to form the patterns. In contrast to optical lithography which uses light for the same purpose, electron lithography used electrons with higher patterning resolution due to the shorter wavelength. The operation voltage is about 10–50 keV.

Fabrication accuracy and process latitude of EBL can be addressed by computation of the energy-intensity distribution in the resist. The electrons from exposure of an adjacent feature spill over into the exposure of the currently written feature, resulting in the blurred boundary and the reduced contrast. Extensive numerical analysis has been carried out to investigate the resist shape after exposure and development [4–6]. The feature size depends on the sensitivities of the e-beam resist, as shown in Table I and [7]. Among these e-beam resists, PMMA (Poly(methyl methacrylate)) is the most popular one, offering extremely high-resolution, ease of handling, excellent film characteristics, and wide process latitude. Many of current nanostructures use PMMA as the sacrificial layer to construct complex suspending structures, as shown in Fig. 1 and [8].

Dedicated e-beam writing systems are often not affordable by most laboratory researchers. Therefore, many currently used EBL systems are built based on a com-

<table>
<thead>
<tr>
<th>Resist</th>
<th>Tone</th>
<th>Resolution</th>
<th>Contrast</th>
<th>Etch resistance</th>
<th>Thickness</th>
<th>Sensitive to white light</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMMA (Poly(methyl methacrylate))</td>
<td>Positive</td>
<td>Very high</td>
<td>Low</td>
<td>Poor</td>
<td>Many dilutions</td>
<td>No</td>
</tr>
<tr>
<td>PMMA-MMA Copolymer</td>
<td>Positive</td>
<td>Low</td>
<td>Low</td>
<td>Poor</td>
<td>Many dilutions</td>
<td>No</td>
</tr>
<tr>
<td>NEB-31</td>
<td>Negative</td>
<td>Very high</td>
<td>High</td>
<td>Good</td>
<td>Several dilutions</td>
<td>Yes</td>
</tr>
<tr>
<td>EBR-9</td>
<td>Positive</td>
<td>Low</td>
<td>Low</td>
<td>Poor</td>
<td>Single dilution</td>
<td>No</td>
</tr>
<tr>
<td>ZEP</td>
<td>Positive</td>
<td>Very high</td>
<td>High</td>
<td>Good</td>
<td>Several dilutions</td>
<td>Yes</td>
</tr>
<tr>
<td>UV-5</td>
<td>Positive</td>
<td>High</td>
<td>High</td>
<td>Good</td>
<td>Several dilutions</td>
<td>Yes</td>
</tr>
</tbody>
</table>
FIG. 1. SEM images of the finished suspended nanostructures after removing PMMA as a sacrificial layer.

mercial scanning electron microscope. With such a system, writing features as small as 20 nm is possible.

The most notable advantage of EBL is the “maskless” nature. The pattern can be designed and revised as often as required using CAD software. Since EBL can break the resolution limit set by the optical diffraction, it is currently used to create the extremely expensive masks used in Deep-UV lithography. Nevertheless, the great potential of EBL is hindered by its weak performance in high-throughput. Since EBL is a serial process in which the electron beam exposes only a very small region (on the order of nanometer) at each moment, it is very time-consuming to pattern a large area. People have attempted to compliment by using multiple electron beam approaches [9] or electron beam cell-projection systems [10]. It is expected when the high resolution and good throughput are integrated, EBL can be extended to a wide variety of applications, such as fabricating direct-etched nanotransistors array.

2.1.3 Focused Ion Beam (FIB) Nanolithography

Besides electron beam, ion beam focused to nanometer diameters is another radical departure from conventional fabrication routine, by presenting the dose of ions as a function of lateral position on the substrate. Similar as EBL, FIB lithography is also a serial process, writing features point by point. The resist exposure sensitivity of FIB lithography is two or more orders of magnitude higher than that of EBL. Moreover, the ion scattering in the resist and low backscattering from the substrate are negligible. These advantages make FIB lithography a feasible technology in production of nanoscale solid state devices.
FIB nanolithography methods have been extensively studied in general for nanofabrication, including the FIB direct milling, resist patterning with light ions, and the dry development of FIB implanted resists. Conventional FIB lithography is pretty similar with EBL. The resist is patterned by using ions, followed by wet development. The resolution of resist patterns exposed by ion beams is determined only by the lateral extent of the ion straggling distance, which is generally much smaller than that of scattered electrons at similar keV energies. The features are transferred to the base substrate by performing dry reactive ion etching (RIE) process. For high aspect ratio structures, bilayer resist may be used [11].

The focused ion beam cannot only pattern the resist layer, but also affect the surface of the substrate. The incident ions lose energy and imbed themselves into the substrate, deforming the atoms lattice by collision. The ions also remove atoms from the surface and induce chemical effect due to the ion–electron interaction. This makes FIB a feasible method to fabricate silicon MOSFET, bipolar devices and resistor structures [12–14], and to implant boron and arsenic [15,16]. Specifically, the focused ion beam milling is advantageous in fabrication of high aspect ratio features. For example, diamond atomic force microscope (AFM) tip with aspect ratio 7 : 1 was demonstrated with a tip radius of typically 30 nm at the apex, as shown in Fig. 2 and [17].

2.1.4 Precision Machining

As miniaturization requirement continues to seek efficient space utilization with more efficient and better quality products, conventional machining technology itself and its variants will become still important in the future. Precision machining specifically refers to machining of structures with dimensions on the order of or less than a couple of hundred µm. It also means machining components that cannot be achieved directly by using conventional techniques [18]. These technologies share
the common features that they extend the conventional scale manufacturing down to micro/nanoscale, while keeping or making minimal modification of the machining schemes and tools.

The most well-known precision manufacturing include electrochemical machining (ECM) [19], electrodischarge machining (EDM) [20], laser beam machining (LBM) [21], and ultrasonic machining [22].

EDM is a good example to show how to adapt these conventional machining techniques in small scale. EDM works by eroding materials in the path of electrical discharges that form an arc between an electrode tool and the work piece. It has been used to cut small or odd-shaped angles, intricate contours or cavities in extremely hard steel and exotic metals. The technology developments in new computer numeric control (CNC) systems and advanced spark generators have helped to improve machined surface quality. Also, the very small process forces and good repeatability of the process results have made micro-EDM the best means for achieving high-aspect-ratio microfeatures. Recently, researchers at University of Arkansas developed a method to precisely carve arrays of tiny holes only 10 nm wide into sheets of gold, by applying electric current through a thin film of oil molecules (Fig. 3 and [23]).

Another possible option needs be noted is to fabricate micro/nanostructures directly with conventional scale manufacturing tools—such as turning, milling or drilling [24,25]. This technology imposes high demands on the machine behavior that currently met by only few ultra precision machine tools. A broad range of materials such as non-ferrous metals or plastics can be machined (shown in Fig. 4 and [25]), producing real three-dimensional structures with an optimal surface quality. It is expected that fabrication of the nanoscale structures will become possible via this method with the continuous developments of the machining tools.

Currently, most technologies in precision manufacturing are mainly confined in micrometer scale. Nevertheless, continuous developments of this field will result in practical solution for fabricating nanostructures for various applications.
2.2 Bottom–Up Methods

2.2.1 Counterparts in Conventional Scale

As the opposite to “Top–Down” fabrication technologies, “Bottom–Up” methods refer to a set of technologies which fabricate by stacking materials on top of a base substrate. These methods are similar in principle to welding and riveting at the conventional scale, in which a different type of material is attached to the base component by melted solder or physical fitting. In welding and riveting, at-
tention is mainly paid to the strength of the contact area in order to maintain the construct as a reliable component for high load application. Similarly, in “Bottom–Up” nanofabrication, the adhesion of the surface layer to the base substrate is also an important concern. There is extensive research on the surfactants to enhance adherence and avoid cracks during the subsequent processing. Research has also focused on autonomous patterning of the surface layer into nanometer scale features since manipulation of nanoscale components is not ever an easy task as compared to that at the conventional scale.

2.2.2 Self-Assembly

Self-assembly is the assembly of molecules without guidance or management from an outside source. Generally speaking, it refers to intermolecular self-assembly, which aggregates individual molecules with desired patterns to form molecular assemblies. The assemblies are normally built on top of a solid surface and patterned in desired structures with nanometer precision. Nanofabrication methods based on self-assembly provide a new strategy for the successful fabrication of next-generation devices. They offer the promise of nanostructured fabrication of devices for novel electronic applications such as quantum computing, sensing, and integration with biotechnology. The current nanofabrication research based on self-assembly focuses extensively on:

1. Self-assembled monolayer (chemical lithography based).
2. Nanophase separation of polymer blends or block copolymers.

Since various configurations and surface chemistry issues of each domain are well beyond the scope of this chapter, we select representative examples for each domain that reflect the latest technical advances, to demonstrate use of self-assembly as a feasible engineering approach to facilitate “Bottom–Up” nanofabrication.

2.2.2.1 Self-Assembled Monolayer. Self-assembled monolayer (SAMs) are homogeneous, highly ordered films of organic molecules covalently anchored to a surface. The typical thickness is about 1–2 nm and the intermolecular spacing about 0.5–1.0 nm [26]. SAMs can be prepared by spontaneous organization of different types of surfactant molecules on different substrates with geometric repeat symmetry at the molecular level. The most widely investigated SAMs are the organic interfaces of the alkanethiols and disulfides on gold surfaces. The properties of the film are largely controlled by the end groups of the molecules comprising the film [27]. In order to pattern desired feature on SAMs with nanometer precision, many conventional and non-conventional techniques have been used. For example, finely focused ion beam lithography (FFIB) was used to generate sub-
micrometer dot arrays [28]; electron beam lithography (EBL) was also used due to its capability of patterning very fine feature (less than 1 nm). Another option is to use AFM. The most famous technology utilizing AFM is Dip-pen nanolithography (DPN), a scanning probe nanopatterning technique in which an AFM tip is used to deliver molecules to a surface via a solvent meniscus, which naturally forms in the ambient atmosphere, as shown in Fig. 5 and [29–31]. This technique offers high-resolution patterning capabilities for a number of molecular and proteins on a variety of substrates, such as metals, semiconductors, and monolayer functionalized surfaces.

2.2.2.2 Nanophase Separation. Nanometer scale structures can also be prepared by mixing different polymer materials. In nanophase separation, the block copolymer consists of at least two chemically different polymer chains end-tethered together. Since the entropy of the copolymer is rather small due to chain connectivity [32], it is difficult to form thermodynamically stable polymer blends. Instead, the homogeneous morphology is changed by phase separation, where several blend components spontaneously demix from their initially homogeneous state to morphology heterogeneity. The thermally stable heterogeneous structure depends on molecular weight, segment size, and the strength of interaction between the blocks. Previous researches revealed the morphology is a function of the Flory–Huggins interaction parameter $\chi$, molecular weight and the composition of the copolymer, such as the volume fraction of the constituent blocks. For a simplest case of block copolymers, AB diblocks (e.g., polystyrene-polysoprene (PS-PI)), the structures in Fig. 6 and [33], are known to be stable at: sphere (body-centered cubic), hexagonal-packed cylinder, bicontinuous cubic, and lamellar. These nanostructures in block copoly-
Thermodynamic equilibrium morphologies in AB diblock copolymers depending on the volume fraction of component A. BCC: spheres, arranged on a body centered cubic lattice; HEX: cylinders, arranged on a hexagonal lattice; GYR: gyroid, biscontinuous phase; LAM: lamellar structure.

A well control external field (e.g., electric field, concentration, temperature, or elastic field) may also regulate the orientation of the structure and macroscopically align the phase-separation morphology of the block copolymers. A typical periodicity is in the range of 10–200 nm [36].

The interactions between the block copolymer and the surface, the confinement geometry, and the scale play dominant roles during the molecular self-assembly. The external interface may selectively attract one of the constituent blocks. The use of the surface field in generation of stable surface patterns has been validated theoretically [37–40].

Mechanical field is also commonly used to the alignment of block copolymer microdomains [41–43]. Keller et al. [44] were the first to report on the molecular dynamics in block copolymers subjected to external mechanical fields. Due to flow alignment, orientation is induced in ordered block copolymers with an initially random distribution of grain orientations. Shear flow is capable of inducing different orientation states in the same material, depending on alignment conditions (temperature, oscillatory shear flow frequency, strain, shear rate, etc.). Macroscopic alignment has been achieved in materials subjected to extrusion, oscillatory [45] and steady shear [46]. Most recently, effect of the elastic field containing arbitrary three-dimensional external loading is investigated with the help of a surface stain field (Fig. 7) [47,48], which provides a method of strain field design to make various monolayer patterns for nanofabrication.

Electric field is another feasible tool for aligning the phase-separation morphology of the block copolymer. Despite that the aligning force of electric fields is relatively weaker than that of the mechanical field, it offers the advantage of local alignment control by application of spatially specific electric fields. Let us consider a simplest case by filling a block copolymer in the gap between parallel, planar electrodes. The electrodes are held at a constant potential by a voltage source. In response to the electric field, the materials with high dielectric constant are attracted to regions of high field strength, and vice versa. As a result, the microdomains tend to orient parallel
FIG. 7. The results of the kinetic simulation for a binary monolayer on a solid surface. Five cases have different degrees of surface stress anisotropy.

to the electric field vector thereby lowering the free energy of the system. Because of the high melt viscosities, temperatures close to the decomposition temperature and high electric field strengths are required to achieve high degrees of orientation [49–51].

Besides above technologies, there are some other self-assembly methods for creating nanometer scale structures, which cannot be explained by any existing models such as phase separation or chemical lithography. Lithographically induced self-assembly (LISA) is a notable one among these. In this technology, a mask with protruding patterns is placed in the vicinity above the flat surface coated with a thin homopolymer film. After certain cool-and-heat cycles, periodic polymer pillars array form by rising against the gravitational force and the surface tension. The area of the array is confined by the geometry of the protruding portion (Fig. 8 and [52]). This method, along with other mask-based self-assembly techniques, self-assembles structures with sizes smaller than those of the mask patterns. These methods have great potentials in simplifying the self-assembly based nanofabrication and initiate development of three-dimensional self-assembled structures.
2.2.3 Atom Lithography

Because nanotechnology is concerned with the construction of objects and devices a few nanometer in size, it is dependent on the control of matter on the near-atomic scale. Atom lithography has become a subject of investigation in recent years. This new technique consists of two steps. First is the use of laser cooling methods for the high collimation of the atomic beam. And second is the focusing of atoms through a laser beam in standing wave configuration (light mask). These two steps produce an ordered pattern with a precise spacing determined by the laser wavelength. Atom lithography method is based on the concept that even feeble forces such as the dipole force of laser light acting on an induced atomic dipole moment, and magnetic forces on paramagnetic atoms, can be used for precise control of the motion of atoms in beams. Using atom lithography, researchers have successfully produced arrays of lines and dots using few different atomic species. Moreover, atom lithography of more technologically relevant materials, like Gallium or Indium, that are among the key building blocks of modern semiconductor devices, have been demonstrated. The future development of atom lithography will be coupled with standard deposition methods (e.g., Molecular Beam Epitaxy), with the ultimate goal of industrial-scale fabrication of regularly ordered nanosized structures. The basic principle and general configuration of atom lithography are elaborated below.

2.2.3.1 Laser Cooling and Trapping. Laser cooling is the physical basis of atom lithography. It refers to a technology using light to “cool” atoms to ultra low temperatures, or in other words, using light to reduce the velocities (kinetic en-
ergies) of atoms. In this technology, the frequency of the light is tuned slightly below an electronic transition in the atoms (red-detuned). Since the light frequency does not substantially overlap with the resonant frequencies of the atoms, a stationary atom is transparent to the vast majority of the photons, and thus unaffected by the photons. The atoms, however, are usually not motionless. When an atom moves towards the light source, the light frequency may have overlapping with the resonant frequency of the atom due to Doppler Effect. Consequently, there is a high chance for an incoming photon to “resonate” with the atom and be absorbed. The momentum of the atom is thus reduced by absorption. After a brief period of time, the atom emits an identical photon in some random, unpredictable direction. The interaction with the incoming photons reduces the traveling velocity of the atom, or in another sense, lowers the temperature of the atom. This is where the name of “laser cooling” comes from. This technology can be used to trap atoms spatially with certain velocity ranges. Atoms are trapped to the light intensity maximum for red-detuned light and are repelled from the high intensity region for blue-detuned light.

The concept of laser cooling was first proposed in 1970s [53,54]. This technology has since been extensive investigated and gives rise to a new field of research referred to as “atom optics.” The Nobel Prize in Physics 1997 was granted to three scientists working on laser cooling because of its great potential in study of fundamental phenomena and measurement of important physical quantities with unprecedented precision [55].

2.2.3.2 One-Dimensional Standing-wave Focusing. The atom optical tools utilizing laser cooling can be used to produce patterns of images generated by atomic beam incident on suitable surfaces. The manufacturing techniques in this way are called atomic nanofabrication (ANF). Here, we use a simple example of standing optical wave focusing to demonstrate the principle of ANF.

Fig. 9 shows the schematics of laser-focused atomic deposition. The standing wave is generated by a pair of counter-propagating laser beam [56]. The inhomogeneous optical field serves as an immaterial focusing lens, which interacts with the atomic dipole moment and generates optical dipole forces towards the incident atoms. The lenses focus appropriately directs the atoms of an incident beam toward the optical nodes or antinodes, depending on the sign of the difference between the atomic and laser frequencies (blue detune or red-detuned). The atomic pattern generated on the surface closely resembles stripe pattern of the intensity distribution of the light mask, having a period of $\lambda/2$, with $\lambda$ denoting the wavelength (Fig. 10 and [57]).

2.2.3.3 Complex Patterns. Besides one-dimensional features, more complex structures can be made possible by creating two-dimensional intensity distrib-
FIG. 9. The standing wave is formed by reflection from the mirror of a laser beam of wavelength $\lambda$. The atomic beam is focused into lines with period of $\lambda/2$.

FIG. 10. Atomic force microscope images of laser-focused chromium nanostructures. (a) Three-dimensionally rendered view showing some of the narrowest features created in Si; (b) long-range plan view of 60-nm high features on sapphire.

The solution of multiple laser beams. McClelland et al. passed Chromium atoms through interference configuration of three laser beams intersecting with each other at 120° [58]. The incident Chromium atoms form islands with hexagonal symmetry on the surface by using red-detuned laser (Fig. 11(a) and (b)), and form honeycomb
NANOBIOTECHNOLOGY 55

(a) (b)

Fig. 11. Two-dimensional nanostructures by atom lithography. (a) and (b): Four laser beams intersecting at 90° form quadratic nanofeature array. (c) and (d): Three laser beams intersecting at 120° form hexagonal islands.

lattice structures using blue-detuned laser. Similarly, the quadratic feature array [59] can be formed by two-dimensional standing wave generated by four laser beams interfered at 90° (Fig. 11(c) and (d) [60]). More complex structures are made possible by various configurations of multiple laser beams.

Attempts have also been done to reduce the feature size. The nanostructures with pitch less than λ/2 were patterned by using standing wave with polarization gradients. Gupta et al. constructed parallel chromium lines with λ/8 period using one-dimensional standing wave in “lin-perp-lin” configuration [61]. Brezger et al. used a two-dimensional hexagonal light mask to generate a chromium pattern with λ/3 feature spacing [62].

Furthermore, even more arbitrary structures can be constructed using atom holography. Fujita et al. constructed patterns of different letters by passing a beam of ultra-cold metastable neon atoms through a computer-generated hologram that encodes the Fourier transform of a desired atomic pattern (Fig. 12 and [63]). The diffraction of the atomic beam by the hologram reconstructs the pattern, in a manner analogous to optical holography.
2.2.4 AFM Direct Writing

“Bottom–Up” methods also include direct writing by manipulating tip-surface interactions using probe microscopes, usually atomic force microscopy (AFM) or scanning tunneling microscopy (STM). A common feature of these nanofabrication technologies is using atomically sharp tips to transfer and deposit atoms or molecules onto dissimilar substrates. Advantages may accrue from creating nanostructures by reshaping an initially flat substrate surface. For example, nanostructures made without interfaces are not as susceptible to thermal stress. Also, modifying, erasing and regenerating nanostructures, important to data storage applications, may be more easily accomplished using a single substrate material.

AFM anodic oxidation lithography is representative example. With the help of the voltage bias between the tip and the work surface, not only geometrical properties of the surface but also the local electrophysical properties of the sample surface can be changed. The local surface under the probe tip can be oxidized by application of voltage to the conductive cantilever (Fig. 13 and [64]).

The AFM tip scanning in the contact mode can produce nanometer-sized parallel ridge structures in polymer surfaces, perpendicular to the scan direction of an AFM tip [65–68]. A simple fracture-peeling mechanism was proposed to model tip-induced pattern formation in polymers such as polyacetylene films [68]. The fracture-peeling model assumes that the stress produced by the scanning tip in the polymer is released through a crack propagation, which is controlled by an energy balance. The scanning tip peels the surface when the applied force exceeds a critical
resistance force, which is determined by the polymer properties and geometry of the opening crack. When the shearing force becomes smaller than the resistance force produced by accumulation of the displaced polymer mass, the tip no longer peels the surface, but slides over the accumulated polymer in front of the tip. Fig. 14 shows the nanobundles formed by scanning of an AFM tip.

2.2.5 Growth of Carbon Nanotubes

Growth of nanotubes and nanowires shows yet another type of “Bottom–Up” method. Here we take Carbon Nanotubes (CNT) as an example to demonstrate this method. Tubular carbon structures (shown in Fig. 15) were first discovered in
With remarkable mechanical, electrical, chemical and thermal properties, they have since been used in various applications, including conductive and high-strength composites; energy storage and energy conversion devices; sensors; field emission displays and radiation sources; hydrogen storage media; and nanometer-sized semiconductor devices, probes, and interconnects.

There are two main types of nanotubes available today. Single-walled nanotubes (SWNT) consist of a single sheet of graphene rolled seamlessly to form a cylinder with diameter of the order of 1 nm and length of up to centimeters [70,71]. Multi-walled nanotubes (MWNT) consist of an array of such cylinders formed concentrically and separated by 0.35 nm, similar to the basal plane separation in graphite [69]. There are a variety of ways to fabricate carbon nanotubes. The major synthesis methods include arc-discharge [70], laser-ablation [72], chemical vapor deposition [73] and ball milling [74].

The arc-discharge method creates carbon nanotubes through arc-vaporization of two carbon (graphite) rods placed end to end (Fig. 16). As the anode move towards the cathode until they are less than 1 mm apart, a current of about 100 A passes through the gap filling with inert gas (such as helium or argon) and a plasma is created between the electrodes. The temperature of this plasma typically reaches temperatures of 4000 K. At this temperature, the carbon on the anode is vaporized and deposits onto the cathode. During the deposition, the distance between the electrodes needs to be kept constant. The diameter distribution of the nanotubes depends on the mixture of the inert gases.

The laser-ablation method was first reported in 1995 [75]. A laser beam is used to vaporize a target of consisting of a mixture of graphite and metal catalysts, such as cobalt or nickel, in an oven, which is filled with helium or argon gas (Fig. 17). Due to the high temperature (1200 °C) and high pressure (500 Torr) in the oven, a very hot vapor plume forms. As the vapor expands, the temperature rapidly goes down. As a result, small carbon molecules and atoms condense to form larger clusters.
these initial clusters, tubular molecules grow into nanotubes until the carbon can no longer diffuse through or over the surface of the catalyst particles.

Chemical vapor deposition (CVD) synthesis is usually a catalyst assisted process (Fig. 18 and [76]). It is based on the concept of decomposing hydrocarbons over a catalyst. The hydrocarbon enters the reactor with an inert gas at high temperature. As the hydrocarbon decomposes, carbon deposits onto the catalyst, which is supported by a material such as alumina. The most common catalysts used for CVD are iron, nickel or cobalt.
Ball milling is another method for the production of carbon nanotubes [77]. The nanotubes are produced from the nanoporous microstructures during the thermal annealing process. Essentially the method consists of placing graphite powder into a stainless steel container along with four hardened steel balls. The container is purged, and argon is introduced. The milling is carried out at room temperature. Following milling, the powder is annealed under an inert gas flow at a high temperature. Research has shown that this method produces more multi-walled nanotubes and few single-walled nanotubes.

2.3 Replication Methods

2.3.1 Nanoimprint

Nanoimprint refers to a group of replication technologies. In this process, a master template is generally required, which has complimentary features of the required features. The template is put into contact with the sample substrate. The features are transferred to the resist coated on the substrate or directly transferred to the substrate itself during the mechanical contact. Unlike conventional lithography, the nanoimprint process itself does not use any energetic beams so that the resolution is not limited by the effects of wave diffraction, scattering and interference in the resist. It is more of a physical process rather than a chemical process, similar in principle to conventional scale techniques like forging and casting. In some processes, a sequential anisotropic etching process is needed to further etch down into the substrate to form three-dimensional structures.

According to the principle of nanoimprint, the resist layer or the sample substrate needs to be sufficiently soft and plastic so that it can be readily deformed into the shape of the template. This requirement hastens the birth of many variants in nanoimprint lithography technologies, such as thermal nanoimprint lithography (T-NIL), photocurable nanoimprint lithography (P-NIL), step and flash nanoimprint lithography (SFIL), laser-assisted direct imprint (LADI) and etc.
2.3.1.1 Thermal Nanoimprint Lithography. Thermal nanoimprint lithography is the most widely used method. In this process, the resist is made soft by heating above its glass transition temperature [78]. The resist material becomes a viscous and flowable liquid at such temperature and can easily be deformed by the mechanical press. In order to reduce the thermal strain of the resist, materials with small coefficients of thermal expansion (CTE) (e.g., PMMA) are used. After imprinting, reactive ion etching (RIE) is used to remove the residue resist in the compressed region and further record the pattern in the underlying base substrate. The high aspect ratio resist structures were successfully achieved, as shown in Fig. 19.

2.3.1.2 Photocurable Nanoimprint Lithography. Photocurable nanoimprint lithography utilizes UV exposure to harden the resist layer after contact. The resist generally contains multiple components with at least one with high etch resistivity to the following plasma etch, and one photoinitiator which initiates polymerization upon UV exposure. Therefore, a tightly bonded polymer network will be formed after UV exposure and retain the deformed features within the resist. In this process, the template is transparent to UV light to allow the UV illumination and rigid to minimize distortion of the pattern when pressure is applied. It should also be able to release from the resist layer without generating particles. The study on minimum size capabilities shows that P-NIL is advanced than T-PIL and structures with pitch of 14 nm were successfully fabricated, as shown in Fig. 20 and [79].

2.3.1.3 Step and Flash Imprint Lithography. Step and flash imprint lithography is a bilayer lithographic scheme on P-NIL, as shown in Fig. 21 and [80]. The substrate is pre-coated with an organic planarization layer, known as “transfer layer.” A thin layer of UV sensitive resist is coated on top of this layer and brought
FIG. 20. SEM image of cured P-NIL polymer patterned by a mold with a 14 nm pitch grating with 7 nm linewidth.

FIG. 21. The SFIL process flow.

into contact using minimal pressure with the master template having low aspect ratio features. The template is then illuminated with UV light thereby crosslinking the UV-resist (etch barrier) at room temperature. After template removal, the feature is molded into the top layer. A sequential RIE process is applied to etch down into the bottom transfer layer, thus forming high aspect ratio structures. According to the principle, the transfer layer must remain adhered to the substrate and etch barrier during the separation process while being immiscible in the etch barrier solution. The transfer layer also must have sufficient reactive ion etch selectivities relative to the
substrate in order to act as an etch mask. SFIL has an advantage that a master template with low aspect ratio features can be used to make high aspect ratio structures into the transfer layer or the substrate.

2.3.1.4 Laser-Assisted Direct Imprint. Laser-assisted direct imprint uses a single XeCl excimer laser pulse passes through a quartz mold and melts a thin surface layer of the silicon substrate within picoseconds [81]. The molten silicon layer is then embossed by the quartz mold. After the liquid silicon solidifies, the mold is separated from the imprinted silicon for the next imprint. A variety of structures with resolution better than 10 nm (Fig. 22 and [82]) have been imprinted into silicon using LADI, and the embossing time is less than 250 ns.

2.3.2 Hot Embossing

Hot embossing is another replication technology widely used for micro/nanofluidic applications [83]. It is also known as Hot embossing Nanoimprint Lithography (HE-NIL). It essentially refers to the stamping of a pattern into a polymer softened by raising the temperature of the polymer just above its glass transition temperature. The embossing master can be a wafer, glass, electroplated nickel mold or other stamp with microfeatures. This technique is used primarily for defining fluidic components, such as micromirrors, microgrooves and lens cavities for fiber communications. Recently it has also been applied in chemistry and life science to form optical sensors and biochips.
2.4 Functional Nanomaterials

2.4.1 Introduction

In nanofabrication, structures and devices generally comprise of a base substrate and the materials stacking on the top of the base substrate. This is especially true in the cases of “Bottom–Up” technologies. The structures and devices are composed of different materials, whose dimensions are within nanometer scale. The unique characteristics of the materials allow the nanodevices to carry specific tasks and interact with the neighboring materials. In this section, we introduce several commonly used materials in nanodevices, with emphasis of their unique characteristics and applications.

2.4.2 Nanoparticles

Nanoparticle is so far the most widely used nanomaterial. It refers to a microscopic particle whose size is measured in nanometers. It is generally accepted that nanoparticles have dimensions less than a couple of hundred nm. Each individual particle is a cluster of a few hundred to a few thousand atoms. The nanoparticles are produced by means of colloid synthesis, although epitaxy and e-beam lithography may also be possible.

Due to the sizes of these nanoparticles, they attract great research attention to bridge between bulk materials and atomic structures. For example, nanoparticles and clusters have discovered the feasible material to form photonics crystals that consist of material patterned with periodic dielectric constants. The difference in dielectric constants forms photonics band structures, providing the opportunities to shape and mold the flow of light for photonic information technology [84]. The chemical sensing is another promising application for nanoparticles. Due to the small feature size, nanoparticles usually have a superior total surface area, making them extremely sensitive in detecting various chemicals, especially vapors and gases. For example, ZnO nanoparticles impregnated with Pt and/or doped with Co show very good sensitivity in hydrogen detection [85].

Latest development in nanoparticle technology allows people to construct nanoscale features using nanoparticles. This is so-called nanoparticle-based nanofabrication. These nanoparticles can be assembled to build larger features by means of optical [86], magnetic [87], or mechanical [88] approaches. Shimotsuma et al. demonstrated Rugby-ball-like asymmetric induced structures inside Ag\(^{+}\)-doped silicate glass by using femtosecond pulse laser. The laser forces the Ag nanoparticles aggregate at the depth of the focal point. This method is useful in the fabrication of the three-dimensional symmetric micro- and nanostructures for many monolithic photonic devices and microelectromechanical system applications. Alternatively,
NANOBIOTECHNOLOGY

FIG. 23. Field-emission scanning-electron microscope images of the enclosed silica nanoparticle channels.

Nanoparticles can be patterned using magnetic forces. For example, superparamagnetic particles and clusters can self assemble into two-dimensional (2D) or three-dimensional (3D) spontaneously, driven by the isotropic interaction between particles. This is similar in principle as non-magnetic particles where the pattern is driven by isotropic van der Waals interactions between the nanoparticles and the steric repulsions due to ligand interdigitation. Direct deposition provides another way for nanoparticle fabrication. The nanoparticles can be deposited onto a template with complementary features to form nanochannels upon removal of the sacrificial layers, as shown in Fig. 23 and [88].

Quantum dots are a certain type of nanoparticles featured by the size of a few nanometers. They are particularly significant for optical applications due to their theoretically high quantum field, and quickly find their role in quantum electronics [89] and quantum dyes for biological applications [90]. Fig. 24 shows the fluorescence of different sized Cadmium Selenide Quantum Dots, which is expected to replace currently used organic dyes [91]. Fig. 25 shows the HD DVD using blue-violet laser enabled by quantum dots instead of red laser used in conventional CD to read and write data [92]. The commercial blue laser has long been thought impossible until the quantum dots break in this field. The shorter wavelength enables a higher precision and allows data to be packed more tightly and store in less space. This technology has currently been used in Playstation 3.
2.4.3 Nanorods and Nanowires

 Besides nanoparticles, quasi one-dimensional (1D) solid nanostructures (nanowires or nanorods) have stimulated considerable interest for scientific research due to their importance and the potential applications in mesoscopic physics studies. Compared with micrometer diameter whiskers and fibers, these nanostructures are expected to have remarkable optical, electrical, magnetic, and mechanical properties. The unique feature of these materials, compared to other low-dimensional systems, is that they have two quantum confined directions while still leaving one unconfined direction for electrical conduction.

 Nanorods generally refer to the structures with aspect ratio 2 to 5, while nanowires are usually with longer aspect ratios. Their unique characteristics are aspect ratio dependant. For example, small change of gold nanorods leads to drastic change of the transmitted colors, as seen in the sample in Fig. 26 and [93].

 The synthesis methods of nanowires and nanorods largely depend on materials and applications [94]. Typical synthesis methods include template methods [95], electrochemical methods [96], photochemical methods [97] and many etc. Similar as nanoparticles, nanowires and nanorods also demonstrated the significant potentials in chemical sensing [98–102]. Recent development in these nanostructures is entering a new domain of highly sophisticated biomedical applications including targeted drug delivery, ultra-sensitive disease detection, gene therapy, genetic screening, and rapid toxicity cleaning.
FIG. 26. The color of gold rods and the respective micrographs. The color changes take place for very small changes in mean aspect ratio.

Above functional materials possess unique characteristics due to their individual sizes, geometries and material properties. In order to construct functional devices, these nanoscale materials usually need to be integrated with bulk substrates, electrodes and micro/nanofabricated structures, as mentioned in previous sections, to form hybrid structures. Conventional methods including deposition, self assembly and lithography are extensively used, and novel technologies are continuously developed. For example, Fig. 27 shows the nanowires assembled on the base substrate using dielectrophoretic forces [103]. Extensive research efforts are and will be increasingly put towards integration of these functional materials within microdevices.

2.5 Emerging Technologies

The increasingly development of nanoelectromechanical systems and the demand of biological and medical application pose a much challenging requirement of fabrication of complicated nanostructures than ever before. Many hybrids methods have emerged by combining two or more of above methods and incorporate various materials to form truly functional structures and devices.

Hybrid methods derived from combination of top-down and bottom-up nanofabrications have the best aspects of both techniques, which are good for massively parallel integration of heterogeneous nanocomponents into higher-order structures and devices. For example, three-dimensional nanostructures were demonstrated by using nanoimprint lithography and layer-by-layer self assembly, as shown in Fig. 28.
FIG. 27. Scanning electron micrographs of assembled and released structures: (a) Carbon tube. (b) Rhodium rod.

FIG. 28. Integrated nanofabrication scheme incorporating nanoimprint lithography and layer-by-layer assembly.
The PMMA nanofeatures were patterned by nanoimprint lithography (NIL), using a rigid substrate as the master template, because of its ability to pattern in the µm and nm ranges. The PMMA template behaved as a physical barrier preventing the formation of a SAM in the covered areas of the substrate. After polymer removal, SAM patterns with nanoscale precision were obtained, shown in Fig. 29 and [104–106].

Nanorobotic approaches show another set of promising methods for nanofabrication. Usually, a piezo-driven AFM tip is used to assemble or stack nanoscale objects with high precision. This allows for construction of nanostructures in an arbitrary order. Nanorobotic manipulation enables this hybrid approach for creating nano-electromechanical system (NEMS) that can attain a higher functionality because they possess more complex structures. Moreover, material and mechanical characterization can be performed after intermediate processes. In situ active characterization can also be performed using manipulation rather than conventional static observations. Fig. 30 shows manipulation of nanotubes with large aspect ratios [107], where the nanotubes were picked up and located to certain position by dielectrophoresis, van der Waals force or contacting a tube with the probe surface or fixing.

For biomedical nanosystems, research interest specifically focuses on polymeric materials due to their unique properties, including biocompatibility, optical transparency and so on. Novel methods are continuously emerging to pattern and fabricate nanostructures on such materials. Since many polymer materials are not compatible with conventional photolithography, the fabrication of polymer materials, more or less, relates to replication approach as it is so far the most economic and feasible way to transfer the nanofeatures to the polymer substrate. In these methods, the replication is combined with micro/nanofabricated probes [108], self-assembly [109], and many others.
2.6 Future Direction

2.6.1 Virtual Reality and Haptics in Nanobiotechnology

Aforementioned, technology development has allowed us to build various nanosstructures and devices with complicated geometries. These structures are sized from millimeters down to the molecular level, and manipulation of such structures can be in the manner of atom by atom. This rises up a question: how can we manipulate (or just monitor) such tiny nanodevices without being able to see them in real time?

The emerging virtual reality (VR) technology may provide a solution. VR techniques are currently explored in nanoscience and biotechnology research as a way to expand the operator’s perception (vision and haptics) to the small ever world. A number of technologies have been developed to provide a better human–machine
interface specific to the nanoworld. The tools for nanoscale perception include STM, SEM, AFM, and scanned-probe microscope (SPM). Optical tweezer also provides the user with a non-contact method for manipulation of nanoscale objects. The working environment is perceived and transmitted by the operator, giving him a feeling of the local environment he is working on. The information is often in a visualized format, reconstructed 2D or 3D pictures.

Fig. 31 and [110] show a 3D VR system to perceive the nanoworld. Preliminary experiments on the user interface show that this system can be used for telenanorobotics applications such as 2D assembly of nanoparticles and biological object manipulation.

AFM based and surface acoustic based tools can provide force feedback with nanoscale precision. This is useful for the study of nanosurgery. A moderate force can thus be exerted to handle the nanoobjects. Otherwise, a too small force may not be able to hold the nanoobjects, while an excessive force applied on the nanoobjects may lead to a non-negligible degree of probe or object deforma-
tion and may destroy the nanoobject or make it flip away. The real-time force-feedback can be performed through 3D vision-based force sensing on a computer display.

The next step of VR system for nanotechnology is to develop tools for nanomotion planning of complex nanotasks. This implies the integration of nanorobotics, 3D sensing, VR systems, and computer-aided design. It requires a quick reconstruction time so as to allow the operator sense the environment and perform actions timely. The integrated haptic device ensures a truly interactive virtual environment that incorporates efficient, dynamic, physically based simulation and motion-planning techniques applicable to complex nanotasks. A couple of successful systems have been developed. Highly Immersive Molecular Modeling (HIMM) team at the Institute of Technical Biochemistry at the University of Stuttgart is conducting research regarding the application of virtual reality technology in the area of molecular modeling, as see in Fig. 32 and [111]. Researchers at the Theoretical Biophysics Group at the University of Illinois at Urbana Champaign are dealing with manipulation of molecules in MD simulations with real-time force feedback and graphical display. Argonne National laboratory also has a VR-based system for studying the docking of a ligand molecule with a protein binding site, as shown in Fig. 33 and [112].

Despite these advances, current VR and haptic systems for the nanoworld are still deficient in the following aspects:

![Image](image_url)

**Fig. 32.** Geometrical representation model for the molecule data.
(1) Time delay.
(2) Lack of “true” and instant force feedback.
(3) Simulation sickness.

With these problems being solved as technology develops, fully functional VR and haptic systems are expected to be developed in the near future and bring us to a “like-real” nanoworld.

2.6.2 Functional Nanodevices

Nanofabrication provides a powerful tool for us to reach the bottom. Currently, most nanofabrication technologies focus on shrinking the sizes of the structures as much as possible. However, there is argued that in most cases, smaller not necessarily means better, but it definitely accompanies with higher cost and fabrication complexity. The future direction of nanofabrication is expected to move towards functional devices. The technologies for tomorrow will be judged by their functionality, rather than the feature size. Since it is hard to imagine that a single and simple structure can conduct functionalities with high complexity, devices consisting multiple nanostructures are of more important. These structures must interact with each other in a coordinated manner. Accordingly, the research focus may shift from fabrication to manipulation, assembly and interconnection. The future nanodevices may also have
in hybrid forms that incorporate living organism in order to increase biocompatibility and to provide energy for forces and motion.

3. Nanobiotechnologies for Sensing and Actuating

3.1 Significance

It is not easy to give a definition of nanobiotechnology. Literally, nanobiotechnology is defined as the nanotechnologies (the technologies we briefed in previous section) being applied into biomedical field. In this sense, the devices or key components used in nanobiotechnology need to be of nanoscale. (The word “nanoscale,” according to peer scientists, generally refers to features equal to or less than 100 nm. The devices and structures with a couple of hundred nanometers in size are generally regarded of sub-micron scale.) This definition is argued by some other researchers who acclaim that nanotechnology should be much broader rather than solely confined by size. In this sense, nanotechnology expands to any technologies directly dealing with or indirectly enabling detection, measurement and actuation of small quantities of physical elements (such as devices that can detect the presence of nM protein, deliver nanoliter liquid, or exert nanonewton forces), even if the devices or the measuring components themselves may be in microscale or larger. In this chapter, we stick with the second definition in order to provide a complete and systematic description of the nanobiotechnology.

Sensing and actuating are two major branches of nanobiotechnology. Biosensors and bioactuators both bridge physical elements and biological information, as shown in Fig. 34. In the path on the left, the biological information of interest turns into physical elements that can be measured directly using known engineering approaches. In the path on the right, certain physical elements are input into the subject biological entities and induce changes in specific biological information, in order to accomplish required functionalities.

![Fig. 34. The roles of biosensors and bioactuators.](image)
3.2 Nanobiosensors

Biosensors are the name of devices that can obtain certain signals from biological entities. Biosensors usually couple a biological recognition element with a physical transducer. The devices contact (not necessary physically) with the target entities in order to collect biological information and convert it into certain forms of signals that are measurable by means of mechanical, electrical, magnetic, optical and so on approaches. In some biosensors, detection can be continuously performed so that the response (or evolution) of the biological information upon certain physiological process or internal/external stimulations are collected.

According to the working conditions, the biosensors can be categorized into \textit{in vivo} biosensors and \textit{in vitro} biosensors. \textit{In vivo} biosensors are implanted into living bodies of animals or human. These biosensors usually collect real-time signals, and reflect the effects of the living condition to the biological information of interest. Since these biosensors are generally made of artificial materials (e.g., plastic, metal and polymer), they may induce inflammation (the response of the immune system to infection and irritation) once being implanted \textit{in vivo}. Also, the wound made during implantation is another cause of the inflammation. As the results, the leukocytes sent by the immune system rapidly aggregate around the point of implantation and may block the measurement. The inflammation also induces symptoms such as redness, fever and organ dysfunction. These symptoms may eventually affect the living condition of the subject and make the measurement unreliable. In order to reduce the inflammations during implantation and measurement, study of \textit{in vivo} biosensor focuses on miniaturization. It is generally accepted that the reduction of the implantation size could reduce the agitation of the natural environment in the subject living body, thus resulting in or less severe or delayed inflammation. This allows for long term measurement without badly disrupting the living condition of the subject.

\textit{In vitro} biosensors deal with biological entities that have been isolated out of the living bodies. These entities are usually maintained intentionally at certain \textit{in vitro} conditions, in order to provide an economically and ethically better alternative of the animal model. The \textit{in vitro} models allow for quick and highly quantitative detection as they provide a platform for parallel screening. Since the immune system is absence, the inflammation is not the major challenge. Instead, the problem rises to interpret the results obtained from the \textit{in vitro} model and compare them with the results from the \textit{in vivo} model.

Progress in micro/nanotechnology provides power tools to address above issues. By using nanotechnology, the overall sizes of biosensors can be reduced to micrometer domain, which greatly decreases the inflammation during implantation and measurements. The nanometer structures also have superior measuring sensitivities and high spatial resolutions. This allows for integration of a large number of sensing
Nanomechanical Sensors

Nanomechanical sensor is a collective name of nanobiosensors which rely on micro/nanoscale mechanical components to detect biomolecular interactions with great accuracy. The unique geometries of this type of sensors ensure high sensitivities along certain orientation(s). Fig. 35(a) shows the most ubiquitous microcantilever structure. The suspending beam connects with the base substrate via a post.

3.2.1.1 Working Principle.Basic mechanism of the sensors is the bending induced in the cantilever when a biomolecular interaction takes place on its surface. The nanomechanical motion is often coupled to an optical leverage setup, interferometric detection or piezo-resistive read-out system. Cantilever sensors can be operated in static, dynamic, and heat modes. In static measurements, the biological information (e.g., cellular forces and absorption/recognition of biomolecules at the tip of the cantilever) induces the out-of-plane bending. The measuring sensitivities depend on out-of-plane spring constant ($k$), a function of the cross-section ($w$ and $t$) and the length of the cantilever ($L$):

$$k = \frac{E t^3 w}{4L^3},$$

where $E$ is Young’s modulus of the cantilever material. In nanobiosensors, the thickness and, in many cases, the width of the cantilever beam is less than 1 µm.

FIG. 35. Microcantilever as biomechanical sensors. (a) The schematics of a cantilever with rectangular cross-section. (b) The working principle of atomic force microscope.
In dynamic measurements, the cantilever is oscillated externally at its resonance frequency using a piezoelectric actuator. The subject biological information may cause molecules to be immobilized to the cantilever surface, and thus induce a shift from its resonance frequency to a lower value. From this shift and the immobilized locations, the absorbed mass down to picogram can be calculated [113], providing a lot more accurate measurements.

In heat measurements, a multiple layered cantilever is used. The difference in the thermal expansion coefficients of the cantilever materials of different layers cause bending of the cantilever sensor if the temperature is changed. In nanomechanical sensors, temperature changes of $10^{-5}$ K produce cantilever deflections of several nanometers, which can be measured easily. This type of sensor can also be revised to detect other characteristics of the local environments, such as humidity and pH values. The spring constant and the natural frequency of cantilevers with various cross-sections or of multiple layered cantilevers are available elsewhere [114–119].

**3.2.1.2 AFM-based Nanomechanical Sensors.** Atomic force microscopy (AFM) is the most successful and widely used cantilever-based nanomechanical sensors. Fig. 35(b) show the working principle of a typical AFM device. It has been used for over a decade for numerous biological investigations such as characterization of the properties of protein–protein and protein–nucleic acid complexes, determination of stoichiometries and association constants of multi-protein assemblies, observation of dynamic properties of biomolecular complexes, and measurement of intermolecular forces between biomolecules [120].

The use of AFM nanomechanical sensors can be clearly viewed by reviewing a typical example. In the measurements, an AFM tip approaches the biological entities of interest that placed on a base surface, shown in Fig. 36(a). Once a stable contact is established, the tip is withdrawn slowly from the surface, and deflected due to the contact. Therefore, a hysteretic load-deflection curve is resulted. Normally, a vertical force is required to pull the tip off the surface. This curve and the vertical force are later interpreted to certain biological information base on a known relationship. For example, the AFM tip can build up covalent bonds with DNA oligonucleotides once it is put into contact. These oligonucleotides are covalently attached on the other end with a base substrate. The interaction forces between single-strands of DNA can thus be determined from the tip deflection [121]. AFM has also been applied for detection of cell–cell adhesion forces [122], protein–cell interactions [123], protein–ligand interactions [124], protein–membrane interactions [125] and many others.

Another way to use AFM tip as mechanical biosensor is to immobilize certain molecules on the surface of AFM cantilever and tip, as shown in Fig. 36(b). When the tip is approaching the subject surfaces or is placed into the subject atmospheres, the immobilized molecules may bond with certain molecules present on the surfaces
or in the atmospheres. This induces tip deflection towards the surfaces due to a differential surface stress between opposite sides of the cantilever, approximated as [126]:

\[
\Delta z \approx \frac{3(1 - \nu)L^2}{E t^2} (\Delta \sigma_1 - \Delta \sigma_2),
\]

where \( \nu \) is Poisson ratio and \( \Delta \sigma_1 - \Delta \sigma_2 \) accounts for the differential surface stress.

The absorption of the molecules also induces a shift from natural resonance frequency of the tip (Fig. 36(c)). For example, an enzyme-coated cantilever has been applied to detect absorption of glucose molecules by living \( S. \text{cerevisiae} \) cells, during image scanning. The image scanning was conducted in a suspension containing glucose and \( S. \text{cerevisiae} \) cells in ultra-pure water using an enzyme-coated AFM tip. In the initial scannings, cell images failed to appear due to successive deflections of the cantilever, indicating the presence of glucose molecules in all the scanned areas (Fig. 37(a)). After 3 hours of incubation, part of this glucose was absorbed by living cells, and the image of the cells appeared in the point where there was no glucose (Fig. 37(b)). This image of cells was gradually completed after all the glucose molecules were absorbed (Fig. 37(c) and (d)) [127].

The cantilever-based nanomechanical sensors can also be used in an array for detection. Fig. 38 shows the scanning electron micrograph of a cantilever sensor array [128]. The array contains multiple cantilevers side by side, providing the advantages to measure the difference in the responses of sensor and reference cantilevers.
Fig. 37. Scanning of the surface of *S. cerevisiae* cells by an AFM tip. (a) The high concentration of glucose impedes the image scanning of the cell body due to the tip deflection. (b) After 3 hours, some of the glucose molecules are absorbed by the cells, and the scanning can performed in some parts. (c) After 4 hours, scanning can be performed in more area. (d) After five hours, scanning can be performed in the entire body of the cell, indicating that the glucose molecules are completely absorbed.

At least one cantilever in this array is pre-coated with a sensitive layer that exhibits an affinity to the molecules to be detected, while other cantilevers are coated with a molecular layer that does not show an affinity to the molecules to be detected. The undesired artifacts such as thermal drift or unspecific adsorption can be successfully canceled out. The cantilever-based sensors array can also be used to differentiate a pattern of responses. For example, the cantilever sensor array has been applied as an artificial nose to detect simple or complex odors [129] by pre-coating individual cantilevers with different types of molecules.

Fig. 38. Scanning electron micrograph of a cantilever sensor array.
3.2.1.3 Polymer Nanomechanical Sensors. Polymer-based cantilever is another group of nanomechanical biosensors. Due to the relative small elastic modulus of polymer materials, this type of sensors presents even higher probing sensitivities as compared to the silicon peers. Moreover, many polymer structures are found much biocompatible with the biological entities to be detected than the conventional materials. Furthermore, the relative low cost makes possible to dispose the polymer nanobiosensors after one time measurement, which is important because the surfaces of the nanobiosensors contacting with the biological entities is very hard to clean.

As aforementioned, only a few polymer materials are compatible with semiconductor microfabrication. The negative photoresist SU-8 first formulated by IBM is one of them. The cantilever array made of SU-8 has been successfully applied to measure the absorption of ssDNA [130]. The configuration and measuring principle of the SU-8 cantilevers are similar with those silicon cantilevers. The non-photodefinable polymers have also found their use in nanobiosensor. PDMS (polydimethylsiloxane) is the most widely used one. This silicone rubber has been involved in the measurements of cell adhesion forces [131], cell contraction forces [132], and external applied forces to pierce cell membranes [133]. Since PDMS is not compatible with conventional semiconductor microfabrication, these structures are fabricated based on replication methods [134,135]. As a result, these cantilever-like structures are not suspending over the surface like the silicon cantilevers, but standing vertically on the substrate, as shown in Fig. 39. The deflections observed at the tips of these cantilevers are a measure of the forces in the horizontal plane. Attempts have been done to simplify the replication process by using a single template to replicate polymer cantilevers with different aspect ratios [135]. This is accomplished by applying a pressure-assisted micromolding process to force premature polymer fill into the master template. It is noted that due to the small elastic modulus and the surface characteristics of the polymer material, an upper limit is imposed to the aspect ratio of the cantilever to avoid structures failures, such as collapses shown in Fig. 40.

![Fig. 39. Polymer mechanical sensors for detection of (a) contractile forces of cardiac myocytes; and (b) external forces to pierce the cell membranes.](image-url)
3.2.2 Nanoelectrical Sensors

Nanoelectrical sensor is another group of nanobiosensors, which converts biological information into signals that can be electrically detected, when the subject biological entities become immobilized or entrapped on or near the electrodes. These sensors are often referred to as amperometric, potentiometric, conductometric and many etc. They are highly sensitive, rapid, inexpensive, and amenable toward microfabrication.

In nanoelectrical sensors, the micro/nanoelectrodes are required in order to obtain electrical signals from the biological entities. These electrodes are usually made of metals, carbon or conductive polymers, and separated by dielectric materials, such as silicon nitride and silicon oxide. The sensitivities of these biosensors highly depend on configurations of the electrode sites. The sizes of the electrodes are designed to match with those of the subject biological entities.

3.2.2.1 Amperometric Biosensor. The detection principle of amperometric biosensors is based on the modification of electrical current between the electrodes and subject biological entities. The amperometric biosensors are attractive because of its high sensitivity and wide linear range. One representative application of amperometric biosensor is the enzyme electrode. Such device combines the specificity of the enzyme for recognizing a given target analyte, with the direct transduction of the rate of the biocatalytic reaction into a current signal. In an enzyme electrode, the electrode surface is coated with a thin layer of enzyme, which is chosen to catalyze a reaction that generates a product or consumes a reactant. The reaction can thus be monitored amperometrically. The amperometric sensors can also be constructed using CNTs. The direct electron transfer between the single-walled/multiple-walled nanotubes and the proteins has been observed in CNT based enzymatic biosensors, suggesting the immobilization of glucose [136], hydrogen peroxide [137], and DNA hybridization [138].
3.2.2.2 Potentiometric Biosensor. The potentiometric biosensor is to use ion sensitive electrodes to transduce the biological reaction into an electrical potential signal, which is proportional to the concentration of species generated or consumed in the recognition event. A permselective ion-conductive membrane is applied on the tip of the electrodes to yield a potential signal primarily due to the target ions. The most widely used potentiometric biosensor is the pH electrode. It consists of an immobilized enzyme membrane surrounding the probe from a pH-meter, where the catalyzed reaction generates or absorbs hydrogen ions. The reaction occurring at the thin sensing glass membrane causes a change in pH value and consequently in electrical voltage. For example, the microbial biosensors based on modification of glass pH electrode with genetically engineered *E. coli* expressing organophosphorus hydrolase intracellularly and on the outer surface of cells and wild-type overpotential degrading bacteria *Flavobacterium* sp. have been reported [139–141]. A potentiometric oxygen electrode with immobilized *S. ellipsoideus* was also successfully used to produce a microbial biosensor for the determination of ethanol with an extended response range [142].

3.2.2.3 Conductometric Biosensor. The detection of conductometric biosensors is based on modification of the junction between two electrodes after molecule immobilization, which induces a measurable change in the conductance of the device. In conductometric sensors, the sizes of the electrodes need to match with the size of the biological entities to be detected. Therefore, the nanowires have been used owing to their small feature size. Since the surface properties of the nanowires can be easily modified and decorated with virtually any potential chemical or biological molecular recognition units, these units are used to transduce absorption/recognition into change of conductance of the nanowires, while making the wires themselves analyte independent. For example, biotin-modified silicon nanowires were used to detect streptavidin down to at least a picomolar concentration range. CNT is another option for detection of nanosized biological molecules. The electrical conductance of CNT is highly sensitive to their environment and varies significantly with changes in electrostatic charges through the surface adsorption of various molecules. The biocompatibility and the excellent catalytic ability of CNT also add the advantages to these biosensors.

3.2.2.4 Hybrid Impedimetric Biosensor. Besides above nanobiosensors, there are a lot more biosensors have and being developed with a variety of configurations, different nanomaterials and various data read-out methods. Fig. 41 shows a nanoswitch for detection of DNA hybridization [143]. It combines nanoparticles and the nanofabricated electrodes. The electrodes in the active area are coated with capture DNA. The nanoparticles modified with target DNA are specifically
FIG. 41. The active area of a biosensor with nanogap: (a) Optical image and SEM micrograph. (b) Experimental scheme for the electric detection of DNA hybridization using a nanoparticle nanoswitch.

attached to the electrodes, resulting in a much higher electrical conductance than those without the nanoparticles in the gap. In other words, the increase in the electric conductance can be interpreted to be the result of gold nanoparticle immobilization between the electrodes gap, suggesting the existence of the target DNA in the sample solution.

It should be noted that these bionanosensors are often very small in size. Therefore, the effective area for the opposite nanoelectrodes is limited, deteriorating the performance of the sensors, especially when the analytes are with low concentration or little target molecules. Therefore, many of the hybrid nanobiosensors are made into interdigitated fashion, using a pair of parallel electrode fingers facing each other for detection, with an aim to maximize the current or the impedance change. Using such interdigitated nanoelectrodes, the detection of bacteria [144] and DNA immobilization [145] have been successfully accomplished.

The planar patch clamp is yet another example of the hybrid impedimetric biosensor. The patch clamp is well known as an electrophysiological technique to detect the transmembrane current of the cells. With the help of microfabrication, the patch clamp is revised into a planar fashion, in order to measure the electrical current caused by a single channel or the alternatively currents from entire small cells. Fig. 42 shows the schematics of such a device [146]. A microscale aperture is made in the substrate (e.g., glass, silicon or PDMS). And the subject cell is immobilized above the aperture and fits tight in. The size of the microaperture beneath the cell allows only single or a few ion channel sites. With appropriate surface modification, the cell forms high resistance sealing with the substrate and therefore, the current change between the two electrodes is solely due to the subject ion channels.
The biosensors can also be made in the form of nanocircuitry, where the biological events to be detected influence one or more electrical input within the circuitry. This information can be monitored from the output. For example, field-effect transistors based on semiconductor nanowires and nanotubes have been applied for highly sensitive detection of biological molecules. Such device combines nanowires/nanotubes and microfabricated electrodes. Since the depletion or accumulation of charge carriers are caused by binding of the charged biological molecules onto the surface of the nanowires or the nanotubes, they can affect the entire cross-sectional conduction pathway of these nanostructures. Fig. 43 shows a simple circuitry [147] where two silicon nanowires are constructed, suspending over the base substrate. One of the nanowires is modified with anti-influenza type A antibody (blue), and the other with anti-adenovirus group III antibody (red). When a virus particle binds to the antibody receptor on a nanowire, the conductance should change from the baseline value; when the virus unbinds, the conductance should return to the baseline value. For a p-type nanowire, the conductance should decrease (increase) when the surface charge of the virus is positive (negative). The two wire configuration enables simultaneous conductance measurements of adenovirus, influenza A, or a mixture of both viruses.
3.2.3 Nanooptical Sensors

Nanomaterials have been applied in biological imaging due to their photostability. This offers advantages over the conventional organic dyes that undergo photobleaching. In addition, most organic dyes only fluoresce in the visible wavelength region with low quantum yields. On the contrary, high quantum efficiency fluorescence can be easily achieved with tunable wavelengths because the dimensions of the nanomaterials can be finely controlled, as previously shown in Fig. 26.

3.2.3.1 Plasmon-Resonant Nanoparticles. Some nanoparticles, such as Au and Ag, are quench resistant and have very high light-scattering power. The optical properties of the nanoparticles depend on the spacing within the polymer aggregate. For example, a color change was observed due to the aggregate scattering properties and the interaction between particle surface plasmons as the distance between the nanoparticles varied [148]. These nanoparticles are so-called plasmon-resonant particles. This phenomenon has been successfully applied to label biological molecules, e.g., nucleic acid [149] and protein [150], and can scatter visible light of certain wavelength upon the incident light. For example, in a colorimetric detection base on evanescent-wave-induced light scattering, 50 nm Au nanoparticles modified with oligonucleotide detection probes are introduced into a solution of the single-stranded target oligonucleotide. A polymeric network of nanoparticles is formed and a color change can be detected visually. These nanoparticles scatter
3.2.3.2 Biobarcode. The biobarcode assay is a recently developed optical technique for biological molecules detection, as shown in Fig. 44 and [151]. The biobarcode assay has a magnetic probe and an Ag nanoparticle that both are specific to the target bimolecules. The gold carries with hundreds to thousands of DNA sequence barcodes that are chosen to be specific to the target of interest. Detection of the released biobarcodes can be done using fluorescent or microarray methods. It can also be detected using electrochemical approaches.

3.2.3.3 Contrast Enhancement. One notable challenge in biological optical sensing is due to the poor contrast between the analyte and the background. This problem becomes extremely serious when detecting nanoscale objects. Study of contrast enhancement has been a focus of recent research. For example, superparamagnetic nanoparticles have been applied in magnetic resonance imaging as contrast agents. These nanoparticles have an inorganic core of iron oxide coated with polymer outer layer. These nanoparticulate contrast agents are currently used for imaging of tissue for diagnostic applications [152].

3.2.4 Other Types of Nanobiosensors

Besides the above mentioned mechanical, electrical and optical approaches, many other physical phenomena have also been applied to make the nanobiosensors. For example, magnetic immunoassay techniques have been developed in which the magnetic field generated by the targets labeled by magnetical nanoparticles is detected directly with a sensitive magnetometer. Antibodies labeled with magnetic nanoparticles give magnetic signals on exposure to a magnetic field. Antibodies bound to targets can thus be identified as unbound antibodies disperse in all directions and produce no net magnetic signal.
Before we ending up this section of biosensors, we would emphasize that in practical applications, one often encounters the cases where more than one type of nanotechnologies need to be used to develop a hybrid nanobiosensor. The hybrid nanobiosensor may work on multiple physical basis and consists of nanomaterials, microstructures and conventional scale parts. The sensor cannot be readily grouped into any of above categories. However, they may indicate the trend of the development of nanobiosensors. The nanobiosensor of tomorrow will become much smarter, more sensitive and multidisiplinary.

3.2.5 Biological Entities as Nanosensors

Biosensors detect certain pieces of information from biological entities, reflect changes within the subject object and reveal the mechanism of biological events. As trained in engineering, we tend to think it in a reverse way: if it is possible to use the biological entities themselves as the sensing components to detect the changes of the local environment? This thinking is based on the fact that many biological events and information vary drastically upon the changes of local environment (such as temperature, humidity, vapor pressure, pollutant and etc.). Since many biological entities are sized in nanoscale, they are naturally good nanosensors with extremely high spatial resolutions and sensitivities.

The answer to this question seems to be positive. For example, DNA has been proposed to work as a screening device for the rapid bioanalysis of environmental pollution. These DNA biosensors show the ability to distinguish in 11 minutes low, medium and high contaminated soils with good correlation with well established techniques. The DNA-based devices have also been used as electrochemical sensor for rapid, high sensitive and simple detection of aromatic amine compounds, a very important class of environmental pollutants. Cellular biosensors were also used for testing and monitoring the effectiveness of drugs and therapies [153,154]. Moreover, the microbial biosensors were used for monitoring microbiologically influenced corrosion of metallic materials [155].

Despite that these advances show the capabilities and potentials of biological entities in sensing of environmental parameters, some key puzzles need to be clarified:

1. Advantages of the environmental biosensors over those made of conventional non-biological materials.
2. The cost efficiency and production complexity.
4. The life time for the environmental biosensors.
3.3 Nanobiomotors

3.3.1 Nanorobots for Medicine Application

Besides sensing applications, nanotechnologies can also be applied to create nanosized machines, which are comparable in size with an atom or atom clusters. These nanomachines are made of engineering materials that have been modified to have minimal inflammatory effect. Due to the sizes of these nanomachines, they may be able to help the living organism against diseases and deliver certain drug/medicine locally to the points of interest. They can also be used as tools to study the behavior of living cells or even perform as the living cells. For example, Fig. 45 shows the imaginary schematics of the in vivo surgery conducted by a nanorobot on the red blood cells [156]. Such nanorobot is expected to perform local drug delivery, real time assay and repairs.

Despite that a wide variety of eye-ball attracting designs and promising prospects have been proposed, the fabrication and construction of the nanorobots are still in the cradle. The first step of making such a nanorobot is to create and to analyze possible designs for nanoscale mechanical parts that could, in principle, be manufactured. Given the size limitation and fabrication technology constraint, molecular bearings and gears are perhaps the most convenient class of components to design because their structure and operation are fairly straightforward. These components are constructed by stacking a large number (normally hundreds to millions, depending on the functionalities and configurations) of atoms of different species (carbon, silicon, oxygen, nitrogen, etc.) in a required manner. The concept has been validated by molecule simulation, as shown in Fig. 46 and [157]. With the help of the molecular pumps and rotor components, it is possible to build up complicated nanomachine at nanometer dimension. A typical example of molecule based nanomachine is the

![Fig. 45. The nanorobots are performing surgery on red blood cells. (Imaginary picture.)](image)
artificial cells, which refers to the mechanical parts that can carry out functionalities of the living cells. For example, Freitas et al. proposed an artificial respirocyte with active pumping powered by endogenous serum glucose, shown in Fig. 47 and [158]. According to the design, this artificial cell is able to deliver 236 times more oxygen to the tissues per unit volume than natural red cells and to manage carbonic acidity. This nanorobot is made of 18 billion atoms precisely arranged in a diamondoid pressure tank that can be pumped full of up to 3 billion oxygen (O₂) and carbon dioxide (CO₂) molecules. Another example is “microbivore,” the artificial mechanical white cell proposed by the same group, shown in Fig. 48 and [159]. The device may consume up to 200 pW of continuous power while completely digesting trapped microbes at a maximum throughput of 2 µm³ of organic material per 30-second cycle, which is large enough to internalize a single microbe from virtually any major bacteremic species in a single gulp. The nanorobots would be about 80 times more efficient as phagocytic agents than macrophages in terms of volume/sec digested per unit volume of phagocytic agent, and would have far larger maximum lifetime capacity for phagocytosis than natural white blood cells.
3.3.2 Protein Nanomotors

Research has also been moving on by incorporating biologic entities as active components into these nanomachines. One advantage of using biologic entities is to provide motion. Motion is an essential concept of all living organism. On a macroscopic scale, animals and microbes have to be able to move towards food and away from poison and predators. Plants turn their leaves toward their energy source, the sunlight. On a molecular scale, motion has been involved into many chemical reactions and synthesis processes, helping to convert the energy from one form to another, as required by the living organism (in most cases, from chemistry energy to kinetic energy). Since these biological entities used in nanomachines are normally sized in nanometers or even smalls, motions with very high precision can be obtained, thereby enabling actions that cannot be accomplished by conventional scales actuators. Within cellular systems, motion often refers to transportation of molecules by DNA and proteins. Researchers at New York University developed a robust, controllable, mechanical device [160]. This device can be manipulated without affecting others in a larger array. The extent of the motion varies from 0.4–4 nm, and the motions as large as 35 nm have been observed. Other researchers focus on the use of motor proteins. It is well known that the living cells are naturally good motors made by god. Different compartments are transported along the rail by molecule transporters: the myosin, kinesin and dynein motors. Taken the interaction of myosin for example, the transport is accomplished by using repeated cycles of coordinated binding and unbinding of myosin light chain with actin filaments, powered by energy derived by hydrolysis of ATP.

Currently, the study of DNA and protein based motors lies mainly in computational analysis and validation. Since the motor events are inherently stochastic, the outcomes are statistical. The overall effect, a directional movement of motor proteins, is only statistically the same, but each molecule performs its own unique “dance” while moving. The motion of the motor proteins (kinesin and myosin) of-
ten accompanies with mechanical forces (on the order of pN). These forces are the natural power source of muscle fibers, and can potentially be applied to drive hybrid nanomotors. Until most recent, some experimental results appear, including direct observation of DNA rotation by branch migration of Holliday junction DNA [161] and construction of a free-running DNA motor driven by Nicking enzyme [162]. It is expected to have more emerging biological entities based nanomachines above the horizon with continuous development of theoretical foundation and experimental approaches.

### 3.3.3 Cell-Based Robots

DNA and protein motors become inappropriate for the devices which need to be driven with relative large mechanical forces, or need to have large motions. Instead, the “whole-cell” was used as a single motor component. Scientists at the University of California, Los Angeles, have recently created ‘living’ robots by using heart cells [163]. The cells were grown on microscopic silicon chips, which are less than a millimeter in length. The assembly can move by themselves upon the cell contraction. Their structure is silicon-based, with mechanical hinges; allowing them to move and bend, to move the whole assembly. It is envisioned a number of applications including musclebots that deliver drugs directly to the cells and tiny machines to convert muscle motion into electric power for microcircuits.

### 4. Nanobiotechnology for Drug Delivery and Therapeutics

Another application of nanotechnologies into biology is the drug delivery and therapeutics. The concept of targeted drugs delivery dates back to 1906. The challenge of drug delivery has been on three fronts: finding the proper targets for a particular disease state; finding a drug that effectively treats this disease; and finding a means of carrying the drug in a stable form to specific sites while avoiding the immunogenic and non-specific interactions that efficiently clear foreign material from the body [164].

Many drugs discovered in the past could not be used in patients because a suitable method of drug delivery was lacking. Since the nanoparticles are sized into nanometers and can be modified to carry active drugs, they express a great potential to serve as the drug carriers. Moreover, the development of nanotechnologies makes it possible to produce nanoscale drug carriers made of biodegradable materials, effectively lowering the risk of immunogenic reactions. As the results, regulatory authorities have increasingly approved the nanoparticle base products for clinical use [165]. Most are anticancer therapies, i.e., liposomes, polymer-coated liposomes, polymeric drugs, antibodies and antibody conjugates, polymer–protein conjugates.
Developments in nanomedicine are truly threatening to the traditional paradigm in drug discovery, which is predicated on finding druggable compounds. An increasing use of nanobiotechnology by the pharmaceutical and biotechnology industries is anticipated. Nanotechnology will be applied at all stages of drug development, from formulations for optimal delivery to diagnostic applications in clinical trials. It already brings clinical benefits, and it will fit in with the concepts for the integration of diagnostics and therapeutics to develop individual-based therapy, which is based on the idea of using individual patient’s genotype as a factor in deciding on treatment options.

5. Concluding Remarks

Nanobiotechnology is no doubt one of the most significant technologies in the 21st century, which has already brought and will continue bringing profit to our communities. What nanobiotechnology promises is a more significant, deeper and longer term impact on the biomedical and clinical research. In the future, we could imagine a world where medical nanodevices are routinely implanted or even injected into the bloodstream to monitor health and to automatically participate in repair of living organisms. The continuous advances in the field of nanobiotechnology are founded on multidisciplinary collaboration of researchers with complementary expertises. Nevertheless, it has never been an easy task to find a collaboration point that can make both sides feeling comfortable, mainly due to the vastly different understanding of researchers in different fields. We often receive complaints from engineering scientists that they are regarded as “tool makers” and from biologists that they are thought doing “kitchen table type of work,” during their collaboration with the other side. Their frustration lies right across the way and may tremendously delay the developments in this field. It may take time before a mutual agreement can be reached, while we believe the increasingly developing start-up companies may serve as a lubricant and promoter in between. With the profit-making products in mind, these companies have to figure out ways to adapt nanotechnologies and applied into biological and medicine applications, which encourages them to have seamless contact with both sides. This will eventually blur the boundaries and pose the need for personnel with expertise from both sides, nano- and bio-.

REFERENCES


Abstract

The development of highly sensitive, selective, reliable, and compact sensing systems to detect toxic chemical and biological agents is of great importance to national security. This chapter examines the best such naturally occurring sensing system, the sense of smell or “olfaction,” as well as artificial sensing systems built to emulate the nose. The goal is to combine lessons learned from natural and artificial olfaction with opportunities presented by advances in nanotechnology, in order to further the development of nose-like sensing systems integrated on the nanometer scale. The olfactory processes are reviewed here in some detail. Dense arrays of olfactory neurons, acting as ultra-small, non-specific sensors, use molecular recognition to perform highly parallel molecular sensing. The sensory signals so generated are identified by the brain using a spatio-temporal coding scheme. In this way the olfactory system recognizes, with great accuracy and sensitivity, a broad range of chemical stimuli. The principles of olfaction have been applied to developing artificial noses that are composed of arrays of cross-reactive gas sensors of various types. Artificial noses based upon conductivity-change devices, mass-change devices, and fluorescent optical fibers are reviewed here. The smallest artificial noses at this time are devices that incorporate micron-scale sensing elements and are comparable in size to a credit card. To more closely approximate the capabilities and compact size of the natural nose, it will be necessary to shrink the individual sensor size even farther, integrating nanometer-scale sensors into systems. Individual nanometer-scale devices, such as carbon nanotubes and nanowires, already have been demonstrated...
to function as gas sensors, and their applicability to nose-like sensing is dis-
cussed. At this point in time, however, no complete, nose-like nanometer-scale
sensing system has been developed. This chapter concludes by presenting for
consideration two proposals for electronic noses based upon nanowires.

1. Introduction ........................................ 104
2. The Physiology of the Sense of Smell ........................................ 108
   2.1. The Olfactory Epithelium ........................................ 109
   2.2. The Olfactory Bulb ........................................ 115
   2.3. Odor Identification ........................................ 119
   2.4. Characterizing Odor Space ........................................ 122
   2.5. Key Operational Principles of Natural olfaction .... 124
3. Electronic Noses: Chemical Sensing Systems ........................................ 126
   3.1. Electrochemical Sensors ........................................ 129
   3.2. Mass-Change Sensors ........................................ 136
   3.3. Optical Sensors ........................................ 139
   3.4. Other Sensors ........................................ 141
   3.5. Next Steps in Artificial Noses ........................................ 143
4. Nanosensors ........................................ 143
   4.1. Carbon Nanotube Sensors ........................................ 145
   4.2. Nanowires and Other Nanostructures ........................................ 147
5. Designing a Nanometer-Scale Nose-Like Sensing System ........................................ 151
   5.1. Consideration of Nanotubes, Nanowires, and Nanobelts for Electronic Noses .... 151
   5.2. New Designs for Electronic Nose Systems Integrated on the Nanometer Scale .... 154
6. Summary ........................................ 157
   Acronyms ........................................ 158
   Acknowledgements ........................................ 159
   References ........................................ 159

1. Introduction

The purpose of this chapter is to examine the sense of smell and how it works, with
the goal of applying lessons learned from the operation of natural and artificial noses
to the development of sensing systems integrated on the nanometer scale. Ensuring
the security of the nation requires the development of fast, reliable, comprehensive,
and compact sensing systems to provide early warning of attacks by terrorists using
toxic chemical or biological agents. The sensing system with the best sensitivity,
speed, and reliability is provided to us by nature in the nose and the sense of smell, or
"olfaction." This natural sensing system has been imitated, over the past decades, by
a variety of artificial nose-like sensing systems, but a wide gap still remains between
the capabilities and sizes of artificial and natural noses. Advances over the past few years in building nanometer-scale devices, however, particularly the development of nanosensors, offer us the opportunity to make new progress towards closing this gap. Combining our understanding of olfaction with these recent developments in nanometer-scale devices and systems offers exciting new prospects for fielding sensors with truly nose-like sensitivity, accuracy, and reliability.

The olfactory system is a sensing system. At the same time, however, it is a highly efficient processing system. Most olfactory processing is performed in the pre-brain or the brain. As will become evident in the following sections, each component of the olfactory system, including the body of neurons in the nose that simply report initial contacts with odorants, performs some kind of data processing on the incoming odorant signals. Some of these processes include signal separation and enhancement, massively parallel processing, combinatorial coding/decoding, signal inhibition, normalization, convergence and feedback processing. All this, and much other processing that is not yet fully understood, takes place before the signals are compared and matched with stored odor memory data. Thus, the olfactory system can be viewed as an application-specific computing system with a sensing front-end. Artificial nose systems emulate much of the known processing performed by the olfactory system, but with far fewer sensors. As the number of sensors in the front-end of artificial systems grows, it is increasingly necessary to apply the data processing model of olfaction, which is the best chemical/biological sensing system known to us.

Thus, in this chapter we will explore the physiological processes of the biological sense of smell, presenting new insights gained over the last few years through genetic research. We will examine how researchers have translated the model of the olfactory system into artificial noses using a variety of sensing mechanisms, looking specifically at examples of very small artificial noses. We will look also at recent developments in nanometer-scale sensors and discuss how these “nanosensors” might serve in artificial noses. Finally, we present two proposals for integrating nanometer-scale sensors into nose-like sensing systems.

The model for nanosensing employed in this work, the sense of smell, is the oldest of the senses, existing at a time when only primitive multi-celled creatures inhabited the oceans, without sight or hearing. Smell has been the primary sense for many animals, helping them navigate through their environment, directing them to food, finding them mates, and alerting them to danger. Even in humans, for whom vision and hearing have replaced smell as the primary sense, the nose remains a highly sensitive instrument. The olfactory system is also very small. The sensing region in the nose is the size of a large postage stamp, with individual sensing elements on the nanometer scale. These two aspects of the sense of smell, its remarkable effectiveness and its very compact size, are without parallel in any other natural or artificial
sensing system. This makes olfaction an ideal model to guide research and development toward advanced ultra-dense nanosensing systems.

Although the basic physiological processes associated with smell have been known for some time [1], it is only recently that science has begun to understand the more detailed processes at the cellular and molecular level. As a result of a series of relatively recent remarkable discoveries [2–13], we now know that the olfactory system is a sensor system that uses molecular recognition and a combinatorial coding process to recognize a broad range of chemical stimuli. Within just the last five or six years, we have been afforded a detailed view of the olfactory processes from initial molecular recognition and transduction within a neuron [2,14,15], to transmission of the signal all the way to the brain [7,8,11,13,16]. The exploration of olfaction spans many disciplines and is being conducted by scientists from many fields, including biology [6,9,10], physiology [16], biochemistry [2,4,11], bioengineering [17], neuroscience [3,5,8], computer science [18], and mathematics [19].

The detailed insights gained in recent years into the sense of smell might be of mostly academic interest were it not for the new awareness, especially after the terrorist attacks in the Fall of 2001, of the grave threats that chemical and biological weapons pose to our national security. Since that date, increased attention has been focused on developing sensor systems for early detection and warning of toxic agents in the environment [20]. The need is for sensor systems that are at once highly sensitive to minute quantities of individual toxic gases or bioagents, while at the same time capable of detecting and discriminating among a broad range of potentially harmful agents. Sensor size is also of importance, as it would be desirable for a chemical/biological sensing system to be small enough to be placed anywhere without interfering with other activities or the operation of other devices in which it might be embedded. Very small sensors would permit dense sensor placement which could provide populated areas with nearly ubiquitous air testing for toxic agents. Ideally, such a sensor system would be so small that it could be incorporated into everyday objects such as books or clothing, including, for example, a soldier’s fatigues. A soldier could therefore have the protection of an effective sensing system every time he or she gets dressed! Such a sensing system also might be able to recognize humans by their unique odor, as dogs do, providing new capabilities for personal identification. To permit the design of truly nose-like sensor systems such as these, the need for an understanding of the processes of olfaction has gone beyond academic interest to being of fundamental practical importance.

Work on developing larger-scale “artificial noses” or “electronic noses,” as they are commonly known, has been in progress since the early 1980s. (In artificial noses the sensing signal is generally transduced into an electrical signal, leading to the term “electronic noses.”) Initial research focused directly on recreating human odor perception, such as making quality assessments in food, beer, wine, and
perfumes. In the years since then, applications of electronic noses have greatly expanded to include such tasks as monitoring air quality, testing for the presence of explosives and landmines, making medical diagnoses, and, in recent years, detecting toxic chemicals and biological agents. Artificial noses have used a variety of fundamental physico-chemical approaches to sensing. Among them are measurement of conductivity changes in metal oxides and polymers, frequency changes in piezoelectric materials, and color changes in fluorescent optical fibers.

Over the years advances have been made toward developing smaller, more effective artificial noses [21,22]. Today’s electronic nose is much smaller than the gas chromatograph or mass spectrometer used in the past; sensor systems that once were desk-sized objects may now be hand-held devices. Although advances have been made in sensing capabilities, these remain relatively primitive when compared with those of the natural nose. Attempting to enhance these capabilities by more closely approximating natural olfaction, however, would result in significant growth in system size, even if today’s smallest, micron-scale, sensing elements were used. The only way artificial noses can emulate the natural nose more closely, while at the same time becoming smaller, is to develop and apply nanometer-scale sensing. This will require integrating sensors which are only a few nanometers in dimension—i.e., on the molecular scale—into extended but still very small and very dense sensing systems. The application of recent advances in nanometer-scale devices to sensing systems promises the future deployment of highly sensitive, selective, reliable, and ultra-small sensing systems for early warning of chemical or biological weapons attacks. Research in developing nanometer-scale sensing systems, therefore, is of great importance to the national defense, and it is one of the areas of research being sponsored by the National Nanotechnology Initiative (NNI). In addition, the Defense Advanced Research Projects Agency (DARPA) has recently initiated a 5-year program called the MoleSensing Program, which has the goal of producing ultra-dense, ultra-small micron-scale sensing systems with the equivalent of $10^{11}$ sensors per square centimeter!

In subsequent sections of this chapter we examine the operational processes of the premier natural sensing system, the sense of smell, as well as those of existing electronic noses, in order to identify operational principles and lessons for developing artificial nanosensing systems. To that end, Section 2 presents a review of current research into the physiology of the sense of smell. A brief overview of olfaction in mammals is followed by a more detailed look at the constituent parts of the olfactory system and the processes which take place in each component. This section concludes with a summary of “lessons learned” about the key operating principles of natural olfaction that seem to be of particular importance for the design of artificial nose-like nanosensing systems. Section 3 contains a survey of sensing systems...
which attempt to emulate olfaction using various sensing mechanisms. Of particular interest are the smallest electronic noses which could be identified in each category.

Individual nanosensors already exist, and Section 4 identifies some current research and development efforts in the area of nanometer-scale sensors. Recently, extended systems of nanowires also have been developed. These nanowire systems do not perform nose-like sensing, however, but function as nanomemories or as fluidic sensors. No extended systems of nanosensors for nose-like sensing have yet been built. Finally, the chapter concludes, in Section 5, with a discussion on the suitability of individual nanometer-scale sensors for use in gas sensing systems, as well as two proposed designs for electronic noses with nanometer-scale sensing components.

2. The Physiology of the Sense of Smell

To make use of olfaction as a model for nanosensing systems we must understand its components and processes. We will explore these in some detail, but here we begin with a high-level description of what happens when odorant molecules, or what are perceived to be “smells,” arrive at the nose.

The smell sensing process begins when an intake of air sends odorant molecules on a journey up the nasal passages and into the olfactory system. (See Fig. 1.) At the top of the nasal passage the odorant encounters a small region, the epithelium, densely covered with millions of olfactory receptor (OR) neurons. These neurons perform the sensing function; when a receptor neuron is stimulated by interaction with an odorant molecule it generates an electrical signal. The pattern of stimulated receptor neurons effectively forms a “signature” by which an odorant may be identified.

The electrical signals generated by the receptor neurons are processed in several stages; the signals travel first to the olfactory bulb, located in the pre-brain, then to the olfactory cortex, and finally to other regions of the brain. Processing at all of these stages contributes to an analysis of the odorant’s signature based upon the particular set of receptor neurons that have been stimulated. This analysis, along with comparisons with stored memories, results in final odor identification [5,23,24].

Research over the last fifteen years has greatly expanded our insights into what happens at each of these stages of olfactory processing. The following subsections explore, in more detail, the current understanding of each of the components and processes of the olfactory system.
FIG. 1. The human olfactory system. Odorant molecules make contact with the olfactory receptor neurons. The signals from those neurons are sent to the olfactory bulb for pre-processing, and then on to the olfactory cortex and other regions of the brain for further processing. See Fig. 2 for a detailed view of the olfactory epithelium and the olfactory receptor neuron.

2.1 The Olfactory Epithelium

2.1.1 The Neurons of the Epithelium

The olfactory process begins in the olfactory epithelium, where the initial encounter with odorant molecules takes place. The epithelium, which in humans is about 4 cm², contains approximately 10 million neuron cells that are specific to the olfactory system. These olfactory receptor (OR) neurons, or simply “receptor neurons,” host the receptor sites for odorant molecules. (See Fig. 2.) The OR neurons are bipolar neurons with a cell body that is approximately 5 µm in width and that has specialized functional projections on each end. On one end, 10 to 30 hair-like cilia project from the epithelium into the mucus layer. The olfactory receptors, where odorant-receptor binding occurs, are located on the surface of the cilia. At the neu-
FIG. 2. Inside the olfactory epithelium. Receptor sites for odorant molecules are found on the surface of the cilia. Odorant binding with receptors initiates neuronal firing, sending signals along the axons to the olfactory bulb.

ron’s other end, an axon, a long, thin process about 0.2 µm in diameter, projects from the epithelium to the first signal processing center, the olfactory bulb [1,25].

Advances in the techniques of molecular genetics have been especially instrumental in revealing a number of the cellular and molecular mechanisms for olfaction in humans and other animals. Researchers have gained a range of insights into the operation of the olfactory neurons indirectly, by studying the genes which code for their receptors. In 1991, Linda Buck and Richard Axel at Columbia University, in a breakthrough discovery for which they won the 2004 Nobel Prize in physiology or medicine, identified the genes encoding the olfactory receptors [2]. In this study, Buck and Axel found that about 1000 mammalian genes encode 1000 different olfactory receptors, and a single gene may be expressed by thousands of receptor neurons [23]. Since the gene pool in mammals is comprised of approximately 35,000 genes, olfactory genes account for about 3% of the gene pool, making them the largest gene family in the genome [26].

Early in 2002, Xinmin Zhang and Stuart Firestein of Columbia University published the results of their study, using the Celera genome database, of the OR gene family in the mouse [10]. They found 1296 mouse OR genes, of which about 20% were pseudogenes, or non-functioning genes. Thus there are approximately 1000 intact mouse OR genes, each encoding a different receptor. Doron Lancet and his team at the Weizmann Institute in Israel have been working on classifying human OR
genes, and have found only 347 fully intact genes in the set of about 900 human OR
genes [27]. Comparisons of the mouse and human gene repertoires indicate, how-
ever, that the human olfactory system covers much of the same “receptor space” as
that of the mouse. That is, humans are able to perceive the same broad range of odor-
ants, but do not necessarily have the same capability for fine discrimination as does
a mouse [10].

Each receptor neuron hosts many receptor sites on the surface of its cilia. For a
time it was not known if receptors on a single receptor neuron could be of different
types. In 1995, Buck and Axel were able to establish that every receptor neuron has
only one type of receptor [23]. Each receptor neuron recognizes not just a single
odorant, however, but a small range of odorants. In 1998, Firestein and his team
demonstrated that each olfactory receptor responds to a restricted set of odorants
which have similar molecular structures [3]. The set of odorants that are recognized
by the receptor comprises its response profile. The Firestein study also provided
additional support for the one neuron-one receptor-type hypothesis. It was found that
forcing an increase in the number of receptor neurons expressing a particular gene
led to an increase in sensitivity to odorants within the receptor’s response profile, but
not to other odorants.

The relationship between odorants and receptors is many to many. Just as one
receptor neuron responds to a range of odorants with similar molecular structures,
so a single odorant molecule can be sensed by a number of receptor neurons of
different types. Integrating the information from the various receptor neurons which
sense an odorant molecule helps identify that particular odorant. Even small changes
in an odorant’s molecular structure will cause a change in the set of receptors that
bind with it. A change in odor concentration will also cause a change in the set of
activated receptor neurons [5].

In addition to receptor neuron cells, the olfactory epithelium also contains stem
cells and supporting cells. The role of the supporting cells is limited to structural
support of the epithelium and to contributing secretions to the mucus. The stem cells,
also known as basal cells, are the source for new generations of olfactory neurons.
Unlike most other neurons, olfactory neurons have a lifespan of 30 to 60 days before
they are replaced by successor neurons. The stem cells undergo mitosis to generate
new olfactory receptor neurons [1].

Olfactory receptor neurons with the same receptor type are scattered about the
epithelium, but the scattering is not totally random. Separate experiments conducted
by Axel [28,29] and by Buck [30,31], then at Harvard University, show that the ol-
factory epithelium is not monolithic, but can be divided into four distinct regions or
zones. Each of the four zones hosts a different set of olfactory receptor types; recep-
tors with similar amino acid structures tend to be clustered in the same zone [5,16].
Each receptor is randomly distributed across its zone. Therefore sensory information
is divided into 4 subsets of data with individual receptor neurons highly distributed across about 25% of the epithelium.

2.1.2 Transduction

At the heart of the sensing process is the transduction of molecular recognition into an electrochemical signal which is sent to the brain. (See Figs. 3 and 4.) Since the initial 1991 study by Buck and Axel, in which they identified both the olfactory receptor proteins and the genes which encode them, there has been much progress in understanding the molecular interactions which are responsible for transduction \[2,14\].

Incoming odorant molecules first dissolve in the mucus layer covering the epithelium, where they bind with small, water-soluble proteins. These odorant-binding proteins help transfer the odorant through the mucus layer to the odorant receptors, which are located on the surface of the cilia. The odorant receptors are membrane proteins called “7 transmembrane domain G-protein coupled receptors” (GPCRs) \[23,24\]. (They derive their name from passing through the neuron cell membrane...)

![Fig. 3. Sensory transduction in the sensory neuron. Odorant binding with receptors starts a chain of chemical reactions inside the cell. These reactions cause the difference between interior and exterior cell voltages to reach a threshold value, which triggers neuronal firing. This figure is reproduced with permission of Professor Tim Jacob of the University of Cardiff, from his website: “Olfaction: A Tutorial on the Sense of Smell,” http://www.cf.ac.uk/biosi/staff/jacob/teaching/sensory/olfact1.html.](http://www.cf.ac.uk/biosi/staff/jacob/teaching/sensory/olfact1.html)
Odor molecule travels through mucus layer and binds with a G protein-coupled receptor on surface of cilium

Odorant-receptor binding activates an olfactory specific G protein, $G_{olf}$, on the interior of cilium

$G_{olf}$ activates adenylyl cyclase (AC)

AC catalyzes conversion of adenosine triphosphate (ATP) to cyclic AMP (cAMP)

cAMP, the “second messenger,” travels throughout the cell to open ion channels in the cell membrane

Entrance of positive ions into the cell reduces the potential across the membrane (depolarization)

If depolarization reaches threshold, an action potential is generated

**Fig. 4.** Flow chart of the transduction process.
seven times.) When an odorant molecule binds to a receptor, a G protein on the interior of the cilium, $G_{olf}$, is activated. The $G_{olf}$ protein, in turn, activates the enzyme adenyl cyclase (AC), which catalyzes the conversion of the intracellular molecule adenosine triphosphate (ATP) into the neurotransmitter cyclic adenosine monophosphate (cAMP). cAMP, which is called a “second messenger,” now travels throughout the cell, binding with and thus opening cyclic nucleotide-gated (CNG) ion channels in the cell membrane. The open ion channels conduct positive ions, or “cations,” such as Na$^+$ and Ca$^{2+}$, into the cell. In a resting state, the interior of the neuron is negatively charged with respect to the exterior; the voltage difference across the membrane is about $-65$ mV. As the cations flow into the cell, the local cell interior becomes less negative, or “depolarized.” The threshold for signaling is reached when the voltage difference has decreased to about $-45$ mV. At that point, an “action potential” is generated; the local depolarization causes the depolarization to threshold of adjacent resting membranes. In this way the action potential is propagated along the length of the axon into the olfactory bulb. There, the signal is transmitted to the secondary neurons for processing [1,25,32].

This process, from odorant binding to activation of the cAMP second messenger system to generation of an action potential, serves to amplify the signal created by one or a few odorant-receptor binding events. One bound receptor can activate tens of $G_{olf}$ proteins. Each of the $G_{olf}$ proteins activates an AC molecule, and each AC molecule can cause the production of 1000 molecules of cAMP per second. It takes three cAMP molecules to open a CNG channel, and once open, hundreds of thousands of ions can cross the channel into the cell. This amplification process can therefore result in a neuronal firing caused by a single molecular binding event [25]. (It should be noted, however, that a single neuronal firing may not be perceptible in the brain.) The second messenger system also serves to integrate binding events occurring within the same neuron over a short period of time. Thus, it acts as a “counter” for molecular binding events, and in this way provides a measure of odorant concentration [25].

A secondary amplification mechanism also operates to increase cell depolarization. The calcium ions entering the cell activate other ion channels which permit passage of negatively charged chloride ions, Cl$^-$, out of the cell. This egress of negatively charged ions from the cell serves to steepen the voltage gradient, and allows the neuron to reach its firing threshold sooner [25].

There is also a regulatory mechanism that works in conjunction with the second messenger system to enhance receptor sensitivity. This mechanism provides negative feedback to the ion channels. As the concentration of calcium within the cell increases, the calcium acts on the ion channels to desensitize them to cAMP, resulting in the requirement for a stronger odorant stimulus, and consequently more cAMP, to open the channels. This adaptation response allows continuous sensitivity to small changes over a broad range of concentrations [25].
2.2 The Olfactory Bulb

2.2.1 Signal Pre-Processing

Once molecular recognition has occurred and a receptor neuron has fired, its axon transmits this signal to the olfactory bulb (OB), the first olfactory processing center in the brain. As illustrated in Fig. 5, the axons, bundled together in groups of 10–100, penetrate a thin bone, the cribriform plate, before entering the OB. There, the axons project onto the glomeruli, fibrous knots about 50–200 μm in diameter. A glomerulus is made up primarily of the bushy endings, called “dendrites,” of the various neurons which meet there. Within the glomeruli, the axons of the olfactory receptor neurons form synapses with the dendrites of the secondary neurons, the mitral and tufted cells [1,16,25,32]. One of the functions of the OB, therefore, is to act as a relay station for signals going from the epithelium to the brain.

While receptor neurons of the same type are distributed randomly about a region of the epithelium, order emerges as the axons make their way to the olfactory bulb. Peter Mombaerts and Fan Wang, both then members of Axel’s group at Columbia University, were able to use gene targeting to visualize axons projecting from the epithelium to the bulb. In separate studies, they showed that all neurons expressing a

![Fig. 5. Inside the olfactory bulb. The electrical signals generated in the epithelium travel down the axons to the olfactory bulb. In the bulb, the axons make contact, within the glomeruli, with secondary neurons, which transmit the signal further into the brain.](image-url)
given receptor gene project their axons onto the same two glomeruli, symmetrically opposite each other on the two halves of the olfactory bulb [4,33]. This creates a fixed sensory map in the olfactory bulb for each set of stimulated receptor neurons, and thus for each odorant.

All neurons expressing a given receptor gene project to a single glomerulus, and they are the only ones projecting to that glomerulus. Kensaku Mori and fellow researchers at the University of Tokyo were able to show that each glomerulus receives input from only a single type of odorant receptor, and that each receptor binds with a limited range of odor molecules with similar molecular structures [16]. Therefore each glomerulus, as well as the mitral and tufted cells which send their dendrites to that glomerulus, is “tuned” to specific molecular features, and these features comprise the molecular receptive range for the glomerular cluster. As each glomerulus is stimulated, it “identifies” particular features of the odorant being sensed, and the pattern of activated glomeruli presents a composite picture of the odorant’s characteristics [16].

There are about twice as many glomeruli as there are different types of receptor neurons. The mouse olfactory bulb, for example, contains approximately 900 types of receptor neurons and 1800 glomeruli [23]. There are thousands of neurons of each receptor type, and their axons all converge onto two glomeruli, each with only 5–25 mitral cells which receive and pass on the signal [25]. This high degree of signal convergence serves two functions; it amplifies sparse signals, increasing the sensitivity of the nose to low concentrations of odorants, and it increases the signal-to-noise ratio in a noisy odor environment.

Additional signal processing is also performed by the olfactory bulb’s local neuronal circuitry. Periglomerular cells and granule cells form connections between the mitral and tufted cells of one glomerular module and those of neighboring modules [1]. The periglomular cells have short bushy dendrites which spread throughout a glomerulus, and a short axon which extends to a radius of about 5 neighboring glomeruli. Granule cells are axonless and form interconnections between mitral and tufted cells. The functions of this local circuitry are less well known, but it does appear that they synchronize neuronal discharges from cells that belong to different glomerular modules, which may converge onto a single target neuron in the olfactory cortex [16,34]. The synchronized discharges may serve as a mechanism for combining signals, generated simultaneously or within a small time interval, from two different neurons in the olfactory bulb.

Different theories have been proposed about other functions of the interglomerular neurons. Mori and others have hypothesized that these neurons are responsible for lateral inhibition, a process in which neuronal activity in an individual cell is inhibited when neighboring cells are activated [16]. If the molecular receptive ranges of neighboring neurons are closely related, lateral inhibition would amplify local dif-
ferences, the equivalent of edge enhancement in vision processing. This would serve to enhance the tuning specificity of individual neurons.

Gilles Laurent, at the California Institute of Technology (Caltech), challenges this interpretation, positing instead a systems oriented viewpoint [34]. Laurent proposes that lateral inhibition may serve to eliminate redundant information and possibly perform other data optimization, thus sharpening the tuning curves of the system as a whole rather than those of single neurons. Laurent points out that there is no convincing evidence that the position of a neuron is a factor in olfactory processing, as it is in vision, where light pixels close to dark ones may make an increase in contrast desirable. Instead, it is the identity of the neuron which is of importance. This calls into question the benefit to be gained by more strongly distinguishing the signal of a neuron from those of its close neighbors. Additional investigations into the processes of olfactory bulb local circuitry will be needed before its functions can be clearly established.

The olfactory neurons, unlike almost all other neurons, are short-lived. Individual neurons are replaced every 30 to 60 days, and the axons of replacement neurons must find the way to their respective glomerulus, one among thousands, on a daily basis [1,14,35]. In 1998, Axel and others speculated that finding the path to the right glomerulus was an inborn characteristic [4]. They hypothesized that the OR gene is not only expressed on the cilia, providing molecular binding sites, but may also be expressed on the axon, where it would serve to help guide the axon to its designated glomerulus. Additionally, the OR gene may need to be complemented by other guidance receptors in order to form a complete set of directions which takes the axon to its destination.

This hypothesis was verified in a 2001 study of the Drosophila (fruit fly) olfactory system by Liqun Luo at Stanford University [7]. Luo and his team demonstrated that the information which leads pairs of neurons to make the correct intersections, or synapses, appears to be part of their programming from birth. The axons of all neurons born with the same OR type found their way to the same glomerulus, no matter where in the epithelium they were located. The OR gene thus controls not only the type of receptor expressed on a neuron, but also the destination for that neuron’s signals. This result, along with other recent research, suggests a high degree of genetic programming, or hardwiring, in the olfactory system [8].

### 2.2.2 Zones in the Olfactory Bulb

The zonal organization of the epithelium (discussed in Section 2.1.1) is carried forward to a corresponding zonal structure in the olfactory bulb; axons maintain zonal segregation as they project into the olfactory bulb [16]. Just as neurons with similar amino acid structures tend to be clustered in the same epithelial zone, glomeruli
which have overlapping specificities also tend to be clustered in the same area within a bulb zone. Thus, the spatial arrangement of activated glomeruli creates a characteristic map in the olfactory bulb.

Recent studies also have shown that the positions of the glomeruli are fixed within a species; therefore all animals in a species will have the same brain activity pattern in response to a particular odorant [4]. Since each glomerulus receives input from a limited set of receptor neurons, and the positions of the glomeruli are topologically fixed, this generates in the olfactory bulb a two-dimensional map of stimulated receptors [8]. A “combinatorial” coding scheme for identification of odorants seems to be operational here; the particular combination of glomeruli and receptors that have been stimulated form an identification code, or signature, for that odorant.

It is important to recognize the distinction between the physical location and the “identity” of a glomerulus, as defined by its input receptor neuron type. In certain cases, physical location may be important. All animals use odor data to reveal their environment, and the speed with which that information is decoded may be of critical importance for survival. Thus, it may be that certain odors activate glomeruli whose location matters. For example, these glomeruli may provide greater speed in transmitting certain signals from the nose to the brain to reduce reaction times. In general, however, there is no data to show that the position of a glomerulus, as opposed to its identity, is of importance in the downstream processing of odor information. As Laurent points out, whether and how the brain makes use of receptor and glomerular position is still an open question [34].

2.2.3 Temporal Aspects of Signaling

The role of the timing of neuron signals in odorant identification has been the subject of numerous studies in recent years [9,16,34,36–46]. This temporal aspect of neuronal activity includes fast oscillatory synchronization of pulses as well as slow patterning, which occurs over hundreds of milliseconds. Information required for the identification of at least some odorants appears to be contained in the timing of these action potentials. In particular, it appears that the olfactory system uses synchronization and patterning of pulses in at least two ways: to determine odor concentration, and to separate constituents within a blend of odors [45].

The temporal aspects of olfaction have been of particular interest to Laurent, whose laboratory at Caltech has been a center of research in this area. In 1997, Laurent and his colleague Mark Stopfer demonstrated that synchronization was necessary for odor identification. Specifically, they found that induced desynchronization of signals impaired the ability of the olfactory system to discriminate between similar odors [39].

In studies of the olfactory system of the zebrafish, Laurent and his colleague Rainer Friedrich found that odorant stimulation is represented in the olfactory bulb
by a pattern of activity across many mitral cells [9]. The mitral cell activity pattern for an odor thus changes continuously over the stimulus period. Laurent and Friedrich found that the slow temporal patterning in the zebrafish olfactory bulb reduced the similarity in patterns for related odors. Over time, this made each odor’s representation in the olfactory bulb more specific, and served to “tune” the responding mitral cells as an ensemble rather than as individual neurons.

In 2000, Thomas Christensen and John Hildebrand of the University of Arizona were able to show, in a study of the moth olfactory system, that neuron firing synchrony depends strongly on contextual variables such as odor intensity and different pulsing patterns used to inject the odor stimulus into the olfactory system under study [42]. Their study also showed that the temporal behavior of the neurons for a blend of odors could not be predicted from the behaviors of the neurons for the constituent odors. This means that the reaction of the olfactory system to an odorant blend is not simply a “sum” of the reactions to the individual odor components.

Synchrony and timing of odor signaling is not as well understood as the functioning of the individual olfactory receptor neurons. Temporal aspects of odor processing are clearly important, but a better understanding of their function within the olfactory system awaits additional investigation. The following section presents the hypotheses of various researchers about how spatial (which refers to neuronal receptor type) and temporal processing, as currently understood, might contribute to odorant identification.

### 2.3 Odor Identification

After the olfactory bulb, the next step in odor identification is signal processing within the olfactory cortex, which is in the brain. Here, the complexities of brain functioning make further explorations extremely difficult. Nonetheless, the understanding that has emerged most clearly is, as explained above, that the olfactory system makes use of a spatio-temporal coding system for odorant identification. Buck and her team at Harvard Medical School investigated the spatial aspects of the code, and in March 1999, Buck and Bettina Malnic were able to prove that different odorants are recognized by different combinations of olfactory receptors [5]. Even slight variations in the odorant or in its concentration changed the make-up of the set of receptors which recognized the odorant. This finding demonstrated that the brain makes use of a combinatorial coding scheme, in which each odorant has an associated receptor code or signature.

In November 2001, Buck and Zhihua Zou were successful in marking the pathway of neurons all the way from the nose to the brain, and showed that there is predetermined order in the projection of neurons from bulb to cortex [8]. Zou, along with Lisa Horowitz, inserted a transneuronal tracer called barley lectin (BL) into two
odor receptor genes, M5 and M50, expressed in two different zones in the olfactory epithelium. The BL marker is transferred across synapses to connecting neurons, labeling chains of connected neurons. This allows tracing of an entire neuronal route.

What Buck’s team found was that marked neurons projected a route to one or two glomeruli symmetrically located on each side of the olfactory bulb, and from there to several clusters of neurons in the olfactory cortex. Signals from a particular receptor neuron cause the excitation of specific clusters of cortical neurons, demonstrating the existence of a stereotyped, or fixed, sensory map in the olfactory cortex. Furthermore, the neuron clusters in the cortex were not randomly distributed. The distribution pattern for a given receptor type was the same for all mice with the engineered M5 and M50 genes. This similarity in the cortical sensory map across a species also offers an explanation for commonality in odor perception; skunk odor is repellent to everyone, while most people enjoy the smell of lavender or roses.

The research of Buck et al. also yielded other significant results, including indications of both the divergence and convergence of signals into the cortex. It appears that signals from different receptor types first converge in the olfactory cortex. Initial signals from the epithelium to the olfactory bulb are segregated by receptor type. Neuronal projections from the bulb to the cortex do not maintain this segregation, however, since inputs from different olfactory receptors map onto partially overlapping neuronal clusters in the cortex. Individual cortical neurons seem to receive inputs from up to 50 different types of receptors. Since an odorant is recognized by a specific combination of receptor neurons, the convergence of signals from different receptors in the cortex may reflect the initial integration of the various signals which make up the “code” for that odorant.

The convergence of signals in the olfactory cortex leaves open the possibility that much information from individual neurons is lost. If this were the case, however, it would render useless the previously described careful segregation of signals by receptor type, and that would run counter to nature’s tendency to maximize the efficiency of such highly evolved systems. The solution to this puzzle may lie in the second component of the coding scheme, which is temporal signaling. If we include neuronal timing, synchronization, and patterning of an ensemble of signals in the coding scheme, along with neuronal identity, then the coding scheme becomes capable of representations of complex information using far fewer neurons at higher processing levels than were required at lower levels.

Laurent and his colleagues at Caltech have conducted numerous studies on the timing patterns which emerge as olfactory signals are sent from the bulb to the brain [9,36–39,41,43]. As discussed previously, early in their investigations they demonstrated that the synchronization of signals was essential for fine odor identification [39]. They also demonstrated a convergence of signals in the cortex, as they found that information encoded in the timing of spikes across an assembly of
neurons converges onto single neurons in the cortex [37]. From these findings, Laurent infers that the brain can “reconstruct” an odor from the information contained in the spike trains of this neural assembly.

A 2002 study of the insect (locust) olfactory system by Laurent and Perez-Orive confirmed again the convergence of signals in the cortex [43]. They showed that while signals in the antennal lobe (the equivalent of the olfactory bulb in mammals) were dense and seemingly redundant, they were sparse in the mushroom body (equivalent to the cortex) and carried by more selective neurons. There may be advantages to a sparser representation of odors, including a reduction in overlaps between individual odor representations, and simpler comparisons between stimulus-evoked patterns and stored memories.

In the systems viewpoint adopted by Laurent and others, olfactory processing is modeled by positioning the odorant within a coding space defined by the features of the odorants [38]. Odorants are mapped to a position in such an “odorant space” based on their spatio-temporal pattern of activated neurons. One round of processing performs “decorrelation,” in which the overlap between representations of related odors is reduced. This is the equivalent of spreading out the representations within the odor space. A concurrent or closely following set of processing is responsible for “sparsening,” or compression of the representations into only a few active neurons. The latter results in an increase in specificity for individual neuronal responses.

Final transmission of olfactory signals to the frontal cortex regions suggests the brain may perform parallel processing of olfactory inputs. The olfactory cortex is composed of several anatomically distinct areas which may have different functions. Buck and Zou found that the labeled neurons projected to clusters in most of these cortical areas [8]. Since each area in the olfactory cortex can transmit information to different frontal cortex regions, this suggests that information from the same olfactory neuron eventually may be transmitted to different regions of the cerebral cortex. The divergence of signals in the cortex would allow the same input data to be organized and processed in several different ways, aiding in correct identification of the odorant.

In summary, the research just discussed suggests that the odor processing appears to rely initially on careful segregation of signals from individual neurons. Individual information streams are maintained for each neuron from the epithelium to the olfactory bulb. Additional “temporal” information, timing and synchrony of signals, appears to be generated within the bulb. This temporal information may be based upon the incoming “spatial” information, which conveys the identity of the particular combination of stimulated neurons. The additional temporal information appears to be incorporated into the data flow as the signals travel from the bulb to the olfactory cortex. Within the olfactory cortex, spatial data are compressed, as multiple signals from the bulb converge onto single neurons in the olfactory cortex. If infor-
mation is not lost at this stage, it seems likely that temporal data now contain some of the information which would otherwise be lost in convergence of spatial data. Finally, information is sent from the olfactory cortex to many other regions of the brain, which appears to perform parallel processing to arrive at final odor identification. The high-level steps within the process of olfaction are summarized within the flow chart of olfactory processes, presented in Fig. 6.

### 2.4 Characterizing Odor Space

If we view olfactory identification as a representation of an odor within an odor space, and consider the odor space to be defined by a set of odorant features, the question remains what those features might be. The difficulty can be seen when olfaction is compared with the visual system or the auditory system. In vision, retinal cells are activated by light over a range of wavelengths. The wavelength of the incident light elicits a particular response in the photoreceptor. Similarly, audio receptors are activated by a range of sound frequencies. There is no similar continuum of physical descriptors which can be used to characterize odorants and their receptors [6,25,44].

One approach scientists have used to characterize the molecular receptive range of a receptor is pharmacological, based upon medicinal chemistry. Since each receptor binds with a small range of odorant molecules, one could establish the molecular characteristics that are common to that set of odorants. Stuart Firestein and Ricardo Arenada at Columbia University used this approach when they attempted to define the molecular range of a particular olfactory receptor known as the “I7” receptor [6]. They found an odorant molecule binding with the I7 receptor must have two specific characteristics: it must have an aldehyde carbonyl, and it must be of length between 8 Å (one Å, or “angstrom,” is 10⁻¹⁰ meters) with at least seven carbons in the backbone, and 12 Å with no more than eleven carbons in the backbone. They found odorants binding to I7 subject to very strict constraints on molecular structure at the carbonyl head of the molecule, while allowing a wide range of variations at the tail end. Extrapolating this result to other receptors would indicate that a receptor may be highly specific for a particular structural component, while being largely indiscriminate for other components of an odorant molecule. This provides us with a general model of receptor affinities: a narrow spectrum of specific characteristics combined with a wide tolerance band for other characteristics. The brain then integrates input from the receptor neurons to generate a representation of an odorant’s molecular composition and structure. As higher-order centers of the brain integrate input from a number of receptor neurons, they would be able to distinguish a wide variety of odors, while at the same time they are able to discriminate among odors with only subtle differences [25].
An odorant molecule binds with a receptor protein within the olfactory epithelium

The binding event is transduced into an electrochemical signal (See the Transduction Flow Chart, Figure 4)

The combination of stimulated receptor types forms an identifying “signature” for the odorant; signals segregated by receptor type transmit the signature to the olfactory bulb

Within the Olfactory Bulb:
Signals from receptor neurons are relayed to secondary neurons
Other intra-bulb processing occurs, possibly converting some spatial into temporal data

The processed signature is transmitted by the secondary neurons to the olfactory cortex, where the signals converge onto far fewer neurons

Odorant signature is transmitted to other regions of the cortex; this parallel processing includes signature comparison with stored memories

The odor is recognized and identified

FIG. 6. Flow chart of the olfactory process.
2.5 Key Operational Principles of Natural olfaction

The olfactory system is a remarkably effective sensing system. With organs that have, in humans, dimensions of only a few centimeters, and using the nanometer-scale transduction process of molecular recognition, the olfactory system is able to recognize and distinguish, with great accuracy, many thousands of odorants. In the preceding subsections we have examined, in some detail, what is now known about the components and operations of the natural nose and sense of smell. Here we present a summary of that examination, and conclude by identifying what appear to be the major operational principles of natural olfaction that have import for the design of artificial analogs.

The model the olfactory system presents has three main components: the sample handling system, the sensing system, and the signal processor. Sample handling in the nose is performed by the physical nasal structure as it ensures an even airflow of odorant molecules across the epithelium, and then allows an influx of fresh air to perform a “wipe-clean” function, as molecules initially adsorbed to the surface are desorbed again. This clears the receptors, and thus resets the system [24].

The other two components of olfaction, sensing and signal processing, are not wholly separable. As discussed earlier in this chapter, information required for odorant identification is contained in assemblies of stimulated neurons and their interrelationships, rather than within a single neuron [34]. This means that the sensing function itself also provides initial signal processing, as, for example, in highly parallel neuronal sensing wherein the convergence of signals improves signal-to-noise ratios. In spite of this overlap it is possible to characterize broadly both the sensing and signal processing systems. The odorants arrive at the epithelium in an air flow, and are sensed, via molecular recognition, by a multitude of tiny neuronal sensors. Each neuron hosts many protein-receptors, all of the same type. The receptors bind with a limited range of odorant molecules that have similar molecular structures. In this way, each neuron acts as a sensor for a set of molecules that have one or more common molecular features. An incoming odorant will bind with a diverse set of neurons, each of which registers a particular feature of the odorant’s molecular structure. Thus, the range of molecules to which a neuron is receptive will generally overlap with the ranges of many other, different, neurons. The sensing function of the epithelial neurons results, finally, in the activation of a set of neurons which provides an initial signature for that odorant.

Olfactory signal processing is highly complex, and not yet well understood. Researchers are in agreement, however, that data processing makes use of a spatio-temporal combinatorial system and also draws on stored memories in order to complete odorant identification. Signal processing begins with the sensing process, as an encounter with an odorant produces a spatial “map” of activated neurons in the
epithelium. The neuronal signals then are projected into a corresponding map of activated glomeruli in the olfactory bulb. This glomerular map preserves the signature generated by the activated epithelial neurons. Further processing in the bulb and the cortex is not well understood, but it appears that intra-bulbar neurons may refine the signature received in the bulb through temporal (signal timing and synchrony) processing. In subsequent data transmission to the olfactory cortex, signals from different glomeruli, and hence from different neurons, are combined. Signal convergence in the olfactory cortex may serve as a step to providing the composite picture of the odorant required for identification.

Although the path of olfactory signals from epithelium to olfactory bulb to olfactory cortex has been identified, other areas of the brain also participate in olfactory signal processing [1]. Signals are sent from the olfactory bulb to several different areas of the olfactory cortex as well as to subcortical limbic structures, whose functions include memory processing. In addition, signals may be projected from one region of the olfactory cortex to another, from the olfactory cortex to the neocortex, and even from the olfactory cortex back to the olfactory bulb. The processing of these olfactory signals in various regions of the brain is not yet understood.

Thus, from our current knowledge of olfaction we can identify the following important features of the sense of smell, which appear to be among the most significant operating principles of olfaction:

- **Operation of vast numbers of very small, densely spaced, sensing neurons.**
- **Sensing neurons which are non-specific sensors, in that each neuron is sensitive to a small range of odor molecules that have one (or a few) common structural features.**
- **The sensing range of an individual neuron overlaps with those of other neurons; each neuron senses different features of a particular molecule.**
- **Repetition of the same sensing neuron many thousands of times within the epithelium (sensing platform), serving to increase the likelihood of odorant binding with receptive neurons and amplifying sparse signals.**
- **Odorant identification using, along with stored odor memories, a spatio-temporal combinatorial system, in which activated sensor-neurons provide a “signature” which can be used to identify the odor.**

Although our knowledge of the sense of smell and its operations is far from complete, these features of the olfactory processes are likely to serve us well as we move on to consider artificial analogs of the nose and its sense of smell.
3. Electronic Noses: Chemical Sensing Systems

Now that we have completed a review of the operations of the natural nose, we will examine some of the diverse artificial chemical sensing systems which have been built in an attempt to imitate the nose. Artificial noses, usually known as “electronic noses,” have been designed to sense and identify odorants, or volatile organic compounds, in ambient air. Electronic noses differ from other sensing systems in that they are meant to operate in the noisy sensory environment of open air rather than in the laboratory. Sensing the presence or absence of a particular substance within a controlled sample of air (or liquid) in a laboratory is a much simpler problem than detecting the presence of that substance in ambient air or water. Sensing elements that react to a particular chemical may react, as well, to many other substances with similar chemical structures. Thus, a single sensing element, viewed in isolation, is likely to produce many false positive readings, and may not correctly identify a particular chemical when it is part of a mixture. The sensing system of the natural nose, in contrast, is not only highly sensitive, but it is capable also of remarkable discrimination among competing odors. Someone cooking dinner, for example, can smell the odors of the herbs and spices and oils of a simmering stew, but at the same time is able to isolate the odor associated with natural gas which warns of a gas leak. A wine connoisseur can pick up subtle differences in wines just by smelling them. Thus, the model provided by the natural nose is ideal for creating artificial sensing systems which operate effectively in the noisy natural environment.

Research on artificial noses is not new, but has been ongoing since the early 1980s [47]. Initial research efforts were geared toward artificial noses for use in the food and cosmetic industries, where they are still widely used to provide quality assessments in food production, flavor control, and quality grading of wines and beers. Electronic noses are used also in environmental monitoring for identification of toxic wastes, testing of ground water for hazardous chemicals, and monitoring of air quality and industrial emissions. Recently, progress has been made also in applications to health monitoring and medical diagnostics [48].

Electronic noses, as do natural noses, perform sample handling, sensing, and signal processing. The sample handling function provides an air-flow to the sensing elements, and, after sensing has occurred, must “wipe-clean” the sensing elements and bring them back to a known resting state. Some electronic noses may revert to a resting state simply by exposure to air, as the molecules initially adsorbed to the surface are desorbed again. Others may require exposure to a “washing gas” such as alcohol vapor, followed by a reference gas which drives the sensors back to their resting state [21].

Natural olfaction uses vast numbers of sensing neurons—tiny, non-specific sensors—to perform the sensing function. Thus, the olfactory sensing mechanism
employs what can be considered to be a multitude of tiny gas sensors. The electronic noses reviewed here follow this model by using arrays of cross-reactive (non-specific) gas sensors. (Although there are also electronic noses using a single gas sensor, such as gas chromatography columns, only electronic noses using arrays of sensors will be considered here.) These electronic noses may be classified by sensor type, which results in the following major categories [21,49,50]:

- Electrochemical sensors:
  - conductance-based sensors: metal-oxide and conducting polymers;
  - potentiometric sensors: chemically sensitive field-effect transistors (FETs).
- Mass-change (piezoelectric) sensors:
  - quartz crystal microbalance (QCM);
  - surface acoustic wave (SAW) devices.
- Optical sensors:
  - fluorescent optical fibers;
  - colorimetric.
- Other artificial nose systems.

Table I presents a survey of electronic noses already developed or under development in each of the above-defined categories.

As described in Section 2, olfactory signal processing makes use of both spatial information (neuron type) as well as temporal information (timing and synchrony of signals). Today’s electronic noses, in contrast, apply a simplified version of this data processing model, using only spatial information to make identifications. The type of sensor element that has been activated by an odorant, as well as the strength of its response, become inputs to the data model. Analysis methods commonly used include multivariate data analyzes, such as principal component analysis (PCA) or discriminant function analysis (DFA), or a non-parametric method such as artificial neural networks (ANNs) [50]. Olfactory-like signal processing for electronic noses has been an area of interest to many researchers, both here and abroad, for several years [18,19,51–58].

A review of the literature was conducted to identify examples of electronic noses using each type of gas sensor, with particular attention to electronic noses whose size is on the micrometer or nanometer scale. The search was conducted primarily within the Journals and Proceedings of the Institute for Electrical and Electronics Engineers (IEEE), as found online in IEEE Xplore journal and conference paper repository, the journals of the American Chemical Society, Science, Nature, Sensors and Actuators A and B, other databases of scientific journals, the Handbook of Machine Olfaction [50], and on the wider Internet. It should be noted that the specific examples of
# Table I

## AN OVERVIEW OF SELECTED R&D INTO ELECTRONIC NOSES

<table>
<thead>
<tr>
<th>Sensor type</th>
<th>Description</th>
<th>Researcher/Developer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conductance-based metal oxide</td>
<td>Sixteen tin dioxide sensors on a 4 × 4 mm² chip, for sensing of all gases except nitrogen and the noble gases</td>
<td>Joachim Goschnick et al., Karlsruhe Research Center, Karlsruhe, Germany</td>
</tr>
<tr>
<td>Conductance-based metal oxide</td>
<td>Ten tin dioxide sensors on a 1.1 × 1.2 cm² substrate; senses combustible and explosive gases</td>
<td>Dai-Sik Lee, Duk-Dong Lee, et al. Telecomm. Research Institute (ETRI) Taejon, Korea</td>
</tr>
<tr>
<td>Conductance-based carbon black composites</td>
<td>Commercial production of hand-held devices using carbon black-composite sensors; can be used for sensing chemical and biological weapons agents</td>
<td>Smiths Detection, Pasadena, Inc. (formerly Cyrano Sciences, Inc.) 73 N. Vinedo Avenue, Pasadena, California 91107</td>
</tr>
<tr>
<td>Potentiometric—ChemFETs</td>
<td>Research toward development of a chemical sensing “nose on a chip”</td>
<td>Julian Gardner, University of Warwick; Tim Pearce, University of Leicester; Alister Hamilton, University of Edinburgh</td>
</tr>
<tr>
<td>Mass-change surface acoustic wave</td>
<td>Array of eight cantilevers, each approximately 500 × 100 × 1 µm³, for gas sensing in ambient air and biosensing in solution</td>
<td>Christoph Gerber, et al. IBM Research Division, Zurich Research Laboratory</td>
</tr>
<tr>
<td>Mass-change surface acoustic wave</td>
<td>Array of cantilevers on a chip approximately 8 × 2 mm² for biosensing in ambient air</td>
<td>William Hunt, et al. Georgia Institute of Technology</td>
</tr>
<tr>
<td>Mass-change surface acoustic wave</td>
<td>Commercial production of hand-held devices using SAW sensors; used for chemical weapons agents and other toxic chemicals</td>
<td>Microsensor Systems Inc. 62 Corporate Court, Bowling Green, KY 42103</td>
</tr>
<tr>
<td>Optical</td>
<td>Fluorescent beads, each 3 µm in diameter, in microwells on the tips of optical fibers; demonstrated for sensing of gases, including nitroaromatic compounds</td>
<td>David Walt et al. Tufts University, Tim Pearce, University of Leicester</td>
</tr>
<tr>
<td>Optical</td>
<td>Optical fiber sensors demonstrated to detect TNT in landmines</td>
<td>John Kauer, Joel White, Tufts University of Medicine</td>
</tr>
<tr>
<td>Optical</td>
<td>Commercial production of biosensors in a BeadArray™, used for genotyping and gene expression profiling</td>
<td>Illumina. Inc. San Diego, California</td>
</tr>
<tr>
<td>Optical-colorimetric</td>
<td>Color change in metalloporphyrins upon exposure to toxic gases</td>
<td>Neal Rakow, Keneth Suslick University of Illinois at Urbana-Champagne</td>
</tr>
<tr>
<td>Combination of mass-change, capacitive, and calorimetric sensors</td>
<td>Sensor system on a 7 × 7 mm² chip, integrates 3 different gas sensors, a temperature sensor, plus microelectronic and micromechanical components; demonstrated to sense ethanol and toluene</td>
<td>A. Hierlemann et al. Physical Electronics Laboratory Swiss Federal Institute of Technology, Zurich</td>
</tr>
</tbody>
</table>
TOWARD NANOMETER-SCALE SENSING SYSTEMS

Electronic noses presented here are not an exhaustive list of all such sensing systems, but are representative of the smallest electronic noses found in each category.

3.1 Electrochemical Sensors

Electrochemical sensors demonstrate detection and recognition of a chemical by a change in conductance, resistance, or electrical potential. There is little practical difference between sensors that measure conductance and those which measure its inverse, resistance, so these two sensor types will be discussed together as “Conductance-Based Sensors.” Field-effect transistors (FETs), such as metal-oxide-semiconductor FETs (MOSFETs) and other chemically sensitive FETs (ChemFETs) are the most common chemical sensors that operate via measurement of change in the potential [21,49].

3.1.1 Conductance-Based Sensors

Conductance-based sensors were among the earliest sensor types developed by pioneer researchers in this field in the early 1980s [21]. In conductance-based sensors, an active material, which may be either a metal oxide or a conducting polymer, is deposited between two metal contacts, as in Fig. 7. Binding of a target agent, a volatile organic compound, with the sensing platform causes a change in resistance between the metal contacts. This change in resistance is proportional to the concentration of the organic compound, and thus the sensor provides an indication of both presence and quantity of the target agent.

3.1.1.1 Metal Oxide Sensors.

Metal oxide-based sensors use a variety of different oxides, most commonly tin, but also zinc, titanium, tungsten, or irridium, and they are usually doped with platinum or palladium. These sensors also require the incorporation of a resistive heating element, because they operate at temperatures between 200 and 400 °C. The sensing platform can be designed using different combinations of metal-oxide and dopant to respond to specific organic compounds. Once the sensor is built, the response selectivity can be adjusted further by changing the temperature. The sensitivity for metal oxide-based sensors ranges from 5 to 500 parts per million [21].

The high operating temperature of metal oxide sensors is their biggest drawback, because it adds a requirement for a heating element and contributes to heat dissipation problems as well. The baseline sensor response also tends to drift with time, and the sensors can be rendered useless by irreversible binding with sulfur compounds. These drawbacks are offset, however, by the relative ease and low cost of their manufacture. Thus, metal oxide sensors have been the most commonly used gas sensors.
Electronic nose research began with metal oxide sensors in the early 1980s [21], and since that time the sensing elements in these devices have been shrinking steadily down to the micron scale. Joachim Goschnick and a team from the Karlsruhe Research Center in Karlsruhe, Germany, have been working for several years on a metal oxide electronic nose they call the “KAMINA,” which stands for “Karlsruhe Mikro-Nase” (Karlsruhe Micro-Nose). KAMINA, which is shown in Fig. 8, is designed to monitor indoor air-quality by sensing gases such as formaldehyde, carbon monoxide, and ammonia [59–61]. KAMINA features 38 or 16 sensors in micro-arrays of dimensions of $10 \times 11 \text{ mm}^2$ and $4 \times 4 \text{ mm}^2$. A micro-array is produced by partitioning a monolithic tin dioxide ($\text{SnO}_2$) or tungsten trioxide ($\text{WO}_3$) layer with very narrow parallel electrode strips made of platinum. Differentiation between the individual segments in the array is caused by temperature gradients and by the difference in thickness, from 2–20 nm, of a gas-permeable silicon dioxide ($\text{SiO}_2$) membrane applied over the metal oxide layer. KAMINA was able to detect some gases, including carbon monoxide (CO), at less than 1 ppm. A KAMINA sensing system that can be used for smoke detection, air quality monitoring and detection of nearly all gases, is available commercially from SPECS Scientific Instruments, Inc. of Sarasota, Florida [59]. Earlier versions of KAMINA were about twice the size of a cell phone. The present version of KAMINA incorporates a sensing chip and all processing elements in a device approximately half the size of a credit card.

Another research effort is that at the Electronics and Telecommunications Research Institute (ETRI) in Taejon, Korea, where a team led by Dae-Sik Lee and Duk-Dong Lee built a tin dioxide sensor array to sense combustible and explosive
gases such as methane, propane, butane, and CO [62]. The ETRI team enhanced the sensitivity of the SnO$_2$ used as the base material in the sensors by employing calcium (Ca) and platinum (Pt) catalysts to reduce the size of the SnO$_2$ grains to 8 nanometers, and thereby increased the overall surface area and reactivity. This material then was enhanced with 10 different additives, modifying the sensitivity spectrum of the base material to create 10 different sensors which were placed on a 1.1 × 1.2 cm$^2$ alumina substrate. The 10-element sensing array was used in combination with a neural network analyzer to classify the kind of gas detected, and a neurofuzzy network to determine gas concentration values. This sensing system accurately recognized target gases and determined continuous concentration levels, at target concentrations of several hundred to several thousand parts per million, with error rates of less than 5%.

3.1.1.2 Conducting Polymers. Early versions of metal-oxide based sensors were built with macro-scale sensing arrays, and only more recently were they shrunk down to the micron scale. For conducting polymers to work, however, the sensor size must be on the micron scale. The two electrodes which experience the change in resistance are separated by a gap of only 10 to 20 µm, and the entire sensor surface area is less than 1 square millimeter [50]. Conducting polymer sensors have sensitivities of 0.1 to 100 ppm [21]. As with metal oxide arrays, varying the composition of individual sensors leads to a different sensitivity spectrum for each sensor.
Conducting polymer sensors operate at room temperature, so there is no need for heating elements; this makes for easier manufacturing. On the other hand, the fabrication process itself is more complex. The polymer bridges between the electrodes are formed in layers using a process called “electropolymerization,” in which there is cycling of the voltage between the electrodes. The voltage sweeps act to deposit layers of polymer, and by varying the voltage sweep rate one can create a variety of different active materials, each engineered to sense a particular volatile organic compound (VOC). Unlike with metal-oxide sensors, however, the surface morphology of the conducting polymers is not predictable. As a result, the surface conductivity and, therefore, the sensing function is not entirely reproducible from batch to batch [21,49].

A number of naturally conducting polymers such as polypyrrole and electroconductive conjugated polymers such as polyaniline, polythiophene, and polycarbonate are used in electronic noses. All of these polymers are highly sensitive to gases and vapors, and experience a marked change in conductance upon the binding of the target analyte. It is also possible to make non-conducting materials such as silicon and polystyrene conductive by adding nanometer-sized particles of carbon called “carbon black” [50]. Unlike other nanomaterials, carbon black is very inexpensive and available in large quantities, which make polymer-carbon black composites attractive sensing elements.

The biggest operational drawback of conducting polymer is its high sensitivity to water vapor. This means that excess humidity can mask responses to VOCs. These sensors are also prone to drift over time, and in addition, they may encounter VOCs which penetrate the polymer base. Removing these VOCs from the polymer stretches out the recovery time, thus lengthening sensor cycle time [21].

Caltech has been a center of research in the use of conducting polymers in microsensors for electronic noses. Chemist Nathan Lewis and his team members began their research in 1992 with the exploration of organic conducting polymers for use in nose-like sensors [63]. The drawback of the polyacetylenes they used originally was that, upon exposure to air, they lost their conductivity in only a few hours. Michael Freund, one of Lewis’ postdocs, then experimented with polypyrrole, which did not deteriorate in the same way. Freund found that if he mixed different kinds of insulators into the polypyrrole, the resultant mixtures had different sensitivities to various solvents such as methanol, ethanol, acetone and benzene. Finally, Lewis and fellow chemist Robert Grubbs realized that they could achieve the same sensing effects by turning the process around and mixing conductors into insulating polymers. It turned out that carbon black particles function well as the conducting element, and Lewis continued his work building resistor arrays composed of insulating organic polymers into which carbon black conductors are dispersed. More recently, Lewis and his team members have designed these carbon black composite sensing arrays to detect explo-
sives and chemical agents such as sarin and soman [64]. They have also used them in medical diagnostics, where they “smell” the air near a patient’s head, called the “headspace,” in order to diagnose certain illnesses without bodily contact [65].

This and other research on conductive polymers and carbon black composites has been applied successfully to commercial electronic noses by a company in Pasadena, California originally known as Cyrano Sciences, Inc. The company was acquired by Smiths Group PLC in March 2004, and is known now as Smiths Detection, Pasadena, Inc. This company markets a hand-held electronic nose, the Cyranose® 320, which houses a chip containing an array of 32 polymer carbon black composite sensors [66]. The Cyranose® 320 can be customized for specific uses. The unit is “trained” by exposure to and measurement of vapors common to the client’s application processes. The sensor activation patterns of vapors encountered in subsequent use are identified by comparison with signatures stored in a database of vapor patterns. The range of applications includes chemical detection, food production and process control. More recently, it has been extended to detection of chemical and biological weapons agents as well as medical diagnostics. Testing of the Cyranose® 320 in applications to medical diagnostics showed that the device was able to identify, from headspace, several types of bacteria causing ear, nose, throat, and eye infections. It was even able to discriminate among subspecies of the same bacteria [48].

Smiths Detection, Pasadena also applies carbon black composite technology to produce customized products it terms NoseChips™, which are complete sensors on a dime-sized chip of less than 5 cm³ in volume. NoseChips™ are made to order for specific sensing applications, and can be used as sensing nodes for facilities monitoring, for personal badge detectors, or incorporated into larger systems. NoseChips™ also can be combined to form intelligent sensor networks for distributed monitoring of air quality in industrial settings or other large enclosed spaces [67].

Lewis and his team at Caltech continued their research into carbon black composites, and in 1998 they compared the performance of arrays of carbon black polymer composites, conducting organic polymers, and tin oxides in distinguishing among 19 solvent vapors [68]. Some of these solvents, such as methanol and benzene, had very different chemical properties, and others, such as n-pentane and n-heptane, were very similar. They found that the carbon black polymer composites significantly outperformed the tin oxide and the conducting organic polymers. In the Lewis study, carbon black-based systems proved best, on average, in pair-wise resolution of the 19 analytes tested. They also had the advantage of most accurately resolving the most difficult-to-resolve pairs of analytes. The performance of the carbon black composites in producing the largest mean statistical separation of response patterns for the analytes was approximately 5 times better than that of tin oxide and at least 8 times better than conducting organic polymers. The tin oxide detectors, however, displayed much faster response times. They consistently achieved steady-state responses in un-
der 7 seconds. The carbon black composites and conducting organic polymers had longer and more varied response times, reaching steady-state responses in 20 to 200 seconds. The tin oxide array also displayed response magnitudes about 10 times greater than the carbon black composites and 15 times greater than conducting organic polymers.

At least for the particular analytes tested in Lewis’s study, carbon black composites and metal oxides seem to outperform conducting organic polymers in resolution ability, speed, and response magnitude. Carbon black composites demonstrated better resolution than tin oxide for the analytes tested, while the situation was reversed for speed and response magnitude. The results of this study would seem to point toward the use of carbon black composites or metal oxides as conductance-based sensing elements. Choice of sensor material for an electronic nose might depend, to a certain extent, on a tradeoff between better resolving abilities or greater speed and response magnitude. Other factors such as operating temperature requirements, sensitivity to humidity, and cost and manufacturability, also must be considered.

This same study by Lewis et al. also attempted to identify an optimum number of sensors in the array. Previous discussion of this question centered around two main viewpoints. The first maintained that a fairly small number of sensing elements are needed to span odor space, and that the addition of more detectors adds to noise without significantly adding to classifying ability. The second, more closely following the paradigm of olfaction, held that one should incorporate as many sensors as possible. Doleman and Lewis found that their study supported the latter view; array performance increased as the number of different detectors increased. All three sensor types showed increasing resolving power with larger numbers of detectors, although there was a leveling off of performance as one approached the full complement of sensors.

3.1.2 Potentiometric Sensors

Field-effect transistors (FETs) also can be used as sensors by making the gate sensitive to the presence of a gas [21]. Normally, a charge applied to a transistor gate opens the gate and allows current to flow. When the gate is modified by application of a sensing layer, incoming volatile organic compounds produce a reaction in the sensing layer. The reactants diffuse into the gate and cause the physical properties of the gate to change. The gate threshold voltage is altered, thus changing channel conductivity. Metal-oxide-semiconductor field-effect transistors (MOSFETs), for example, may have their gates coated with a noble metal catalyst such as platinum, palladium, or iridium. More recently, conducting polymers such as polyaniline have also been used as the gate’s sensing layer [49].

Potentiometric sensors have a natural advantage over conductivity-based sensors in that the magnitude of the signal they generate does not depend on the size of the
sensing area [49]. They are transistors, and thus amplify small signals. This should make such chemical FETs (ChemFETs) good candidates for miniaturization. However, miniaturization of chemical sensors has not kept pace with that of conventional electronics; the size of today’s transistor is now below 1 µm, and that of a ChemFET remains at approximately 5 µm, as it has been since the mid-1970s. This is because there are specialized requirements for chemical sensing arrays which are over and above those for conventional microelectronics. These requirements include thicker gate insulators which can withstand harsh chemicals. Gold or platinum must be used as conductors, because their inertness and electrochemical qualities make them preferable to the less expensive aluminum or polysilicon. One of the biggest drawbacks to micron-scale potentiometric sensing devices is that in order to measure a change in potential they require a reference component for comparison, analogous to the reference electrode in macroscopic potentiometric analytical devices [21]. A reference ChemFET would need to be shielded from exposure to the target gas. For an array of broadly sensitive ChemFETs, this adds the requirement for development of microchannels which conduct the target gas to the sensing ChemFETs, while preventing the reference ChemFETs and other chip electronics from being exposed to the target agent. This adds greatly to the complexity of device manufacture.

Among the groups attempting to develop an electronic nose using gas-sensitive FETs is a consortium of researchers at universities in Great Britain, where investigations into developing electronic noses first began [21]. Julian Gardner, a professor of engineering at the University of Warwick, is a pioneer in the field of electronic noses. He began his investigations in the early 1980s using metal oxide sensors. Later in the decade the Warwick team also began to use conducting polymers, and by 1993 they had developed an electronic nose for testing beer quality, using an array of 12 different conducting polymers [69].

More recently, Gardner has been collaborating with researchers at two other British universities, Tim Pearce at the University of Leicester and Alister Hamilton at the University of Edinburgh [70]. Their goal is to develop a neuromorphic analogue sensor on a very large scale integration (VLSI) chemical-sensing chip. This would be, in essence, a complete “nose on a chip.” (See Fig. 9.) The sensors being used are an array of polymer-gated FETs, each sensitive to different odors. They are to be combined with a diffusion microchannel for odor delivery to the FETs. The FET array will be combined with the signal processing components on a single chip that is approximately a square centimeter in size. The team is proposing to mimic the olfactory system much more closely than prior efforts have, in that the new system will model temporal as well as spatial processing. An encounter with a group of target molecules will cause trains of voltage spikes to be generated with a frequency proportional to the concentration of the molecules [71]. This attempt to emulate the temporal as well as the spatial aspects of olfactory processing is the only
FIG. 9. The Warwick/Leicester/Edinburgh nose-on-a-chip. A design for an electronic nose using ChemFET sensors. Both the sensing and the processing components are meant to mimic the components of the olfactory system. This figure is reproduced with permission of Dr. Tim Pearce of the University of Leicester, from his website “Silicon olfactory system implementation,” http://www.le.ac.uk/eg/tcp1/avlsi/.

such research effort in electronic noses encountered by the author in the course of reviewing the literature for this report.

3.2 Mass-Change Sensors

Piezoelectric devices, which can be used to measure a variety of physical phenomena, are used in electronic noses to measure changes in mass [21]. Piezoelectric crystals have the interesting property that they naturally resonate under an applied voltage. The resonance frequency of a particular crystalline structure is dependent
upon its mass. When the mass changes, the resonance frequency changes, and this frequency change can be used to report the presence of gases.

There are two types of piezoelectric devices used as sensors: quartz crystal microbalance (QCM) and surface acoustic-wave (SAW) devices. A QCM is a thin, polymer-coated resonating disk. Adsorption of gas molecules to the polymer coating increases the mass of the disk, thus changing the resonance frequency. QCM devices have been used by the military for several years to detect explosives and other hazardous compounds. They can measure mass changes to an accuracy of 1 picogram ($10^{-12}$ grams), which is the equivalent of sensing less than 0.01 ppm [21].

In surface acoustic wave (SAW) sensors, the waves travel primarily over the surface rather than throughout the device. A signal applied at an input transducer creates an acoustic wave which travels across the piezoelectric surface to an output transducer, having undergone a phase shift in that distance. The phase shift is due in part to the distance traversed, but also depends on the mass of the substrate. As gas molecules are adsorbed to the substrate its mass changes, and the frequency shift and phase shift in the traveling wave can be used to sense the presence of the molecules. Since they are planar, SAW sensors can be fabricated using microelectromechanical systems (MEMS) technology, and this makes their manufacture in large quantities relatively inexpensive [21]. A drawback is that SAW devices are very sensitive to temperature. A change in temperature also contributes to a change in resonance frequency, producing an ambiguous sensor response [72]. Another drawback is that as the SAW devices get smaller they get noisier. When SAW devices shrink, the surface-to-volume ratio increases, and this causes instabilities in the surface processes, which lead to degradation of the signal-to-noise ratio [21]. Reference sensors generally are included with a SAW sensing array so that temperature effects and noise can be subtracted from the sensor readings.

In 2003, a team from two German universities reported in the journal IEEE Sensors on the design and fabrication of a miniaturized QCM sensing system for liquids [73]. As of this time, however, QCM sensors still have not been miniaturized into arrays for gas sensing. Miniature SAW devices, on the other hand, have been the subject of several fruitful research efforts [72,74–78]. In the 14 April 2000 issue of the journal Science, Christoph Gerber, H.P. Lang, M.K. Baller, J. Fritz, and others at IBM Research Division, Zurich Research Laboratory, reported on the development of an electronic nose using a micromechanical cantilever array, in which each of eight cantilevers is coated with a different sensor layer [76]. This system is depicted in Fig. 10. The individual cantilevers are 1 µm thick, 500 µm long, and 100 µm wide, with a pitch of 250 µm and a spring constant of 0.02 Nm$^{-1}$. Adsorption of a compound onto the cantilever results in a change of mass, causing a deflection of the cantilever. The amount of deflection is measured using an optical beam deflection technique. Alternatively, the cantilever array can be actuated by a piezoelectric drive
controlled by a phase-locked loop (PLL) unit, in which case the mass change can be sensed by measuring changes in cantilever resonance frequency. Using this sensor in ambient air, the IBM team was able to demonstrate detection of various analytes, including ethenes, alcohols, natural flavors and water vapor [74,75]. They also used the cantilever sensor in a liquid cell, where the cantilevers were functionalized with a range of different biomolecules. With that modification they were able to demonstrate detection of a single base pair difference between DNA strands [76].

Using gas sensors to recognize biomolecules can be difficult. This is because of the low vapor pressure of biomolecules, such as antibodies, and because they must be in an aqueous environment in order to maintain their structure and function [77]. In 2002, however, William Hunt, Desmond Stubbs, and Sang-Hun Lee at the Georgia Institute of Technology successfully demonstrated recognition of biomolecules in the vapor phase using an electronic nose with SAW sensors [77]. Using a layer of antibodies as the substrate, they were able to observe a baseline frequency shift when the analyte presented was the antigen for the immobilized antibody. To ensure that the frequency shift they were observing was, indeed, due to antibody/antigen binding, they performed an independent check using a confocal laser scanning microscope. With this instrument they were able to identify the locations of the attached analytes,
verifying that the SAW gas sensor was observing molecular recognition. The size of the SAW chip used was $8.10 \times 1.98 \text{ mm}^2$, with individual resonating electrodes 1.5 µm in width.

As is the case for metal oxide and carbon-black composite sensors, a hand-held device using SAW sensing elements is available commercially. Microsensor Systems Inc. of Bowling Green, Kentucky, has combined millimeter-scale SAW gas sensors with electrochemical cell sensors in their HAZMATCAD Plus™ chemical agent detector [79]. This device is $5.8 \times 20.0 \times 24.9 \text{ cm}^3$ in size, and detects chemical weapons agents such as nerve, blood, and blister agents. Nerve agents can be identified at 0.04–0.14 ppm in 20 seconds when operating in fast mode, or at 0.01–0.03 ppm in 120 seconds in high sensitivity mode.

### 3.3 Optical Sensors

Optical fibers also can be used as the sensing elements in artificial sensors that simulate noses [21]. Optical fibers are composed of an inner ring, called the “core,” and an outer ring, called the “cladding” [80]. A slightly higher refractive index in the core than in the cladding allows light to be transmitted for long distances through the fiber. These fibers can be turned into sensors by attaching sensing material either to the fiber’s end, or by removing the cladding and coating the sides with it. The sensing materials used are polymers that contain chemically active fluorescent dyes. The presence of a target agent causes a change in the polarity of the fluorescent dyes, which, in turn, causes a shift in wavelength. The sensor is read with a pulse of light from an external source; activated fluorescent dyes then emit light with a shift in emission spectrum [21]. The polymers react differently to different agents, and the resulting fluorescence changes over time create a pattern which acts as a signature for that agent.

For the past ten years David Walt and his team at Tufts University have been conducting leading research on optical fiber sensors. As early as 1996 they devised an artificial nose composed of 19 individual optical fibers [81]. Their sensor was able to identify components within mixtures, and also could characterize test compounds on the basis of chemical characteristics such as functional groups and relative molecular weight. Those fibers, at a diameter of 300 µm, were relatively large. By 1999, Walt’s team had modified this approach by using as sensing elements thousands of spherical fluorescent beads, each only 3 µm in diameter [82]. In place of the larger, single-core fibers, they used optical image guides, 500 µm in diameter, which enclosed approximately 5000 closely packed optical fibers, each of diameter 3.5 µm. The microbeads were then distributed randomly into chemically etched microwells on the faces of the optical fibers. The beads, which were of three different types, were encoded so that once in place, each bead could be identified. The sensing system was “trained”
To identify the sensor bead type at each location, and was then successfully used to detect the presence of four different vapors, methanol, dichloromethane, toluene, and acetone.

This approach, using large numbers of sensors of a few discrete classes, has several advantages. Since there is significant redundancy in microbeads across the array, combining the signals from same-type beads serves to amplify the signal and enhance the overall sensitivity of the array. Summing the responses from a number of microbeads also serves to improve signal-to-noise ratio, which is of particular importance when only low concentrations of a target agent are present. In addition, the random dispersal of microbeads obviates the need for the precise placement of sensing elements inherent in other manufacturing methods [82]. The downside of the random sensor placement is that effort is shifted from manufacturing to data processing, as each optical fiber sensor must be trained individually before use.

In a more recent effort, Walt’s group at Tufts University and Pearce’s group at the University of Leicester collaborated in testing an optical bead array with six different bead types, each replicated over 250 times [83]. This sensing platform was tested for its ability to discriminate between six different complex odors, including acetone, toluene, 1,3-dinitrotoluene (1,3-DNB), and three coffee types. The system was able to fully discriminate all analytes, achieving 100% correct classification at the highest relative concentration levels (e.g., 9000 ppm each of acetone and toluene and 0.9 ppm of 1,3-DNB), and better than 85% correct at the lower concentration levels (e.g., 450 ppm each of acetone and toluene and 0.1 ppm 1,3-DNB).

Optical fiber-based artificial noses also hold promise for detecting explosives, particularly landmines. The toluene and 1,3-DNB compounds successfully detected by Walt and Pearce’s optical fiber sensors are nitroaromatic compounds similar to low-level explosives. When dogs are used to hunt for landmines, they most likely smell DNT (dinitrotoluene), a byproduct of the explosive TNT (trinitrotoluene) in the landmine. Two other scientists at Tufts University, neuroscientists John Kauer and Joel White of the Tufts University School of Medicine, both active in studying natural and artificial olfaction, conducted a study in 2001 investigating the substitution of an optical fiber sensor for a dog’s nose in detecting landmines [84]. Since a dog can detect 1 part per billion (ppb) or less of TNT, this is a difficult challenge. In a test conducted in a special chamber, the artificial nose was not yet that sensitive, detecting only 10 to 15 ppb. In a field test, the artificial nose was able to detect the presence of landmines but could not locate them very precisely. Kauer suggested this could have been due to interference from other odors in the environment. Other factors, such as wind or lateral diffusion of the odorant, also may have contributed to the failure to locate the source precisely. The problem of accurate sensing in a complex or interfering environment is one that will continue to challenge sensing systems of all types.
Tufts University has licensed its optical bead sensing technology to Illumina, a San Diego-based company founded in 1998 [85]. Illumina now builds biological sensors using BeadArray™ technology on two substrates. One of these, the Array Matrix, is composed of 96 arrays of fiber optic bundles. Each bundle consists of 50,000 individually etched fiber optic strands. Microbeads are deposited into the etched wells, which are 6 microns from center to center. The Array Matrix and the BeadChip, in which microbeads are assembled into etched wells in a slide-shaped device, are sold commercially for genotyping and gene expression profiling. (See Fig. 11.) Illumina also is in partnership with Dow Chemical and Chevron on a project to develop chemical sensing systems using this technology. The oNose™, their optical electronic nose currently in development, will be a handheld device about the size of a calculator.

3.4 Other Sensors

Most sensing mechanisms used in artificial noses fall into one of the three categories discussed above: electrochemical, mass-change, or optical fiber. A different
An optically based method is that of the “colorimetric” nose, created in 2000 by Neal Rakow and Kenneth Suslick at the University of Illinois at Urbana-Champaign [86]. Rakow and Suslick observed that many of the most toxic vapors are excellent ligands for metal ions, an area which had been little investigated in previous artificial nose research. They also observed that metalloporphyrins are good candidates for detection of metal-ligating agents, as they provide excellent binding sites and undergo large spectral shifts upon ligand binding. So they created a sensor array of metalloporphyrin dots deposited on a reverse phase silica surface, and then used a flat-bed scanner to obtain color images of the dots before and after exposure to an analyte. The color change in dots was measured by using a computer to subtract the pixels in the “before” image from the “after” image. The array response was then compared to a library of color “fingerprints” for identification of the analyte. Using this method, the researchers were able to identify, at below 2 ppm, a wide range of gases, including alcohols, amines, ethers, phosphines, phosphites, thioethers, thiols, aromes, halocarbons, and ketones. However, the operational requirement for a scanner would make it difficult, at least for the present, to make significant size reductions in the colorimetric sensor (e.g., down to the micron or nanometer scale).

Another effort by Baltes, Brand and their colleagues incorporates three different types of gas sensors plus a temperature sensor to create a sensing system on a chip [87]. The sensors used were mass-sensitive (a micro-machined cantilever), capacitive (differential signal between a polymer-coated sensing capacitor and a reference capacitor), and calorimetric. The calorimetric sensor works by detecting changes in enthalpy, or energy content per unit mass, upon adsorption or desorption of an analyte on a sensitive polymer film. Adsorption or desorption of molecules causes heat to be released or absorbed, which is recorded by an array of polysilicon/aluminum thermocouples. The temperature variations are translated into a voltage change. The temperature sensor was included to calibrate the sensors for detection of analyte concentration, since adsorption of volatile organic compounds on polymers is temperature dependent. These sensing elements, together with analog-to-digital converters and a digital bus interface for transmission of data to off-chip recording units, were integrated on a 7 × 7 mm² chip, which then was used to detect the presence of ethanol and toluene. The three sensors displayed varying responses to the two analytes; ethanol caused a capacitance increase, for example, while toluene caused a capacitance decrease, and the cantilever responded more strongly to toluene than ethanol, because ethanol has a lower molecular mass. From these results it can be seen that future artificial noses need not be composed of sensors of only one type; an array of sensors of different types may broaden the spectrum of agents that are detectable.
3.5 Next Steps in Artificial Noses

As can be seen from the survey presented above, investigators and developers of artificial noses have attempted to follow the model of olfaction in devising sensor arrays with broadband sensing elements, using many different materials and sensing methods. Each sensor array element reacts differently to different target agents, both in absolute terms (activated/not activated) and in the degree of activation. Thus, these sensing arrays are designed to produce “signatures” for chemical agents in much the same way that activated olfactory neurons produce an odor’s signature. The number of different sensing elements in artificial noses has not yet exceeded the two or three dozen of the KAMINA micronose [59], however, while the number of different sensing neurons in humans is at least an order of magnitude greater [27]. The total number of sensing array elements is also much smaller than the number of neurons in the epithelium; Walt’s fiber-optic sensors have several thousand sensing elements (of only three different types) compared with several million epithelial neurons [1,82]. All of these artificial noses, therefore, fall short of emulating another characteristic of the olfactory system, which is that it contains millions of small, densely-spaced sensors in an area only several square centimeters in size.

To come close to matching the sensitivity, selectivity, and discriminatory abilities of natural noses, it will be necessary to increase the numbers and kinds of sensing elements in artificial noses. Even the individual microsensors used today, however, when replicated thousands of times, would result in sensing systems much larger than hand-held devices, let alone ultra-small sensing systems which could be incorporated into clothing. To make this leap down the size scale, we will need to use sensing elements that are themselves orders of magnitude smaller than the sensing elements in today’s smallest artificial noses. This means that the sensing elements in artificial noses will have to be nanometer scale structures and devices. Research on individual nanometer-scale sensors is already underway. More recently, efforts to develop prototype systems of nanosensors have begun, as well [88–90]. The next section of this chapter presents a review of several research efforts to explore and develop nanometer-scale sensors, or “nanosensors.”

4. Nanosensors

This section reviews research and development upon individual nanometer-scale sensors and towards entire artificial nose-like sensor systems integrated on the nanometer scale. In this consideration of nanosensors, we attempt to apply the
lessons learned in the review of natural noses and nose-like microsensor systems that appeared in the foregoing sections of this chapter. In reviewing the literature for very small artificial noses, it was possible to find examples of micron-scale sensing systems of all three major types (conductance-based, mass change, and optical), some already in commercial production. No operational artificial nose-like sensing systems of any type were found that are integrated on the nanometer-scale. Research and development is being conducted upon individual nanometer-scale sensors, however [91–109]. Investigations also are ongoing in chemical sensing using large numbers of nanometer-scale sensors in thin films, meshes, or in other ordered arrays [88,89,97, 110–117], as well as in the integration of nanosensing elements into microelectronic systems [118,119].

The results of several studies on the use of individual nanometer-scale devices as sensors are presented in the following sections. In particular, these discussions concern research on carbon nanotubes [91–93], nanowires [101,102], and other nanostructures [120] as nanosensing elements. An overview of the research efforts discussed below is presented in Table II.

### Table II

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
<th>Researcher/Institute</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbon nanotube</td>
<td>Semiconducting-SWNTs sense NH₃, NO₂ and H₂ by undergoing a conductance change</td>
<td>Hongjie Dai et al. Stanford University [91,93,121]</td>
</tr>
<tr>
<td>Carbon nanotube</td>
<td>SWNTs show high sensitivity to exposure to oxygen or air, changing their electrical properties</td>
<td>A. Zettl et al., University of California at Berkeley [92]</td>
</tr>
<tr>
<td>Carbon nanotube</td>
<td>MWNTs, which generate very high electric fields at their tips, are used in an array to form a miniature gas ionization sensor</td>
<td>Pulickel Ajayan et al., Rensselaer Polytechnic Institute [94]</td>
</tr>
<tr>
<td>Nanowire</td>
<td>Silicon nanowires in solution sense pH changes, presence of streptavidin and Ca²⁺, and antibiotin/biotin binding with conductance changes</td>
<td>Charles Lieber et al., Harvard University [101,122]</td>
</tr>
<tr>
<td>Nanowire</td>
<td>Tin oxide nanowires sense N₂ and CO with conductance changes</td>
<td>Martin Moskovits et al., University of California at Santa Barbara [102]</td>
</tr>
<tr>
<td>Nanobelt</td>
<td>Tin oxide nanobelts sense CO, NO₂ with a change in conductance</td>
<td>Zhong Wang et al. Georgia Institute of Technology G. Sberveglieri et al., The National Institute for the Physics of Matter, Brescia, Italy, and Università di Brescia [120,123]</td>
</tr>
</tbody>
</table>

SWNT = Single-walled carbon nanotube.
MWNT = Multwalled carbon nanotube.
4.1 Carbon Nanotube Sensors

Carbon nanotubes are ultra-small, ultra-strong, tubular molecules of pure carbon, which have unique electrical properties [121,124]. They can behave either as a semiconductor, a narrow bandgap semiconductor, or a metal, depending in large part on their chirality or “twist.” The fact that they are conductors makes them good candidates for conductance-based sensing, and a number of investigations over the past few years have shown that both single-walled carbon nanotubes (SWNTs) [91–93, 95,96,100] and multi-walled carbon nanotubes (MWNTs) [94,97–99] can be used as chemical sensors.

For example, a team led by Hongjie Dai at Stanford University has been investigating the use of carbon nanotubes as sensors. In 2000, Dai and Jing Kong showed that the electrical resistance of semiconducting SWNTs changes dramatically upon exposure to nitrous oxide (NO\textsubscript{2}) or ammonia (NH\textsubscript{3}) [91]. (See Fig. 12.) Dai and Kong placed a SWNT approximately 1.8 nm in diameter and 3 µm in length between two metal electrodes, each consisting of a 20 nm-thick layer of nickel covered by a 60 nm layer of gold. Exposure to NH\textsubscript{3} caused a decrease in SWNT conductance by approximately two orders of magnitude, with a response time (the time required for the resistance to change by 1 order of magnitude) between 1 and 2 minutes for 1% NH\textsubscript{3}, and about 10 minutes for 0.1% NH\textsubscript{3}. An increase in SWNT conductance of approximately three orders of magnitude resulted from exposure to NO\textsubscript{2}, with response times of about 2 to 10 seconds for 200 ppm, 0.5 to 1 minute for 20 ppm, and 5 minutes for 2 ppm.

Carbon nanotubes also are highly sensitive to oxygen, however, which may have an impact on their application to sensing in ambient air [92]. Philip Collins and A. Zettl at the University of California at Berkeley found that exposure to oxygen, whether pure dry oxygen or air, caused an increase in conductance by 10 to 15%. They also found that the effects of oxygen exposure were increasingly irreversible with decreasing temperatures, and that once the SWNTs were exposed to oxygen, it was not possible to deoxygenate them without heating them in a high-vacuum environment to 110 to 150 °C for several hours. Exposure to oxygen also changes other characteristics of the SWNTs, including electrical resistance, the local density of states, and their thermoelectric power (µV/K). Such exposure can convert a semiconducting carbon nanotube into one that is metallic. Thus, the properties of a SWNT, including its conductance, are functions of its gas exposure history as well as its diameter and chirality.

The carbon nanotubes tested by the Stanford and UC Berkeley teams were unmodified SWNTs. Such carbon nanotubes, while sensitive to NO\textsubscript{2}, NH\textsubscript{3} and oxygen, are not sensitive to many other target molecules such as hydrogen (H\textsubscript{2}) and carbon monoxide (CO) [93]. Functionalizing the SWNTs, however, by decorating the sidewalls with desired molecular groups, can change the sensitivity of the SWNT
Fig. 12. Single-walled carbon nanotubes (SWNT) as conductive gas sensors. Hongjie Dai’s group at Stanford University has demonstrated that individual semiconducting SWNTs, as seen in (A), behave as sensors for the presence of nitrous oxide (NO$_2$) and ammonia (NH$_3$). SWNT exposure to NH$_3$ causes a decrease in conductance, as shown in (B), while (C) shows increased conductance due to exposure to NO$_2$. Reprinted with permission from Jing Kong, Hongjie Dai, et al., “Nanotube molecular wires as chemical sensors,” Science 287 (28 January 2000) 623.

© 2000 AAAS

for specific target gases. In 2001, Kong and Dai were able to create carbon nanotube hydrogen sensors by decorating SWNTs with palladium (Pd) nanoparticles [93]. Deposition of approximately 5 Å of palladium over an entire SiO$_2$ substrate containing the SWNTs resulted in a thin, non-continuous layer of palladium deposited on the nanotubes. The gaps in the palladium layer ensured that the electrical conduction path in the modified SWNT was not through the palladium particles, but remained through the decorated SWNT. Exposure of the decorated nanotubes to a flow of air mixed with 400 ppm H$_2$ then resulted in a decrease in conductance by approximately
a factor of 2, with ready reversibility when the flow of hydrogen was stopped. The response time (i.e., the time during which half the resistance change occurred) was about 5 to 10 seconds, with approximately 400 seconds required for full recovery. Similar results were obtained with concentrations of 40 and 4 ppm H₂. This experiment demonstrated that the functionalization of SWNTs can expand the spectrum of gases to which they are sensitive.

The experiments described above examined the effects of the interactions of different gases with the surfaces or sidewalls of carbon nanotubes. In contrast, Pulickel Ajayan’s group at Rensselaer Polytechnic Institute (RPI) looked, instead, at reactions occurring at their ends [94]. In a 2003 study, Ajayan, Ashish Modi, and others found that multi-walled carbon nanotubes (MWNTs) could be used as gas ionization sensors. Ionization sensors work by identifying gases by their unique breakdown voltage, which is the electric field at which the gas is ionized. Ajayan and his team created an ionization sensor using an aluminum plate cathode, separated by 150 µm from the anode, which was a film of vertically aligned MWNTs on a SiO₂ substrate. The MWNTs were about 25 to 30 nm in diameter and 30 µm long, with a separation of 50 µm between nanotubes. The sharp tip of a MWNT generates a very high non-linear electric field, and the billions of MWNTs which compose the anode contribute to an increase in the electric field. This allows breakdown at voltages that are several times lower than those of traditional electrodes. Several gases were tested by releasing them, one at a time, into a high vacuum chamber containing the sensing device, and each gas exhibited a unique breakdown voltage. The breakdown voltage of air was observed to be 65% lower (346 vs. 960 V) in the vicinity of the MWNT electrode than that required near conventional metal electrodes. Other gases tested showed similar results. In addition, the discharge current at ionization also increased approximately 6-fold. Since discharge current is proportional to analyte concentration, higher discharge currents allow detection of dilute concentrations and thus increase sensor sensitivity. Gases tested and successfully detected were helium, argon, nitrogen, oxygen, carbon dioxide, and ammonia, as well as air. Unlike other ionization sensors, which must be used with a gas chromatograph, this sensor can be used directly with gas mixtures. Lower voltage requirements also translate into lower power requirements. Thus, a sensor using this approach could be fairly compact. The present design requires millions of MWNTs spaced tens of micrometers apart, which results in a device size in the millimeter range; future designs may be able to provide further size reductions.

4.2 Nanowires and Other Nanostructures

Semiconducting “nanowires” have shown themselves to be particularly promising for applications as nanosensors and for the development of dense systems of
such sensors [101,102,120,122,123,125,126]. Nanowires and other nanometer scale structures with promise for sensor applications have been synthesized from a number of materials, including zinc, tin, indium, cadmium, silicon, and polysiloles (silicon-polymer combinations). Two research efforts to develop and perfect nanowire sensors for gases and biological agents, plus an investigation into “nanobelt” sensors, are described below [101,102,120].

Early in 2001, Charles Lieber and his group at Harvard University demonstrated that semiconducting nanowires can be assembled to build nanometer-scale electronic devices, including passive diodes, transistors, and inverter-like digital circuits [122]. They fabricated their devices and circuits from boron- and phosphorus-doped silicon “wires” approximately 20 nm in diameter. Subsequently, they tested the ability of these nanowires to perform four different sensing functions in aqueous solution [101]. They used boron-doped silicon nanowires (SiNWs) that were functionalized by different chemical substituents to make the SiNWs sensitive to four different target agents, and measured the conductance change upon exposure to the target analytes. Using amine- and oxide-functionalized SiNWs, Lieber’s group observed constant conductance for a given pH, stepwise increases in conductance with discrete changes in pH from 2 to 9, and sensor reversibility for increasing or decreasing pH. (See Fig. 13.) SiNWs modified with biotin were able to detect ligand-receptor binding of biotin-streptavidin with an increase in conductance. However, this process was not reversible. Lieber’s group also tested the reversible binding of monoclonal antibiotin (m-antibiotin) with biotin, and found a well-defined drop in conductance upon introduction of m-antibiotin solution, and an increase to approximately the original conductance value upon addition of pure buffer solution. Finally, the team used SiNWs functionalized with calmodulin to detect Ca$^{2+}$. The decorated SiNWs displayed a drop in conductance upon encounter with a 25 µM Ca$^{2+}$ solution, and a conductance increase when a Ca$^{2+}$-free solution was introduced. Undecorated SiNWs did not react to the presence or absence of Ca$^{2+}$, indicating that the calmodulin was, for this purpose, an essential element of the nanowire detector.

Lieber’s work showed that nanowires could behave as nanosensors for chemical and biological agents in solution. In 2003, Martin Moskovits and a team at the University of California at Santa Barbara demonstrated the applicability of nanowires to gas sensing, as they successfully used nanowires to detect the presence of CO and O$_2$ [102]. Moskovits’ group measured the conductance change in individual tin (SnO$_2$) nanowires, each approximately 60 nm in diameter, connecting timonium/gold (Ti/Au) electrodes. The experiment was performed at operating temperatures between 120 and 300 °C. It was found that the nanowires were good conductors in the absence of oxygen, and they were converted into insulators in the presence of oxygen. Nanowires exposed to CO experienced an increase in conductance. The experimenters found that alternating pulses of N$_2$ + 10% O$_2$ and CO, delivered at
approximately 10 minute intervals, caused a sequence of sharp conductance changes in the nanowires. (See Fig. 14.) Response to CO was dependent on temperature and concentration, with larger conductance changes demonstrated at higher temperatures. Pulses of CO caused an increase in conductance that was linearly proportional to CO concentration, and pulses of oxygen decreased the conductance back to a baseline value for nanowires in an oxygen environment. The response time (time for full conductance change) upon exposure to CO was approximately 30 seconds.

Moskovits and his team point out that for bulk SnO2 at 500 K, electron exchange between the surface states and the bulk takes place in a layer that is approximately 43 nm thick. This means that for a 60 nm diameter SnO2 nanowire, the “surface layer” encompasses the entire structure. Thus, adsorption or desorption of analytes on the surface alters the bulk properties of the entire nanowire. This makes these SnO2 nanowires conductance switches whose conductivity is determined entirely by reactions on their surfaces. Moskovits proposes that a large array of differently func-
Fig. 14. Tin dioxide (SnO$_2$) nanowires as gas sensors. Martin Moskovits and his group at the University of California at Santa Barbara demonstrated that SnO$_2$ nanowires behave as conductivity-based sensors for carbon monoxide (CO) and oxygen (O$_2$). Part (a) of the figure shows CO binding with oxygen at the surface of the nanowire, freeing electrons to form a conducting channel through the nanowire. Part (b) shows the effect upon nanowire conductance of exposure to alternating pulses of CO and O$_2$, with CO concentration decreasing with successive pulses. Reprinted with permission from Kolmakov, Moskovits, et al., “Detection of CO and O$_2$ using tin oxide nanowire sensors,” Advanced Materials 15 (12) (17 June 2003) 999. © 2003 WILEY–VCH Verlag GmbH & Co KGaA, Weinheim

Tin dioxide nanowires could be integrated to create a parallel sensing system similar to the olfactory system.

A different kind of nanostructure was synthesized in 2001 by a group at the Georgia Institute of Technology led by Zhong Lin Wang [123]. These structures, called “nanobelts,” are ultralong, ribbon-like nanostructures with a rectangular cross section. Typically, they are between 30 and 300 nm wide, with width-to-thickness ratios of 5 to 10, and lengths in the tens or hundreds of micrometers, sometimes extending to millimeters. Wang’s team successfully synthesized nanobelts of semiconducting oxides of zinc (ZnO), tin (SnO$_2$), indium (In$_2$O$_3$), and cadmium (CdO). These were found to be pure, structurally uniform, single crystals that are relatively free of defects and dislocations. They concluded a 2001 report on their findings in Science by suggesting that the nanobelts could be used as nanosensors.

The following year, 2002, G. Sberveglieri and colleagues at the University of Brescia, in Brescia, Italy, in collaboration with Wang and his team, actually used tin nanobelts as sensors [120]. They built a sensing device by placing SnO$_2$ nanobelts
atop a platinum interdigitated electrode structure on an alumina substrate. A platinum heater on the reverse side of the substrate kept the working temperature at 400 °C. The nanobelts successfully registered the presence of carbon monoxide (CO), ethanol, and nitrogen dioxide (NO₂) by a significant change in conductance. The sensor response, defined as the ratio of the change in conductance to the initial conductance (ΔG/G), was found to be +0.9 for CO, +41.6 for ethanol, and −15.5 for NO₂. Thus, tin dioxide nanobelts were shown to function as effective gas sensors.

5. Designing a Nanometer-Scale Nose-Like Sensing System

As discussed earlier in this chapter, the development of artificial noses which truly emulate the remarkable sensing abilities of the natural nose will require the integration of many ultra-small, nanometer-scale sensors into systems. In the preceding section, we presented an overview of several different research efforts for developing individual nanosensors. This section will consider how one might use the information above to integrate a number of components to design a nanometer-scale nose-like sensing system. In order to do that, we will consider, in Section 5.1, the advantages and drawbacks of each of these devices—carbon nanotubes, nanowires and nanobelts—as candidate sensing elements in such an ultra-small electronic nose. Based on these considerations, we recommend that such a system be built using nanowires, as discussed below in Section 5.2.

5.1 Consideration of Nanotubes, Nanowires, and Nanobelts for Electronic Noses

Carbon nanotubes, nanowires, and nanobelts all have high surface-to-volume ratios and all have been demonstrated to react to the presence of certain gases and other analytes with a change in conductance. Thus, each of these devices could conceivably be used as a nanosensing element within an electronic nose. The strengths and weaknesses of using these devices for this purpose are discussed below.

Carbon nanotubes show strong sensitivities to several gases, and can operate at room temperature. Using carbon nanotubes as sensors, however, presents several difficulties [101]. Large numbers of nanotubes with uniform, predictable characteristics would be required for building arrays of nanosensors. Using presently available techniques, carbon nanotubes always are synthesized in mixtures of metallic and semiconducting nanotubes. Often, the product mixture contains single- and double-wall nanotubes with a wide range of diameters and chiralities. Researchers have put
forward different methods for separating nanotubes, including a recent proposal by Ralph Krupke and Frank Hennrich at the Karlsruhe Research Center [127], and another by a team at DuPont Central Research and Development [128]. The MITRE Corporation also has obtained a December 2003 patent for a method to perform bulk separation of SWNTs based on chirality [129,130]. At the present time, however, sorting and selecting carbon nanotubes of a specific kind remains difficult and time consuming.

Functionalizing or decorating carbon nanotubes also presents challenges. Dai’s group successfully functionalized carbon nanotubes by depositing a layer of palladium on the SWNT. In general, however, flexible methods for the functionalization of carbon nanotubes are not easily available [101], because the long tubular surface of the molecule presents few ready binding sites for acceptor molecules. Finally, the strong, essentially irreversible, binding of carbon nanotubes to oxygen may change their characteristics. This raises questions about the predictability of their sensing behavior following prolonged exposure to ambient air.

Nanowires and nanobelts of tin oxide also have been demonstrated to be effective gas sensors, and both are considered together here as “nanowires.” Individual metal and semiconducting nanowires can be fabricated predictably and uniformly, even in very large arrays, as was demonstrated in 2003 by James Heath and his group at Caltech [131]. The characteristics of nanowires can be fixed during the fabrication process by controlling the dopant types and concentrations. Thus, their sensing behavior is predictable [101]. In addition, the use of metal oxides and conducting polymers in electronic noses over several decades has resulted in some familiarity with chemical modifiers for metal and silicon oxide surfaces [50,101]. This existing body of knowledge provides a natural starting point for development of the sets of nanoscale sensors that are required for olfaction-like sensing: ultrasmall sensors with broad, overlapping sensitivity spectra. The primary disadvantage of metal oxide nanowires is their requirement for high operating temperatures. For example, Moskovits’ team measured response curves for their SnO$_2$ nanowires at 200 to 280°C [102]. The need for a high temperature operating environment for tin oxide nanowires is similar to the requirement for metal oxide macro- or microsensors. This requirement has been met on the micron scale, as demonstrated by the KAMINA micronose and its micron- to millimeter-scale heating element [59]. It is possible that a heating element analogous to that employed in such a microsensor (see Section 3.1.1.1) could be implemented for a nanosensing system as well. Since the nanosensing system also will require close integration with a data processing system, a portion of the heat required for sensing might even be derived from the dissipation arising from an integrated nanoprocessor or nanomemory.
At present, the uniformity, predictable behavior, and well characterized materials interactions exhibited by nanowires offer clear advantages over carbon nanotubes when selecting building blocks for a nanosensing system. Metal and silicon nanowires also appear to share with larger metal oxide sensors a characteristic required for electronic noses: broad-based chemical sensing capability. The amazing sensing abilities of the mammalian nose are not due to highly specific sensing elements, but to vast numbers of semi-specific sensors, whose combined input leads to identification of the target agent. This feature of olfaction is one that has been successfully implemented in large-scale electronic noses, and should be emulated in nanometer-scale gas sensor systems as well. Lewis, Walt, and others have pointed out that a system with “lock-and-key” sensing elements, each specific to only one target agent, would be very difficult to implement for chemical sensing in unknown environments [22]. Specific sensors work only when background and interfering agents can be controlled. In an unknown environment, such specific sensors are likely to respond to agents with molecular composition similar to that of the designated target. Binding with these similarly-structured molecules would result in a positive sensor reading even in the absence of the target agent. The resulting false positives would reduce greatly the reliability of the system.

From the discussion above, we conclude that nanowires appear at present to be the most reasonable choice for the sensing elements in a nose-like nanosensing system. Both Lieber at Harvard [101] and Moskovits at UC Santa Barbara [102] have demonstrated that nanowires can behave as sensors. The experiments conducted by Lieber’s group were conducted with silicon oxide nanowires in solution, and those by Moskovits’ team were conducted with metal oxide nanowires in air. For the purpose of designing a nose-like gas sensing system to be used in ambient air, it seems reasonable to start with the metal oxide nanowires already demonstrated to work as gas sensors.

Recently, Andrei Kolmakov at Southern Illinois University took a significant step toward realizing ultra small, nanowire sensing systems, using nanowires approximately 100 nm in diameter, and somewhat larger mesowires, about 1000 nm (a micron) in diameter [132]. Kolmakov and his team built an electronic nose using pristine tin oxide (SnO₂) nanowires, surface nickel-doped (Ni)–SnO₂ tin oxide nanowires, and mesowires of indium oxide (In₂O₃) and titanium oxide (TiO₂). The result was an array of four nano-/mesowires with overlapping but different sensitivities. The team was successful in using pattern recognition techniques commonly applied in larger-scale artificial noses in order to detect and differentiate between the presence of hydrogen (H₂) and carbon monoxide (CO). While not yet entirely composed of nanoscale sensors, this device is a “proof-of-concept” that demonstrates that nanowire sensors can be integrated successfully into ultra-small nose-like systems.
5.2 New Designs for Electronic Nose Systems Integrated on the Nanometer Scale

We propose, therefore, based on the considerations outlined in Section 5.1, that very small electronic noses might be built that use nanowires as sensing elements. Two specific concepts are presented here; the first uses an array of differentially sensitized nanowires, and the second employs a nanowire crossbar array, essentially a functionalized nanomemory.

5.2.1 A Nanowire Sensing Array

We present here a concept for a nanowire array electronic nose. The sensing elements are core–shell nanowires composed of a platinum-doped SnO2 core surrounded by a SiO2 shell. Such a nanowire sensing element is shown in Fig. 15. The SnO2 nanowire core, 60 nm in diameter, is similar to the nanowires demonstrated to function as gas sensors by Moskovits [102]. Here we sketch out how such nanowires might be assembled to produce a nose-like sensing system.

In order to incorporate these nanowires into a “nanonose,” it is necessary to differentiate them. Following the example of the sensing neurons, the range of sensitivity for a particular nanowire should overlap but not be identical to that of other nanowires. Goschnick and his Karlsruhe team have accomplished this differentiation in their KAMINA micronose by applying a layer of SiO2 of graduated thickness over a segmented sheet of SnO2 [59]. The variance in SiO2 thickness from segment to segment is responsible for differences in the sensing range of each segment. We will employ the same concept in the nose-like sensing system we propose here. Over the last few years, core–shell nanowires of various materials have been developed successfully by several research teams; a few of those teams are referenced here [133–137]. We will use individual SnO2/SiO2 core–shell nanowires, with SiO2 shell

![Diagram of a nanowire sensing element](image URL)

**Fig. 15.** A proposed nanowire sensing element. This core–shell nanowire sensing element synthesizes the ideas of two researchers. The 60 nm diameter core consists of Pt-doped SnO2, as in the nanowire sensors demonstrated by Moskovits et al. [102]. The SiO2 shell will vary in thickness, from wire to wire, from a minimum of 2 nm to a maximum of 20 nm. Application of a SiO2 layer is the method used by Goschnick et al. [59–61] to differentially sensitize SnO2 to different chemicals.
FIG. 16. A proposed nanowire “Nose-Like” sensing system. Each nanowire sensing element in this sensing system has a SnO$_2$ core and a SiO$_2$ shell, as in Fig. 15 above. SiO$_2$ shell thicknesses of 2 to 18 nm result in sensing elements with diameters of 64 to 96 nm. As proposed in this chapter, wires of different diameters would react to different chemicals such as formaldehyde, carbon monoxide, and ammonia.

thicknesses that vary, from wire to wire, from approximately 2 to 20 nm. Such a nanowire is pictured in Fig. 15. The sensing range of each nanowire will depend on the thickness of its shell. Differentiated nanowires will be integrated into a nose-like sensing system in which the nanowires function much like the individual segments of the KAMINA micronose. Therefore, this nanosensing system, in combination with a heating element and a data processor, could be expected to detect also the same range of agents detected by the KAMINA system. This includes most non-inert gases, particularly formaldehyde, carbon monoxide, benzene, ammonia, acrolein, and sulfur dioxide.

Differentiated core–shell nanowires of various diameters, arranged in parallel, would form a “memory array,” with each array element reporting an encounter with a target agent by a change in conductance/resistance. The particular combination of nanowires reporting conductance changes would provide the target agent identification. Figure 16 is a sketch of such an electronic nose, composed of six individual nanometer-scale sensors in a $6 \times 1$ memory array.

5.2.2 A Nanomemory Sensing System

A second concept for a nanowire electronic nose uses nanowires in a crossbar array. This design builds upon nanowire crossbar memories [138,139], which already have been developed [131,140]. Research efforts by Lieber’s group at Harvard [141] and by a Hewlett-Packard/Caltech team lead by R. Stanley Williams and James Heath [142,143] have succeeded in fabricating addressable nanowire crossbar arrays. The junctions of the crossbars function as the “memory bits” and store the 0/1 bit information. To use such a device for nose-like sensing, the junctions could be decorated with a variety of different binding agents. An encounter with a target agent at junction $(x, y)$ would flip the value of bit $(x, y)$ from 1 to 0, or 0 to 1. Thus, each memory bit
is a sensing element, and a simple memory read operation, by identifying activated memory locations, would provide a complete status report of all sensor elements. An illustration of a conceptual nanomemory sensing system is provided in Fig. 17.

Such a nanowire crossbar sensing array could be used either for specific or semi-specific, i.e., nose-like, sensing. The sensing scheme would be determined by the choice of binding agents used to functionalize the crossbar junctions. Highly specific binding agents would provide a direct read-out for presence of their targets. On the other hand, binding agents that are not strongly discriminating would provide sensor status data that, as in olfactory modeling, can be used to identify the agents encountered.

While attaching a binding agent to an individual crossbar junction \((x, y)\) might be a complex task, functionalization of a small region of junctions, \(\{(x \pm m, y \pm n)\}\), most likely would be easier to realize. A small region of like sensors provides redundancy in sensor elements, and, as in the nose, this would serve to both enhance signal-to-noise ratio and increase the probability of encounter with a target agent.
6. Summary

Recent explorations of the olfactory system have begun to fill in the blanks in our knowledge of the sense of smell and how it works. What was pure speculation in the middle of the last century has been transformed, especially over the past 15 years, into detailed insights about the functions and processes of the individual cell chemistry, neuronal interactions, and information processing that make olfaction work. Advances in genomics have allowed researchers to identify the genes responsible for the sense of smell, and this has led to broad new understandings of olfactory detection processes in the tissues and cells of the nose and brain. New technologies also have allowed researchers to follow the path of olfactory signals into the brain, providing exciting new insights into how the brain responds to olfactory stimulation. This expanded body of knowledge has presented us with a more complete picture of a natural sensing system which is broadband in its range of application while being remarkably sensitive to individual odorants. This new knowledge also presents us with a picture of an integrated system, in which the sensing function is performed by large ensembles of non-specific receptors, and information is processed in parallel and sequentially in several stages. It is a system that is remarkably successful in using sparse information to make precise identifications.

Although much progress has been made in creating artificial analogues to the olfactory system, electronic noses still fall short of duplicating the amazing abilities of the natural nose. In part, this is due to the fact that many questions still remain to be answered about how the nose works so well. However, the shortfalls in the performance of artificial noses also are due to the fact that what we do know about the functioning of the nose is not necessarily easy to emulate. Although we know that odorant identification utilizes a spatio-temporal combinatorial coding system, it is not yet clear how to characterize molecules by artificial means in such a way that the resulting code would permit identification of a wide range of toxic chemical or biological agents. In addition, present-day artificial noses make use of a variety of techniques for molecular recognition. However, these techniques cannot be made to work yet at the density of epithelial neurons or within the size constraints of the human olfactory system, let alone those of the far more sensitive rat or mouse nose.

It seems likely, though, that nanotechnology can make significant contributions to electronic noses and ultra-small gas sensing systems in terms of reduction in size and increase in sensor density. Nanometer-scale sensing devices such as carbon nanotubes and nanowires already have been demonstrated. Nanomemories with more than a gigabit of memory per square centimeter have been demonstrated [143]. The next step will be integration of nanoscale sensors into sensing systems. Two proposals for integration of nanowires into electronic nose systems are presented here. Researchers certainly will propose many more designs in coming years. For exam-
ple, several investigative groups in the U.S. have begun development of nanosensor systems under the auspices of the Applications of Molecular Electronics (MoleApps) R&D Program that was recently initiated by DARPA.

National security concerns since the terrorist attacks of September 11, 2001 have made the design of small, fast, sensitive, broadband sensors for chemical and biological agents a matter of high importance and great urgency. Our growing knowledge of the olfactory processes and their translation into electronic noses provide much encouragement for this effort. Further research offers the potential for the development of the high performance, truly nose-like nanosensing systems we seek to build.

**Acronyms**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Å</td>
<td>angstrom ($= 10^{-10}$ m)</td>
</tr>
<tr>
<td>ANN</td>
<td>artificial neural network</td>
</tr>
<tr>
<td>RDA</td>
<td>advanced research and development activity</td>
</tr>
<tr>
<td>ATP</td>
<td>adenosine triphosphate</td>
</tr>
<tr>
<td>cAMP</td>
<td>cyclic adenosine monophosphate</td>
</tr>
<tr>
<td>BL</td>
<td>barley lectin</td>
</tr>
<tr>
<td>ChemFET</td>
<td>chemically sensitive field-effect transistor</td>
</tr>
<tr>
<td>cm</td>
<td>centimeter ($= 10^{-2}$ m)</td>
</tr>
<tr>
<td>CNG</td>
<td>cyclic nucleotide-gated</td>
</tr>
<tr>
<td>DARPA</td>
<td>Defense Advanced Research Projects Agency</td>
</tr>
<tr>
<td>DFA</td>
<td>discriminant function analysis</td>
</tr>
<tr>
<td>1,3-DNB</td>
<td>1,3-dinitrotoluene</td>
</tr>
<tr>
<td>DNT</td>
<td>dinitrotoluene</td>
</tr>
<tr>
<td>ETRI</td>
<td>Electronics and Telecommunications Research Institute, Korea</td>
</tr>
<tr>
<td>FET</td>
<td>field-effect transistor</td>
</tr>
<tr>
<td>G</td>
<td>electrical conductance</td>
</tr>
<tr>
<td>Golg</td>
<td>G protein specific to the olfactory system</td>
</tr>
<tr>
<td>IT</td>
<td>information technology</td>
</tr>
<tr>
<td>KAMINA</td>
<td>Karlsruhe Mikro-Nase (Karlsruhe [Research Center] Micro-Nose)</td>
</tr>
<tr>
<td>m</td>
<td>meter</td>
</tr>
<tr>
<td>MEMS</td>
<td>microelectromechanical systems</td>
</tr>
<tr>
<td>mm</td>
<td>millimeter ($= 10^{-3}$ m)</td>
</tr>
<tr>
<td>MOSFET</td>
<td>metal-oxide-semiconductor field-effect transistor</td>
</tr>
<tr>
<td>mV</td>
<td>millivolts</td>
</tr>
<tr>
<td>MWNT</td>
<td>multi-walled carbon nanotube</td>
</tr>
<tr>
<td>µm</td>
<td>micrometer ($= 10^{-6}$ m)</td>
</tr>
<tr>
<td>N</td>
<td>Newton</td>
</tr>
</tbody>
</table>
TOWARD NANOMETER-SCALE SENSING SYSTEMS

Toward Nanometer-Scale Sensing Systems

nm nanometer (= 10\textsuperscript{-9} m)
NNI National Nanotechnology Initiative
OB olfactory bulb
OR olfactory receptor
PCA principal component analysis
pH hyprogen power, or the degree of acidity or alkalinity of a solution
PLL phase-locked loop
ppb parts per billion
ppm parts per million
QCM quartz crystal microbalance
RPI Rensselaer Polytechnic Institute
SAW surface acoustic wave
SiNW silicon nanowire
SWNT single-walled carbon nanotube
TNT trinitrotoluene
V volts
VLSI very large scale integrated
VOC volatile organic compound

ACKNOWLEDGEMENTS

The author wishes to thank Dr. James Ellenbogen and Dr. Carl Picconatto of the MITRE Nanosystems Group for their technical guidance and helpful comments, as well as their encouragement during the writing of this report. Also contributing to the report were Divya Nettimi and Eliah Shamir, who worked in 2002 on a taxonomy of sensors and investigations into molecular recognition; Laura Plummer, who in 2003 reviewed the literature concerning carbon nanotubes and nanowires; and Mark Taczak and John Whitenack, who provided additional assistance in the literature search and in final editing. Artist Mickey Stinnett’s help in drawing the illustrations is much appreciated. This research was supported by the U.S. Intelligence Community’s Advanced Research and Development Activity (ARDA) Molecular Information Technology (IT) Program, as well as by the Nano-Enabled Technology Initiative (NETI) of the Intelligence Technology Innovation Center (ITIC).

REFERENCES


Simulation of Nanoscale Electronic Systems

UMBERTO RAVAIOLI

Beckman Institute and Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign
Urbana, IL 61801
USA
ravaioli@uiuc.edu

Abstract
As modern silicon integrated devices have reached nanometer scale, device simulation has evolved to provide designers with physical tools that account for particle and quantum features of the transport. At the same time, the ultimate scaling limits of silicon structures are being approached and new material systems have been investigated as candidates to complement or possibly replace silicon technology in the future. Simulation has an important role for the exploration of new molecular device systems, but the increased model complexity presents also many new computational challenges. This chapter will present an overview of simulation approaches for nanoscale device systems with application examples ranging from traditional silicon MOSFETs to molecular and bio-inspired structures.

1. Introduction ................................................................ 168
2. Simulation Hierarchy for Semiconductor Devices ............... 173
3. Simulation Issues in Nanoscale Silicon Devices ................... 188
4. Organic Molecular Devices ........................................ 197
5. Simulation of Molecular Conduction ............................... 200
6. Carbon Nanotubes .................................................. 208
7. Ionic Channels ........................................................ 220
8. Conclusions ................................................................ 241
   Acknowledgements .................................................... 242
   References .............................................................. 242
1. Introduction

Since the invention of the transistor and the introduction of the integrated circuit, solid-state semiconductor devices have powered the spectacular growth of electronics and computing. The law of growth extrapolated by Moore in 1965 [1] has remained amazingly true for several decades, and the sizes of integrated devices has shrunk to the point where nearly a billion transistors can be packed today on a single integrated circuit.

Many factors have created the conditions for such an extraordinary progress. The development of semiconductor processing technology certainly deserves a large share of the merits. While a number of materials have been investigated for device applications, silicon has emerged as the semiconductor of choice to pursue extreme scaling and integration [2–6]. The ease of oxide growth on top of a silicon layer has been the essential feature upon which the processing steps of planar fabrication technology are based. The metal-oxide-semiconductor field-effect transistor (MOSFET) is the most common device used today in integrated circuits [7]. At the core of its function is a thin layer of silicon dioxide (SiO$_2$) sandwiched between the conducting channel and the gate electrode, which forms a capacitive structure controlling the flow of charge in the channel when one varies the voltage applied to the gate (Fig. 1). In modern commercial devices the effective length of the channel has shrunk to about 50 nm and the thickness of the oxide layer can be on the order of just 1 nm. One could argue that the MOSFET is the most common and the most engineered nanotechnology product available today but few expert would have been willing to predict the possibility of fabricating such small working silicon devices just a decade or two ago. If one goes back in history a little farther, it was even generally doubted whether a MOSFET structure could function reliably at all as a transistor.

It turned out that the main problem plaguing the quality of early MOSFETs was contamination of the oxide. Until the need for absolute cleanliness in the fabrica-
tion laboratory was understood, contaminants like sodium ions from sweaty hands would find their way into the gate oxide layers. Sodium is extremely mobile under the intense electric fields of the gate and created havoc in the behavior of the device capacitance, making devices uncontrollable over time. Fabrication quality issues were resolved early on in the history of semiconductors, still, few would have thought that silicon was the best candidate for high-speed computing. Silicon was perceived as a “transitional” material and research was very active to find a replacement, particularly in the 1980s, to the point that fundamental research in silicon devices was not considered worthy of funding by US agencies at that time.

The main issue lies with the carrier mobility in silicon. Electrons move somewhat slowly under the application of an electric field, because conduction electrons in silicon experience a relatively heavy effective mass and suffer frequent scattering events due to vibrations (phonons) of the atoms forming the crystal structure. Other III–V semiconductor materials like GaAs have a smaller effective mass and electrons can move much faster in a certain range of electric fields with a much higher mobility and drift velocity, as shown in Fig. 2. For comparison, the figure shows also the drift velocity for a semiconducting carbon nanotube, which will be discussed in detail.

![Drift velocity vs Electric Field](image)

Fig. 2. Comparison of drift velocities as a function of electric field for bulk GaAs (circles), bulk Si (squares) and for a zig-zag semiconducting carbon nanotube (continuous line).
later. It should be natural to think that for high-speed computing it is necessary to adopt materials where transistor switching can take place with an intrinsically faster mechanism. In addition, it was found almost thirty years ago that the velocity of carriers, on the plane of the heterojunction between materials like GaAs and the alloy AlGaAs, could be increased by orders of magnitudes, leading to the creation of a completely new class of high electron mobility transistors (HEMT) [4]. The appeal of these new materials was even greater because of high compatibility with optoelectronics devices that could have allowed the integration of high-speed transistors with optical interconnects. It was widely expected then that the rich variety of possibilities afforded by band gap engineering of new semiconductor alloys and heterostructures would quickly bring a wealth of new applications characterized by ultra-fast devices. Also, due to the higher carrier velocity and lower scattering rates, it was hoped to realize devices virtually immune from phonon scattering, working in the so-called “ballistic” regime.

Although new semiconductor materials have found an important niche in microwave applications and optoelectronics, the predicted computing revolution based on them never materialized. Cost and reliability issues have certainly been important factors, since silicon growth, processing and fabrication have enjoyed over five decades of development up to now and it is very difficult for any new technology to become commercially competitive. Still, silicon was destined to remain unchallenged in the integrated circuit arena precisely because of material properties which were considered to be its weaknesses. In the race towards miniaturization, it is very important for a device to retain its behavior essentially unchanged when scaled, so that one does not need to modify radically fabrication approaches and system architectures along the way. The carrier transport behavior of a material like GaAs is strongly dependent on the actual electric fields established inside the conduction channel. While electron mobility is very high at low fields, the saturation velocity at the fields that would be established in a practical conduction channel is actually slight lower than in silicon, due to the intervalley scattering mechanisms to upper valleys in the conduction band, which have a much higher effective mass and lower mobility [8]. To really harness the potential of high mobility, devices need to be sufficiently small so that it is possible to operate in the so-called overshoot regime, where electrons traverse the channel before intervalley transfer events can take place. But this also means that by scaling further the device, a regime of quasi-ballistic transport can be established more easily where quantum effects and coherence of the electron wave need to be dealt with. Silicon has a heavier effective mass and it has essentially retained until today most of its classical behavior as far as transport along the channel axis is concerned. Size quantization effects in the cross-section of the channel are becoming more influential but the nature of electronic transport has not changed substantially even in the nanometer size range. Generations of device designers have
been able to gradually adapt their approaches to make new devices behave acceptably as they have been scaled, but the process has been an evolutionary rather than a revolutionary one.

Another winning point for silicon technology has been the C-MOS architecture [4,5], based on a basic inverter structure which only conducts current when the switching between two logic states takes place. For this, the insulating nature of the readily available SiO₂ gate is of fundamental importance. The realization of comparable gate oxides in alternative material systems is in general impractical or not possible at all. When millions of transistors are packed in the same integrated circuits it is crucial to minimize power consumption to avoid the possibility of thermal failure. To-date there is simply no other technology that can approach, even remotely, what can be achieved with present C-MOS silicon technology for large scale device integration.

It is conceivable that practical MOSFETs can be designed and fabricated with gate lengths of only several nanometers, but this is only part of the requirements to maintain scaling trends for commercial applications. It is progressively more difficult to control the leakage currents in the off-state for extremely scaled devices, leading to serious power dissipation issues. This problem could be addressed by power management strategies that disconnect idle devices to eliminate leakage sources, but at the same time, fluctuations and imperfections in the manufacturing process are becoming limiting factors in the reliability of devices. New circuit design approaches will be required to tolerate imperfect or non-uniform device behavior, while still providing the expected output from a processor [9]. Keeping up with Moore’s law will become increasingly difficult and one could argue that for the long term the law is surely unsustainable if one simply invokes fundamental physical limitations. A device could not be smaller than a single atom, after all. A slowdown in the industry is already noticeable and the race for smaller, faster MOSFETs has been deemphasized in recent times in advertisements of personal computers. It is certain that designers and technologists in the future will attempt to exploit any possible margin afforded by degrees of freedom in the design of new architectures, to alleviate the growing problems connected with deep scaling of devices. Nevertheless, it is also expected that in a relatively short time silicon technology, as it is known today, will mature completely and its scaling will come to an end.

The outlook may seem bleak to some, but others may see this as an exciting opportunity to explore alternatives that could one day complement or even replace today’s semiconductor technology with new ones for large scale applications. Scientists have naturally looked at molecular systems, as possible candidates for the exploration of new nanotechnologies. One can find in nature a variety of stable and robust molecules that could be produced or harvested with nearly perfect uniformity of shapes and properties and which exhibit device-like behavior under external stimuli. In certain
areas of electronics, organic materials have already made promising inroads, particularly for the realization of optical devices and displays [10]. Conduction through organic molecules is being studied with the goal to realize elementary computational or memory elements while a great deal of attention is also being paid to carbon nanotubes, which yield solid-state structures dominated in their behavior by the specific molecular footprint achieved during growth and which may allow a rich variety of new applications [11]. There is also interest in biomolecular systems. DNA molecules, which have long wire shapes, have been suggested as possible structures for computing applications [12]. Biological ionic channels that act as natural nanoscale devices and display a range of behaviors, are investigated for possible applications as switches, sensors and actuators [13]. Finally, semiconductor materials with magnetic properties are investigated to provide the ability to select the spin quantum number of carriers injected into a device [14]. Since the time decay constant of a given spin state can be much longer than characteristic device operation times, the possibility to encode information in spin states provides a new computational paradigm full of intriguing possibilities. Fundamental research on quantum computation in general is currently receiving considerable attention [15] but practical applications are not yet in sight and this area of work will not be examined here in detail. The main limitation today is due to the fact that extremely low temperatures are still necessary to demonstrate working quantum computing applications of any kind.

While molecular systems are the obvious object of attention in the attempt to realize nanoscale device elements, realistically one cannot expect a rapid technological development that may quickly provide an alternative to semiconductor devices. Large scale silicon integrated circuits still provide an unparalleled success story of ingenuity and innovation which has gradually revolutionized modern society by introducing the power of computation in almost all aspects of everyday life. One should expect that when intrinsic scaling limits are reached, silicon technology would continue to be refined, playing a fundamental role in any electronics application. Even if totally new devices become practical, we are still likely to see for a long period of time the emergence of hybrid technologies where new systems, made of nanoscale building blocks that realize specialized tasks, will be implanted onto host systems based on traditional semiconductor technology approaches.

Systems completely realized with molecular technologies are still difficult to envision. Many efforts are focused on finding a one-to-one replacement for the MOSFET but, arguably, completely new architecture and computational paradigms are necessary to realize the promise of new potential molecular approaches. Biological or artificial (biomimetic) nanochannels may find uses as single molecule detectors in novel sensing applications, possibly operating as a combination of analog and digital computing elements.
Going back to the factors that have seeded the rapid growth of semiconductor technology, computation has played a rather remarkable role, evolving alongside the evolution of the computer itself. Computer aided design (CAD) is now an essential stage in the design and fabrication loop and design of large systems would be inconceivable today without the use of CAD tools for circuit simulation [16]. While once it was possible to fine tune a device structure by trial-and-error in the laboratory, at the nanoscale regime of today one has to rely more and more on process and device simulators to optimize and calibrate design and fabrication stages to achieve the necessary performance. The exploration of new device concepts relies heavily on physical simulation to probe parameters that cannot be measured directly in systems that are approaching rapidly the atomic scale. One can see that there is a positive feedback, where the realization of more advanced computers provides the tools to invent the new technologies necessary to sustain the pace of innovation. The main goal of this work is to discuss the current status of computation in device simulation and to outline the challenges ahead as research moves on to consider new material systems to overcome the roadblocks of the future.

2. Simulation Hierarchy for Semiconductor Devices

Semiconductor device integration has progressed enormously since 1990, often surpassing predictions and roadmap objectives. With the reduction of device size and increase of operating frequency, the needs for device simulation have also undergone major changes. In the conduction channel of short silicon devices very large fields are established, even if every effort is made to reduce bias voltages, and the energy distribution of carriers is appreciably out of equilibrium. In particular, the high energy tail of the carrier distribution becomes of great importance to understand the details of transport and device behavior. The presence of these so-called hot carriers is typically linked to reliability problems. When carriers reach an energy threshold (approximately equal to the band gap energy) impact ionization can take place with generation of electron–hole pairs that induces a substrate current able to trigger parasitic bipolar action [17].

Extremely energetic carriers may also surmount the oxide interface barrier and can contribute to damage of the insulator [18,19] although this injection phenomenon also performs a beneficial function, when properly controlled, being at the basis of floating gate programming in flash memory devices [20,21]. In device fabricated with III–V compound materials like GaAs, hot carrier effects had to be contended with very early on, due to the typically small effective mass at the bottom of the conduction band, which leads to markedly non-equilibrium energy distributions at much lower fields than in silicon. The conduction band structure of most III–V compounds
of practical interest exhibits satellite energy valleys with much larger effective mass, which usually can be reached by hot carriers at relatively low fields [22]. Intervalle
transfer leads to a degradation of drift velocity which is detrimental for high speed transistor operation. The same phenomenon, known as Gunn effect [8], has of course beneficial applications when applied to create traveling carrier domains for the generation of high frequency signals in Gunn diodes. Oxide injection in MOS devices is analogous to what is usually called, in III–V compound heterostructures, real space transfer, taking place when hot carriers in a conduction channel defined by heterojunctions can transfer from the smaller bandgap channel region into the surrounding layers with larger bandgap [23]. Since the energy barrier to overcome is quite smaller than in the Si/SiO₂ system, the transfer phenomenon can be very pronounced and is at the basis, for instance, of transistor action in a number of structures that have been proposed both for microwave amplification and high frequency generation [24,25].

The starting point for analysis of carrier transport in a given material is usually the dependence of drift velocity versus applied electric field in steady-state bulk conditions. At low fields, the drift velocity increases linearly with the electric field, according to a slope that corresponds to the low field mobility of the material. At high fields, because of a balance established between carrier acceleration and scattering losses, a saturation of the drift velocity is normally observed. In the intermediate field range, the velocity behavior depends largely on the detail of the energy band structure and on the scattering rates. In devices with long conduction channels, the electric field parallel to the channel tends to be slowly-varying at normal drain bias voltages, so that the steady-state drift velocity is established locally as in a bulk situation at the same electric field. When devices are scaled and electric fields exhibit both high strength and rapid variation along the conduction channels, a bulk-like steady-state cannot be established locally. The velocity in excess of the reference bulk value at the given local field is called velocity overshoot. In III–V devices the overshoot phenomenon is quite pronounced while it is more modest but detectable in silicon devices [26]. However, since drift velocity is a quantity averaged over the local carrier distribution, it only gives an indication of the global current flow. Drift velocity and local average energy are not necessarily good indicators of hot carrier phenomena and reliability in highly scaled devices. Complete transport information requires also the knowledge of the actual carrier distribution in energy or momentum space. It is, therefore, useful to survey the hierarchy of physical simulation for semiconductor transport and device simulation, giving an indication of the limits of applicability and approximations underlying the various approaches. The main focus is on the relevance of the approaches for the simulation of hot carrier effects in deeply scaled devices.

A schematic diagram for the complete hierarchy of approaches that can be applied to device simulation is shown in Fig. 3. At the top of the hierarchy we have
approaches based on a quantum description of transport, an area which is computationally challenging and is still undergoing considerable development. The levels below are instead well established and based on the Boltzmann Transport equation or its simplifications, with a semi-classical rather than quantum description of transport. All the simulation levels, except for the compact approaches at the bottom of the hierarchy, involve the solution of a set of coupled partial differential equations, where the transport equation can actually be an integro-differential equation as is the case for the Boltzmann equation and for quantum formulations including scattering. Analytical approaches are based on integral solutions of such semiconductor equations, suitable as building blocks for the solution of complete circuits. In order to attain a self-consistent solution, the equations of the transport approaches are nearly always coupled to Poisson equation. The general form of the equation is

\[ \nabla \times \varepsilon \nabla V = -\rho = q(n - p + N_A^- - N_D^+), \]  

where \( V \) is the electrostatic potential, \( q \) is the magnitude of the electron charge and \( \varepsilon \) is the dielectric permittivity. The space dependent charge density \( \rho \) is obtained from the knowledge of the density of ionized donor \( N_D^+ \) and acceptor \( N_A^- \) dopants, and the density \( n \) and \( p \) of mobile electron and hole carriers. Since the carrier densities provide the coupling with the transport equations, the implementation of Poisson equation for the various approaches differs in the determination of \( n \) and \( p \), which
may be defined in terms of wave function, distribution function, quasi-Fermi levels, and so on.

All simulation approaches in the hierarchy are differentiated by the band structure model used to describe the dispersion relation between energy \( E(k) \) and momentum \( k \). The most complete approach would incorporate the complete dispersion relation from the full band structure determined numerically, where the particle group velocity is obtained from \( v = \nabla_k E(k)/\hbar \). The elementary approach, valid only at low energies, uses just a simple analytical parabolic band \( E(k) = \frac{\hbar^2 k^2}{2m^*} \) characterized by a fixed effective mass \( m^* \) and velocity \( v = \hbar k/(2m^*) \). An intermediate analytical approach extends the validity of the bands by adding a non-parabolicity coefficient \( \alpha \), as \( E(k) + \alpha E^2(k) = \frac{\hbar^2 k^2}{2m^*} \) where now the velocity becomes \( v = \hbar k/(m^*[1 + 2\alpha E(k)]) \). Based, on the non-parabolic band form, various approximations of the full band can be constructed, by using sets of conduction valleys that are coupled by the transport equations through appropriate scattering rates.

The quantum approach can in principle describe transport phenomena in solids with the most complete formulation, but a full quantum model can almost never be implemented with the computational resources available today, and quantum device simulation is rather applied to restricted physical situations. The most useful application of quantum models in devices has been perhaps the analysis of size quantization, which involves the solution of the coupled time-independent Schrödinger equation and the Poisson equation, successfully applied to 2D electron gas confinement, quantum wires and quantum dots [27–29]. Non-parabolic energy dispersion can be used with some approximations of the operator [30,31] and a numerical approach with a general dispersion relation is also possible in 3D problems, using a spectral technique [32]. These are, however, all static applications. In the most common transport problem, the Schrödinger equation for the envelope wave function is solved under the assumption of effective mass [33,34]. A simple parabolic band structure is in most cases the only practical alternative, because of the complications in formulating the Hamiltonian for a more general band. In addition, it is extremely difficult to include scattering in a practical manner, and often the simulations are restricted to low temperature cases or extremely small structures, where ballistic transport can be assumed. More sophisticated approaches involve the use of the density matrix [35,36] or the Wigner distribution function [37–40]. The Wigner function is the Fourier transform of the density matrix and it represents the quantum version of the classical carrier distribution function, whose transport is governed by the Liouville equation. Scattering can be included in a simple way in these models by adding phenomenological relaxation terms which are not always completely satisfactory. In all cases, inclusion of a detailed band structure remains a problem that limits the range of practical applicability of quantum approaches. An alternative approach based on non-equilibrium Green’s functions has been become more practical in recent times
Within the formalism it is possible to include a more detailed band structure, for instance defined with a tight binding formulation. However, the complexity of the problem increases considerably when going from simple ballistic transport to a model including scattering.

The solution of a complete device in traditional terms still remains a formidable task. The most analyzed device is the resonant tunneling diode, for which a 1D approach is sufficient [37–41,45]. With present scaling trends, interest in quantum models is growing also for silicon devices. It is not well understood, yet, what the real scalability limits of MOSFET structures might be, but for the practical structures of today there is a prevalence of simulations that combine semi-classical hot carrier transport and size quantization features via quantum corrections of the potential. For some other devices, quantum effects may be strong but localized. An example is the MOSFET with silicide contacts where injection from the source region into a conduction channel takes place through a Schottky barrier, with tunneling modulated by the gate voltage [46,47].

A quantum hydrodynamic approach has also been developed where quantum potential corrections are introduced in a formalism similar to the classical hydrodynamic approach based on moments of Boltzmann equation, described later on. This approach has been applied to resonant tunneling diodes [48] and very small heterojunction transistors [49]. The goal is to include the essential features of hot carrier and quantum transport, with a numerical formalism that has already been widely explored, and may serve as a useful engineering tool to bridge the gap between semi-classical and complete quantum approaches.

The next levels of the hierarchy relate to approaches based on the semi-classical Boltzmann Transport equation. At this level, the carrier population is represented by a gas of classical particles, with definite position and momentum, described by a distribution function. The equation has the form

$$\frac{\partial f}{\partial t} + v \cdot \nabla f + \frac{eF}{\hbar} \cdot \nabla_k f = \sum_{k', \lambda} \left\{ S_{\lambda}(k', k) f(r, k', t) \left[ 1 - f(r, k, t) \right] - S_{\lambda}(k, k') f(r, k, t) \left[ 1 - f(r, k', t) \right] \right\},$$

where $r$ is the position, $\lambda$ denotes scattering mechanisms, $f(k, t)$ is the distribution function, $v$ is the group velocity, $F$ is the electric field, $S(k, k')$ is the transition probability between the momentum states $k$ and $k'$, and $[1 - f(k', t)]$ is the probability of non-occupation for a momentum state $k'$. The summation on the right-hand side is the collision term, which accounts for all the scattering events.

The solution of Boltzmann equation has been attempted in many different ways, due to the fact that an approach by direct numerical discretization is very challenging. Stochastic Monte Carlo approaches based on particle simulation have been the most
popular and successful, and for this reason a specific entry for Monte Carlo methods is indicated in the hierarchy of Fig. 3. A number of alternative numerical approaches, have also received increasing attention in recent times but since Boltzmann equation is an integro-partial differential equation, direct solution of the equation remains very difficult for a realistic device structure, even in the approximation of parabolic bands. Since the distribution function is defined on the complete phase space, a multidimensional solution involves an enormous level of complexity. The only practical device applications, where the Boltzmann equation is solved directly, involve an expansion of the distribution function in spherical harmonics in momentum space, to reduce the dimensionality of the system [50]. The coefficients of the expansion become a new set of unknowns, which depend on the physical space coordinates. An advantage of this approach is that the accuracy of the spherical harmonic expansion can be controlled by selecting an arbitrary number of terms, with the possibility of very efficient computational applications. Simulations have been reported for MOSFET structures [51,52]. The model can be formulated for multiple conduction bands, at the cost of increased computational complexity. The main limitations are in the approximations that need to be made on the scattering model to express the right-hand side of the BTE in a manageable form, and in the intrinsic loss of information arising from the substitution of the original distribution function with a spherical expansion.

Other approaches have been proposed to solve the Boltzmann equation with the goal to overcome some of the limitations of direct solutions while still offering computational efficiency. Besides the methods discussed below, there are other interesting approaches using deterministic particle methods [53] and Chambers Path Integrals [54] for which the reader is referred to the references for details.

In the Scattering Matrix approach, carrier transport is viewed as the transmission and reflection of carrier fluxes within a semiconductor [55]. The simulation domain is subdivided into thin slices (1D) or meshes (2D), so that these regions are sufficiently small to assume constant doping and fields within. Transport across each region is described by a matrix equation which relates the incident carrier fluxes to the emerging fluxes, through the transmission coefficient of a scattering matrix. To resolve the distribution function, the flux is discretized into a number of bins in momentum space (modes), and the elements of the scattering matrix can be computed by Monte Carlo simulation of thin slabs at constant fields. In device analysis, scattering matrices for each region are connected and some known fluxes are injected from the contacts. The resulting fluxes through each device region are then computed iteratively, until they converge. The position dependent distribution function is constructed from the magnitude of the fluxes within each mode. This technique can be computationally efficient even in 2D, if the implementation is done carefully, and the resulting distribution functions are essentially free of noise. Of course, this technique depends on other methods (like Monte Carlo) to pre-calculate the scattering
matrices. While the memory storage can become quite large if very fine resolution is required, a full band structure can be included without substantial modifications of the algorithm.

The Cellular Automata approach stems from similar work developed to solve fluid dynamics problems. The non-local integro-differential Boltzmann equation is converted into a cellular automaton with only local (nearest neighbor) interaction in position space by discretization over a virtual lattice in phase space [56,57]. Attached to each position space cell $x$ there is a set of momentum states $k$, over which the distribution function is defined. The sites $x$ and $k$ are populated by fictitious particles obeying non-deterministic dynamical rules. Such rules consist of transitions between the discrete $k$ cells at a given lattice site, representing scattering events in momentum space, and transitions between $k$ cells on neighboring lattice sites $x$, corresponding to carrier motion. If these rules are carefully selected, the cellular automaton effectively represents a boolean equivalent of Boltzmann equation. The cellular automaton approach simulates particles as in Monte Carlo techniques, but since these are virtual particles that mimic Boltzmann equation, rather than the actual behavior of physical particles, the marching algorithm evaluating the transport is not affected by randomness and noise as much as in Monte Carlo approaches, and can be very effectively implemented on supercomputing platforms.

The transport process can also be resolved by Evolutionary Algorithms using a mathematical optimization techniques which is inspired by natural evolution of species, based on creation of new individuals and selection of the fittest [58]. The natural evolution can be transferred to mathematical optimization problems by viewing each vector $V$ in a search space $S$ as a possible individual of the species, in the environment specified by an objective function $F$. The value of the objective function for each vector $F(V)$ is the equivalent of the fitness of these mathematical individuals. An initial set of vectors, called population, is the starting point of the mathematical evolution. At each following iteration, a new generation of individuals is created, first by selection of a subset of the old population, then by creation of new individuals through modifications of the vectors of the subset, and finally by selection of the next generation population from both old and new vectors. In the case of semiconductor transport, the optimization procedure is applied to the distribution function of carriers, where the evolution at each step is accomplished by a physical mutation operator which was developed to mimic the effect of an ensemble Monte Carlo step looking at the behavior of the global distribution function, rather than single particles. Besides the physical mutation operator, the electron distributions are modified by two random mutation processes (one increases or decreases the tail of the distribution function, the other changes the distribution smoothly around a randomly selected energy) and one recombination process that is achieved by exchanging sections of two distributions. The random mutation operators and the recombination process are
purely mathematical operators and do not contain physical information. A physical quantity, that could be for instance the result of a measurement, is set as the goal and is used to define the fitness of the iterated distribution, by comparison with the corresponding calculated quantity. In the case of a MOSFET, where one wants to find the high energy tail, the substrate current measured at a specified bias could be selected as the basis for the objective function. The Evolutionary Algorithm methods, besides being computationally very efficient, can be performed backward in time and are therefore very suitable to infer distribution functions in correspondence of specific experimental results.

For several decades now, practical solution of Boltzmann equation for semiconductor transport has been accomplished by means of particle Monte Carlo methods [22,59–62]. The name of Monte Carlo method is usually given to a variety of stochastic techniques which use random number generation. In the case of particle transport, the Monte Carlo method is used to solve statistically the Boltzmann equation, without making assumptions on the distribution function. The appropriate probability distributions for the observables involved are obtained from uniform random number sequences. This is achieved through numerical experiments which simulate the actual motion of particles inside a semiconductor, described as a sequence of free flights interrupted by scattering events. The trajectories and the momentum variations are calculated using classical mechanics, while the scattering rates and the scattering angles are calculated according to quantum mechanical rules. Self-consistent device applications require the simulation of an ensemble of particles, chosen to be a sufficiently representative sample of the device population, and the solution of Poisson equation at frequent time intervals to update the forces acting on the particles. Over the last decade, device simulation including the complete band structure has become practical [62,63,21,64]. In this approach, the band structure data is made available to the simulation through large tables for energy, velocity, and density of states.

The Monte Carlo approach is equivalent to the Boltzmann equation in the limit of infinite sampling. However, because Monte Carlo simulations are able to resolve physical fluctuations, in many respects the approach can go beyond the limits of validity of the Boltzmann equation when it is properly implemented, at least in its ensemble self-consistent form. The main drawbacks of Monte Carlo is in the large computational resources necessary and in the statistical noise that arises from the fact that only a limited sample of particles (typically 10,000 to 50,000) are simulated. Statistical methods to reduce the variance of averaged quantities are useful in eliminating large part of the noise, thus providing a way to attempt an estimate of rare hot carrier events like injection into the oxide of a MOSFET [65–67]. It could be argued that a Monte Carlo approach may simulate naturally occurring noise in devices, and therefore a well set up Monte Carlo model could yield more realistic
results for a single device than the actual Boltzmann equation, where the results are implied to be an average over a large number of similar devices.

There is a hierarchy also within Monte Carlo approaches, which are differentiated both in terms of band structure and scattering models. While for III–V compound devices an analytical non-parabolic multivalley band structure is often adequate, in the case of silicon under high field conditions a more detailed band is desirable. Besides a full numerical band structure represented by large tables, a combination of many valleys that mimic the actual numerical band have been attempted in various forms. Most device simulations are based on energy dependent scattering rate models, where the global density of states is used, while the anisotropy of the scattering rates over a given energy shell in momentum space is neglected. In models that implement a detailed band structure, such anisotropy is accounted for at the level of random selection of the final states after a scattering event. The main difficulty (and source of uncertainty) in the determination of the scattering rates is the selection of factors called deformation potentials which, although measurable in some particular cases, enter the rate formulations as adjustable parameters. In addition, deformation potentials can be anisotropic within a given energy shell.

For practical device simulation purposes, the energy dependent scattering model is normally considered to be sufficient. Improvements of the scattering approach have been attempted, although usually limited to applications for bulk simulation [68–70] or non-self-consistent device models with approximated bands [71]. To capture fully the anisotropy of the bands, transition rates between all possible momentum states are necessary, leading to enormous tables, even after taking advantage of the symmetry of the Brillouin zone. One of the difficulties is to formulate an appropriate physical model for the calculation of the transition rates. Since the band structure must be calculated, the same parameters could be used for the evaluation of transition rates, thus eliminating the need to find the deformation potentials which have to be determined empirically for most of the scattering rates used in standard Monte Carlo approaches. Therefore, the unified theory for band structure and transport parameters can be based only on several material constants, reducing the most considerable source of uncertainty in the Monte Carlo scattering models that are based on a fit of deformation potentials. Work in this area was done by using empirical pseudopotentials [68,69] and \textit{ab initio} pseudopotentials [70]. The electron–phonon scattering model is based on a rigid pseudo-ion approach. The empirical pseudopotentials are assumed to be sums of spherical potentials moving rigidly with the crystal atoms, in electron–phonon calculations. In the \textit{ab initio} density functional approach, ionic pseudopotentials are assumed to be spherical and rigidly transferable from the atom. The total Kohn–Sham equation contains contributions from the self-consistent crystalline charge density under distortion. The deformation potentials are obtained directly from the theory, and depend on the pseudopotentials used to calculate the
band structure. Therefore, scattering rates and band structure are consistent. The unfortunate aspect of this approach is that some tunable degrees of freedom are lost, leaving less room for calibration. The quality of the transition rates is therefore dependent on the complexity of the theory, and there is still work to be done to achieve a precise quantitative match. While for the empirical pseudopotential approach in [69] average deformation potentials are assumed on a given energy shell to define the scattering rates, in the \textit{ab initio} case the deformation potentials are a result of the model and it is shown in [70] that they may vary quite strongly as a function of momentum and energy.

The lower levels in the hierarchy do not resolve the distribution function in detail, rather the transport model is based on the moments obtained by integrating Boltzmann equation after multiplication by a suitable variable (velocity, energy, etc.). The various transport variables are represented through average densities that are obtained as moments of the distribution functions. For instance, the electron density \( n \), the average velocity \( v \) and the energy density \( E \) are given by the following expressions

\[
\begin{align*}
n &= \int f(r, k, t) \, d^3k, \\
nv &= \int v(k) f(r, k, t) \, d^3k, \\
E &= \frac{m^*}{2} \int v(k) f(r, k, t) \, d^3k.
\end{align*}
\]

One can find in the literature a wide variety of approaches that originate from the moments of Boltzmann equation, and which differentiate themselves mainly in the approximations or simplifications made to obtain practical implementations [72–77]. We will examine in detail the two formulations that have more relevance today, and that best exemplify the capabilities of the moment method. The two approaches also provide an illustration of practical approximations of the collision integral in the Boltzmann equation, in terms of relaxation times. In the first approach, the so-called \textit{Hydrodynamic model} [72] the moments are initially performed and then average relaxation times for the momentum and energy continuity equations are introduced. In an alternative approach, the \textit{Energy Transport model}, the collision term in the Boltzmann equation is first formulated in terms of a \textit{microscopic} relaxation time for the distribution function and then the moments are based on this approximated equation [73–76].

The transport equations which express conservation of particles, momentum, and energy for the Hydrodynamic model may be obtained from the moments of the Boltz-
mann equation as [72]
\[
\frac{\partial n}{\partial t} + \nabla \cdot (vn) = \left( \frac{\partial n}{\partial t} \right)_{\text{coll}},
\]
\[
\frac{\partial P}{\partial t} + \nabla \cdot (vP) = -qnF - \nabla (nk_BT) + \left( \frac{\partial E}{\partial t} \right)_{\text{coll}},
\]
\[
\frac{\partial E}{\partial t} + \nabla \cdot (vE) = -qnv \cdot F - \nabla \cdot (nk_BTv) - \nabla \cdot Q + \left( \frac{\partial W}{\partial t} \right)_{\text{coll}},
\]
where \( v \) represents here the average velocity, \( P \) is the momentum density, \( E \) is the kinetic energy density, \( T \) is the carrier temperature tensor, \( Q \) is the heat flow vector, and \( F \) is the electric field. No assumption has been made on the nature of the collision terms, in the above equations. One set of equations can be written for each valley in the conduction band of the semiconductor considered, with different sets of valley-specific variables. The three equation above are in the order the carrier continuity equation, the momentum balance equation, and the energy balance equation.

Since the hydrodynamic equations are a truncated series of moments, a fourth equation that defines \( Q \) is necessary for closure. Heat conduction has been traditionally assumed to obey the classical Franz–Wiedemann law \( Q = -\kappa \nabla T \), where \( \kappa \) is the heat conductivity of the electron gas. However, it has been recognized in recent times that this approach is not correct for semiconductors, particularly in correspondence of junctions, where very large and unphysical velocity overshoot peaks are established by a model that is based just on Franz–Wiedemann law. In a simple approach, this problem can be moderated by treating the factor \( \kappa \) as a tuneable parameter, modified to match as closely as possible more accurate calculations. This procedure does not give good results with scaling, and recalibration of \( \kappa \) is always necessary for different structures. A more rigorous approach would involve, for instance, the extension to higher-order moments of Boltzmann equation [78]. This leads, however, to obvious model complications and still requires closure of the moments at a higher level. Another possibility is to modify the heat conduction equation to have a more physical heat transport model dependent on current flow. In [79] the following new form is suggested for practical applications:
\[
Q = -\kappa \nabla T + \frac{5(1 - r)}{2} \frac{k_BT_0}{q} J,
\]
where \( T_0 \) is the lattice temperature, \( J \) is the current density and \( r \) is a tuneable parameter less than unity. A more rigorous and complete assessment of thermal effects can be found in [80]. From the practical point of view of simulation, it is fair to say that the closure problem has not yet been resolved in a fully satisfactory way. The temperature is a measure of the average thermal energy per electron, and often
its true tensorial nature is not considered, so that the temperature is related to an average total kinetic energy $w$ as

$$ w = \frac{E}{n} = \frac{3}{2} k_B T + \frac{m^* |v|^2}{2}. \tag{6} $$

The momentum density is also simply related to the average (drift) velocity according to $P = m^* n v$. As mentioned earlier, the collision terms are approximated after the moments of the Boltzmann equation have been taken. This leads to the need for different phenomenological relaxation times $\tau_m$ for momentum and $\tau_e$ for energy.

For the simple case of a single band with only intravalley scattering we have

$$ \left( \frac{\partial P}{\partial t} \right)_{\text{coll}} = -\frac{P}{\tau_m}, $$

$$ \left( \frac{\partial E}{\partial t} \right)_{\text{coll}} = -\frac{n w}{\tau_e} - n \frac{w - 3k_B T_0 / 2}{\tau_e}. \tag{7} $$

A more detailed band structure can be introduced by using multiple analytical bands, at the cost of a full set of conservation equations for each valley. This also involves the definition of additional relaxation times accounting for intervalley scattering. The relaxation times for momentum and energy can be determined by analytical scattering models or can be extrapolated from Monte Carlo transport simulations. It should be pointed out that in a device simulation, where highly non-uniform conditions are established, relaxation times, particularly for energy, are not necessarily single valued functions, since the local relaxation does not only depend on the local value of the average energy, but rather involves the global transport state of a conduction channel. This is a problem that is always encountered when average relaxation times are employed, and which becomes more acute when very rapidly varying fields are established in small devices.

One of the difficulties in using the hydrodynamics approach is in the numerical nature of the equations. When the average velocity values exceed a certain limit, the conservation laws become hyperbolic and the solutions behave similarly to supersonic flow, with the formation of numerical shock waves that require very careful algorithmic implementation. The Essentially Non-Oscillatory (ENO) scheme, which allows for adaptive discretization stencils to resolve numerical shocks to an arbitrary degree, is perhaps the most natural approach, where the steady state problem is solved by iterating the time dependent equations [81]. More traditional approaches based on Newton-method solution of the non-linear steady-state problem have also been reported [82–84].

In the Energy Transport approach a microscopic relaxation time $\tau$ for the distribution function is assumed in the Boltzmann equation [73,74]. By decomposing the distribution function $f$ into its even and odd parts, $f_0$ and $f_1$ [75,76], and using the
microscopic relaxation time approximation \((\partial f_1/\partial t)_{\text{coll}} = -f_1/\tau\) the odd part of the steady-state Boltzmann equation becomes

\[
f_1 = \frac{q\tau F}{\hbar} \nabla_k f_0 - \tau v \nabla_r f_0.
\]  

The zeroth and second-order moments of the steady—state Boltzmann equation give the carrier continuity and energy balance equations:

\[
\nabla \cdot J = -q(G - R),
\]

\[
\nabla \cdot S = F \cdot J - n \left( \frac{\partial E}{\partial t} \right)_{\text{coll}},
\]  

where the current density \(J\) and the energy flow \(S\) are defined by

\[
J = -qn v = -q \int vf d^3k,
\]

\[
S = n \langle Ev \rangle = \int Ev f d^3k
\]

and \((G - R)\) is the carrier generation–recombination rate. Substitution of the expression for the odd part of the distribution \(f_1\) into the expressions for \(J\) and \(S\) yields tensorial equations. It is worth noticing that the constraint \(f_1 \ll f_0\), which is the basis of the Legendre polynomial expansion approach in the work by Stratton [73] or in other perturbative approaches, is not necessary in the Energy Transport model. For materials like Si, where the mean kinetic energy \(m \langle v^2 \rangle / 2\) is much smaller than the mean thermal energy \(m \langle v^2 \rangle / 2\), a correction from the field-induced anisotropy is generally not important. Therefore, \(f_0\) can be approximated by an isotropic distribution of energy and the tensorial transport coefficient may be approximated by scalar functions of the mean carrier energy. In a simple approach, one can assume a relaxation time of the form \(\tau = \tau(E) \propto E^{-p}\), where \(p\) is determined by the dominant scattering mechanism and ranges from 0.5 to 1 according to bulk Monte Carlo simulations. A non-parabolic band can be introduced fairly easily in this approach, as well as other more subtle details, like non-Maxwellian effects in the distribution \(f_0\).

With the above assumptions, the Einstein relations may be assumed to hold not only for carrier mobility \(\mu\) and diffusivity \(D\) but also for analogous parameters for energy, \(\mu^E\) and \(D^E\), as \(D = k_BT_m\mu/q\) and \(D^E = k_BT_m\mu^E/q\) where the equivalent carrier temperature \(T_m = (1 + \gamma)T_e\) is related to the mean carrier energy by

\[
\langle E \rangle = \left( 1 + \gamma + \frac{5}{2} \alpha k_B T_e \right) \frac{3}{2} k_B T_e \simeq \left( 1 + \frac{5}{2} \alpha k_B T_m \right) \frac{3}{2} k_B T_m.
\]  

Here, \(\gamma\) is the non-Maxwellian parameter that indicates how \(f_0\) deviates from a Maxwellian distribution at an elevated carrier temperature \(T_e\). Then, we can arrive at
the model transport equations for current density and energy flow

\[ J = q \left( -\mu n \nabla \psi + D \nabla n + n \frac{\partial D}{\partial T_m} \right), \]

\[ S = -C_e k_B T_m \left( -\mu n \nabla \psi + D \nabla n + \frac{n}{T_m} \frac{\partial (D T_m)}{\partial T_m} \nabla T_m \right), \]  

(12)

where \( \psi \) is the electrical potential and \( C_e \) is the energy flow factor

\[ C_e = \frac{\mu E}{k_B T_m \mu} = \left( \frac{5}{2} - p \right) \left( 1 - \frac{\alpha k_B T_m}{2} \right), \]  

(13)

\( C_e \) is in general a smoothly varying function of the mean carrier energy, and for nearly parabolic bands we have \( C_e \simeq \frac{5}{2} - p \). It is important to notice that in this approach a law for heat flow, like the Franz–Wiedemann law, does not need to be invoked, as it is necessary in the conventional hydrodynamic method to obtain a closure of the moment equations. In this energy transport model, the necessary transport parameters \( \mu(T_m), C_e(T_m) \) and \( \tau E(T_m) \) can all be determined, for instance, from bulk Monte Carlo simulations.

The energy transport approach is appealing for practical simulation applications, because the numerical procedure developed for drift-diffusion problems can be easily extended to accommodate this approach. Some of the numerical difficulties encountered with hydrodynamic models are not an issue because the momentum continuity equation is not explicitly used.

Drift-diffusion approaches represent an approximation of the lowest order obtainable from the moments of Boltzmann equation [85]. Along with Poisson equation, only the continuity equations for electron and hole carrier densities enter the model

\[ \frac{\partial n}{\partial t} = \frac{1}{q} \nabla \cdot J_n + U_n, \]

\[ \frac{\partial p}{\partial t} = -\frac{1}{q} \nabla \cdot J_p + U_p, \]  

(14)

where \( U_n \) and \( U_p \) indicate the net generation–recombination terms. It is useful to derive an expression for the currents starting from Boltzmann equation, to clearly identify the relationship of drift-diffusion with the other approaches in the hierarchy. For simplicity, the analysis is limited to the 1D electron case. The 1D current density can be redefined as \( J(x) = -q \int v f(v, x) dv \) in terms of distribution function with an explicit dependence on velocity. This definition of current can be related to the Boltzmann equation by taking the moment over velocity \( v \). If we assume steady-state and parabolic bands, and approximate the collision integral with a term \( (f_{eq} - f) / \tau \), where \( \tau \) is an average relaxation time and \( f_{eq} \) is a local equilibrium distribution
function, the result of the moment is

\[ J(x) = -\frac{q \tau}{m^*} F \int v \frac{\partial f}{\partial v} \partial v + q \tau \frac{d}{dx} \int v^2 f(v, x) \, dv. \]  

(15)

The first integral be equal to \(-n(x)\), while the second integral on the right-hand side cannot be solved without the knowledge of the distribution function, and can be formally rewritten as \(n(x)(v^2)\) by using the definition of the average of \(v^2\). The resulting current equation

\[ J(x) = qn(x) \frac{\tau}{m^*} F + q \tau \frac{d}{dx} \langle n(x)(v^2) \rangle \]  

(16)

provides correct information on hot carrier transport, within the limitations of the model assumptions. In order to obtain finally the standard drift-diffusion equation, the low field mobility \(\mu = -\frac{q \tau}{m^*}\) and the diffusion coefficient \(D = -\mu k_B T_0 / q\) are introduced. The term \(\langle v^2 \rangle\) is replaced by the corresponding average equilibrium value \(k_B T_0 / m^*\) in order to have an explicit expression. This approximation, however, removes all the thermal effects in the model. The resulting drift-diffusion current

\[ J(x) = qn(x) \mu n F + q D n \frac{dn}{dx} \]  

(17)

is therefore only valid in the region of low fields where the velocity is linear, the low field mobility being the slope of the curve. The validity of the drift-diffusion equations is empirically extended by introducing field dependent mobility \(\mu(F)\) and diffusion coefficient \(D(F)\), usually extrapolated from measurement data. The simple derivation above should give a feeling of the limitation of drift-diffusion approaches. One should keep in mind that even with the introduction of field dependent quantities, velocity overshoot is not included in the transport model.

An extension to the overshoot regime can be accomplished in 1D by adding an extra term that depends on the space derivative of the electric field [86–88]:

\[ J_n(x) = qn(x) \left[ \mu_n(F) F + \mu_n(F)L(F) \frac{\partial F}{\partial x} \right] + q D_n(F) \frac{dn}{dx}. \]  

(18)

The velocity overshoot information is contained in the length coefficient \(L(F)\), which can be obtained from Monte Carlo calculations [87] or analytical models [89]. The inclusion of velocity overshoot information without the addition of another equation comes at the cost of strong non-linearity, since the field derivative is proportional to the carrier density. The extension to the multidimensional case is not straightforward [90], leading to cumbersome numerical implementation of the augmented drift-diffusion method for general simulation applications. In certain cases, the equations can be simplified by involving the quasi-Fermi level to express the electric field, as introduced empirically in [91] and formally proven in [90].
With the name of compact approaches, one may define a broad category of analytical or semi-analytical methods that are formulated to provide a simple solution for device behavior, whenever possible in closed analytical form. It is beyond the scope of this review to go into specific details and the discussion here is therefore limited to general observations on how compact approaches relate to the detailed physical approaches presented above. The point of departure of compact approaches is often a spatial integration of a drift-diffusion formulation, that yields terminal current values. The integration can be performed if simplifications are made in the geometry and doping distribution of the model structure. The limitations of the physical model are the same of the drift-diffusion formulation adopted. Therefore, the computational advantage in the transition from a numerical drift-diffusion simulation to an analytical formulation, comes at the cost of loss in resolution of the structure features. In order to capture the details of terminal current characteristics as verified in experiments, compact models are modified through the addition of adjustable terms that must be calibrated appropriately. The limitation of this approach is in the lack of scalability. Since the empirical fitting has to be performed in a statistical sense, but ignoring the underlying non-linear physical transport, new sets of calibration parameters are necessary when dimensions and other variables (doping, bias voltages) are changed. Analytical or integrated formulations are still of enormous value, because fast approaches are essential for efficient circuit design. The main problem, when devices are scaled into the deep-submicron regime, is in the rapidly soaring costs of experimental verification for model calibration. This stresses the importance of detailed physical approaches that are scalable and can provide sufficiently quantitative answers, to supplement or replace experimental verification for the purpose of calibration of simple models. Advanced compact approaches may require a large number of calibration parameters. It is possible to limit the number of these parameters if a physical model which is more accurate than drift-diffusion is used as a starting point. For instance, the augmented drift-diffusion formulation is suitable for such applications, since some of the essential features of hot carrier transport can be introduced with a relatively modest complication of the current equation [92,93].

3. Simulation Issues in Nanoscale Silicon Devices

Although drift-diffusion simulation is based on a rather simplified model, it is still of great utility because the mobility can be altered semi-empirically to fit the behavior of a specific technology. Once the model is set to reproduce acceptably current–voltage terminal characteristics, drift-diffusion satisfies most of the designers’ needs. Speed of computation makes up for any unphysical behavior of detailed
local variables, like potential distribution and carrier density, while a combination of experimental data and physical simulation can be used to calibrate the mobility model.

For exploratory analysis of scaled structures and new devices, physical simulation has an important role when the goal is to compare different structures and to pinpoint problems early in the development of structure design. However, it is important to have a hierarchy of models since no single approach is complete. A full quantum simulation is in principle appealing but there are a number of unresolved issues for quantum transport simulation, besides the mere computational cost. It is still problematic to reconcile the continuum and statistical nature of electron wave functions with the granular nature of transport in a very small device. In principle, a time-dependent particle model provides the best way to capture device behavior, since with the shrinking number of carriers at any one time inside the device channel, individual particle configurations are not representative of the distribution function locally and only a time-average of such configurations can yield a meaningful distribution. Therefore, even a steady-state quantum solution will suffer the same problems of all continuum models when applied to a granular (rarefied) system. Basically, a continuum model represents flow as a fluid of virtual particles of infinitesimal size, weight and charge because in essence an infinite particle averaging is already applied before electrical self-consistency is considered. This is just the nature of a steady-state picture. Averaging of the actual time-dependent self-consistent problem may yield a different results of a purely steady-state model when the system is appreciably granular. From an electrical point of view this can be expressed by saying that the overall average charge screening established in the steady-state problem is not necessarily representative of the local time-dependent electrical and charge behavior. The conclusion is that a steady-state quantum solution for transport may remain qualitative exactly for a range of interesting problems or conditions where the quantum features become important and prevail over semi-classical features. The most desired approach would be a true time-dependent quantum model, but there is still an undeniable ambiguity in representing actual particles with mathematical objects like wave packets. Quantum molecular dynamics methods of chemistry could be useful to represent better the overall physics in a system with marked atomic feature but they remain unbearably expensive an impractical for any device application.

Advanced MOSFET devices have approached a regime where transport consists of an admixture of ballistic, quantum, and semi-classical features. Hydrodynamics and energy transport approaches are becoming clearly deficient [94], because the basic assumptions underlying the models are mostly invalid and at the same time a calibration procedure, as one could apply to drift-diffusion, would perhaps serve little practical purpose. Particle Monte Carlo models, despite a certain computa-
tional cost and their intrinsic noise, are instead ideally suited to capture behavior at the intersection of semi-classical and ballistic transport in a very small structure.

There is a clear gap in the hierarchy between Monte Carlo and quantum models. Semi-classical particle approaches lack features of quantum coherence and available quantum models are deficient in terms of band structure and scattering models. Practical device simulation has focused on the task to augment Monte Carlo simulation by adding quantum features. A typical device consists of a conduction channel between two contacts and it can be viewed as an electronic waveguide where quantum effects can be separated in the cross-section and in the direction of transport. Size quantization in the cross-section can be dealt with by solving the time-independent Schrödinger at many locations along the channel to resolve transverse subbands induced by the confinement. This is the most important effect to include in a transport model of advanced MOSFETs, since size quantization alters the potential energy of particles and tends to push away charge from the oxide interface. One possible implementation is to include the detail of the subbands, assigning a population of electrons to each of them and using the corresponding wave functions for charge assignment [95–97]. Each subband is associated with specific scattering rates, since the shape of the wave function yields different results for scattering calculations. In addition, new scattering rates must be introduced to account for inter-subband transfer mechanisms. The number of subband may be quite large in Si devices, because of the large effective mass and a larger number of scattering tables for intra- and inter-subband mechanisms must be developed. The advantage of a detailed subband model is that the true momentum quantization with discrete values for the transverse momentum component can be accounted for. The major complication is in the fact that the reduced dimensionality region, where size quantization is introduced, must be coupled to regions where transport has essentially a 3D classical behavior. Also, very energetic particles in the channel would also be in quasi-continuum states that cannot be easily described by the quantum solution, creating the need for a special scattering rate between quantized and quasi-continuum energy states.

An alternative is to introduce size quantization with a quantum correction that modifies the electrostatic potential profile to mimic the global behavior of quantum repulsion at the interfaces. Particle are still simulated as semi-classical objects with scattering rates and equations of motion as used for the 3D continuum, but the quantum corrections force the flow of charge to assume a pattern resembling the one followed by the quantum system, although the detail of the subbands and the correct transverse momentum quantization are not explicitly applied to the particle transport. The main advantage of a quantum correction approach is that there is a seamless connection with regions where transport is predominantly classical as in the contacts.
This approach, however, is expected to become inadequate for extremely thin layers or for narrow wire-like channel, when the energy separation between quantized transverse states and the quantization of momentum cannot be explicitly ignored. A quantum correction approach only adds small overhead to the computational cost of a semi-classical Monte Carlo, therefore it is preferable for device applications, since a model with full subband details is much more onerous in terms of implementation and computation being more suitable for the analysis of transport parameters rather than practical device simulation.

Several approaches have been pursued for quantum correction of particle simulation. One of the earliest is based on the effective potential, obtained by integrating over the cross-section coordinates the product of the electrostatic potential $V$ with a Gaussian function associated with a spread $a$ related to the particle size according to [98]

$$V_{\text{eff}}(x) = \int V(x) e^{-\frac{(x-x')^2}{2a^2}} dx'.$$

(19)

This correction is simple to implement and it works best for smooth, symmetric potentials because of the use of a Gaussian weight. The parameter is treated as a fitting length and it is possible to obtain a good result for the average repulsion of charge by a potential interface without sensitivity to the noise of particle simulation since the electrostatic potential is normally smooth. However, the details of the actual charge distribution is not quite accurate, since the density near the interface may be grossly underestimated by this model. Improvements can be obtained by using a weight functions that more closely resembles the asymmetric distribution of wave function for a specific heterojunction structure or by using a more advanced and more computationally demanding formalism which depends explicitly on the wave vector and does not require a fitting parameter [99].

An alternative approach uses the Wigner representation of quantum transport, based on the observation that the Wigner function can be intended as the quantum analogous of the distribution function in the Boltzmann transport equation. Indeed, it is possible to rewrite the quantum transport equation for the Wigner function as the classical Boltzmann equation, augmented by an infinite series representing the quantum contributions. If the series is truncated, resulting transport equation resembles the Boltzmann form with a quantum corrected term replacing the electric field force [100]. This model is formally very elegant and leads to a hierarchy of possible approximation. However, when coupled with Monte Carlo to obtain a size quantization correction in the channel of a MOSFET, there are several difficulties. Since the local carrier density enters the expression for the correction term, convergence is affected by the noise in Monte Carlo simulation. In addition, for the specific case of the
MOS structure, one needs to pick a carrier density at the interface as a fitting parameter to get the correct spatial carrier density, which is otherwise more realistic than in the case of the original effective potential formulation. An similar approach uses the Bohm representation leading to an equation for the effective conduction band edge by assuming a certain functional relationship between density and corrected potential (e.g., an exponential). The results presented in the literature are strictly valid under thermal equilibrium and take into account the effect of carriers only occupying the quantized ground state [101] but the algorithm is of simple implementation and not particularly affected by noise.

Another approach uses directly the solution of the time-independent Schrödinger equation at various channel cross-sections, to formulate the potential correction through a direct analysis of the quantum states corresponding to the transverse electrostatic potential as it evolves during the Monte Carlo simulation [102]. The energy states and the wave functions obtained from the solution of the quantum eigenvalue problem provide a reference distribution of the quantum density, \( n_q \), close to the interface. If a quasi-equilibrium relationship is assumed between the potential and the reference quantum distribution, which is reasonable in the direction transverse to the channel axis, the quantum correction is simply obtained by inverting that relationship. The result provides the appropriate shape of the quantum correction, which depends on \( n_q \) and on the original potential distribution, and which just needs to be shifted by using an appropriate reference potential corresponding to a location in the cross-section where the quantum correction is zero or negligible. The main advantage of this method is that no fitting parameter is necessary and by design the Monte Carlo charge distribution follows very closely the pattern of the corresponding quantum problem that we want to approximate, while it is not sensitive to Monte Carlo noise, since the particle density is not directly involved in the determination of the result. The algorithms requires repeated solutions at a number of channel cross-sections, but by using an efficient eigenvalue solver the overhead in a typical Monte Carlo simulation is fairly small and well worth the gain in accuracy.

Figure 4 shows examples of Monte Carlo simulations for a bulk MOS capacitor structure. The Schrödinger based correction represents the space dependent density very accurately, as expected even is no adjustable parameter needs to be used. Examples for double gate structures are shown in Fig. 5. As a rule of thumb, simulations indicate that a quantum correction approach is reasonable as long as the silicon layer thickness is 3 nm or more. For thinner layers, momentum quantization becomes more important due to the wider spacing between energy eigenvalues and a detailed model including individual energy subbands becomes necessary.

A major problem in scaled MOS devices with nanoscale gate length is the difficulty in controlling leakage current between source and drain contacts when the
device is OFF. In standard planar MOS structures doping profiles may be used to mitigate in part this problem, but the drawback of increased channel doping is a reduction of mobility. Another problem is the need to scale also the oxide thickness to maintain device performance, but once thicknesses below 1.5 nm are used, tunneling gate current increases beyond tolerable levels. One trend is to consider new dielectric materials with a larger dielectric constant (high-κ dielectric), to yield similar interface field with larger thickness at the same voltage. This solution has also problems, however, since experiments show that in general the use of these dielectrics is associated with a drastic drop in mobility, related to remote optical interaction which has little effect in regular silicon dioxide.

Another issue of great importance for device designers is to improve current drive of scaled devices, because this helps in limiting the bias voltage necessary for normal operation. The main current limitation in a traditional MOSFET device is due the carrier–interface interaction that lowers the mobility with respect to bulk material. This is difficult to avoid, since the physics of the MOS system establishes
FIG. 5. Electron concentration in a double-gate MOS capacitor calculated with the Schrödinger-corrected Monte Carlo and self-consistent Schrödinger–Poisson methods over a range of body thicknesses.

the conduction channel at the oxide interface to start with. The velocity overshoot phenomenon introduced earlier, that establishes over short distances a drift velocity which is higher than in corresponding bulk material with the same electric field, is only of limited efficacy in silicon. Considerable interest is now placed on the use of strain applied to the channel region, because strain modifies the band structure by changing the local structure of the crystal and if properly engineered can lead to lower effective mass and higher mobility [103,104].

Advanced device design has focused on highly confined channels to improve source-drain isolation. Typically, a double gate MOS structure is considered, where a narrow slab of silicon has a similar gate on the top and bottom interfaces. While this is a useful design template, a planar double gate device is of difficult realization. Practical structures utilize instead a vertical configuration to realize a configuration that approaches a double gate, for instance the FinFET [105] as shown schematically in Fig. 6. Such devices in general provide good source-drain isolation, but offer somewhat modest current drive. There are a number of possible reasons for this.
A major structural feature of the FinFET family is that instead of an overlap between the gate and the contact regions they present an extension region (or underlap) separating physically gate and contacts. The current drive properties of the channel degrade with increasing extension length, because too much of the source-drain applied voltage drops there rather than in the channel under the gate. Figure 7 shows the results of particle Monte Carlo simulations for FinFETs with increasing extension length from 5 to 20 nm for a 20 nm gate length. The voltage applied to drain and gate is 1.0 V for this example. One can see how the potential become flat in the channel for large extension lengths, with degradation of the velocity that consequently leads to poor current drive.

The use of high-$\kappa$ dielectric materials for the gate raises also some issues in the realization of a narrow conduction channel. When a particle approaches the dielectric gate structure, an image charge is formed which generates a force. In a self-consistent simulation, the Poisson equation is solved and it is difficult to isolate the image force behavior to assess the impact of the dielectric choice. To investigate this issue, a particle steering model was implemented, where a particle is dragged at different channel locations. The 3D Poisson equation is solved for each position, obtaining by integration the potential energy that needs to be overcome to enter the channel. Figure 8
Fig. 7. Band edge potential profile, 1D equivalent line charge electron density, electric field distribution and average velocity along the channel for a series of FinFETs with same gate length of 20 nm and increasing extension lengths, simulated with a particle Monte Carlo model.

shows results for a square channel structure of decreasing diameter for silicon dioxide (permittivity of 3.9) and for a high-κ dielectric (permittivity of 30). One can see that the potential barrier for silicon dioxide increases considerably for narrower channels, reaching several multiples of kT. Conversely, for a high-κ dielectric with permittivity larger than that of silicon (about 11.9) the image force changes sign, resulting in an attractive rather than repulsive field that would tend to draw the particle into the channel. These results apply to an empty channel for a single particle attempting traversal, therefore they are representative of OFF or subthreshold condition and suggest how naturally the channel opposes leakage currents when few particles would enter the channel. Obviously, a high-κ dielectric material does not favor isolation with respect to regular silicon dioxide.
4. Organic Molecular Devices

The replacement of transistors by molecules is appealing for scaling, since a silicon integrated devices require a volume involving from $10^6$ to $10^9$ atoms while a molecular structure may involve up to several hundred atoms. Therefore, there is the potential for enormous scaling gains. A sufficient number of key experiments and theoretical calculations have shown the feasibility of molecular devices in principle, but many challenges are still on the way for practical commercial applications. The enormous progress undergone by computational chemistry in recent years has provided the needed theoretical background for investigation and refinement of possible applications via computational exploration.

The first suggestion of using molecules in place of standard electronic devices was advanced by Aviram and Ratner [106] who studied a molecular complex acting as the equivalent of a $p$–$n$ junction diode. This rectifier structure consisted of a donor system and an acceptor system, separated by a sigma-bonded (methylene) tunneling bridge. Unambiguous experimental confirmation of molecular rectification came only about two decades later [107] but this work planted the seed for the now rapidly expanding and interdisciplinary field of molecular electronics. Rectification by molecular monolayers was shown in [108,109]. The results of Ref. [109] are particularly important, because they demonstrated the use of simple phenyl-based molecular systems. Measurements for conductance of a single benzene-1,4-dithiol molecule between gold contacts was performed by Reed and coworkers [110] using a clever arrangement where the contact tips geometry may allow to contact one single
molecule of a self-assembled layer. Further work on ethylene-phenylene molecular structures led to measurements of very strong negative differential resistance [111]. A number of other structures have been studied after these early successful demonstrations. Several reviews provide information on the recent state-of-the-art experimental developments [112–115].

The standard device paradigm is to consider the conducting properties of a molecular structure embedded between two metallic terminal contacts. Aromatic structures based on Benzene (C₆H₆) have received particular attention, because of their versatility and simplicity [110]. By removing one of the hydrogen atoms (deprotonation), one obtains the phenyl group (C₆H₅), which can be bonded to another molecular component. With the removal of two hydrogen atoms, one obtains a phenylene group (C₆H₄) with two binding sites, so that a chain of similar groups can be formed (polyphenylene) as shown in the schematic diagram of Fig. 9. These chains are usually called “Tour wires,” after J.M. Tour, one of the pioneers in this field. It is also possible to modify the behavior of the molecular chain by inserting other groups, e.g., acetylene spacers.

In order to realize a device, a single molecule or a chain of molecules is terminated by metal contacts. Joining directly a metal surface to a molecule usually provides an imperfect contact characterized by a Schottky barrier. This is not always desirable if one wants to realize ohmic contacts, since the ability to pass a current is greatly reduced by the barrier. A thiol (–SH) group may be added at the end of the chain in order to provide a way to attach the molecular structure to the metallic contacts, typically gold (Au), as illustrated by the diagram in Fig. 10. The thiolate is normally terminated by hydrogen until contact with the gold is established. The insertion of this group provides tight bond to the metal and it facilitates conduction, thus overcoming the Schottky barrier problem. The element which performs the transition between the metal surface and the molecule is colorfully dubbed in the literature as “molecular alligator clip,” in analogy with the metal clips used to connect wires to electrodes in the laboratory. A molecular alligator clip allows the connection, via a covalent bond, of a molecular system with a macroscopic inter-

![Fig. 9. Diagrams of polyphenylene chain of benzene molecules also known as “Tour wire.”](image-url)
face, usually a metal tip/cluster. The thiol on gold does not necessarily form the best electrical contact for devices, because the geometry of the sulfur orbitals does not permit the conjugated $\pi$-orbitals of the organic chain to interact strongly with the gold conduction orbitals, creating a potential energy barrier and additional resistance [112]. However, other metal contact may be affected by surface oxidation or side reaction with the organic molecules, and a practical alternative has not been found, yet.

Working with these building blocks it is possible to realize resistive elements, rectifying diodes, and even resonant-tunneling diodes. As an illustration, Fig. 11 shows schematically the typical behavior of a molecular rectifying structure. Measurements and simulations tend to give qualitatively very similar curves for current and conductance, but in general the theoretical currents [116] are much larger than the measurements [110] for the structure in Fig. 10, even by more than two orders of magnitude. The behavior of the contact resistance is certainly of crucial importance for the explanation of this discrepancy.

Viable simulation approaches for molecular systems have been limited to the analysis of single devices. Molecular structures are not rigid and when one envisions large scale circuits the first issue that comes to mind is whether and to which extent close proximity with other devices may alter the individual electrical response, due to short-range interactions. If it is found that this issue is important, then one may question whether a traditional logic design approach is efficient or if one should rather pursue other approaches that rely on cooperative behavior of neighboring devices for computation. The QCA idea would be a step in this direction.

Fig. 10. Depiction of a benzene molecule connected to gold (Au) contacts via sulfur (S) transition atoms.
5. Simulation of Molecular Conduction

The conduction through molecular structures has been studied experimentally, but the exact mechanisms that control electronic transport are not yet completely understood. A number of approaches have been developed to create a simulation framework that can be used to address fundamental questions. A major difficulty in validation and calibration of simulations for devices based on organic molecular structures is the fact that it is still very difficult to conduct measurements of conduction in single molecules or single molecular strands. If the molecular layers are sandwiched between two metallic contacts, measurements are practical mainly for relatively large self-assembled layers consisting of many molecular elements in parallel. This is quite understandable, due to the very small size of molecules under consideration. Beyond the exciting demonstrations of molecular-based devices that have been realized over the last several years, it is still necessary to develop critical characterization methodologies that may allow direct comparison of simulations and experiments. Both fabrication and metrology issues must be pursued to create reproducible and verifiable molecular electronic transport systems to serve as prototypes for simulation benchmarks.

Another important measurement approach for molecules is Scanning Tunneling Microscopy (STM). In this case, the molecular layer is arranged on top of a conduct-
ing substrate but now the contact with the other end is provided by the STM tip in close proximity. Of course, this set-up is not practical to realize circuits but it has the great advantage of addressability of individual molecules or strands, thus providing a flexible and powerful tool to investigate fundamental behavior of molecules.

Simplified simulation models are very suitable to provide a qualitative picture to interpret experimental results but they need calibration of a number of fitting parameters using measurements in single structures. More advanced simulation models, including \textit{ab initio} ones, provide a more physical “predictive” picture but numerical accuracy could only be possible if absolutely all of the physical details were included in the models, which remains a challenging task. While advanced models are invaluable to guide or even suggest directions for experimental work, one needs to retain in the model some calibration parameters to be tuned with experimental data.

Since measurements for arrays of molecular structures are available, one could extrapolate results for a single molecular conduction path. However, there are still many unknowns. In particular, it is not known to which extent interaction of molecular strands with neighboring ones is important. Calibration of simplified models would need to take into account the coupling between neighbors with additional terms in the model. These terms would need to be formulated with the aid of analysis from more advanced simulation methodologies, which are still too expensive to attempt computationally on a large scale. A detailed comparison between carefully conducted measurements and a hierarchical system of simulation methods of increasing complexity is then necessary to reach a complete understanding. In order to complete the hierarchy, even simpler analytical models would be very important for the experimentalists, to get a description of the behavior of basic physical or circuital phenomena in larger molecular structures.

A variety of intrinsic transport mechanisms across the molecular system could be responsible for the overall conduction process, and fitting analytical models may provide a simple avenue to identify the relative importance of different processes. One should consider, for instance, elementary mechanisms like direct tunneling, Fowler–Nordheim tunneling, thermionic emission and hopping conduction. The approximate relationships between the current density $J$ and voltage $V$ or temperature $T$ are known from standard device theory. Thermionic emission and hopping conduction are sensitive functions of temperature, while tunneling processes are practically insensitive to temperature changes. Direct tunneling and hopping conduction give a current density which is approximately linear with applied voltage, while for thermionic emission one has that the natural logarithm of the current density is approximately inversely proportional to the applied voltage. For Fowler–Nordheim tunneling one has the approximate relationship $\ln(J/V^2) \sim V^{1/2}$. A set of measurements under various voltages and temperatures may show which process is prevalent in different regimes.
A simplified band structure capturing the qualitative behavior of molecular systems is also very important. Typically, a simple two-band model is adopted, which is analogous to the valence band and conduction band diagrams used for semiconductors. Here, one considers the gap between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO). For a chain of molecules, the HOMO and LUMO levels form energy bands. Figure 12 depicts the simple energy level diagram for a single molecule connected to metal contacts with a positive bias applied. When a molecule is directly put in contact with metal layers, there is a jump in potential energy at the interface that may be accounted for by a Schottky barrier between the metal Fermi level $E_F$ and the HOMO or the LUMO levels. In the analysis of tunneling measurements, a simple dispersion relation for the attenuation coefficient of the evanescent wave inside the barrier can be used for this elementary two band model. The overall relation matches together two approximately parabolic dispersion curves starting above the HOMO and below the LUMO. In such a model, Schottky barrier, effective masses and attenuation coefficient become tunable parameters, while the position of the Fermi level must be postulated, usually in the middle of the HOMO–LUMO gap. This simple picture is very practical to compare experiments and to characterize the system response qualitatively. Of course, there
are many uncertainties that one has to be aware of, when trying to apply a more detailed physical picture. The model of the contact, and to some extent the computational approach adopted, influence the actual placement of the Fermi level across the structure with respect to the molecular levels. The amount of charge transfer between the molecular structure and the contact is also an important factor in determining the placement of the Fermi level. The HOMO and LUMO themselves are only references for the band gap which are also influenced by the way contacts are realized, while there are rich details of electronic level structure above the LUMO and below the HOMO, which are simply crudely approximated in an elementary picture. Finally, a simple transport model assumes a linear voltage drop across the molecule, treated as a purely 1D system, which in all likelihood is also a rather crude approximation. A self-consistent approach is necessary to determine accurately the potential distribution.

The most basic issue is the understanding of the resistance of a molecule. Let us consider as a prototype system a phenylene chain terminated by gold contacts. In general, one observes that there is an exponential increase of the resistance with the number of benzene rings in the chain and that there is a sensitive dependence on the relative orientation of the rings, as well as the bonding between them. At the most fundamental level, one may adopt the 1D conduction approach exemplified by the Landauer formula, which provides the resistance as [117]

$$ R = \frac{h}{2e^2} \frac{1}{T(E_f)}. $$

Here, $h$ is Planck’s constant, $e$ is the electronic charge and $T$ is the transmission function between the contacts. As pointed out in the literature [118] this expression does not include inelastic processes that may occur in the molecule. One should expect a tight coupling between electrons and molecular vibrations that may make this approach not accurate. However, in order to understand the basic principles, the simplicity of this model makes it ideal for a first-order evaluation of the phenomena.

The actual location of the Fermi energy $E_f$ plays an important role in the determination of the resistance. The basic energetic picture for molecules considered in practice presents a set of occupied level separated by a gap of several eV from a set of unoccupied levels. The Fermi level in the contact normally lies somewhere close to the center of this gap. A number of approaches could be adopted to evaluate the transmission function $T$. The Green’s function approach is perhaps the most popular, due to its versatility in handling the open boundaries associated with a model assuming semi-infinite contacts. The Green function of the molecular system can be written as

$$ G = (ES - H - \Sigma_1 - \Sigma_2)^{-1}, $$

(21)
where $E$ is the energy, $H$ is the Hamiltonian of the system, $S$ is the overlap matrix, and $\Sigma_1$ and $\Sigma_2$ are the self-energies corresponding to the coupling of the molecules to the gold pads. The matrix $S$ is simply the unity matrix for an orthogonal basis of states. If one denotes with $g_1$ and $g_2$ the Green functions of the isolated gold pads, and $\tau_1$ and $\tau_2$ the coupling matrices for the molecule/gold transition, the self-energies are expressed formally by

$$\Sigma_i = \tau_i^\dagger g_i \tau_i.$$  \hspace{1cm} (22)

After the Green function has been evaluated, the transmission function is given by a trace operation

$$T(E) = \text{Tr}[A_1 M A_2 M^\dagger],$$ \hspace{1cm} (23)

where the matrix $M$ is given by

$$M = \tau_1 G \tau_2^\dagger$$ \hspace{1cm} (24)

and $A_1$ and $A_2$ are the spectral functions of the isolated gold pads. The calculation of these spectral functions for the gold atoms is rather involved, since the $s$, $p$, and $d$ levels contributing to the band structure of gold must be included. However, such evaluation needs to be performed only once and the results can be stored for use in all subsequent computations.

The crucial element in the overall model is the representation of the Hamiltonian of the molecule. This is accomplished by combining a number of steps. First, the structure of the molecule needs to be defined with the aid of a molecular mechanics approach. Available software packages provide the capability to do so, for instance HyperChem (http://www.hyper.com). Once the molecule structure is defined, the goal is to obtain the electronic distribution, which corresponds to solving the Schrödinger equation. A range of approaches is possible to formulate the Hamiltonian of the system, ranging from empirical to $ab\ initio$.

The choice of approximations in empirical approaches is somewhat dictated by the structure of the molecules. Benzene-based molecules have extensive $\pi$-bonding systems, which are not well described by classical valence bond theory. Such molecules are approximately planar and their composition is determined essentially by $\sigma$ and $\pi$ bonding. The $\pi$ bonds are considered to be delocalized over a large region of the molecule and much of the basic chemical analysis has focused on the corresponding $\pi$-electrons to explain molecular behavior. The $\sigma$-electrons together with the nuclei determine a potential distribution which is approximately constant. The Hückel model, introduced since 1931 [119], has been very successful because it has found application for many molecular structures. This model considers only the $\pi$-electrons so that each carbon atom contributes only one electron to the system via
an orbital of appropriate symmetry. One starts from the eigenvalue problem for the molecular orbitals

\[ \hat{H} \psi_i = E_i \psi_i. \]  

(25)

The procedure seeks to expand the orbitals in terms of basis functions \( \chi_i \)

\[ \psi = \sum_{i=1}^{n} c_i \chi_i. \]  

(26)

The expansion coefficients are obtained from an eigenvalue problem

\[ H c_i = E_i S c_i, \]  

(27)

where the matrix elements of \( H \) are

\[ h_{ij} = \int \chi_i(r) \hat{H}(r) \chi_j(r) \, d\tau \]  

(28)

and the elements of the overlap matrix are

\[ s_{ij} = \int \chi_i(r) \chi_j(r) \, d\tau. \]  

(29)

In the Hückel model, the basis functions are atomic orbitals, which are assumed to be normalized and orthogonal, so that the overlap matrix is a unity matrix. In addition, it is assumed that diagonal elements of the Hamiltonian matrix depend only on the nature of the atom at which the corresponding basis function \( \chi_i \) is centered. Off-diagonal elements are zero for non-bonded atoms. For bonded atom pairs, the off-diagonal elements depend only on the types of atoms involved.

The Hückel model has of course limitations. Electron repulsion is not treated in an explicit sense and it is present as some kind of average, in the spirit of Hartree–Fock theory. However, it is evident that in the Hückel \( \pi \)-electron theory the orbital energies are added to give the \( \pi \)-electron energy. However, there is a discrepancy here, since in Hartree–Fock theory orbital energies add differently. A number of modifications have been introduced over the years to add the effect of repulsion between the valence electrons, but a detailed overview is beyond the scope of this work.

A propaedeutic introduction with a number of essential references can be found for instance in [120]. Another issue is that the overlap matrix should not be unitary, since the ordinary atomic orbitals are not orthonormal in the molecular structure. Despite the simplicity of the overall model, the Hückel approach as extended in [121] has the advantage to provide an intuitive framework for the understanding of basic principles in molecular conduction [118]. A detailed comparison analysis presented in [122] gives a good quantitative evaluation of the approximations involved with Hückel-type models within a Green’s function formalism.
Ab initio approaches instead seek to treat the problem from first principles. This implies that given a certain Hamiltonian and a basis function set, all intermediate calculations are carried out rigorously, without the use of empirical adjustments or evaluations derived from experiments. This does not mean that sizable approximations are not made within such theories. Usually, the computational cost of the approach is indeed related to the level of approximation used, for instance, to describe the system Hamiltonian. Existing approaches applied to molecular devices have treated single molecules ab initio, using Density Functional theory (DFT). At the basis of DFT is the idea to describe the system of interacting electrons in terms of its density rather than the many-body wave function. The use of density rather than wave function has profound implication for computational applications. The wave function for N electrons requires $3N$ variables while the density only depends on the space coordinates, no matter what the particle number is. The reduction in computational complexity is enormous making it possible to explore large-scale systems. The main issue is that while an energy density functional must exist, its form is not known a priori and the most common strategy is to formulate an approximation.

The original seminal ideas for DFT deal with the fact that the ground state of the system is defined by the electron density distribution that minimizes the total energy [123]. All the properties of the ground state were shown to be functionals of the ground state electron energy. The corresponding Hamiltonian equation, the Kohn–Sham equation, resembles the time-independent Schrödinger equation but here the potential is also a functional of the electron density. In fact, the Kohn–Sham equation looks like a standard Hartree–Fock equation where the exchange term is replaced by an exchange-correlation potential of unknown form.

The simplest approach is to follow the Local Density Approximation (LDA) where the contribution of each volume element to the total exchange correlation energy is assumed to be the same of an element of homogeneous electron gas with the density at the point under consideration [124–126]. In the Generalized Gradient Approximation (GGA), contributions to each volume element of the total exchange correlation energy are estimated considering the magnitude and gradient of the electron density in the element. The GGA tries to correct the main shortcoming of LDA arising from the fact that in LDA the exchange term does not exhibit a correct asymptotic behavior. The work on GGA has led to a model usually dubbed Beck-3 Perdew–Wang 91 (B3PW91) which is very versatile for computational chemistry applications [127–129]. The functionals provide exchange through the Beck-3 model and correlation through the Perdew–Wang 91 model.

Practical simulation approaches couple the Green’s function methodology with DFT [130,131]. In constructing the model of the molecule, the structure can be augmented by one or two Au atoms at each contact ends, as in [130] or three Au atoms at each end, as in [131], to simulate the alligator clip connection. A detailed DFT
study of the behavior of small Au clusters interfaced with various realizations of alligator clips is presented in [132]. Many groups implement DFT calculations for the structures by using the Gaussian-98 code (http://www.gaussian.com) one of the most established standard tools of computational chemistry. In [130] DFT is also applied to study the Au contacts within a linear combination of atomic orbitals (LCAO) approximation with periodic boundary conditions, using the Crystal-98 suite of codes (http://www.ccwp.ac.uk/ccwp/cg_crystal.html). For an exhaustive theoretical introduction to the self-consistent coupling of Green’s functions and DFT, the readers is referred to Ref. [133], which contains a large number of technical and computational details.

Another approach used in the literature is the density-functional tight-binding (DFTB) method [134]. DFTB is based on a second-order expansion of the Kohn–Sham total energy with respect to charge density fluctuations. The zeroth-order approach would be equivalent to the standard non-self-consistent tight-binding (TB) scheme [135]. The standard TB method expands the Hamiltonian eigenstates into an orthogonalized basis of atomic-like orbitals. A parameterized Hamiltonian matrix represents the many-body Hamiltonian and the matrix elements are fitted to the band structure of a reference system. Results from a standard TB calculation depends strongly on the parameterization scheme and it is problematic to transfer the scheme to various scale systems. Extended TB schemes have been introduced to avoid the standard parameterization while preserving the efficiency of the approach [136–139]. Among these schemes, the DFTB [134,140,141] is based on a two-center approximation for the Hamiltonian matrix elements. The matrix elements are explicitly calculated within a non-orthogonal basis of atomic orbitals. The TB total-energy expression is modified according to a second-order expansion of the Kohn–Sham total energy with respect to charge density fluctuations, ensuring a proper distribution of the charge and overcoming the requirement of local charge neutrality. The DFTB introduced in [134] is particularly useful for the analysis of structural and energetic properties of molecular system, because of the introduction of self-consistency. The electronic charge density is expanded as

\[ n(r) = n^0(r) + \delta n(r). \]  

(30)

Here, \( n(r) \) is a reference density that can be given by the neutral atoms and \( \delta n(r) \) is a density fluctuation. The Kohn energy functional up to second order in the density fluctuations can be written as

\[ E_{\text{tot}}[n] = \sum_i n_i \langle \Psi_i | H^0 | \Psi_i \rangle + E_{\text{rep}}[n^0] + E^{(2)}, \]  

(31)

where the first two terms on the right-hand side are the zeroth-order contributions. The last term is the second-order correction to the Hartree and the exchange-
correlation potential (the first-order corrections vanish)

\[ E^{(2)}[\delta n] = \frac{1}{2} \iint \left[ \frac{1}{|r - r'|} + \frac{\delta^2 E_{xc}}{\delta n(r)\delta n(r')} \right] \delta n(r)\delta n(r') \, dr \, dr'. \quad (32) \]

The modified Hamiltonian for the Kohn–Sham equations is obtained after application of the variational principle to the energy functional

\[ H_{\mu\nu} = H_{\mu\nu}^0 + \frac{1}{2} S_{\mu\nu} \sum_k (\gamma_{ik} + \gamma_{jk}) \Delta q_k \quad \forall \mu \in i, \, \nu \in j, \quad (33) \]

where \( \Delta q_k \) is the net excess charge on atom \( k \), evaluated with Mulliken charge analysis [142], and the matrix elements \( \gamma_{ik} \) contain the evaluation of the second-order correction with normalized charge densities of atoms \( i \) and \( k \). A self-consistent procedure is required since the atomic charges depend on the one-particle wave functions.

The DFTB formalism has been coupled with the non-equilibrium Green function approach to study currents in molecular structures bridging Au contacts, ranging from a DNA fragment [143] to alkeno-thiolate molecules [144]. Specific issues related to the implementation of NEGF in a DFTB code are discussed in [144].

6. Carbon Nanotubes

Carbon nanotubes (CNT) represent a new material form of carbon with a wide range of properties that have numerous potential applications as electronic devices. The CNT configuration is equivalent to a finite two-dimensional graphene sheet rolled into a tube. Mechanical properties of nanotubes are extraordinary by typical material standards. With a Young modulus over 1.0 TPa and estimated tensile strength of about 200 GPa, CNTs are ideal candidates as elements for nanoelectromechanical systems (NEMS) applications. CNTs are also electrically versatile. Depending on their configuration they can be metallic or semiconducting and could be used as building blocks for future nanoelectronic structures. The CNT potential in electronics has been demonstrated in prototype Field Effect Transistor structures [145,146]. In another report [147], hybrid devices have been realized by encapsulating single walled CNTs in semiconductor heterostructures of GaAs/AlAs, with contact realized using (Ga, Mn)As, a ferromagnetic semiconductor, with possible implications for future applications in spintronics. CNTs may find applications in a number of important areas. In electronics, the application that has already the potential to quickly become widespread is the use of CNTs as field emitters in flat-panel display technologies. The CNTs provide minuscule conductive points allowing a
field-emission display to work more efficiently and this technology might pose a serious challenge to liquid-crystal and plasma displays for large flat panels because of the lesser complexity and manufacturing costs.

CNTs were discovered by Ijima at NEC Laboratories in Japan in 1991 [148]. Typical growth processes produce both single-walled and multi-walled CNTs. High pressure processes have been developed which are able to produce only single-walled CNTs in batches of several gram [149]. We will consider here the specific case of single-walled CNT, which are more interesting for device applications. Typically, these nanotubes are just a few nanometers in diameter and may be grown in lengths of several microns. The structures are classified according to the way tubes are rolled from a graphene sheet. A complete tutorial on the subject is beyond the scope of this work and only essential information will be given here. The reader is referred to a classic text for an introduction [150]. For the following discussion it is sufficient to mention here that the CNTs are classified in terms of chirality which indicates at which angle the graphene sheet has been rolled in terms of lattice vector, as shown in Fig. 13.

The lattice of a graphene sheet has a honeycomb matrix of hexagons. Indicating with $a_1$ and $a_2$ the two lattice vectors, one can imagine to roll a lattice point with coordinates given by a vector $\mathbf{v} = (n_1 a_1, n_2 a_2)$ to coincide with the origin, at the chiral angle specified by the vectors $a_1$ and $\mathbf{v}$. The integers of the pair $(n_1, n_2)$ are called chiral indices and they completely specify the CNT, since they also determine

![Fig. 13. Diagram for the folding of a graphene sheet into carbon nanotubes. The illustrated example shows the vector configuration for a (6, 5) nanotube.](image)
Due to symmetry of the graphite lattice, the range of the chiral angle for all possible CNTs is between $0^\circ$ and $30^\circ$. For the extreme values of the chiral angle, the open edges of the tube assume very specific shapes which are used to label the structure, so that for $0^\circ$ chiral angle we have a zigzag CNT and for $30^\circ$, an armchair CNT (Fig. 14). Very important for electronic device applications is the fact that the band structure of the CNT depends on the chirality. All armchair tubes are metallic with a zero band gap. Tubes with $n_1 - n_2 = 3n$ (with $n$ any positive integer) have a small band gap and quasi-metallic behavior, while with $n_1 - n_2 \neq 3n$ they have a finite band gap with semiconducting behavior. Experimentally it is quite difficult to distinguish between metallic and quasi-metallic behavior, particularly at finite temperatures and in the presence of contact resistance; therefore, the practical distinction is normally just between metallic and semiconducting CNTs.

A CNT is a system that has many similarities with traditional crystal materials, because of the periodic lattice structure but at the same time it also can be though of as a molecular chain. A remarkable feature with respect to traditional solid-state structures is the fact that CNTs, like ordinary molecules, deform under external forces and with that electrical properties may change; in particular, the band gap may vary. The effect is observed by examining the conductance in tubes deformed, for instance, by the tip of an atomic force microscope (AFM) [151].

Simulation of small CNT devices may be conducted with methodologies in some respects similar to the ones described earlier for organic molecular devices. A quantum simulation approach requires a similar determination of the system Hamiltonian.
and a detailed understanding of the metal–CNT transition in the contacts. However, simulation of relatively long CNT structures may be carried out using approach from other levels of the simulation hierarchy.

Practical electronic devices to-date have been realized with relatively long CNTs. It is appealing to develop for such structures a semi-classical theory, where electron transport is described by phonon scattering mechanisms and interaction with the electrostatic field. Such theories, suitable for the development of a device simulation framework, have been developed only recently. A complete model can be found, for instance in [152].

The interest for device realization has usually focused on semiconducting nanotubes, where the room temperature band gap is significantly larger than the thermal energy at room temperature. In [152], for instance, zig-zag CNTs with index \( n_1 = 10 \) to 59 are considered. One should keep in mind that the CNT band gap decreases with increasing diameter, since it is inversely proportional to the index \( n_1 \). For the range of \( n \) given above, the band gap goes from about 1.1 eV (diameter 0.8 nm) to about 0.1 eV (diameter 4.6 nm) [152].

The main difference when one transitions from a graphene sheet to a narrow CNT, is that electron and phonon eigenvectors are significantly confined in the cross-section with discrete wave vector values. Electrons are essentially delocalized on the cross-section and one may use only the electric field along the axis to describe particle motion with a 1D model. The band structure relevant for transport presents a number of subbands, which are the result of folding of graphene antibonding \( \pi \)-orbitals. An example for a zig-zag \((10, 0)\) CNT is shown in Fig. 15. Phonon scattering may cause intrasubband, intrasubband-intravalley and intrasubband-intervalley transitions, following conservation of the crystal momentum and total energy selection rules.

Semi-classical particle simulation can be performed with a Monte Carlo method, where classical mechanics is applied to accelerate particles along the CNT axis under the influence of the field. Quantum mechanics determines the energy band structure and the phonon scattering rates. A complete simulation accounting for complete details of subbands, phonon scattering, and charge–charge interaction, remains a very challenging task due the complexity of the model and computational cost, even if 1D semi-classical transport is considered. The approach in [152], which represents to-date the most complete semi-classical set of calculations for transport in zig-zag CNTs, still contains for practical reasons a number of simplifying assumptions. In this work, only longitudinal acoustic phonon scattering is considered, while charge–charge interaction is not included. This allows one to adopt a computationally efficient single-electron Monte Carlo method, where independent particles are followed in the structures under a homogeneous field. Since transport is limited to intrasubband scattering, the absence of electron–electron interaction should not be a
concern if the calculations have the objective to investigate the electron drift velocity at relatively small fields. Intrasubband scattering due to electron–electron interactions in one dimension corresponds to indistinguishable initial and final interacting electron pairs that should not affect the drift velocity. The likelihood of intersubband-intervalley electron–electron mechanisms increases for larger diameter tubes, since band gap and band separation decrease in energy.

The peculiar difficulty of 1D Monte Carlo simulation is in the fact that typically scattering rates exhibit large peaks at the energy threshold for transition. One needs to take care of adjusting the time steps for simulation of the electron drift to be much smaller than the inverse of the scattering rate. It is also well known that in a 1D system the semi-classical density of states diverges at the subband levels, with obvious complications in the computation of scattering rates. The experience accumulated with 1D transport in quantum wires provides a good template to overcome these difficulties by including quantum mechanical broadening within the Fermi golden rule formalism used to determine the scattering rates [153,154].

The most interesting result presented in [152] from the point of view of device realization is the fact that drift velocity results present the negative differential mobility behavior typical of transferred-electron effect from one subband to another. This phenomenon is more likely to take place in practice for the smaller diameter tubes, where the lowest subbands are well separated, so that their energy gap is large.

FIG. 15. Example of band structure for a zig-zag (10, 0) carbon nanotube.
compared to the thermal energy. For larger diameter wires, this is not necessarily the case.

The semi-classical transport calculations in [152] indicate that in a perfect semi-conducting zig-zag CNT drift velocities as high as $5.0 \times 10^7$ cm/s may be reached at large homogeneous fields in steady-state. These are high velocity values by typical solid-state device standards suggesting the capability to achieve fast switching devices. Figure 2 introduces earlier a comparison with velocity-field curves for common semiconductor bulk materials, like Si and GaAs. The observed theoretical negative differential mobility is also suggestive of microwave applications, as for the classical Gunn diode. As the technological capabilities are developed to grow, select and manipulate readily CNTs of specific chirality, such semi-classical level simulations are perhaps the most effective device level tools for predictive studies, since they offer a good compromise between model complexity and computational cost, which is essential for the exploration of a vast parameter space.

Because of their small size, CNTs are potentially interesting for high-frequency device applications. However, theoretical work in this direction is still in need of fundamental development [155]. To make progress in this field one needs not only to understand transport in traditional device simulation terms, but also circuit properties of CNT systems. Both the a.c. impedance of these 1D systems as passive elements and the gain properties as active transistor elements need to be characterized. To assess the maximum frequency at which current gain could be achieved, one may use the general formula

$$f_T = \frac{g_m}{2\pi C_{gs}},$$

(34)

where $g_m$ is the transconductance and $C_{gs}$ is the gate-source capacitance. Estimates of the transconductance range from experimental values of 20 µS [156] to theoretical predictions of 60 µS [157]. With capacitances that could be on the order of 50 aF/µm, predictions suggest nanotube transistors with cutoff frequencies in the THz range for nanoscale gate length, exceeding the capabilities of standard solid-state semiconductor devices [155].

An obvious extension of this semi-classical methodology is the study of field effect transistors. Long CNT based transistors as demonstrated in the literature [52,53] do not lend themselves to practical particle simulation, because of obvious computational cost limitations. On the other hand, fabrication of nanoscale length CNT transistors is still problematic. Theoretical exploration allows one to examine alternative structures to probe other possible scenarios. For instance, studies have been conducted to understand the performance of a conventional MOSFET with nanoscale gate length when CNTs are placed in the device channel.

In a recent approach [158] a MOSFET with 0.1 µm gate length was considered and compared with a similar structure where zig-zag CNTs were placed between
source and drain to form the channel, analogous to the experimental devices reported in [146]. The case with \( n = 10 \) (0.8 nm diameter) and \( n = 59 \) (4.6 nm diameter) were studied to compare I–V curve with the conventional MOSFET structure. Using the same Monte Carlo approach as in [158] to determine the transport properties of the CNTs, a modified drift-diffusion simulator was then used to study the overall structure. The drift-diffusion equations were revised to handle the CNT-silicon heterostructure. Besides the Poisson equation and the carrier continuity equations, this model contains also the Schrödinger equation, to evaluate quantum confinement effects, and the lattice heat equation. The effect of quantum confinement introduces a spatial dependence on the intrinsic carrier concentration (similar to an effective band gap broadening). The current equation is modified to reflect the effects of quantum correction. The model simulations in [158] indicate that the CNT MOSFETs would exhibit improved current drive, when compared with a conventional MOSFET. Interestingly, while the narrower CNT is associated with a lower peak drift velocity as evaluated with Monte Carlo simulation (about \( 3 \times 10^8 \) cm/s, as opposed to \( 5 \times 10^8 \) cm/s for the wider CNT) the corresponding CNT MOSFET shows higher currents. This probably reflects a more favorable band gap difference with respect to the silicon material. Subthreshold characteristics are also favorable with a lower turn-on voltage, which is attractive for low power applications. This good isolation behavior between source-drain is maintained both at low and high drain voltage. Despite the outstanding technical challenges standing on the way of practical realization, these calculations provide again a good reference point to direct further experimental efforts.

The attention of quantum level approaches has been directed mainly towards small CNT structures that can be practically simulated. For a short CNT, the assumption of ballistic transport is usually made. Narrow CNTs are truly 1D systems where phonon scattering may only result in final states with forward or backward directed velocity, and in a relatively short tube it should not be an influential process. As in organic molecular structures, a complete quantum approach should involve an \textit{ab initio} treatment of the system Hamiltonian including the metallic contacts. However, due to the large number of carbon atoms even in short samples, this is not a feasible approach, yet. An entire device description can only be accomplished with simple semi-empirical orbital models as a suitable basis for the transport model [159].

In the outer shell of the carbon atom one finds four orbitals. The \( s \), \( p_x \) and \( p_y \) orbitals are usually not relevant for transport, since they originate bands sufficiently above or below the Fermi level. The bands linked to the remaining \( p_z \) orbital are largely uncoupled to the other ones and are of interest for transport. Therefore a minimal model may use a real space basis with one \( p_z \) orbital for each carbon atom. As a consequence, the number \( N \) of carbon atoms in the system determines the size of the Hamiltonian matrix describing the nanotube. A typical system will have several
thousand atoms, which still is challenge for band structure calculations. A viable approach is to use the tight-binding approximation for interaction between atoms and truncate the coupling to the nearest neighbor only, although this may be a crude approximation in certain cases. Assuming an appropriate coupling parameter for $p_z$ orbitals (e.g., 2.5 to 3 eV) the $N \times N$ matrix describing the Hamiltonian, $H$, is assembled taking into consideration the structure configuration that depends on the chirality. A banded block tridiagonal matrix is obtained. The quantum model for the CNT is completed by specifying the self-energy matrices, $\Sigma_S$ and $\Sigma_D$, for the source and drain end contacts. These matrices simply specify open boundary conditions for the Schrödinger equation. The size of the self-energy matrix is the same as for the Hamiltonian matrix, but the only non-zero elements are in one $n \times n$ block corresponding to the $n$ atoms in communication with the contacts.

Ballistic transport in the CNT ideally connected to open contacts is then completely described by the retarded Green's function

$$G^r = \left[ \left( E + i\delta \right) I - H - \Sigma_S - \Sigma_D \right]^{-1}. \quad (35)$$

While direct inversion of this matrix is in general a very expensive proposition, the block tridiagonal structure allows one to use cheaper recursive algorithms [160]. It is a great computational advantage to assume ballistic transport, because one only needs the columns of the Green's function matrix in correspondence of the $n$ atoms in communication with the open contacts on each side. The local density of states are recovered from the knowledge of the Green's function, as well as the charge density [160], with the established NEGF techniques. The Poisson equation can be coupled to the system for a self-consistent treatment, iterating with the NEGF equations under specified potential boundary conditions that determine the relative positions of the Fermi levels in the contacts.

The source-drain transmission $T(E)$ is evaluated from the Green’s function with a trace operation, similarly to the procedure illustrated for molecular systems. The current flowing in the system is then simply estimated from

$$I = \frac{4e}{h} \int T(E) \left[ f_S(E) - f_D(E) \right] dE, \quad (36)$$

where $f_S$ and $f_D$ are the energy distributions assumed in the contacts.

While the description of ballistic transport via the Hamiltonian in a regular CNT structure is well resolved by the theory, a proper simulation of the contacts is still an open question. In the prototypical CNT transistors, metal contacts for source and drain may be directly connected to the carbon structure and when the gate voltage is varied, the Schottky barrier at the metal–carbon interfaces should be modulated by it. A full quantum mechanical treatment would be necessary to treat the physical contact, but present methodologies usually adopt a phenomenological approach.
where the simple open boundary self-energies introduced above are augmented by terms that must contain the height of the Schottky barrier and the density of metal induced gap states. A description of this phenomenological approach applied to the study of CNT FETs with Schottky barrier contacts can be found in [159]. Since the simple semi-empirical atomistic model for the CNT is adequate to resolve transport involving thousands of carbon atoms, a more advanced model of the contacts may involve an *ab initio* treatment only for a small portion of the CNT ends coupled with a number of metal atoms. A continuum resistive model may be sufficient to represent the remaining portions of the metal contacts. The overall procedure is multi-scale [161] and it remains a great computational challenge.

Although simplified, the atomistic approach outlined above remains very expensive, due to the large matrix size which is on the order of the number of atoms in the system. The so-called mode space approach, demonstrated for instance in nanoscale MOSFET simulation [162] is a candidate for reducing the size of the Hamiltonian matrix. Because of the symmetry in the radial direction, one may expand transport in terms of transverse modes determined by periodicity on the circular cross-section. Transverse modes are uncoupled as long as the CNT structure is uniform, so one may perform a transformation of the problem into as many 1D mode space lattices as the number of transverse modes. The computational advantage is even greater if one considers that matrices associated to the 1D problems are tridiagonal, allowing one to apply very efficient algorithms [160]. Further reductions in computational cost come from considering only the modes which are relevant for transport, as also done in electron wave guide applications of Green’s functions [163].

The unique structure of the CNT allows one also to conceive a coaxial structure for the gating. One can easily see that for zig-zag CNT the transverse modes remain plane waves satisfying the periodic boundary conditions, so that there is an exact mapping between the complete real space approach and the mode space approach. Readers interested in the technical details for the implementation of NEGF approaches are invited to consult the work in [160], which provides a very complete account of all the algorithmic and model development steps involved, in the more general context of nanotransistor simulation.

The nature of the contacts on a CNT are of great importance for device operation and detailed simulation. A number of specialized papers in the literature discuss specific issue. The conductance through CNTs with disorder is studied with a Green’s function approach in [164] and similarly the transmission through CNTs terminated by polyhedral caps is studied in [165]. The intrinsic current carrying capacity of CNTs coupled to ideal contact in ballistic conditions is analyzed in [166].

The role of contacts in transistor operation is analyzed in [167]. This work shows that fabricated CNT devices operate as Schottky barrier transistors, arguing that contact resistance modulation by the gate is primarily responsible for transistor action.
Detailed models of CNT coupling to metallic contacts are investigated in [168]. This work shows theoretically that conductance scales with contact length, as observed also in experiments and it also shows that chirality influences the threshold value of the Fermi wave vector below which coupling is not significant. In [168] the coupling of Au/Ag contacts to armchair and to zig-zag CNTs is also discussed and a further study [169] concludes that wave function details explain how metallic zig-zag CNTs are superior to armchair ones as nanowires, by modeling the metal–CNT interface.

To fulfill their promise as elements for integrated circuits, CNT devices need to be scaled down to nanometer scale if they are to compete with silicon devices. Computer simulation is an ideal tool to explore possibilities in this direction. The analysis of ballistic CNT transistors in [157] indicates an intrinsic better performance of the coaxially gated nanotube device with respect to MOSFETs. However, one needs consider other device regimes, particularly the subthreshold characteristics that affect device turn-off and leakage current in the off-state. A number of theoretical investigations have shown that a CNT FET has different behavior, when compared to conventional transistors [143,170]. A recent analysis has introduced also the complex band structure effects, including the wave vector evanescence in the conduction band gap [171]. This is an important issue, because one needs to characterize the tunneling between source and drain for a nanoscale device, in order to assess leakage and subthreshold slope. The work in [171] considers only the first subband which is the most influential in subthreshold regime. Transport is studied using the NEGF approach for coaxially gated CNT structures, considering 10 nm channel zig-zag CNT with indices (17, 0) and (10, 0) and assuming a contact obtained with heavily doped CNT ends embedded in the metal contacts, with typical bias of 100 meV. This work is interesting because it performs a comparison between quantum calculations and a ballistic classical model. In the on-state, quantum and classical simulations yield comparable results, but when approaching the off-state the quantum calculation with complex band included shows higher subthreshold currents as one should expect with tunneling. Since the (10, 0) CNT has a wider band gap, the quantum model shows better leakage and subthreshold swing behavior than in the (17, 0) CNT where the band gap is smaller.

CNTs are extremely promising for applications in nanoelectromechanical systems (NEMS) because of their extraordinary mechanical and electrical properties and for the potential to be utilized as sensor elements. An exhaustive survey is beyond the scope of this work but several interesting research directions involving large scale simulation will be introduced here. There is also a growing literature on experimental applications and only some examples will be cited to give the reader an idea of the field, knowing full well that many equally exciting results will be neglected. For instance, a strategy for the controlled growth of long single walled CNTs for reliable integration into electronic and electromechanical devices is presented in [172].
The method develops patterned CNT growth to bridge predefined molybdenum electrodes, and allows the realization of suspended nanotubes which are electrically conducting, hence ideally suited for nanoscale electromechanical applications. An important result is the discovery that molybdenum is the only compatible metal for the growth in the high temperature CVD system used in [172]. Another recent report [173] describes a process where self-aligned mechanical attachment of CNTs is realized to silicon dioxide structures by a CVD process. This is very appealing for the possible integration of nanomechanical CNT devices in standard silicon processing.

Functionalization is an approach where incorporation of molecules in the wall or mouth of the CNT can be used to tune the electrical properties. In this way one may develop extremely sensitive nanosensing elements. Both inorganic molecules like NH\textsubscript{3} and NO\textsubscript{2} [174,175], and organic molecules like amines [176] have been used in experiments to demonstrate the extreme sensitivity of CNTs to molecular species. The goal is to tune charge carriers in single-walled CNTs simply by adsorption of molecules.

For the development of sensors, particular in biological and environmental applications, it is also relevant to consider the CNT system in a wet environment. A very interesting demonstration of CNT transistors gated by an electrolyte solution is presented in [156]. Such devices show very high mobility and transconductance and are attractive for electronic applications as well as chemical and biological sensors. From the point of view of wet sensing it is also interesting to determine how water and ions in the electrolyte may transport through the CNT pore. This is very close to the topic of biological ionic channels, discussed in the following section. CNTs could be used to imitate the behavior of such biological systems by appropriate functionalization.

The understanding of CNT functionalization requires various levels of simulation approaches. Classical molecular dynamics has been used to investigate the mechanical properties and deformation of CNTs functionalized by covalent chemical attachments [177,178] finding that the maximum compressive (buckling) force is reduced by 15% regardless of CNT radius or chirality. Molecular dynamics was also used to study the diffusive flow of gas molecules in CNT [178,179]. A comprehensive review of CNT functionalization may be found in [180].

A computational study of electronic transport properties in functionalized CNT is presented in [181]. Here, a Green’s function/density-functional/tight-binding approach is used, similar to the method already adopted by the same group for molecular devices, as reviewed earlier. This study considers single-walled CNTs reacting with C\textsubscript{6}H\textsubscript{4} (benzyne) molecules. The CNT domain is divided into three sections: a central finite region on which benzyne molecules are adsorbed, connected to two regions on either side assumed to be semi-infinite contacts. The computational method monitors the ballistic current flowing in the contact, during a molecular dynamics simulation of the central region reacting with one or more benzyne molecules. The
steady state current is strongly reduced when molecules are adsorbed compared to the unperturbed CNT system. An interesting finding is that the current reduction depends strongly on the orientation of the molecules with respect to the CNT, which affects the process of phase coherence breaking for the transmitted electronic wave when interaction occurs.

Molecular dynamics is an approach suitable to study the mechanical behavior of NEMS, but it may be rather expensive, as is density functional theory for the study of the system Hamiltonian. An overview on large-scale computational nanotechnology applications with CNT can be found in [182,183]. An issue relevant for NEMS is the variation of electrical properties when the CNT is deformed. In [182], the band structure is analyzed using a first-principle total energy pseudo-potential calculation for flattened single-walled CNTs, using DFT within the LDA approximation. This interesting work predicts a novel semiconductor–metal–semiconductor transition increasing the degree of flattening deformation, on a (8, 0) CNT. These results extend the earlier findings in [184], where a semiconductor–metal transition was found from first-principle calculations on a (10, 0) CNT.

For more practical design level applications, continuum theories can be used to combine electrical and mechanical analysis of switching elements like cantilevers realized with a CNT. In [185] an electromechanical analysis of a CNT cantilever over a ground electrode is performed by accounting for three coupled energy domains: elastostatic energy, electrostatic energy and van der Waals energy. When a potential difference is created between the cantilever and the ground electrode, charge distributions are generated in both conductors. An electrostatic pressure deflects the cantilever but in addition one also needs to account for van der Waals forces at the nanoscale to compute the equilibrium deflection. Self-consistency is reached when elastic, electrostatic and van der Waals forces are balanced. In the absence of extensive experimental data for validation, the analysis in [185] provides comparisons with molecular dynamics simulations. Since multiphysics phenomena need to be taken into account for NEMS design, a continuum approach is very desirable for computational efficiency, and it may also be envisioned as one level in multiscale simulation approaches, where a more accurate and computer intensive method like molecular dynamics may be applied only to a limited region of the domain under study.

For sensor applications, one also needs understanding of the behavior of CNTs in wet environments. The presence of water makes simulation of such systems enormously complicated. The main issues will be discussed in the following section, in the context of ionic channels. From the point of view of CNTs, the fundamental question regards the actual ability of water to enter the nanotube and how functionalization affects the transport of ions through the tube. Using molecular dynamics in various CNT structures, molecular dynamics simulation in several studies [186–188]...
shows that confinement can induce unusual behavior in the properties of water under normal ambient conditions of pressure and temperature. Water molecules appear to be ordered because of the interplay between narrow confinement and charge distribution on the water molecule. Water generally shows dynamics that slows down with respect to bulk inside the CNT, and for some critical diameter it forms a rigid structure resembling ice [187,188]. These structures are very interesting to support proton conduction, since proton should readily transport in ordered water by hopping conduction [186,188]. Other studies implement a practical tight-binding approach to rapidly determine polarization charges on the wall of finite length CNT, in the presence of polar water molecules or ions inside the pore [189] and couple this model to molecular dynamics simulation [190] for a self-consistent treatment of polarization effect. With present computational resources this type of calculation is possible because the regular nature of the CNT structure simplifies the calculations, while a similar approach is still prohibitive for more complex natural structures. Molecular dynamics calculations also show that an ion inside a finite length CNT may oscillate inside the pore at frequencies on the order of the THz [191], which suggests intriguing possibilities for new types of detectors or electromagnetic sources composed of arrays of CNTs in a membrane.

### 7. Ionic Channels

The fundamental principles of biological activities in all living beings are based on electrolyte fluid flow in vessels and channels that involve a wide spectrum of length and time scales. Biological organisms function as complex systems hinging on a miraculous feedback balance between these fluid flows, which are seamlessly coupled going from the macroscopic size of organs and limbs down to the nanoscale size of cellular membrane channels.

From an engineering point of view, the organization of organisms is set up to accomplish a large variety of computing, sensing and actuating operations, where the regulation is accomplished by charged fluid flow signaling, in a way that resembles the connectivity of digital computers regulated by electrical currents. However, the solid-state counterpart is based on a much simpler binary logic and on circuits constructed with a relatively small palette of elementary electronic devices as building blocks, accomplishing basic switching, memory, and connectivity functions.

The *biological machine*, in contrast, presents a much richer variety of elementary components and behaviors that could be roughly described as a mixture of multi-logic analog/digital circuitry. In addition, the biological systems are self-assembling, self-healing and fault-tolerant. For instance, the computational ability of the brain is
based on a still slightly understood massively parallel architecture, which can perform certain tasks like pattern recognition extremely fast even if the inherent clock speed is very slow compared to what can be achieved today with electronic circuits.

The desire to simulate in detail biological systems may have many motivations. Detailed computational models that provide a deeper understanding of biological functions at different scales have the potential to impact the ability to treat diseases by suggesting new targeted drug synthesis or therapeutic approaches. In addition, unlocking even the simplest mechanisms of biological function may provide the ability to realize synthetic structures with biomimetic properties suitable for novel sensor devices, for prosthetic applications, and for interfacing organisms with solid-state electronics systems.

The interest in biomimetic devices is growing steadily, motivating diverse communities of scientists that range from electrical engineering to chemistry. Ionic channels provide a compelling biological template for the conception of biomimetic structures, because of the wide range of device-like functions that they exhibit. Ionic channels are a class of proteins found in the membranes of all biological cells. Each channel consists of a chain of amino acids folded in such a way that the protein forms a nanoscopic water-filled tunnel, controlling ion transport through the otherwise impermeable membrane. An essential feature of proteins is that the side chains of the amino acids residues are ionizable via the addition or subtraction of protons. The ionization state of a given residue depends on the pH and salt concentration of the solution in which the protein is immersed. Thus, every ion channel carries a strong and steeply varying distribution of permanent charge, which depends on the particular combination of channel type and prevalent physiological conditions. The charge distribution residing on the channel plays a critical role in determining the permeation characteristics of the open channel. The conduction mechanisms in channels are key to understanding the detection mechanisms at the basis of the complex feedback and control systems in living organisms, optimized by millions of years of evolution. A schematic depiction of ionic channel is shown in Fig. 16.

From a physiological standpoint, ionic channels regulate the transport of ions in and out of the cell, and in and out of compartments inside cells like mitochondria and nuclei, thereby maintaining the correct internal ion composition that is crucial to cell survival and function. Many channels have the ability to selectively transmit or block a particular ion species and most exhibit switching properties similar to electronic devices. Malfunctioning channels cause or are associated with many diseases, and a large number of drugs act directly or indirectly on channels [192].

From a device point of view, ion channels can be thought of as elements of an electrical circuit, behaving like resistors, diodes or batteries, depending on the specific channel and environmental factors. Some channels perform specialized functions re-
Fig. 16. Schematic depiction of an ionic channel regulating ion flow across a membrane.

Evolution is certainly responsible for the variety of shapes, compositions and function of ionic channels, which can be classified as a class of membrane glycoproteins. While the presence of pores in cellular membranes was advanced since the 1800s, to explain phenomena like osmosis, only over the last twenty years ionic channels have been unequivocally identified. The landmark voltage clamp work of Hodgkin and Huxley on squid giant axons [198] showed for the first time the role of ionic selectivity in permeation of nerve cells, using as an explanation the presence of voltage-gated Na and K channels. Progressively, ionic channels have been recognized as key players in the regulation of electrophysiological responses, often organized in systems of great complexity. In cardiac cells, for instance, it is believed at present that there are over twenty pathways for current conduction [199].

There are a number of experimental methodologies employed to study various aspects of ionic channels with the aid of the tools from different biological disciplines. Sequencing, cloning and mutagenesis are the typical methods of molecular biology, while a range of microscopy and imaging techniques are employed in structural biology to determine the crystallographic structure of the protein channel. In electrophysiology some of the most exciting experiments are performed, where the behavior of the cell or of the isolated channels are measured in conditions of voltage/current clamp or solution concentration jump. From the point of view of constructing a simulation model, the most important information is provided by the crystallographic structure, which allows one to build the topology of the actual channel. For specific channel types, single channel conductance measurements under voltage clamp or
concentration jump are the most valuable to test and calibrate theoretical models and simulations.

Several known properties of ion channels are important and influence their wide range of biological functions. Selectivity is the ability of ion channels to discriminate what type of ion can permeate the pore. The size or the sign of the charge may be factors in determining the selection. Ion channels also are associated to conductivity. Since an ionic channel consists of a nanoscale constriction on a membrane, it is not easy for charged ions to go through the opening, and in many cases permeation is a rare lucky event out of many unsuccessful attempts. The average rate of ion traversal obviously influences the conductivity. In addition, if an ion is already in the channel, it will tend to repel carriers of the same species that might be able to go through. Therefore, how rapidly an ion can permeate the channel is also a factor in determining the conductivity. The ion channel may possess also a sensitivity to external stimuli that causes a modulation of the conductivity. Factors affecting the sensitivity may be the composition of the electrolyte environment, the transmembrane voltage, the surface tension of the membrane, or chemical interaction between the ions and the pore walls that could, for instance, lead to binding of an ion at a certain site.

Ion channels exhibit a wide range of biological functions that potentially have enormous significance for the design of artificial biomimetic structures. First of all, channels accomplish signaling and computation tasks. Channels are the current carrying and regulatory molecules of the nervous system, with tasks including sensory input, cognition, coordination of motion, and other general functions that control the well being of organisms. Obviously, however small, these structure cannot compete with the computational speed of solid-state devices. Large throughput is achieved by neuronal organization which may perform certain specialized tasks very efficiently, greatly surpassing the ability of computers, as in pattern recognition and coordination of complex motions.

Ionic channels may work as triggers for cellular events. Many cellular processes are triggered by electrical signals generated at the cell surface by ion channels, from muscle contraction to glandular secretion. In specialized organisms, ion channels may be organized to generate electrical power. For instance, it is well know that some species of fish have electric organs that produce large pulses of electrical energy used for hunting. The generation mechanisms uses ion channels in membranes that are stacked in series with each other, similar to elements in a battery. Using selectivity, ion channels act as molecular batteries that transduce the chemical potential of the ionic species for which they are selective into an electromotive force that can drive electrical current. Energy transduction can also be accomplished by proton channels, transducing energy from light into a form that can be used for nutrient synthesis in plants. Proton channels are equally important in the exchange of energy between
ATP and other forms of useful chemical energy in the cell, with processes which are nearly 100% efficient.

The structure itself of the membranes permeated by channel pores is ideal for fluid pumping and filtration tasks. The universal mechanism for moving fluid and electrolytes across boundaries in organisms is via osmotic pressure created by regulated ion movement through channels. For example, channels regulate the flow of water across the airway epithelium to keep it moist, and across the intestinal epithelium to regulate water content in the bowels. Another example of filtration is the function of the kidney. Chemical sensing is another important feature, which stems from the ability of channel proteins to bind with certain chemical species at particular sites. As a reaction to the binding process, the channel conduction may be gated. Chemically sensitive ion channels in the tongue giving the sensation of taste are a good example. Finally, certain ion channels are responsible for mechanosensitivity in cells. It is known that some bacterial channels open and close in response to surface tension changes in the membrane in which they are embedded. In other cases, the opening and closing of ion channels which sense mechanical deformation are mediated by changes in cytoskeleton tension, rather than directly in the membrane.

Although we have only started to understand and catalog the wide range of possible device-like functions, ionic channels have already generated a great deal of interest in the engineering community. On the one hand, it is appealing to exploit the functionality of naturally occurring channels in traditional devices and circuits, e.g., for extreme miniaturization of sensors and to achieve single molecule detection. One approach is to harvest channels from biological samples and synthesize artificial membranes incorporating them. Hybrid structures could then be designed implanting these membranes of biological origin onto solid-state substrates, where standard circuitry can be easily integrated, to accomplish specialized functions.

On the other hand, a clear understanding of channel operation may provide a template for the design of functional elements based on synthetic molecular nanotube systems. The use of molecular elements for nanoscale integration would have the distinct advantage of perfect structure duplication and self-assembly, while the solid-state device counterparts tend to be strongly affected by statistical fluctuations and defects at the nanoscale. The methods of molecular genetics and biology often allow the control of channel proteins with atomic resolution. By replacing or deleting one or more of the amino acids, many channels can be mutated, altering the charge distribution along the pore [200,201]. Engineering of natural channels, with specific conductance and selectivity properties, is therefore conceivable. As the features of integrated devices continue to shrink, the incorporation of biological ion channels in the design of novel bio-devices becomes an increasingly appealing and compelling possibility for futuristic applications.
One approach to the synthesis of artificial biomimetic channels consists in stacking organic ring molecules to form a pore. Function may be realized by attaching chemical compounds to the structure, forming for instance flaps that close the pore under specific external stimuli or in the presence of other chemical agents. Similarly, one may utilize single-walled CNTs, which can be selected with different molecular properties in terms of diameter of the pore and periodic molecular folding (chirality). As discussed earlier, CNTs have the additional property to be conducting or semiconducting materials, depending on the chirality. CNTs can then be functionalized in principle by binding certain molecules on the mouth or on the side of the nanotube. Nanopores may also be formed by etching holes in ultra-thin solid-state membrane layers. This arrangement offers the easiest way to realize structures directly in planar integrated circuit technology, although there is perhaps less design flexibility from the point of view of chemical functionalization. Because of the randomness of the etching process, the walls of the nanopore are also not as regular or controllable as with the other synthesis approaches discussed above. Nanopore structures of this kind have been proposed for applications in DNA sequencing, for instance.

The numerical challenges in simulation of biological and biomimetic systems arise from the multiscale nature of the conduction processes influenced by the physics of ionic and molecular motion. The time-scales of interest in a complete description of such systems are summarized in Table I. For the specific example of charge conduction through ionic channels, ion traversal through the channel pore is in most cases a rare event, in the sense that an actual ion traversal is the result of very many unsuccessful attempts with trajectories reflected back into the electrolyte reservoir of origin. The measured currents are on the order of picoamperes, with on the average one ion traversal approximately every $10^{-7}$ sec per picoampere. A single-channel current measurement is the result of long-time averaging, at least over milliseconds, with respect to the actual ion traversal times, which may be on the order of $10^{-10}$ sec once a favorable trajectory is initiated by a lucky ion. Simulation of bio-

<table>
<thead>
<tr>
<th>Time scale</th>
<th>Phenomena</th>
</tr>
</thead>
<tbody>
<tr>
<td>Femtoseconds</td>
<td>Bond stretching, breaking and bending</td>
</tr>
<tr>
<td>Picoseconds</td>
<td>Hydrogen bond lifetimes</td>
</tr>
<tr>
<td>Nanoseconds</td>
<td>Diffusion times for small ions to move by nanometers</td>
</tr>
<tr>
<td>Microseconds</td>
<td>Statistically meaningful ion transport to define the state of the system</td>
</tr>
<tr>
<td>Milliseconds</td>
<td>Nanoscale molecular assemblies undergo significant conformational changes</td>
</tr>
</tbody>
</table>

Table I

TIME SCALES IN BIOLOGICAL PROCESSES
logical systems is clearly a time-multiscale problem. All possible length scales are also involved in the complete description of a biological system. A summary is presented in Table II. From a numerical point of view, the main space-scale limitations in comprehensive simulation of such structures arise from the size of grids that may be handled by a computational tool. Coupling of two or more space-scales may require stitching together domains covering different regions characterized by different grid sizes and possibly altogether different representation of the physical observables in the simulation approaches.

The goal of the following sections is to give a summary of ionic channel simulation, focusing on methodologies rather than known physiological function, for which the reader is referred to the excellent review by Tieleman et al. [202].

Ionic channels are immersed in a water environment and free ions originating from salts dissolved in the water are the vehicle for charge conduction through the pore. This is in contrast with the molecular structures discussed earlier, where electronic currents are controlled to generate function. Conduction in a solid-state device implies that carrier motion is regulated by a relationship between momentum and energy, captured by the band structure. Once this dispersion relation is known, a simulation model treats carriers as quasi-particles that obey the band structure as if moving in a vacuum but now an effective mass is associated to the carrier, which depends on the curvature of the energy–momentum relationship. While the band structure is determined by assuming a rigid and perfectly periodic crystal matrix, the effect of vibration of constitutive ions from their equilibrium (zero temperature) position is introduced with the carrier–phonon interaction. In a wet environment, the crystal matrix is replaced by polar water molecules in constant motion and characterized by a neutral but non-uniform charge distribution. Actual ionic particles, as opposed to quasi-particles, are treated by the models. Ions are considerably heavier than electrons, may have multiple ionization charges depending on the valence of the atom, and correspondingly to the large mass have a wide radius which is not negligible.

In semiconductor devices, conduction channels are delimited by depletion layers or potential energy steps at hetero-junctions, the most common being the sili-
con/oxide interface of MOSFETs. An ionic channel is delimited by a highly charged protein membrane creating a pore. Typically, solid-state transistors are based on a planar technology, so that conduction channels may be very thin in one direction normal to the channel axis, but are still very wide in the third direction. There is structural similarity only with recently developed transistors like the FinFET discussed earlier, where a rod of silicon embedded in oxide forms a narrowly confined channel. Since the ion mass is very large when compared to the electron effective mass, size quantization in the transverse cross-section, even in very narrow nanopores, is not an important concern for ion motion and a continuum of transverse states can be considered.

The real disadvantage in scaling solid-state structures is that fluctuations of doping and sizes are difficult to control reliably at the deep nanoscale. Doping becomes problematic because, in restricted dimensionality systems, the defect or excess of just one dopant atom may completely change the behavior of a specific device. As already pointed out before, biological ion channels are perfect replicas of a given structure and stable mutations can be reproduced to modify the behavior of a reference system. Solid-state systems tend to be very robust and may withstand higher temperatures than biomolecular matter, however, high fields may impart sufficient energy to electrons, so that they may become “hot” and cause structural damage. Ions in water instead quickly thermalize, losing energy acquired by field acceleration to the water environment.

Even the smallest ionic channels from bacteria contain an extremely complicated protein structure. A wide variety of molecular groups are present and only for several channels a detailed crystallographic structure is known, while the number of atoms involved is orders of magnitude higher than for the simple organic molecules discussed earlier for device applications. The description of the channel protein by itself with a quantum mechanical model is already an enormous task. Applications are mainly limited to determining the charge distribution on the protein wall defining the pore, which is a necessary input for studying the dynamics of permeating charged particles. Ideally, one would like not only to be able to determine the quantum mechanical charge distribution, but also to follow charge fluctuation and structural changes that may occur when ions enter the pore, changing the local charge. This problem is still out of reach for present computational resources, and just formulating the complete theoretical problem properly is a difficult task. Ion permeation is usually studied by assuming an average fixed configuration of the protein, ignoring changes induced by mobile charge. For proof of concept, simulations are often carried out with a schematic channel structure, defined by a cylindrical surface with varying cross-section, formulated to average local charge properties and achieve a reasonably representative potential profile along the axis. These qualitative models may be very useful to test general hypotheses but are affected by wide uncertainties.
Still, for very complex channel structures this may be the only viable representation possible, since complete structural data is normally unavailable.

Certain ionic channel structures have been studied with particular attention and used as benchmarks for development of simulation approaches. Some examples include gramicidin A, an early form of antibiotic which forms a relatively small channel easier to characterize; KcsA, a bacterial potassium channel with mechanosensitive properties; ompF porin, a resilient triple channel found in the membrane of *E. coli* bacteria, which has a much bigger structure and which can be manipulated in the lab to create controlled mutations. Although the playground of well characterized structures is still limited, there is sufficient information from measurements of current conduction to construct a meaningful hierarchy of simulation approaches that can be compared and calibrated with available experimental data. The same simulation approaches are suitable with minimal modifications to address biomimetic structures, which are inspired by biological systems but may have much simpler complexity and predictable structure.

Ionic channels are important for the development of biomolecular simulation, because they provide the opportunity to relate biological function to the detailed atomic structure of a single protein, since for a number of channels measurements are available for both crystallographic structure and current–voltage behavior. Different levels of detail to be included in the model may be appropriate depending on the goals of the simulation. Most of the available simulations of ionic channels are based on either atomistic or coarse-grained simulation. In an atomistic simulation the majority of atoms in the system are treated explicitly, using a classical approach. With typical computational resources, practical simulations may include on the order of 50,000 atoms for approximately 10 ns of simulated time [203–205]. In a coarse-grained simulation the protein is considered in detail to obtain an appropriate charge distribution. The protein structure is however kept fixed and the environment is modeled by a continuum. The difficulty with this approach is that continuum parameters like dielectric permittivity must be assigned to the various regions of the environment to solve for the fields and this is not a straightforward task. Coarse-grained simulations, however, allow one to extend the observation to much longer times, sufficiently meaningful to compare with actual experimental measurements [206, 207]. There is a variety of possibilities in constructing coarse-grained models and new developments are continuously taking place. The main distinction is between models that treat the conduction ions explicitly or as a continuum fluid. Clearly, there is a qualitative aspect to coarse-grained models, and in order for them to be predictive, calibrations steps are necessary, either by comparing with measurements or with atomic level simulations.

Atomistic simulations are typically based on a classical molecular dynamics (MD) approach. The electrostatic force between two charged particles is obtained from the
classical formula
\[ F_{ij} = \frac{Q_i Q_j}{4\pi\varepsilon_0 |\vec{r}_{ij}|^3} \]  
and the total potential energy is simply obtained as
\[ \Phi = \sum_{ij} \frac{Q_i Q_j}{4\pi\varepsilon_0 |\vec{r}_{ij}|}. \]

where \( Q \) is the charge, \( r_{ij} \) is the interparticle distance and \( \varepsilon_0 \) is the vacuum permittivity. \( Q \) refers both to the charge carried by ionized atoms flowing in the system (always an integer number of elementary electron charges) and to the “partial” charges distributed over the molecular structure (distributed in general as fractional values). An \textit{ab initio} approach must be used to provide an electrostatic potential distribution as generated by the molecular structure. The partial charges are then chosen to reproduce that potential distribution. Although the electronic charge is distributed across the atoms spacing in the molecules, one would normally obtain a certain number of fixed charges placed in correspondence of atomic nuclei.

In addition to the electrostatic potential one must account for the attractive van der Waals interaction and for the repulsive energy of the atomic cores (a consequence of Pauli exclusion principle when atoms get too close). The combined effects are normally treated with the Lennard–Jones potential
\[ \Phi = 4E \left\{ \left( \frac{\sigma}{|\vec{r}_{ij}|} \right)^{12} - \left( \frac{\sigma}{|\vec{r}_{ij}|} \right)^6 \right\}, \]

where the van der Waals interaction is expressed by the second term with exponent 6. Here, \( \sigma \) represents the point at which the potential is zero and \( E \) is the depth of the resulting energy well, which reaches its minimum at distance \( r_{\text{min}} = 2^{1/6} \sigma \). Because of the exponents in the formula, the Lennard–Jones potential is also often called “12-6 potential.” This expression is pertinent to interaction between the same type of atoms, therefore, for each species one has a particular value for \( \sigma \) and \( E \). In order to treat the interaction between dissimilar atoms, it would be necessary to use a different set of parameters for each pair of species. This is obviously difficult when a large number of species is present. An approximation commonly used is to average the parameters for the two species. A geometric average is used for the potential well and an arithmetic average for the \( \sigma \) parameter
\[ E_{ij} = \xi_{ij} \sqrt{E_{ii} E_{jj}}, \]
\[ \sigma_{ij} = \eta_{ij} \frac{\sigma_{ii} + \sigma_{jj}}{2}. \]
This approach is known as the Lorentz–Berthelot mixing rule for two atomic species \( i \) and \( j \). The rule is written above in a generalized form, to indicate that in reality the results may deviate from the simple averages. In the simple approach one would just set the coefficients \( \xi_{ij} \) and \( \eta_{ij} \) to unity.

One complication is that the parameters used for this model of an ionic channel must be consistent with the fact that ions are interacting with water but the problem shows an admixture of situations ranging from aqueous to crystal-like. One has to assign a radius to the ions, and it is not appropriate to simply use the original crystal radius (e.g., for typical salts like KCl or NaCl). An in-depth discussion of this issue is beyond the scope of this work, but the reader may find information in [202]. The detailed treatment of interaction between ions and water is also not straightforward. Because of the mutual interaction, water forms a hydration shell surrounding the ion in the bulk. But inside the channel constriction the protein becomes involved in the interaction, probably within times on the order of nanoseconds [202].

Water itself does not maintain bulk properties within a narrow constriction. The orientation of water molecules may change, so that the overall structure could differ considerably from bulk, and the rate of diffusion and rotational correlations could be altered giving different dynamical properties. MD simulations can be performed to show ordering of the water molecules in constrictions. When the channel is so narrow that only one water molecule may fit in the cross-section, then single file motion takes place [208]. The parameter normally used to quantify water motion is the self-diffusion coefficient. An estimate from simulations can be accomplished by considering the mean square displacement of water molecules in time. A number of standard water models can be selected in commonly available MD codes (e.g., GROMACS, NAMD) and, even in the case of bulk, different values of self-diffusion are yielded by the various models [209]. Within a pore, given a specific bulk water model, a different self-diffusion coefficient is obtained and the simple theory based on mean square displacement should certainly be revised when the cross-section of the channel only allows single-file motion.

Given the complexity of water motion, it is particularly instructive to perform MD simulations in pores of regular shapes. For instance, as illustrated earlier, molecular dynamics simulations of water in carbon nanotubes, forming much more regular pores, show that water effectively freezes in such a confined structure [186–188]. Even simpler situations modeling pores as cylindrical cavities with smooth hydrophobic walls have been considered in the literature [210,211]. When the irregular pore surface of a biological ion channel is considered, although the details are more complicated, one still observes a deviation of water properties from the bulk phase.

The truly unique strength of a full MD simulation is in the ability to map the conformational changes of the protein during the channel functions. Practical simulations are still limited to short observation times, but the potential is there to evaluate
the details of conformational changes when the ions are permeating the pore and to assess the consequences of not including protein structure variation in simplified models. A narrow channel like gramicidin A is a good test case for examining such effects. For instance, it is reviewed in [208] that the choice of a rigid or deformable channel leads to different estimates of cation permeation energetics. The channel protein may also undergo changes in the ionization states of its component molecular elements. The side chains that wrap around the pore contain a number of ionizable states and these chains may interact with each other so that the ionization states could be different from the normal values of the molecular components as individual amino acids. For instance, adjacent sub-units may share a proton, $\text{H}^+$, so that one component is neutral or “protonated” and the other is negatively charged. The pH of the electrolyte solution and also the presence of permeating ions near or inside the channel may change the protonation states. Such states are actually to be intended as probabilistic, with the proton residing at one or the other chain for a certain fraction of time. Since these probabilities should be sensitive to the dynamics in the electrolyte, in a complete simulation one should incorporate the mechanisms of protonation and deprotonation, particularly in those side-chain that may be in closest proximity with the permeating ions. This would be particularly interesting near the channel mouth where dynamic changes in ionization states would certainly affect the approach of ions attempting to enter the pore. To-date, the computational cost and implementation difficulties of such simulation details have in general discouraged their inclusion in full MD simulations of ion channels.

The computational limitations of atomistic simulations still prevent the exploration of ionic channel structures to the interesting limits. Over the practical observation times of tens of nanosecond, it is possible to record only a small number of ion permeation events, even for a large channel structure. Much longer observation times are necessary to determine macroscopic properties like current flow. MD simulation is also difficult to cast into a “device” simulation formalism, as it necessary when a transmembrane potential is applied to the structure and the membrane is in a non-equilibrium state. Although the potential is additive, it is not straightforward to add the external potential to the one evaluated from charge–charge interaction in the MD algorithm. One should evaluate the potential distribution for the channel structure empty of charges with appropriate applied bias conditions. This can be readily accomplished with a numerical solution of Poisson, but interpolation to ionic positions is cumbersome and prone to introduce errors in the overall field evaluation.

A coarse-grained methodology applies a higher level description of the physics involved in the conduction process, thus attempting to duplicate current measurements with a more macroscopic focus. The common element in a coarse-grain simulation is that some or all the components of the physical model are described by a continuum theory. The lipid bilayer and the protein are normally replaced by solid regions con-
sidered to be impenetrable. A specific value of dielectric permittivity is associated to the lipid and to the protein. In addition, similarly to what is done in MD simulation, a distribution of partial charges, consistent with \textit{ab initio} calculations, is assigned to the protein walls. Also, water is not simulated in terms of individual molecules but it is assumed to be a background medium for the charged conducting ions, with a specified dielectric permittivity.

A physical equation for the flow of an equivalent charged fluid introduces both the drift originated by potentials and the diffusivity due to fluctuations in density. The ions are therefore treated as a compressible fluid. In classical chemistry the flow equation is known as Nernst–Planck equation, and it gives the current density for each species $i$ of ions with a law of the type

$$J_i(\vec{r}, t) = -q D_i \frac{n_i(\vec{r}, t)}{k_B T} \nabla \phi(\vec{r}) - q D_i \nabla n_i(\vec{r}, t).$$

(41)

Here, $q$ is the carrier charge, $n_i$ is the particle density, $\phi$ is the electrostatic potential, $-\nabla \phi$ is the electric field, and $D_i$ is the diffusion coefficient. Readers should recognize immediately that the Nernst–Planck equation is virtually identical to the drift-diffusion equation discussed earlier for semiconductor modeling. The ratio $D_i/k_B T$ represents the familiar low-field mobility of the charge carriers.

The drift-diffusion equation is very suitable to describe charge transport when carriers are abundant in the structure and do not acquire too much energy from the applied bias, since the implicit assumption behind the equation is for the system to be close to a local equilibrium everywhere. These conditions are well satisfied in relatively large semiconductor devices where the collective motion of electrons and holes resembles a fluid. Drift-diffusion breaks down as a physical law when devices become smaller, but although the predictive power of the equation may be lost, it is still qualitatively acceptable and it is possible to calibrate the transport parameters (e.g., diffusivities) to match experimental results. When calibration is necessary, the drift-diffusion model becomes an “engineering” tool, rather than a physical one, and while the terminal currents represented by the model may be adjusted to be correct, the actual density distribution in the structure represents the behavior of an equivalent ideal fluid that yields the same macroscopic flow rate as the actual carriers. Diffusivities in the channel region become effective transport parameters, but no longer represent an actual physical constant. Therefore, the density distribution of the calibrated drift-diffusion may be suggestive of the physical processes but not of a correct representation of the physical details. Similarly, in ionic channels one cannot expect the Nernst–Planck equation to be anything else but a qualitatively useful model that needs to be calibrated for quantitative applications.

If a numerical approach is followed, the Nernst–Planck equation can be solved for any arbitrary potential profile distribution. In some cases, when the detailed channel
structure is not known, the 1D potential profile may be reverse engineered until the currents obtained with the Nernst–Planck equation match the measured I–V characteristics. An approach of this kind was taken, for instance in [212].

For a self-consistent treatment in a continuum model, one needs to solve the Poisson equation, relating fixed and mobile charge density $\rho$ to the electrostatic potential, $\phi$. For a space-dependent dielectric permittivity $\varepsilon$, Poisson equation has the form

$$\nabla \cdot \varepsilon(\vec{r}) \nabla \phi(\vec{r}) = -\rho_{\text{fixed}}(\vec{r}) - \sum_{\text{ion}} \rho_{\text{ion}}(\vec{r}).$$

(42)

The Poisson equation may be coupled to the Nernst–Planck equation, to describe a self-consistent continuum ionic flow, or simply to an equilibrium statistics to determine the equilibrium continuum charge distribution in the structure. For the latter case, the mobile charge in equilibrium may be approximated by using the exponential Boltzmann factor, so that for each ionic species one has

$$\rho_{\text{ion}}(\vec{r}) = Q_{\text{ion}} n_{\text{ion}} \exp\left[-\frac{Q_{\text{ion}} \phi(\vec{r})}{k_B T}\right],$$

(43)

where $Q_{\text{ion}}$ is the ionic charge and $n_{\text{ion}}$ is the density at a reference potential $\phi = 0$. The non-linear Poisson equation is normally referred to as the Poisson–Boltzmann model in the biophysical literature. This equation is very appealing, because it is not too expensive to solve and community codes are also available for the complete 3D solution [213]. The Poisson–Boltzmann model has therefore become a very popular tool for model calculations [213–219].

With the addition of the Boltzman statistics, Poisson equation resembles now the familiar non-linear form used routinely for low-doped semiconductor structures. There are a number of assumptions limiting the validity of a Poisson–Boltzmann model for applications to ionic channels that have similar parallels in semiconductors. The theory is strictly valid for equilibrium of a charge fluid, assumed to be a continuum, and the statistics is applicable only at relatively low concentrations. In a semiconductor the Fermi–Dirac statistics needs to be used at high concentration, to account for the Pauli exclusion principle of electronic states. The limit on concentration appears to be more stringent for ions, since it is acknowledged that the Boltzmann statistics may be only valid in the millimolar range while concentrations on the order of one molar or more are common. However, it is not clear what alternative statistics could be used for ions to improve the approach. In non-equilibrium conditions the model is not able to provide an accurate relationship between local charge and potential [220,221] because a complete flow equation should be invoked. This is a well known issue in semiconductor models, where the explicit reference potential (Fermi level) appears in the statistics. Under bias, one may still identify quasi-equilibrium Fermi levels in the contacts, but inside the active regions where
conditions deviate significantly from bulk equilibrium one may not describe the concentrations of mobile positive and negative charges with a single potential value. Semiconductor models solve this issue in various ways, for instance by defining a fictitious “quasi-Fermi level” for each species, but it is clear that flow equations are mandatory to reach the steady-state concentrations of charged carriers which are consistent with the electrostatic potential distribution.

Because of the similarities of the non-linear Poisson and Nernst–Planck equations with the system of semiconductor equations, it is very appealing to take advantage of the robust numerical approaches developed for solid-state device simulation to solve electrolyte flow in biological structures. However, there are some basic differences in the physical systems that should be pointed out because they need to be taken into consideration for model implementation. In a pure semiconductor, there is a temperature dependent intrinsic concentration of mobile electrons in the conduction band and of holes in the valence band. Valence electrons overcome the band gap energy to reach the conduction band, leaving behind an equal number of positively charged mobile holes. When appropriate dopant impurities are added, with energy levels close either to valence or to conduction band, an imbalance of positive or negative mobile charge is created, with a corresponding opposite fixed dopant ion charge. The doping effect is at the basis of most modern solid-state device applications. This cannot be duplicated in an ionic solution. By increasing the concentration of salt in water, equal quantities of mobile positive and negative charges are generated. In a semiconductor it is not possible to increase at the same time the concentration of electrons and holes by adding both donor and acceptor impurities, because of the well known phenomenon of compensation, so that only the net difference of acceptors and donors contributes to mobile carriers.

For a single type of salt dissolved in water, the bulk electrolyte behaves like an intrinsic semiconductor. The mobile charge concentration in electrolyte is varied simply by modifying the salt concentration, which is equivalent to varying the intrinsic concentration of the semiconductor. If one wants to use solid-state software for simulating ionic flow, it is then important to be able to have access to the input data structure. In the several commercially available device software packages, this may not be possible, since input data for specific materials may be kept hidden from the user. If so, only very limiting cases of ionic channel simulations may be investigated. For instance, if the channel is known to be strongly selective for cations (e.g., Na\(^+\), K\(^+\)) one may assume that the anion species (e.g., Cl\(^-\)) is not mobile but simply provides a fixed charge background to provide the appropriate charge screening. The semiconductor model can be used to mimic the electrolyte by using an acceptor doped material that provides mobile positive holes, now representing cations, and negative fixed acceptor charge for anions. Only one flow equation needs to be solved
and the results of calculations need to be appropriately scaled, to account for the different ion transport parameters.

Obviously, it may be very frustrating to have a sophisticated numerical engine but not the capability to fully control the input parameters. One needs to keep in mind that when 3D simulation is necessary, the realization of a robust and efficient simulator is quite difficult and requires a large development investment. Some of the best numerical engines are proprietary systems marketed to integrated circuits manufacturers. Independent development of a fully capable 3D simulator by small research groups is a daunting process, because of the large amount of expertise needed to realize professional quality software. Some key issues and common pitfalls will be illustrated later. Despite its limitations, the Poisson–Boltzmann theory has been undeniably valuable for a qualitative or semi-quantitative analysis of ion channels, for example the analysis of calcium channels [222].

The continuum approach represents an equivalent fluid corresponding to an average flow of the actual particles. One should always keep in mind that the model is calibrated trying to match terminal currents. Therefore, the actual density and potential distribution evaluated inside a nanoscale constriction, as found inside an ionic channel or a nanopore, should be interpreted with care because a number of physical effects may be lumped in the calibrated parameters. In order to increase the physical descriptive ability of the continuum model, one may provide additional equations or additional terms in the currents, to account for water occupation and ionic size effects [223] or to account for the detailed image force established at the dielectric water–protein boundary [224,225].

It is therefore important to develop a modular computational platform that can be extended to build a sub-hierarchy of continuum models as new theories are introduced, to test and implement them. Continuum models are also very important to build a complete multiscale framework for efficient simulation of large systems, since bulk conditions are well represented by a continuum flow while more detailed and computer intensive physical models may be applied mainly in correspondence of constrictions and interfaces where physical laws deviate markedly from bulk conditions. The continuum flow models is physical under certain conditions (bulk, near equilibrium) and it is valuable to understand the circumstances under which the mathematical model has to transition from predictive to calibrated in order to provide accurate terminal currents. The typical limitation that plagues semiconductor applications is the emergence of hot carrier transport, originating from high field conditions leading to energy distributions of carriers which drastically differ from equilibrium conditions. This is normally not a concern when considering ions in water, since the strong interaction with water molecules quickly thermalizes ions accelerated by the electric field, creating an over-damped regime. The semiconductor case also deviates from the electrolyte one when considering the size of charged
particles. Electrons can be treated as point particles, while one always needs to be aware of the size of ions in restricted dimensionality systems, like ion channels and nanopores.

The drift-diffusion or PNP theory written for steady-state does not require the explicit simulation of ion movement in time and the 3D equations can be solved with a variety of numerical methods (finite-difference, finite volume, finite-element, or meshless algorithms). The basic assumption is that the electric field is calculated self-consistently from the average ionic densities. Ion–ion interactions are thus incorporated at a mean-field level.

The construction of a consistent continuum model from molecular system information of biological system is an area still under investigation by the biophysical community. It is therefore important to provide a numerically robust and flexible continuum simulation platform to the community, so that user may focus on testing and refining model building hypotheses, rather than struggling with unstable or unwieldy numerical code. There are a number of interesting issues that deserve attention in order to provide practical physical pictures of biological ion channels and biomimetic devices. In the transition from atomistic to continuum picture, for instance, one has to construct models for water, proteins, lipids, or other components, in terms of continuum media with a certain dielectric permittivity distribution. It is clear than the dielectric permittivity in bulk water differs from the effective value one should use in a narrow channel where water molecules are ordered interact strongly with the walls of the pore. By the same token, a dielectric permittivity needs to be assigned to the protein forming the pore, as well as a charge distribution which results from polarization and ionization effects [226]. One also needs to identify a boundary between water and protein, also considering the role of the ionic radius in relationship to flow in the pore. The most relevant transport issue is in the definition of diffusivity models. With molecular dynamics simulations one may extract local diffusivity profiles that can be used as inputs for a continuum simulator in tabular form. There are no general answers to each of these physical questions, since one has to consider the specific situation at hand. For the development of generic numerical software the path is already delineated because the constitutive equations are well-defined, independently of physical details. The construction of physical inputs instead becomes a true community effort, which may require adjustments as research is conducted and practical applications are developed.

The main limitation of the continuum approach for ion representation is in the fact that the finite size of the carriers is not accounted for. This is not an issue in semiconductor electronics, since electrons and holes may be assumed to be point-like particles. In the case of ionic channels with small cross-section, the ion diameter is certainly comparable with the pore diameter. General limitations are discussed for instance in [216,227–229]. The fluid representation of particle flow leads to certain
consequences mainly in the electrostatic behavior of the narrow pores. Electrical screening is certainly altered when the continuum limit allows smearing of the discrete ionic charge in space. Due to the strong charge in the protein, nothing prevents fluid of the opposite charge to be attracted to the pore walls, reaching local concentrations in excess of physical values allowed by the actual discrete ion charges. The fluid represents a continuum of point particles and the equations do not have provisions to prevent excessive bunching of charge. One approach to avoid this is to add a correction potential in the current equation, with the function to limit the density as allowed by the presence of other ions and water. There are two distinct correction terms, one for ion size and one for charge interaction when ions are in proximity of each other [223]. Full implementation of such models in existing 3D continuum solvers is still subject of investigation, because the evaluation of correction terms is relatively expensive and it introduces additional non-linearity in the system. An simplified implementation to first order can be found in [230].

The next level of coarse-grained models maintains a similar implicit continuum description of channel and water, but retains the atomistic nature of the ions which are treated as individual particles similarly to Molecular Dynamics. The most typical approach is to adopt Brownian Dynamics (BD) for tracking ion trajectories, using a Langevin-type equation for the acceleration, where $M_i$ and $v_i$ are mass and velocity of the ion

$$M_i \frac{d v_i}{d t} = -\gamma_i v_i + F_{\text{el}} + F_{\text{ran}} + F_{\text{SR}}.$$  \hspace{1cm} (44)

The interaction with water is implicitly accounted for by the friction coefficient $\gamma_i$ directly related to the diffusivity, and a random force $F_{\text{ran}}$ due to stochastic random collisions. Since the water ion system is overdamped, it is common to neglect the inertial term on the left-hand side, thus obtaining an explicit formulation for the velocity. The short range force $F_{\text{SR}}$ is added to include ion size in the ion–ion interaction so that particles cannot overlap, and it can be modeled with a Lennard–Jones type potential. The electrostatic force $F_{\text{el}}$ originates from the charges in the system and the applied potential. In approximate treatments, the external potential is accounted for by a uniform field in the channel, corresponding the average effect of the transmembrane potential, plus the coulomb interaction between ions and fixed charges. There is a vast literature on BD applications to ionic channels. Some of the most essential references on the subject can be found in [231–237]

In a realistic structure with variations of relative permittivity and fixed charges embedded in the protein it is not a trivial task to evaluate accurately the inter-charge forces. Image forces should also be included when ions are in proximity of dielectric boundaries. To overcome these theoretical difficulties it is more natural to perform a solution of Poisson equation on the domain, similarly to what is done in the PNP
model. In this case, however, one needs to add a procedure to map the ionic charge onto a grid defined over the system domain. There are a number of possible ways to do so. The most common approaches consist of mapping the charge density onto the grid point nearest to the ion, Near Grid (NG) assignment, or to distribute the charge onto the grid nodes, Cloud in Cell (CiC) assignment [61]. The CiC assignment is more flexible but one limitation of the scheme is in the fact that if the mesh is not regular, a self-force may be introduced when the fields are interpolated to the ion location. Obviously, a particle should not feel itself after it is replaced by a distribution of charges over the nodes of the mesh cell. This is a classical problem of particle simulation, and it has been addressed in [61,238]. The analysis in [238] indicates that for an unstructured grid it is not practical to avoid self-forces.

Another particle approach, also popular with device simulation, plasma physics and neutron simulation in nuclear reactors, is based on particle Monte Carlo algorithms [60–62,239]. Since the term Monte Carlo is used for many different types of calculations with different meanings in the various disciplines, this specific algorithm will be called Transport Monte Carlo (TMC) in the following. In TMC, the particle trajectories are developed as a collection of free flights under the influence of the fields, interrupted by scattering events using a procedure very similar to the one adopted for electron transport in solid-state devices, so that during the free flights the particle obeys classical mechanics motion laws under the influence of the electrostatic forces. For the generation of the distribution of free flight using random numbers, a single equivalent scattering rate is used, related to the diffusivity of the system. After scattering, a new momentum state is selected, providing a random direction of motion. For an overdamped electrolyte system, the random energy state after scattering may be chosen from a Boltzmann distribution, assuming that water interaction provides quick thermalization for the ions. The TMC approach is amenable to efficient implementation for high-performance computing [240].

For self-consistent treatment, in both BD and TMC an ensemble of particles is followed and updated inside the structure to account for injection in and out of contacts and Poisson equation is solved similarly. When the potential is resolved on a grid, one needs to consider the fact that Poisson equation treats only the so-called long-range interaction. When ions are in close proximity there is also a short-range coulomb interaction, in addition to the Lennard–Jones potential. For particles within the short-range radius, usually encompassing two or more mesh intervals, a coulomb force (particle–particle) is computed directly and added to the Poisson force (particle–mesh). Because of the overlap between the short-range domain and the mesh, there is some double counting of the interaction which must be deducted from the total force, for instance using the classical particle–particle–particle–mesh (P³M) scheme [61,241].
The particle simulation only provides a sequence of “snapshots” of the carrier ensemble configurations inside the structure under study. Any of these snapshots by itself does not contain complete physical information. The physical observables of interest need to be constructed through an averaging process. Average distribution of carrier density and potential are obtained by mapping the histories of each mesh in space. The current through the structure under a given external bias, usually the only macroscopic quantity observable by measurements, is recovered by counting the successful crossings for all ionic species in a given simulation time. As stated earlier, such a lucky traversal trajectory is a rare event and the outcome of a very large number of unsuccessful attempts generated by the simulation. While in PNP simulation a smooth continuous occupation density for the particle is obtained directly from the solution of the system of equations, this kind of information requires a lengthy averaging process that tracks particle occupation in the simulation domains. It is typical for a statistical approach like TMC to display noise in the results, reflecting the intrinsic physical randomness of the transport processes for discrete charged particles.

The available computational resources are just now becoming sufficiently massive to allow for estimation of currents from 3D particle simulation, approaching biological time scales. Parallelism is the key to make this simulation process practical. In order to extend the simulation to the needed long observation times, it is possible to apply averaging over an ensemble of time simulations associated with different random number sequences. Exploiting the ergodic principle, in steady-state conditions one may assume that a long observation time is equivalent to the sum of a number of time-segments simulated independently, as long as they are of sufficient length to contain a meaningful number of channel traversal events representing the current of the system. This partition of the overall simulation time into independent sub-intervals is possible because of number of reasons. The system under study is strongly over-damped by water interaction with ions and although electric fields may be of considerable strength locally, the charged ions are never able to gain large energies, since the excess energy acquired during acceleration is quickly lost to water by collisions. The thermalization process is sufficiently fast not to make the distribution function of ions in the system deviate appreciably from local equilibrium. It is reasonable to apply the ergodic principle and state that averaging over ensemble is equivalent to averaging over time. In this case, since the nanoscale system has a small number of particles, one can actually use an ensemble of statistically meaningful observation segments to recover the averaging over a long simulation time, to give an estimate of the current flow under a given bias condition. While the calculation of these processes is expensive, the application of the ergodic assumption is extremely favorable to massively parallel applications, as long as each computational node is powerful enough to carry out simulation of a statistically meaningful time segment.
with acceptable turn-around wall-clock times. The problem at hand can be mapped to
a large cluster with high efficiency and approaching nearly embarrassing parallelism.

As an illustration of coarse-grained particle simulation with transport Monte
Carlo, we consider a prototypical ion channel, gramicidin A, represented in frontal
view in Fig. 17 to show the pore opening. Figure 18 presents a successful permeation
trajectory of a sodium ion in a gramicidin channel, displayed on a section containing
the channel axis. The simulation was performed with the code bio-MOCA, developed at the University of Illinois [242], which is available for on-line computation
at the nano-HUB, the web portal of the Network for Computational Nanotechnology
(http://www.nanohub.org). The current flowing through the ion channel is calculated
by collecting statistics on a large number of trajectories. The compromise between
coarse-grained representation of the electrolyte and detailed electrostatics for a real-
istric protein pore permits a realistic description of the charge permeation processes
on a biological time-scale. For typical conditions and depending on the frequency
of solution of Poisson equation to update forces, using a 2 GHz Intel XEON single
processor, the simulation of 1 ns of evolution may require up to one hour. For a struc-
ture with a rather small pore like gramicidin, iterations are relatively fast because the
cost of the 3D Poisson equation is contained and only a few tens of ions must be
included at typical concentrations. However, simulation cost per unit time is not the
only factor to determine the overall cost of a meaningful simulation. One also needs
to consider that current does not flow readily through this narrow channel, a suc-
cessful traversal trajectory being a rare event. Therefore, longer simulation times are

![Fig. 17. Atomic structure of gramicidin A ionic channel, aligned to show the nanopore opening.](image)
FIG. 18. Trajectory of a successful traversal event for Na\(^+\) ion in a model gramicidin A channel, calculated with a transport Monte Carlo model that uses a scattering model for water–ion interaction. The top plot shows the ion location at different times during the simulation run, indicating a relatively fast traversal process one the favorable trajectory is taken by the particle.

necessary to resolve current flow in gramicidin (at least on the order of milliseconds) than in larger channels where permeation may occur more readily, generating stronger currents, for which several microseconds may be sufficiently representative. Simulation of larger channels involve a larger domain for Poisson equation and may require several hours for 1 ns of simulation [243,244].

8. Conclusions

The needs for simulation of charge transport in electronic structures have changed greatly over time as devices have been progressively scaled. At the nanoscale, a comprehensive hierarchy of models to observe phenomena at different scales in time and space. A model hierarchy also facilitates the process to test and calibrated coarser level models and to interface appropriately with experimental measurements. Many of the simulation tools developed for traditional semiconductor electronics have
found useful applications in the analysis of new systems with molecular nature. For instance, a unified quantum methodology like the non-equilibrium Green's function approach can treat similarly any electronic nanoconductor embedded between two contacts. On the other hand, classical continuum and particle approaches that are only physically valid to relatively large electronic systems, can be used to understand electrolyte flow in nanostructures because of the large mass of ions involved does not require quantum mechanical details for the transport process itself. As the size is reduced, there is an intriguing confluence of disciplines that end up sharing very similar goals and computational approaches. Electronics, chemistry, biophysics and mechanics ultimately need to contend with almost identical problems at the nanoscale, where a range of classical, quantum, and atomistic phenomena coexist and interact across linked multiple scales. Future progress in all of these disciplines is ultimately connected with a sustained growth in computational capabilities since simulation fulfils an increasingly important role, from fundamental physical investigations to practical design. As the end of the traditional silicon technology looms without a clear replacement in sight, one should expect that an even tighter interaction, between the various disciplines that contribute to nanotechnology, will be necessary to overcome future challenges and generate technologies able to fulfil the increasing demands for computation capacity.

ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation through the ITR grant SBC-5011045 and through the Network for Computational Nanotechnology grant EEC-0228390 and by the Army DURINT contract SIT 527826-08.

REFERENCES

This page intentionally left blank
Abstract
Manufacturing materials and systems with components thousands of times smaller than the width of a human hair promises vast and sometimes unimaginable advances in technology. Yet the term nanotechnology has formed as much from people’s expectations as from scientific reality. Understanding the creation and context of this social construction can help us appreciate and guide what may be a burgeoning revolution. This chapter considers what different groups are referring to when they say nanotechnology, how this relates to the science involved, and how the various definitions of this broad field of endeavor might be improved. The ramifications and implications of these seemingly innocuous naming choices are also discussed. Although in many respects nanotechnology has served as cover justification for increased spending in the physical sciences, at present it is the most hopeful route to some of the planet’s greatest problems.
1. Introduction

When Norio Taniguchi first referred to his work as “nanotechnology” [26], it is doubtful he imagined that scholars might one day be pondering the word and considering its implications on humanity, even on the fate of humanity. But this year research on the societal implications of nanotechnology accounts for nearly 10% of direct federal funding on nanotechnology in the United States: 80% of that on environmental and toxicological effects and the remaining on broader sociological studies [15,18,19]. That is not pocket change out of a budget of over 1 billion USD. But how can you study the societal impact of nanotechnology when the term itself has not been clearly defined? In many ways, the label “nanotech” has been thrust on science and engineering externally, formed by society before solidifying as a field in its own right. This timing mismatch has interesting consequences for the shape of progress in nanorelated fields and enterprises.

A nanometer is one billionth of a meter or roughly equal to 10 hydrogen atoms laid side-by-side. Many systems, from biological to chemical to quantum, exhibit fundamentally new behavior at this length regime, often for completely coincidental reasons. For example, the blood-brain barrier often fails to filter particles of nanometer size, while being very selective about larger objects. The nanometer regime also happens to be where the quantum properties of electrons start to dominate in miniaturized circuits. The two have no connection in cause, but they are both considered “new nanoscale phenomena.” This reality of nanotechnology contrasts with the vision of nanotechnology promoted in recent decades by futurists, science fiction writers, and doomsayers, as well as actual scientists. But the reality and the vision of nanotechnology have become entwined culturally despite their almost extreme differences.

Growing excitement in the field often traces back to two people: physicist Richard P. Feynman and researcher K. Eric Drexler. Physicists often justify their work by bowing to a past physics god, of which Feynman is one of the most famous. In his 1959 talk There’s Plenty of Room at the Bottom: An Invitation to Enter a New Field of Physics [8], Feynman laid out the vision of what would one day be called nanotechnology. Simultaneously simplistic and prescient, Feynman explored the vast technical applications that mastering matter on a small scale would create. He gave crude but surprisingly insightful suggestions about how to access and take advantage of all the room “at the bottom.” Although one must be careful of fulfilling Matthew’s Law (automatically attributing good ideas in a field to the most famous person in that
field),¹ Feynman’s talk anticipated manufacturing techniques such as the focused ion-beam etching Taniguchi was later to work on (and which is only now becoming mainstream in university labs), miniaturization in information storage, computation and machines, electron spin electronics (or spintronics), and even hinted at quantum computing.

Drexler is no physics god. But he deserves substantial credit for popularizing and motivating nanotechnology with his 1987 book *Engines of Creation: The Coming Era of Nanotechnology* [4] and later contributions. He falls squarely under the vision category of nanotechnology. The vision of nanotechnology could be described as atomic level systems engineering. Atoms, like so many Lego blocks, can be arranged one-by-one to build microscopic analogues of macroscopic machines or indeed anything one could imagine, from scratch. This is the life analog, the construction block view of both Feynman and Drexler, whom the latter calls machine-phase nanotechnology. From drugs designed from the bottom up to artificial nanomachines that could flow through and oxygenate blood on demand, the vision is dramatic. Drexler imagines assemblers, i.e., nanomachines that build other nanomachines, able to tackle problems from the size of the body to the globe. Take wood, an undisputedly useful building material “grown” by nanoscale/microscale machines. What else could be “grown” if we design the life that grows it? The implications, like those of biotechnology, can be scary. Fortunately, the vision is hard, and the reason this is so is because reality starts to intrude. The novel physics, surface science, and chemistry that begin to emerge in the nanoscale regime makes the construction block paradigms fall apart. The reality is actually much more interesting. Feynman himself pointed out that quantum effects and the complex behavior of many interacting particles—be they electrons, atoms, or molecules—would make this building block strategy difficult. In essence, that difficulty is nanotechnology today. But even he could not have imagined how big and how interesting those complications would be ... and more importantly, how useful.

Let us pause for a brief moment to discuss bias and definitions. Many scientists could claim the authority to explain nanotechnology—biologists, chemists, materials scientists, and physicists. And all would necessarily do an incomplete job, a factor this chapter too will suffer from. Our bias is that of condensed matter and materials physics, the fields from which the transistor, solid-state laser, and silicon integrated circuit were born. Complex terms and definitions, unfortunately, will be commonplace in such a broad discussion of new technologies and their categorization. Get used to being ignorant in this field. Brief definitions will be provided where possible (see Appendix A) but understand that scientists can use one word—say

¹ The Gospel of Matthew (5–7) is where you find the recounting of the Sermon on the Mount. Matthew’s Law is a bit of physics oral tradition; it refers to the tendency of the most well-known person in a field to get credit for the best ideas in that field whether or not he had anything to do with it.
“spintronics”—to represent thousands of technical papers that may cover multiple sub-fields and have subtle meanings. Part of our challenge is that the forward line of the advancement of human knowledge has become so long. With so many individual innovators, complete categorization becomes impossible. You are encouraged to sleuth on your own when you encounter a cool sounding word you do not understand. Call it the joy of finding things out.

1.1 The Scope of Nanotechnology

There are many reports of varying quality that describe examples and applications of nanotechnology, from the perspectives of length-scale, medicine, environment, risk, society, and public policy. We especially recommend [22] and [1]. This chapter will describe the social construction of this term without becoming encyclopedic.

If the social construct of nanotechnology encompasses both the vision and the reality, one must consider the scope of both. (A pragmatic summary of our definition for nanotechnology can be found in Table I.) Oddly, the vision of nanotechnology includes both the fantastic almost science-fiction visions of nanoassemblers and artificial life but also very non-nanovisions including small machines and devices. Compared to a nanometer biological “machines” can be relatively huge; a red blood corpuscle is approximately 7000 nm while bacteria is 1000 nm and a virus is typically 60–100 nm. And key body dimensions are downright enormous. Swallowing a millimeter-sized pill that can take pictures of your insides, though it might be called nanotechnology, may have no nanofabricated components whatsoever. This imprecision is present in both perspectives. In many ways the reality of what is going on in nanotechnology today—as presently named—is broader than even the vision.

At a recent conference on quantum nanoscience in Australia, a physics professor stood up and said something representative of many scientists’ viewpoint: “Can’t we just call it novel-technology?” Maybe we should. In the policy-technology world nanotechnology has become a catchall for all advanced materials technologies, from biointegration to molecular transistors to quantum mesoscopics. The USA National Nanotechnology Initiative is an example of this, as its proponents seek to make nanotechnology a unifying theme across many of the sciences that deal with fabrication and manipulation of matter at small-scales.

Despite these facts-on-the-ground, some have claimed that atomic and molecular engineering alone is nanotechnology [11]. It is simply too late for this narrow and somewhat boring definition. Additionally, there are convincing reasons to lump larger systems under the nanotechnology roof. All the knowledge gained through the study of mesoscopic physics and complex systems might one day help the atomic-level construction technology. The science and technology community has to deal with putting these systems together and with the complex emergent behavior that
implies. What quantum mesoscopics, for example, gives us is a physical window on nature’s immense complication.

It would also be easy to say that the vision or molecular assembly view, as theorized by Drexler’s so-called machine-phase nanotechnology, is the “true” nanotechnology and stop right here. But most of the money being spent on nanotechnology and most of the people who believe they are working on nanotechnology are in the reality camp. The glow that the term nanotechnology exudes not only attracts new students to science but also attracts scientists to adopt the term and become “nanotechnologists.” And it helps with funding. Will nanotechnology just become a rallying cry or does it have any real meaning anymore?

From a societal standpoint, the consequences of this emerging collection of new artifacts, both directly as threats to the environment and our health, and indirectly via the transformative properties of life-changing technologies, have unquestionable importance. But there are limits to the extent of nanotechnology’s embrace. In some situations, e.g., federal funding or patent decisions, more precise categorization can become necessary. For example, the patent office makes a distinction that a new invention just cannot be smaller to be patentable—there has to be something new there (like a fundamentally new manufacturing process). Many of the suggestions Feynman made just took advantage of making things smaller and were not necessarily based on new phenomena at that length scale. In the research world some inventions or lines of discovery, though they might coincidentally take place at the nanoscale, are better classified in a different scientific context. These are often one of disparate yet specific subfields. Specific examples will be presented in Section 3.

The reality of nanotechnology has matured at its own rate as technological manufacturing and measurement techniques have steadily improved over the last 50 years, putting fabrication and control of the nanorealm within reach starting some 15 years ago [6]. It is easy to realize that “seeing” was the key to progress. Indeed, improvements in microscopes such as the electron microscope to the creation of new classes of devices such as atomic force microscopes (AFMs) and many other techniques contributed to advances in small science. Much of this is just the natural progress of science and technology, driven by human curiosity, economic interests like the semiconductor or drug industries, and luck.

The construction block view of nanotechnology is divorced from the daily interests of most scientists and so far away as to be utterly ignorable. In some sense it is a very simplistic idea. “Hey! Let us design things from the ground up!” Let us just make the nanoscopic equivalent of a macroscopic device. Replace the surgeon and scalpel with the minirobot. It might be more difficult to do, for various reasons, but you are not necessarily using any new science to do it. Well, not so fast.

Complexity is an inherent problem in fields from computer science to biology to physics. It is the problem of the next one thousand years. Nanotechnology is not like
building the first bridge. While it is clearly too soon for the vision of nanotechnology to be a wholly engineering endeavor, the beginnings are upon us. Nanoparticles like quantum dots and carbon nanotubes are the quintessential reality of nanotechnology today. They can exhibit new nanoscale phenomena in physics, chemical, or biological manifestations. Even if our theories of quantum physics, chemistry, and biology are complete, how they manifest themselves in complex situations is not. Progress in the nanoscience of mesoscopic and biological-materials systems will be the foundation of larger artificial small systems.

1.2 Threats and Futures and Politics

Fear as much as hope has popularized nanotechnology, from the hysterical visions of “grey goo” to the quite realistic fears of human toxicity. The image of machines invisible to the naked eye wreaking havoc has certainly captured the public’s imagination [3,12]. But are not self-replicating robots that take over the world bad at any scale? In truth, many of the fears of robotics and artificial intelligence that have been around for decades have just been transferred to the invisible realm of nanotechnology, adding another dimension of fear. In any case, we already have replicating killer machines—they are called viruses, and bacteria, and humans. Redoing these with non-biological matter—almost the essence of the vision of nanotechnology—may well be more dangerous, or not, but will certainly be more difficult. We have to deal with those implications now in a plenty frightening way in biotechnology and genetic engineering. We are not sure what purpose calling what some biologists are already doing “nanotechnology” serves, except to say that it expands it to non-carbon based approaches.

Some of the new phenomena that nanoscale systems exhibit can be just as profoundly harmful as useful [2,9,28]. Particularly relevant are the chemical and biological effects. The same reactivity that makes nanoparticles powerful in military-grade explosives can also have dramatic impact on environmental systems [13]. And nanotoxicity is an even more serious concern. Nanoparticle transmission and accumulation through the body is largely unknown. So investment into the study of environmental and health impact is necessary.

The real issue is why there has been so much funding for the societal implications of nanotechnology. The initial reason is to deal with the vision—but the fears of the vision are ridiculous at this early point in time. A reputable scientist would find it laughable. Seriously, where are the nanomachines? Stuck on a chip in a refrigerator in the basement of some university department? Funding the study of futures like this does not seem responsible. Biology today is much more scary. The possible societal harm of nanotechnology pales in comparison to the prospect of an avian flu epidemic,
designer babies, or proliferating suitcase nukes. Even so, the funding of bioethics is currently minimal and of questionable value.

We believe the truth is more indirect. Industry is desperate not to repeat the fiascos of genetically modified organisms (GMO) in food—now banned in Europe—and to a lesser extent nuclear energy. What better way to “educate” the public then to fund a mass of public affairs and ethics professors to study the societal effects of nanotechnology? People tend to be more positive toward ideas when financially dependent on them.

Let us be clear, many of the fears disseminated to the public via science fiction and overzealous technology pundits or “futurists” are irrational; here education is important and worthwhile. Nor are professors and thinkers doing anything untoward. But it is interesting to note that the calls for all-out moratoriums on nanotechnology, ridiculous though they are, come exclusively from independent non-profit groups. So is the government’s largesse enlightened or subtle propaganda? We cannot know, but a prominent political theorist, Langdon Winner, has called on the United States Congress not to create a nanoethicist full employment act [29]. What are important in nanotechnology are science, the progress, and the serious considerations of how to manage the environmental and social changes it creates as it develops.

When it comes down to it, saying no to nanotechnology may be like saying no to tackling the biggest problems of our time. We are faced with epic shortfalls in energy and clean water, and environmental pollution. Nanotechnology related advances in our understanding of ever-interconnected physical, material, chemical, and biological systems are our greatest hope for beating these huge problems. We must not stop science and technology.

1.3 What is in a Name?

When asked about nanotechnology in an interview for the San Francisco Chronicle in 2004 former Intel CEO Craig Barrett made some illustrative remarks that are worth repeating:

“Nanotechnology is a buzzword that you in the press have popularized, and the government popularizes it. The formal definition of nanotechnology is anything below 100 nanometers. Every transistor that we make is below 100 nanometers. So, if you want to know the investment that we’re making in nanotechnology, it’s the total investment that Intel is making. . . . We don’t happen to be the nanotechnology that you popularize with carbon nanotubes and quantum dots and organic molecules that are going to replace CMOS transistors. Most of that is an esoteric and populist impression of what nanotechnology is. The bulk of nanotechnology is what companies like Intel and Texas Instruments and others do today.”
Is Intel really the largest nanotechnology company on earth? Is everything smaller than 100 nm nanotechnology? Or is this just CEO spin? Certainly under our bipolar transition of nanotechnology above—the combination of vision and reality—Intel’s latest and greatest CMOS transistor evolution is not nanotechnology. It is just the logical progression of microtechnology. First of all, companies like Intel are not taking advantage of any new nanoscale phenomena. In fact, quantum effects (like leakage via electron tunneling) are a major problem in the most advanced integrated-circuit design. Nor are they building transistors from the ground up atom-by-atom with molecules. It may be that as the biggest dimension of Intel’s transistors fall below the 10 nm length scale they will have to make use of quantum size effects or switch to another paradigm—then maybe they can claim the title.

Definitions matter in contexts relevant to nanotechnology, including patenting, risk analysis, proper funding allocation, and education. Like it or not, nanotechnology is largely perceived by the public in two ways. One, as Star Trek and most science fiction conveys it: “machines you cannot see”—the more fantastic vision perspective. Nanotechnology has become what we imagine it to be. From the self-cleaning walls, to the nanites of Star Trek, to the grow anything future of The Diamond Age [23]. Second, the public sees the toxicology issue—dangerous particles in food and cosmetic products. This is exaggerated by the wide promotion of the term as marketing tool whether justified or not. The Woodrow Wilson Center has compiled a database of all products that claim to use nanotechnology [30]. But the companies that are actually using the nanotechnology of today—nanoparticles—in situations that may be hazardous (such as in cosmetics) do not advertise this fact.

Nanotechnology has clearly caught the public’s imagination, but they are missing the great stuff—the reality. There is a complete disconnect between what the public perceives and fears and what government agencies are defining as nanotechnology. Is the public seeing more than the toxicology or nanorobot issues? The last century taught us that chemicals could do truly awful things. Can nanotoxicity really be all that worse or is it just a scary name for more of the same? Meanwhile, governments and corporations around the world are funding nanotechnology for the promise of the next big thing and as a unifying theme in materials science development. Understanding how society—through the policy makers, thinkers, scientists, engineers, speakers—are defining and shaping the umbrella that is nanotechnology is our goal.

2. Definitions ad infinitum (and More Politics)

Definitions for nanotechnology abound from government program managers and institutions, corporate entities, individual researchers, and non-profits. Let us collect and analyze some of them here. Since these are the institutions that are driving the
debate and/or funding the evolution of nanotechnology, this is a worthy goal (though somewhat tedious).

The strongest shaper of the direction of nanotechnology is the United States federal government acting through its research funding agencies, primarily the National Science Foundation and various military equivalents (DARPA, ARDA, ARO, NRO, etc.). Mihail Roco, current head of the National Nanotechnology Initiative (NNI) and cheerleader for it even before its inception under President Clinton has given a number of definitions of nanotechnology that are often quoted. Since the NNI is shaping the future of nanotechnology by the power of federal grants [14], this is a good place to start. Roco’s definitions are:

“Nanotechnology is the creation of functional materials, devices, and systems through control of matter on the nanometer length scale, exploiting novel phenomena and properties (physical, chemical, biological) present only at that length scale.” [17]

There is an alternate definition as well:

“The field of nanotechnology deals with materials and systems having these key properties: they have at least one dimension of about one to 100 nanometers, they are designed through processes that exhibit fundamental control over the physical and chemical attributes of molecular-scale structures, and they can be combined to form larger structures.” [22]

And another:

“Nanotechnology is the ability to understand, control, and manipulate matter at the level of individual atoms and molecules, as well as at the “supramolecular” level involving clusters of molecules. Its goal is to create materials, devices, and systems with essentially new properties and functions because of their small structure.” [20]

According to Roco, the NNI definition encourages new contributions that were not possible before:

- “novel phenomena, properties and functions at nanoscale, which are non-scalable outside of the nanometer domain;
- the ability to measure/control/manipulate matter at the nanoscale in order to change those properties and functions;
- integration along length scales, and fields of application.”

Obviously Roco is firmly in the reality camp of nanotechnology, as are most of the definitions to be described. Before we analyze it further let us look at some similar definitions. The United Kingdom, which has been ahead of the game in considering
the implications and risks of nanotechnology, put together a widely cited report in 2004. The Royal Society of London report defines nanotechnology as:

“Nanoscience is the study of phenomena and manipulation of materials at atomic, molecular, and macromolecular scales, where properties differ significantly from those at a larger scale.” [21]

This definition is often quoted outside of the United States.

**TABLE I**

**NANOTECHNOLOGY CIRCA 2006, VISION VERSUS REALITY, A PRAGMATIC DEFINITION**

**The vision**: From the fantastic to the mundane the vision of nanotechnology began over 40 years ago and encompasses the building block view of building matter and machines, and artificial life. Programmable machines you cannot see.

**Examples**: Atom-by-atom or molecule-by-molecule construction of everything from drugs to tiny robots; not necessarily nanometer-sized but also just small (e.g., cameras or surgical machines that can be swallowed); self-replicating nanomachines (or assemblers).

**Threats**: Unknown, but greatly imagined: prolonging human life indefinitely; uncontrollable, replicating nanoassemblers turning the world into grey goo.

**Biggest questions**: Is this a joke? How will modifying or extending the human body change civilization? What about being able to “grow” anything cheaply with minimal materials?

**The reality**: The science and technology behind nanotechnology today predominately defined as new phenomena that emerge in the nanometer length regime in physical, chemical, and biological systems.

**Examples**: Nanoparticles that are (1) many times more reactive than microparticles, (2) have changing properties—like color—with small size changes due to quantum confinement effects, or (3) have unknown biological effects like increased absorption in the body; advanced materials technology; quantum “mesoscopic” physics; everything and anything.

**Threats**: Possible environmental hazard as nanoparticles can be much more reactive than larger particles and may linger longer; potential risk to human and animal health due to unknown effects on the body.

**Biggest questions**: How do we regulate these materials? How do we classify them and determine their risk?

**Nanoenabling technologies**: Fabrication and measurement techniques that aid in the realization of the above two definitions but do not fall within their purview. The reliable manufacturing of devices and structures with at least two dimensions at the nanoscale.
Roco is seeking to ignite a new “man-on-the-moon” scale investment in science and technology in this country, foreseeing a generational evolution of the investment and development (see Table II). He is looking at nanotechnology as a unifying theme to advance all advanced technologies, which are based largely on increased understanding of the sciences made in the past century. His definition is very biased and broad to encourage these opportunities. Business as well has a motivation to make nanotechnology widely used, at least for now, because “cool” sells.

Recently, the elite academic publisher Nature (London) spun off a new journal called Nature Nanotechnology. It is illustrative to find out what they consider acceptable for submission. In Nature’s words, Nature Nanotechnology “is a multidisciplinary journal that publishes papers of the highest quality and significance in all areas of nanoscience and nanotechnology. The journal covers research into the design, characterization and production of structures, devices and systems that involve the manipulation and control of materials and phenomena at atomic, molecular and macromolecular scales. Both bottom–up and top–down approaches—and combinations of the two—are covered.” The topics included are listed in Table III.

Meanwhile, the USA Patent and Trademark Office, trying to reorganize to deal with nanotechnology patents and simplify nanotech investment, has it is own definition:

“Nanotechnology is related to research and technology development at the atomic, molecular or macromolecular levels, in the length of scale of approximately 1–100 nanometer range in at least one dimension; that provide a fundamental understanding of phenomena and materials at the nanoscale; and to create and use structures, devices and systems that have novel properties and functions because of their small and/or intermediate size.” [27]

The FDA, concerned with human exposure, takes a similar approach.

“The FDA defines “nanotechnology” as research and technology or development of products regulated by the FDA that involve all of the following:

(1) the existence of materials or products at the atomic, molecular or macromolecular levels, where at least one dimension that affects the functional

---

Table II

<table>
<thead>
<tr>
<th>Four Generations of Nanotechnology According to NNI: Timeline for Beginning of Industrial Prototyping and Technology Commercialization [20]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st: Passive nanostructures. Example: coatings, nanoparticles, nanostructured metals, polymers, ceramics</td>
</tr>
<tr>
<td>2nd: Active nanostructures. Example: transistors, amplifiers, targeted drugs, actuators, adaptive structures</td>
</tr>
<tr>
<td>3rd: Systems of nanosystems. Example: guided assembling, 3D networking and new hierarchical architectures, robotics, evolutionary</td>
</tr>
<tr>
<td>4th: Molecular nanosystems. Example: molecular devices ‘by design,’ atomic design, emerging functions</td>
</tr>
</tbody>
</table>

---
We see a lot of similarities in all of the above definitions. We also see where Craig Barrett’s confusion comes from. These definitions are so broad as to encompass virtually all developments in materials science and biotechnology. For example, although most of them make a point of requiring a “new” property to be considered, what is new?

It is not clear whether simply making something small (even in only one dimension) is enough to be considered nanotechnology. Consider two additional contrarian definitions. Two of the most vocal non-profits critical of unfettered nanotechnology investment are Greenpeace and ETC Group. Greenpeace’s definition is:

“The most common definition of nanotechnology is that of manipulation, observation and measurement at a scale of less than 100 nanometers (one nanometer is one millionth of a millimeter). However, the emergence of a multi-disciplinary field called ‘nanotechnology’ arises from new instrumentation only recently available, and a flow of public money into a great number of techniques and relevant academic disciplines in what has been described as an ‘arms race’ between governments. Nanotechnology is really a convenient label for a variety of scientific disciplines which serves as a way of getting money from Government budgets.” [10]
The ETC Group definition:

“Nanotechnology refers to the manipulation of living and non-living matter at the level of the nanometer (nm), one billionth of a meter. It is at this scale that quantum physics takes over from classical physics and the properties of elements change character in novel and unpredictable ways.” [5]

The ETC Group definition is focused only on the quantum nature of some nanoscale phenomena, accounting for only a part of the reality side of nanotechnology. The Greenpeace definition is very politically astute. It is interesting though that neither include in their definition anything about risk.

It will prove interesting to check the longevity of these definitions a decade from now. They are evidence of the shotgun approach the professional community is taking toward nanotechnology. There has been some backlash to this from some of the scientific community. One example is a recent Nature commentary by French nanoscientist Christian Joachim. He suggests defining nano as:

“Nanoscience should be reserved solely for the study of a single atom or a single molecule, that is, of one entity at a time, not for groups of such entities where statistics or interactions between them come into play.” [11]

This too narrow definition has good intentions. The ambiguity of the term nanotechnology sometimes attracts students with misperceptions about what they are getting into. Joachim suggests that a student might find themselves in a “physics of microelectronic devices” class instead of one on imagined nanomachines. Of course, one could argue that attracting students to take a physics class is good, misperception or not.

We will give a more targeted definition of what nanotechnology is today in the Conclusions of this chapter.

3. Perspectives from Science

Most nanotechnology definitions are cast in the language of application. This is good, since we are trying to describe a technology that does have applications now and in the future. Indeed, the language of application may be what unites all the subfields that make up nanotech and makes it a viable societal construct, though not necessarily a scientific one. Nevertheless, there are serious reasons not to lump too broad an array of advanced technologies under one banner. Education is one example. Patenting is another. Labeling inventions “nanotechnologies” sounds like a good idea for promoting interest and investment in a new and promising arena, but divorcing a patent from the specific field or principle it derives from may be devastating
to the patent process. Adequately judging prior work, quality, and worthiness may
depend crucially on expertise in a specific field. There exist no experts in nanotech-
nology, nor will there, as society has defined the term.

Here we want to comment on the “new”-ness of “new nanoscale phenomena,”
which is central to so many of the definitions of nanotech above. We also want to
provide some context for the scientific sub-fields that nanotechnology is starting to
claim, and show for a few examples where this is and is not worthwhile.

Condensed matter (CM) physics, formerly known as solid-state physics, is the
largest subfield of physics, the modern six being astrophysics, atomic physics, bio-
physics, condensed matter or solid-state physics, high energy physics, and nuclear
physics. CM physics deals with the study of matter that is condensed, including the
physics of metals, semiconductors, liquids, and emergent phenomena in many parti-
cle systems (such as superconductivity and the fractional quantum hall effect, atomic
gases, as well as “soft” systems as in polymer physics). CM seeks to understand the
non-classical behavior of everything, solid and liquid, that is cold (as compared to
the sun) from the scale of a few atoms to something you can hold in your hand. So
its purview is quite broad and overlaps very much with materials science and phys-
ical chemistry. In the CM community there is very little change of behavior due to
the introduction of the word nanotechnology except indirectly via funding programs.
Here, emphasis on applied and interdisciplinary work has become more prevalent in
recent years.

CM physics has ridden (and driven) the trend in the electronics industry towards
miniaturization to interesting discoveries at the nanoscale, including quantization of
electrical and thermal conductance and coulomb blockade in lateral quantum dots,
though often at extremely low temperatures. When CM physicists consider nanotech
they usually think of one of three things: (1) quantum size effects that emerge at the
nanoscale, (2) new surface physics/energetics/mechanics in nanoscale structures due
to the increased surface to volume ratio, and (3) quantum coherent transport effects
through mesoscopic systems. Largely these occur in metallic and crystallographic sys-
tems, the most common of the latter being silicon, gallium arsenide, and carbon.

There is also a history of studying extremely small life—“nanobiology”—but this
history has been largely forgotten. The nanobiotechnology we see today in electrical
engineering labs that are going “wet” is really a merger of hard inorganic device
(CM) physics being applied to biological matter.

### 3.1 Cold or Hot, Quantum or Not

One could easily fill an office with books on only one topic: silicon. Probably the
most economically important and well-studied material on earth, it has taken almost
100 years (starting with the discovery of quantum mechanics) to almost fully understand this simple periodic structure. (If it has taken this long to master what is basically a simpler version of sand, how long will it take to understand biological systems?) Consider an atom of silicon. It has discrete energy levels or “shells” where electrons can exist. As the atoms come together to form a crystal, the energy levels become energy bands, continuous as a function of the electron’s position within the crystal lattice. Since silicon is a semiconductor, its optical properties are largely governed by a band-gap that separates the valence (or core) electron bands from the conduction bands, where the electrons are much more mobile. Semiconductor technology is quantum, in some sense, as it requires a quantum mechanical description of matter to explain. The progression of a block of silicon crystal from the macroscopic scale down is a great way to understand some of the key manifestations of new nanoscale phenomena, but we could use many systems.

Nanotechnology proponents often bring up new quantum behavior that emerges at the nanoscale. There is nothing new, per se, about this phenomenon—being governed by well-known quantum physics—but only recently could matter be controlled well enough to reach this scale. There are two ways in general to see quantum effects: make a material colder or make it smaller. This is because the properties of materials, from atoms to molecules to metals and crystals, are usually governed by the interaction of electrons. And the properties of electrons are governed by the energy levels where they are allowed to exist. Crudely speaking, temperature blurs these energy levels so that electrons can hop from one level to another more easily. Here is the key point: if we make a material small enough, these energy levels will move farther apart, making this hopping due to temperature much more difficult. So, in some sense, making an object very small makes it “colder.” Other new properties appear as well, such as a change in the optical excitation response—as this is defined by the spacing of these energy levels as well. This phenomena, which happens to take place around the nanoscale (say 10 nm for silicon) is what is often referred to as “quantum size effects.”

### 3.2 The Nano in Nanoparticles

Consider nanoparticles, the nanotechnology of the present. These include nanocrystals (also called colloidal quantum dots in some cases) made of either metal or semiconductor as well as the carbon fullerene family, including buckyballs and nanotubes (the two other isotopes of carbon). Nanowires are classified under nanocrystals since they are solid objects as opposed to hollow nanotubes. All nanoparticles have at least two dimensions in the nanoscale regime, 1–100 nm. They have new properties that appear because of their size that would disappear in the micron regime. These “new properties” can be different in origin and even coincidental.
We categorize them broadly into three bins: (1) new quantum effects which appear due to confinement, which are often called quantum size effects; (2) new surface physics or much increased reactivity, usually attributed to increased surface to volume ratios—this might also include structural changes from adding nanoparticles to a material (like rubber) to make it stronger; and (3) new biological properties which are largely a coincidence of scale, as the body is not attuned to defend against nano-sized particles in many situations.

It would be easy to argue that nanoparticles or quantum dots are the only nanotechnology at this point in time. They certainly constitute most of the risk. Nanoparticles have large surface-to-volume ratios making them extremely reactive. They also are able to penetrate sensitive areas of the body and accumulate there. They can be embedded in larger substrates such as polymer systems or fabrics, which may decay over time and release these nanoparticles into the environment. They are also found in sunscreens and cosmetics. Nanoparticles and (eventually) active nanomachines bridge the materials to biology gap. Evolution in biomarkers for medical imaging and cancer treatment has tracked well with nanoparticle technology. But not all quantum dots are dangerous. For example, lateral quantum dots built in quantum well structures are no more toxic than the crystal they are in (gallium arsenide or silicon, as examples).

3.3 Not Nanotechnology

Technologies are often organized by scientific principle or concept, whereas nanotech can incorporate these just because of a coincidence in scales. Take the case of photonic band gap materials, which apply the theories historically developed for electronic waves in semiconductors to photon waves (light) traveling in periodic index of refraction materials—sometimes called meta-materials. Visible light photonic band gap materials will likely require nanoscale-manufactured devices (< 100 nm 3D structures) but work as well at microwave frequencies in order 1 cm sized structures. The physics does not change whether the features are 1 nm or 1 cm. So the only thing “nano” about some photonic meta-materials may be the fabrication process. Is there any point calling it nanotechnology? Would you send this patent application to be reviewed by the “nanotechnology” patent clerk or the physical optics patent clerk? It should be a no-brainer (the latter).

In our opinion, devices or structures with only one dimension at the nanometer length scale are in general not nanotechnology. Quantum wells are a good example. To make a quantum well you sandwich a thin semiconductor layer between two bigger semiconductor layers of a different type. This allows you to trap electrons in the inner layer. Physically this is not much different from inversion layers, where one material is grown on the other and an electric field traps electrons at
the interface. These kinds of semiconductor devices are well known. Electrons in these 2D sheets exhibit many new and interesting phenomena, often attributed to the nanoscale. But, new nanophenomena in this context could just as easily be called new quantum phenomena or new many-body phenomena. Superconductivity, quantum wells and electron gasses, giant magneto-resistance (GMR) are all examples of quantum behavior misappropriated to the nanofield. This also explains some of the ridiculous money expenditure estimates—in the tens of billions of dollars some of them—which have been touted as nanoinvestment. If you include all the investment in GMR hard drive technology, you get big numbers.

4. Conclusions

Much of nanotechnology is just the natural progression of scientific and technical trends began long ago. Whether it constitutes a real paradigm shift in how we approach the manipulation and utilization of nature is still an open question. We have taken a pragmatic approach by considering the social shaping of the term as it presently stands, introducing the vision and reality viewpoints of nanotechnology. From the perspective of the scientific community, the label nanotechnology does not seem to be doing any harm. Funding via this route has been relatively sane, perhaps better than it would have been otherwise. Certainly investment in the cross-disciplinary developments of the fundamental science in physics, chemistry, materials science, and biology has great promise. It may even be that there is some over-investment occurring. Many states and universities across the country are building nanotechnology centers without enough qualified people to fill them.

We are left with two basic questions to ponder. How should proponents—predominately government and industry—want nanotechnology to be understood by the public? And how should we define the term in the proper sense for risk assessment? In the sole context of risk, we should immediately define a very clear definition of nanotechnology:

Nanotechnology, at present, is nanoparticles and nanomaterials that contain nanoparticles. Nanoparticles are defined as objects or devices with at least two dimensions in the nanoscale regime (typically under 10 nm) that exhibit new properties, physical, chemical, or biological, or change the properties of a bulk material, due to their size. Nanotechnology of the future will include atom-by-atom or molecule-by-molecule built active devices.

It should be evident from the rest of our analysis that this definition is incomplete as compared to society’s definition of the word. (Also note that we explicitly enforce two dimensions being in the nanoscale regime.) But it has the salience of being rigorous and relevant to the questions of nanotechnology most directly impacting the
public. It might be beneficial from a public policy perspective to leave the science and technical communities alone with their developments—with their new transis-
tors and quantum dots and so on—instead redefining nanotechnology separately, in
the context only of direct and novel environmental and human impact. This would
have the benefit of isolating truly worrying nanotechnology-based products from the
bulk of nanoscience research, which is completely innocuous. With recent calls for a
moratorium on all nanotechnology by groups such as ETC, such a redefinition might
have real value.

While the breadth of phenomena that lie within the nanoscale regime points to the
ridiculousness of categorizing technologies based on size, the term nanotechnology
has become embedded in our society, which gives it meaning explicitly. Microtech-
nology, as an historical counterpoint, has referred generally to a specific set of tech-
niques and processes with little ambiguity. But what is defined as nanotechnology
today may not be nanotechnology tomorrow. Taniguchi was considering precision
manufacturing: separation, addition, or removal of materials at the atomic/molecular
scale. Indeed, the mechanical systems that lead Taniguchi to coin the term are now
called MEMS, microelectromechanical systems, and are considered by many today
not to be nanotechnology at all.

ACKNOWLEDGEMENTS

The author would like to thank Wendy Crone, Clark Miller, the other members of
the Wisconsin Nanotechnology and Society Initiative, and the students of his course
at the University of Wisconsin-Madison [24,25] for useful conversations. The author
is supported by a USA National Science Foundation Math and Physical Sciences
Distinguished International Postdoctoral Research Fellowship.

Appendix A: List of Some Technical Terms

*Spintronics*: Devices that use the electron spin instead of the electron charge for
electronics. The spin is a purely quantum property of the electron with two possible
states that may be used to encode information (up = 1, down = 0).

*Mesoscopics*: In physics and chemistry, the mesoscopic scale refers to the length
scale at which one can reasonably discuss the properties of a material or phenomenon
without having to discuss the behavior of individual atoms. For solids and liquids
this is typically a few to ten nanometers, and involves averaging over a few thou-
sand atoms or molecules. Hence, the mesoscopic scale is roughly identical to the
nanoscopic or nanotechnology scale for most solids. This region often exhibits inter-
esting quantum many-body phenomena, especially at cold temperatures.
Quantum computer: A proposed computer that would exploit the quantum mechanical nature of particles, such as electrons or atomic nuclei, to manipulate information as quantum bits, or qubits. Whereas an ordinary bit has at any time a value of either 0 or 1, a qubit can also take on both values at once. Many qubits can be “entangled” (the non-local property of quantum physics) for extended interactions. Because a quantum computer can act on these multiple states simultaneously in an exponential way, it is potentially many times as powerful as a conventional computer if a proper quantum algorithm is performed.

Quantum well: A quantum well is a potential well that confines particles in one dimension, forcing them to occupy a planar region, usually less than ten nanometers.

Quantum wire: In material science, a quantum wire is an electrically conducting wire, in which quantum transport effects are important. Typically, this is due to the small diameter (typically of the order of nanometers) of the wires. Nanowires are a particular example of a quantum wire.

‘Lateral’ quantum dot: A 0-dimensional trap formed inside a semiconductor structure, the natural progression of removing another dimension of electron movement from a quantum wire. Quantum dots can exhibit properties characteristic of atoms—electron energy shells for example—and are often referred to as artificial atoms.

Meta-material: In electromagnetism (covering areas like optics and photonics), a meta material (or metamaterial) is an object that gains its (electromagnetic) material properties from its structure rather than inheriting them directly from the materials it is composed of. This term is particularly used when the resulting material has properties not found in naturally formed substances.

Photonic band-gap material: A meta-material with alternating regions of dielectric constant which can modify, trap, or guide the transport of light.

Superconductivity: Property of particular metals at extremely low temperatures. The electrical resistance of a conductor becomes zero, so that an electric current can flow without loss. The superfluid is an analogous quantum many-body phenomena where the friction in a liquid goes to zero.

Transistor: A three terminal amplifying device, the fundamental component of most active electronic circuits, including digital electronics.

Leakage: In the context of semiconductor electronics, leakage refers to the tunneling of electrons outside of the barriers that define a transistor, for example. The result is usually heat and increased inefficiency.

‘Colloidal’ quantum dot or nanocrystal: Since the term emphasizes the quantum confinement effect it typically refers to the sub-class of nanocrystals that are small enough to exist in the quantum confinement regime, and more typically refers to fluorescent nanocrystals in the quantum confined size range.
Nanoparticles: Particles with controlled dimensions on the order of nanometers. Examples include colloidal gold, magnetite particles, and luminescent semiconductor aggregates that are also known as 'quantum dots.'

Carbon fullerenes and nanotubes: Any of various cage-like, hollow molecules composed of hexagonal and pentagonal groups of atoms, and especially those formed from carbon, that constitute the third form of carbon after diamond and graphite. Concentric shells of graphite formed by one sheet of conventional graphite rolled up into a cylinder. The lattice of carbon atoms remains continuous around the circumference. The nanotubes are designed to be very small in order to store hydrogen.

Photonics: The technology of transmission, control, and detection of light (photons). This is also known as fiber optics and optoelectronics.

Electrons, photons, phonons, magnons, plasmons, polaritons, ...: Examples of fundamental excitations of nature some which are actual particles and some of which are quasi-particles (particles that represent the excitations of a sea of more fundamental particles).

MEMS: Tiny mechanical devices that are built onto semiconductor chips and are measured in micrometers. They are used to make pressure, temperature, chemical and vibration sensors, light reflectors and switches as well as accelerometers for airbags, vehicle control, pacemakers and games.

REFERENCES


FURTHER READING

The Convergence of Nanotechnology, Policy, and Ethics

ERIK FISHER

Center for Nanotechnology in Society at Arizona State University
USA

Abstract
This chapter reviews the role ethical and societal issues associated with nanotechnology have played in the development of national, and particularly US nanotechnology policies. The prominence of nanotechnology as a matter of national policy is significant, as is the attention being afforded to ethical and societal considerations. Notably, there is an emphasis on the early anticipation of societal dimensions of nanotechnology and of collaborative socio-technical integration during research and development activities. While these policies are fairly unprecedented and pose considerable challenges to the societal and technical researchers to whom the tasks of early integration may fall, there is reason to believe that continued efforts aimed at their implementation are likely. The chapter provides a brief history and explanation of the US nanotechnology “ethics policy” in reference to the discourse and concerns motivating policy makers. It then surveys some of the growing body of literature emerging around what has been termed nanoethics. This literature includes a similar emphasis on early and collaborative anticipation of the ethical and societal implications surrounding nanotechnology. Finally, the chapter identifies a list of fundamental tasks that would be necessary to address for the sake of developing the capacity of social and technical researchers to effectively engage in socio-technical integration. These tasks include defining the scope of nanoscale science and engineering subject to the ethics policy, and the interactions, responsibility, participation, and regulation envisioned or implied by the ethics policy under consideration. A few emerging cases of socio-technical collaborations are noted, and key statements made by congressional witnesses and others are cited. Throughout the chapter, an attempt is made to frame the statements within conceptual considerations drawn from the diverse contexts of nanotechnology, policy, and ethics.
1. Introduction

Given the vast amount of differentiation in subject matter, methodologies, and envisioned applications that populate the research agendas and practices of those working at the nanoscale, a comprehensive yet meaningful definition of the term “nanotechnology”\footnote{For convenience—and in keeping with prevalent trends in scholarly and government literature—this term is here employed widely to include nanoscale science, engineering, and technologies.} is both elusive \cite{49} and contentious \cite{17,4}. The term, however, has already become a subject of national policy in numerous countries, both developed and developing \cite{82}. The inclusion of the term in the 2006 Presidential State of the Union address as one of “the most critical basic research programs in the physical sciences” \cite{13} gives some indication of the status nanotechnology has acquired as a matter of national interest in the US. From a national policy perspective, the term is associated with economic growth and competitiveness. Increasingly, nanotechnology is also fast becoming the locus for an abundance of ethical and societal concerns.

The creation of the US National Nanotechnology Initiative (NNI) in 2000 was, on the one hand, a response to an international “funding race” \cite{22} that at the time involved over 30 countries. On the other hand, it also seems to have accelerated the global effort: after 2000, worldwide funding for nanotechnology increased significantly, with several countries adopting national frameworks and programs. A Congressional mandate for the program in 2003 raised its annual budget to nearly one billion dollars and made it in the words of one US Senator “the single largest federally funded, multiagency scientific research initiative since the space program in the 1960s” \cite{1}. While policy makers around the world are investing in nanotechnology for economic advantages and competitiveness, many are also taking pains to avoid the unwelcome prospect of a “technological backfire” \cite{102}—the possibility that public mistrust of the emerging technology’s stewards could limit its envisioned commercial success and public funding base. As a result of concerns and over the uncertain socio-technical consequences of engineered nanoparticles, nanoenabled capabilities, and public perceptions thereof, various national nanotechnology programs
are promoting the largely unprecedented integration of broader societal considerations into the early stages of nanoscale research and development (R&D) activities.

In the US, one of the four formal goals of the NNI is to “support responsible development of nanotechnology” [65, p. 10]. Accordingly, in 2005 the US National Science Foundation (NSF) granted $14.3 million to establish a “series of initiatives,” including two Centers for Nanotechnology in Society, with “national goals” for “50-state outreach programs and stakeholder participation” [68]. NSF funding of numerous smaller projects dealing with broader societal considerations related to nanotechnology date back to at least 2002, and other US agencies such as the Environmental Protection Agency, the Food and Drug Administration, the National Institute for Occupational Safety and Health, and the National Institute for Standards and Technology are looking into environmental, health, and safety dimensions of nanotechnology. Elsewhere around the world, the European Commission has employed language of “responsible research, development and innovation in nanotechnology” [27, p. 5], and governments in the United Kingdom, the Netherlands and Belgium have instituted programs intended to influence the early stages of nanotechnological development trajectories in response to ethical and societal considerations. International organizations with a focus on responsible governance include the International Dialogue for Responsible Nanotechnology R&D [79] and the International Council for Risk Governance [45].

While the early attention being placed on the societal dimensions of nanotechnology and the budget allocations that follow in this name are significant, neither of these phenomena are particularly unique. Emerging technologies of the past have been marked by similar concerns as those now being associated with nanotechnology [56]. Moreover, funding for nanotechnology related societal research does not begin to approach the high watermark set by the Human Genome Project, which entailed ethical, legal, and societal implication research that was funded at 3–5% of the total project budget and that commenced virtually from the beginning of the program. What does make the “ethics policies” that are emerging for nanotechnology in various forms and settings potentially innovative, however, is their emphasis on functionally linking societal research, processes, and considerations with their scientific and technological counterparts. This emphasis is evident in the language used to describe such policies: Societal considerations pertaining to nanotechnology are intended, in the Netherlands, to “broaden the scope of strategic choices” [77, Appendix 2]; in the UK, to “shape scientific priorities” [100, p. 22]; and in the US, to

The term “ethics” is here includes explicitly normative (including societal, political, cultural, and philosophical) issues and the activities and approaches that seek to identify, analyze, and address them as such. The term “ethics policy,” coined in discussions with Carl Mitcham, denotes policy programs or mandates that include an explicit emphasis on issues, concepts, or methods of normative engagement as such.
“influence the direction of research” [44]. Such aims go beyond traditional risk management and product regulation, and hold implications for the broader governance of science and technology development processes themselves. Judging from government, scholarly, and civil sector publications, the goals include influencing research priority setting but also the actual conduct of research, development, and design of nanotechnologies.

In theory, to the extent that scientific, engineering, and technological development activities explicitly or implicitly involve ethical and societal considerations, they afford potential intervention points for ethics policy agendas. Depending upon what is interpreted as an ethical or a societal issue and the means used to address them, such ethics policies could go well beyond traditional scientific, engineering, and professional approaches to ethics and end-of-pipe regulatory approaches. In this respect, they suggest new roles for social scientists, ethicists, scientific researchers, and the public. They also hold implications for long standing de facto science policies, which have hitherto carefully protected scientific and technological endeavors from broader social attempts to control them. As is to be expected, the prospect of integrating ethical and societal research into nanotechnology policies and nanotechnology R&D activities faces significant challenges. For instance, existing ethics frameworks and approaches have been traditionally designed more in relation to identifying and addressing obvious and egregious violations of established norms and codes of conduct. Shaping technological development trajectories towards more socially robust platforms and systems, however, would require more granular fine tuning of research agendas and practices, which have hitherto been shielded from such attempts. De facto science policies have largely assumed the ethical neutrality of techno-scientific innovation and have relied for the most part on market forces to do this work.

As some have argued, traditional ethical frameworks and “rule based reasoning” [6] are not adequate to address macro-level and system-wide societal considerations during the emergence of new technologies. A significant challenge for new ethical and societal frameworks would be to acquire the ability to interface productively with the complex and evolving dynamics of R&D settings—which often move at a highly accelerated and unpredictable rate of change and are sheltered from broader societal considerations—without compromising research integrity or unnecessarily hampering its productivity. This chapter provides an account of the role that ethical and societal concerns have played in the history leading up to the creation of the NNI and the federal legislation that authorizes it. It surveys some of the growing body of scholarly literature on the ethics of nanotechnology, or “nanoethics,” and discusses some of the challenging tasks implied for developing a new capability for collaborative “socio-technical integration” during R&D.
2. Converging Paths

As has been argued [32], the fact that ethical and societal concerns have become a nanotechnology policy issue can be traced to two interrelated factors. On the one hand, a steadily escalating international funding race over the massive economic and other gains projected for nanotechnology products has been used to justify an aggressive US approach to promote rapid technological development and accelerated marketplace transfer. The NSF has estimated an economic impact for the nanotechnology market of over US $1 trillion by the year 2015 [80]. Meanwhile, international funding for nanotechnology increased seven times between 1997 and 2003, with a sharp increase in 2000 that corresponded to the creation of the NNI [78]. The global interest in nanotechnology reportedly caused US government officials to “worry that perhaps for the first time in recent memory, the United States does not have a clear advantage” in the “crucial” field of nanotechnology [43]. Concern about US competitiveness was voiced by others who were in a position to influence policy, and who advised Congress to act “aggressively” [55] in order to avoid economic and “militarily disastrous” [72] consequences.

On the other hand, heightened awareness of the role that public concerns and perceptions can play in the success and adoption of new technologies occasioned policy language intended to assure the successful realization of envisioned nanotechnology benefits, including legislation requiring research on societal concerns to be integrated into nanotechnology R&D. Speaking at an NSF workshop on the societal implications of nanotechnology, a US Department of Commerce administration official stated,

The body politic is susceptible to the virus of fear. When the public catches a public policy cold virus, their elected representatives sneeze. Our democratic institutions are designed to be responsive to the public. To keep technology moving forward, we must prevent fear from taking hold among the public [11].

Prior experiences with public perceptions of technologies and mistrust of regulatory systems, such as genetically modified (GM) organisms in the UK and nuclear power in the US, were cited frequently in policy contexts and used to justify the early consideration of ethical concerns associated with nanotechnology.

2.1 Nanotechnology and Ethics

Ethical concerns, goals, and dilemmas have attended nanotechnology from some of its earliest discursive stages. In 1986, K. Eric Drexler’s *Engines of Creation: The Coming Era of Nanotechnology* introduced the term into popular discourse, stimulating both hopes and fears about the prospect of a “new technology [that] will handle
individual atoms and molecules with control and precision” [20, p. 4]. While Engines offered vivid and, for many, compelling scientific, technological, and societal visions; as well as arguments for the inevitability and imperative of pursuing nanotechnology; it also depicted dangers associated with what could also become “engines of destruction” [20, p. 171]. Several of the conceptions that were presented in the book, both of the promise and the peril of nanotechnology, would resurface later in the popular press and in policy settings, including Congressional hearings. In addition to making a case for nanotechnology’s promotion, Drexler devoted two chapters of the book’s fifteen to the dark side of his subject and to counter-strategies for assuring human survival. One concern in particular—that masses of self-replicating molecular machines might be enabled to run amok, out-competing human beings by eventually consuming natural resources—had, apparently, already been christened the “gray goo problem” [20, p. 172]. The vivid (yet debatable) idea that humankind’s own nanotechnological machines could render themselves obsolete was destined to get further play in the future.

Drexler’s conception of nanotechnology, since termed “molecular manufacturing” and “molecular nanotechnology” (MNT), was more radical and longer-term than what some of the primary supporters of the NNI had in mind by the term. The nanotechnology that came to be publicly funded was by contrast “vastly broadened” [21, p. 21]. In spite of this, ideas derived from molecular manufacturing—from atomic precision to gray goo—were used to support efforts both to promote and, soon afterwards, to control more “mainstream” [49] conceptions of nanotechnology. For instance, a 1999 government sponsored brochure, Nanotechnology: Shaping the World Atom by Atom [66] “drew upon nanoutopian rhetoric” [56, p. 186] and “described and seemingly endorsed” an idea resembling MNT [73, p. 11]. Compelling visions, language, and promises used to advance nanotechnology research policies during the envisioning of the NNI were thus reminiscent of the imagery used to promote Drexlerian conceptions.

The employment of utopian (and dystopian) rhetoric and imagery to raise expectations (and fears) about an emerging technology such as nanotechnology illustrate the role that “imaginaries”—guiding images, symbols, expectations, and the like—can play in science policy. Accordingly, the scholarly attention on nanotechnology has included explicating the role that visions and images seem to be playing in its conception and promotion.³ In light of the power of language and imagery, scholars have called for an ethical “study of the imagination” [7, p. 636] and for assessments

³ There are indications that earlier nanotechnology beginnings that inspired Drexler may have had their own links to an underlying set of imagery. Several scholars have suggested Richard Feynman in his now celebrated talk “There’s Plenty of Room at the Bottom” [30], which is taken by many to be the inauguration of nanotechnology, “drew upon utopian visions proffered by contemporary fiction writers” ([56, p. 181]; see also [58]) such as Heinlein [41].
of the ethical and philosophical implications connected with its use in the case of nanotechnology [50,36].

While the use of provocative imagery may be largely connected to the promotional success enjoyed by nanotechnology, it may also help explain the prominent association it has come to have with ethical concerns and considerations. In April of 2000, Bill Joy published a popular treatment of the gray goo scenario in *Wired* magazine. Referencing Drexler’s work, Joy presented a much more pessimistic assessment of the fate of humans with regard to various scenarios he associated with the convergence of genetic engineering, robotics, and (an MNT conception of) nanotechnology. Joy’s article “provoked widespread discussion” and was seen by promoters as “a serious public relations problem” [73, p. 11]. Due largely to the influence of this article, which recommended “relinquishment” of nanotechnology research lines [46], discourse about the concerns associated with nanotechnology now proliferated alongside continued promotional efforts [5]. Within months, the NSF sponsored a workshop on the societal implications of nanotechnology and released a lengthy report a year later [80]. Meanwhile, promoters who had originally helped create the NNI reportedly sought to distance federally funded nanotechnology from more radical versions of nanotechnology in an attempt to “exclude” them due to its association with scenarios such as gray goo [21, p. 21]. As one MNT proponent relates, “People following the NNI (in 2002) knew where it was headed and that it tried to avoid MNT-related topics” [73, p. 12].

Despite such tactics, numerous developments over the course of the next three years intensified the focus on the gathering cloud of ethical issues and concerns surrounding nanotechnology. Popular science fiction novels about nanotechnology proliferated and explored dystopian MNT-inspired futures. Most notably, Michael Crichton’s *Prey* [16], was widely cited and discussed and was explicitly referred to in Congressional hearings. In 2003, reports by non-government activist organizations such as Greenpeace [2] and the ETC Group [24] prominently raised concerns about the potential environmental and societal downsides of nanotechnology. The ETC Group had played a role in supporting public resistance to GM crops, and it now called for an international moratorium on nanoparticle research [24]. Perhaps as a result of the ETC report, remarks by the Prince of Wales led to renewed references to gray goo [25]. Meanwhile, newspaper reports [28,29] raised environmental and health concerns associated with engineered nanoparticles and researchers [9,15] considered inconclusive results. Increasingly, the climate was shifting, as indicated by opinion pieces such as the following one published in *Nature*:

> For emerging areas of science and technology, public acceptance is a vital precursor to sustaining their development. Scientists and industrialists working with genetically modified (GM) foods have learned this the hard way. There are now signs on both sides of the Atlantic that funders of nanotechnology research are
starting to take the issue of public acceptance seriously and learn from others’ mistakes [69].

In the midst of such statements and developments, the US Congress was preparing to authorize for the first time the NNI as a publicly mandated program.

2.2 Nanotechnology Ethics Policy

In 2002, a review of the NNI by the National Research Council expressed criticism of the program for inadequately addressing societal implications [67]. In 2003, the US Congress held three separate hearings on the two nanotechnology bills respectively being considered in the Senate and the House of Representatives. Explicit references to ethical and societal considerations of nanotechnology occurred in all three of these hearings, and one was devoted exclusively to the topic.

Witnesses for this hearing before the House Science Committee were asked to comment on several questions, including how well the House bill addressed the recommendations made by the NRC review of the NNI. As stated, this report was critical of the attention paid to societal considerations. The argument presented in the report for the importance of addressing these considerations was essentially a commercial one: in order to minimize or avoid the kind of public resistance that was experienced in Europe and elsewhere in response to GM agriculture, considerations of this nature ought to be explicitly raised and addressed. The House and Senate bills, although differing in approach, both required attention to societal considerations. While the Senate bill required an American Nanotechnology Preparedness Center, the House was focused on integrating societal considerations into nanotechnology research. A report in Science characterized the House approach:

In the House, some lawmakers pushed to replicate the approach taken by the US Human Genome Project, which earmarks 5% of its spending for [social, economic, and environmental impacts] studies. But the House Science Committee rejected that path, opting instead to have nanotechnology researchers incorporate societal studies into their technical work [86, p. 27].

In the legislation that ensued, the 21st Century Nanotechnology Research and Development Act of 2003 prescribed the Preparedness Center but also mandated that “societal concerns” be “considered during the development of nanotechnology” [97]. Furthermore, the Act designated various means by which societal concerns should be considered—including, “insofar as possible, integrating research on societal, ethical, and environmental concerns with nanotechnology research and development” [97]. As one House staffer reportedly stated, “the ideal outcome is to integrate social concerns directly into the technology development process” [86, p. 27]. The intended purpose of the integration of societal and technical research is explained by the House Science Committee report that accompanied the legislation:
The Committee stresses the importance of integrating research on environmental, societal, and ethical implications with nanotechnology research and development programs to ensure that the results of the research influences [sic] the direction of ongoing nanotechnology research and development of commercial applications [44].

The “integration” approach prescribed by the Act’s ethics policy thus seeks to mitigate or avoid negative public reactions by early anticipation of potential concerns which, in turn, is intended to shape nanotechnology R&D directions in response to these concerns in order to build or maintain public trust. Rather than relying on market mechanisms or on traditional regulatory frameworks, this policy of sociotechnical integration relies more heavily on scientists and engineers to work in tandem with ethical and societal research programs.

On December 3, 2003, US President Bush signed the 21st Century Nanotechnology Research and Development Act of 2003 into public law [97, P.L. 108–153]. In keeping with one of the primary justifications for promoting federal funding of nanotechnology, the Act emphasized international competitiveness, prescribing “United States global leadership in the development and application of nanotechnology” and “accelerating the deployment and application of nanotechnology research and development in the private sector” [97]. Yet, as noted above, it also prescribed that “societal concerns” be “consider during” and “integrated with” nanotechnology development. As argued elsewhere, these two prescriptions are somewhat in tension with one another, in that they assume differing models of technology development [32]. The tension between the two approaches can be characterized as that between “rapid” and “responsible” nanotechnology development. While rapid development aims largely at achieving economic gains, responsible development aims largely at avoiding economic losses.

Efforts to implement this dual-focused policy in the US have been mixed,5 but are likely to continue, and could intensify, since the pressures driving the policy—

4 Given that responsibility is a problematic term, as argued below, a somewhat more accurate term than “responsible” development is perhaps “responsive” nanotechnology development.

5 In the summer of 2004, the NSF announced a program to fund a Center for Nanotechnology in Society (CNS). In December 2004, the NNIs Strategic Plan identified supporting “responsible development of nanotechnology” as a key goal [65]. In August 2005, two Centers for Nanotechnology in Society were established, one at Arizona State University and one at the University of California at Santa Barbara. Congressional hearings on the safety of nanotechnology were held in November 2005 and January 2006. In May 2006, the NNI held a public workshop on the public participation in nanotechnology. In August 2006, the Food and Drug Administration (FDA) established a task force to “evaluate possible adverse health effects from FDA-regulated products that use nanotechnology materials” [35] and in October 2006 held a public meeting on the topic. In November 2006, the Environmental Protection Agency introduced the first nanospecific regulation in the US.
namely, concerns over the uncertain impacts of the emerging field of nanotechnology and its widespread and rapid pursuit—are not likely to abate in the near future. A public opinion survey conducted a few months after the Act’s passage suggested a correlation between lack of trust and heightened perception of risk [14], and a second survey appeared to confirm this finding [91]. In May of 2004, SwissRe, the world’s second largest reinsurance firm, published a report that recommended a precautionary approach to nanotechnology in order to avoid risk problems such as those which occurred with asbestos [94]. In the United Kingdom, similar concerns as those raised in the US over the public reception of nanotechnology had occasioned early and integrative approaches to prospective public reaction [101]; in July 2004, Royal Society and the Royal Academy of Engineering published a report that recommended public “upstream” engagement of societal considerations pertaining to nanotechnology and recommended taking a “precautionary” approach to nanoparticles [81, p. 5, 6]. Also in 2004, the European Commission mandated “An effective two-way dialogue... whereby the general public’s views are taken into account and may be seen to influence decisions concerning R&D policy” [26].

3. From Convergence to Collaboration

As shown above, there has been an increasingly tighter coupling of scientific, ethical, and governance issues in the case of nanotechnology. While much of the convergence of these issue domains takes place at the level of public policy discourses, there are also implications for the way scientific and ethical research is conceived and conducted. The 21st Century Nanotechnology Research and Development Act’s provision for integrating social and natural science research, however, represents a legal conjunction of ethics and R&D that goes beyond traditional approaches to ethics. Traditional approaches, such as professional codes of conduct and the responsible conduct of research, are aimed at more clearly demarcated norms such as avoiding plagiarism and ethical dilemmas such as whistle-blowing. The goals of influencing and shaping ongoing research and development trajectories and their evolution [44] by means of integrating societal and technical research [97], however, implies a more nuanced and granular approach to bringing ethical and societal considerations to bear on scientific activities. In many cases, this seems to require innovative collaborations in which ethicists and social scientists interact with scientists and engineers.

In fact, the call for social and natural scientists to “work hand in hand” in order that “problems may be solved as they arise, and so that many of them may not arise in the first instance” goes back to a policy report written by the economist John R. Steelman [88] prior to the founding of the NSF. As Guston and Sarewitz [40] point out, this notion was originally articulated in a National Research Council report.
They also observe that, while the NSF did eventually come to include a role for the social sciences, this role has been limited to that of providing “tools for monitoring the science enterprise—but not assessing societal implications” [40, p. 95]. The early inclusion during R&D of social and ethical perspectives and expertise in order to enhance the social benefits of technology has been more recently argued in terms of the co-production [40] of technological outcomes and the co-responsibility [60] of science and society with respect to these outcomes. In the case of nanotechnology, calls for new collaborative models among natural scientists, engineers, social scientists, and ethicists have been made by several educators and scholars of science and technology [3,38,50,76].

One such re-conceptualization of the roles of ethicists and scientists working at the nanoscale was presented before the Senate Committee on Commerce, Science and Transportation in 2003. Baird’s Congressional testimony stressed a collaborative role for ethical researchers in relationship with the research that goes on in the laboratory:

> Productive work on societal implications needs to be engaged with the research from the start. Ethicists need to go into the lab to understand what’s possible. Scientists and engineers need to engage with humanists to start thinking about this aspect of their work. Students need training now that will take their understanding of nanotechnology from laboratory to society. These students today, trained in the right interdisciplinary setting, will become a cadre of scientists, engineers and scholars used to working together, thinking about the societal and technical problems side-by side. Only thus, working together in dialog, will we make genuine progress on the societal and ethical issues that nanotechnology poses [3].

The notion of such socio-technical collaboration in nanotechnology R&D contexts is being implemented in several ongoing “experiments.” For instance, several case studies involving university nanoscale science and engineering research projects have already been initiated or completed [19,33,37,38,48,65,93]. And in the private sector, companies such as DuPont and Unilever have partnered with the non-governmental organizations Environmental Defense and Demos, respectively [54,52]. Sweeney suggests including among the “grand challenges” for nanotechnology that of “encouraging prospective and current nanotechnology researchers to engage—in a thoughtful and critical manner—with [ethical and societal] issues as an integral part of their research endeavors” [93, p. 442].

---

6 A panel devoted to this topic was featured during the 2006 annual meeting of the Society for the Social Studies of Science in Vancouver, British Columbia, Canada.
3.1 Collaboration Tasks

Societal and ethical concerns about nanotechnology may, in many cases, turn out not to be unique to nanotechnology [6,53,56]. As Wynne implies, however, this observation is not in itself sufficient grounds for waiting to address such concerns until only after research trajectories and products have been largely stabilized and completed, or for dismissing them altogether, as has occurred in the past [103]. Further, government reports [67,81], Congressional witnesses [102,70], and societal researchers [14,63] have all implied that such reasoning would most likely exacerbate public risk perceptions and lack of trust in regulatory regimes. Rather, while public risk perceptions can be widely inaccurate [85,92], anticipating and responding to potential societal concerns, whether “real” or “perceptual” [70], is simply a matter of prudent policy.

Insofar as early anticipation and response to emerging concerns is to be done by collaborative integration during R&D, several challenging tasks can be identified. These include defining the scope of nanotechnology research, and the nature of ethical inquiry, responsibility, participation, and regulation.7

3.1.1 Scope of Nanotechnology Research

A key task is to define the scope of nanotechnological research that may be covered by the cited ethics policy. As commentators have pointed out, the definition of nanotechnology holds implications for the societal discussion and governance of it [96,99]. For, the definition of nanotechnology that is employed, or assumed, will influence the identification of ethical issues to be addressed [50]. Yet, as numerous more commentators have pointed out, defining nanotechnology is no easy task [62]. For one thing, use of the term has encouraged the blurring of various boundaries, including those between scientific disciplines [49], top–down and bottom–up methodologies [93], and basic and applied science [56]. While advances in microscopy opened up fundamentally new capabilities for imaging and manipulating matter at the atomic scale and led to discoveries of novel properties of matter at the nanoscale, these “revolutionary” developments and their implications may be more the exception than the rule. For instance, “mainstream nanotechnology” can be seen largely as an extension of the chemistry and material science that had already been going on for much of the twentieth century [49]. On the other hand, MNT supporters distinguish as “near term” the nanotechnology currently being funded by public programs from what they envision as “long term” nanotechnology [72]. Similarly, another permeable boundary seems to be that between what counts as science and what counts as science fiction [4,58].

7 These tasks are somewhat related to the broader call for what Nordmann has termed “disentanglement” of conceptual and social “entanglements” in the “nanoworld” [71].
An example of the attempt to clarify definitional questions, and the related compounded difficulties, is Khushf’s attempt to mark out three definitional approaches to nanotechnology. In addition to MNT, he characterizes a second approach as “a grab bag of unrelated research” [50]. This approach is in part a result of what several commentator have observed as a strategy to acquire or secure research funding (e.g., [73,90]). As one observer puts it, “nanotechnology became a buzzword that some say was just a trendy rebranding of chemistry” [23]. From another perspective, given the trend in the microelectronics industry towards miniaturization and closing in on the gap between microscale and nanoscale worlds, much of the commercial interests driving nanotechnology descends from envisioned improvements in computational power as a result of smaller and smaller components and processes. Thus, definitions of nanotechnology can often include microscale components such as microelectromechanical systems (MEMS).

In distinguishing a third approach, Khushf describes the (now abandoned) NNI definition of nanotechnology as “the middle way” between MNT and the “grab bag” phenomenon. The NNI definitional approach Khushf had in mind was distinct from that of the other two and more focused than Keiper’s “mainstream” nanotechnology in that it includes the notion of exploiting novel properties of matter at the nanoscale. As Khushf notes, this NNI approach was more of a characterization than a definition, for it sought to develop scientific and engineering capacities in directions that are not entirely known. Thus, while the “old” NNI definition may have been more focused, it was still broad enough to encompass a vast range of techniques, approaches, and applications.

3.1.2 Scope of Socio-Technical Inquiry and Interactions

Defining the scope of ethical and societal inquiry for nanotechnology will no doubt be the subject of ongoing discussion. For instance, the NNI divides “responsible development of nanotechnology” into two categories: “environment, health, and safety” [EHS] implications,” and “ethical, legal, and all other societal issues” [65, p. 10]. With respect to EHS implications, the President of Lux Research has suggested “Two Distinct, Equally Important Classes of Risk Impact Nanotech” [70, p. 2]. These are “real risks” [70, p. 2] and “perceptual risks” [70, p. 4]. A growing

---

8 This “old” NNI definition of nanotechnology included three criteria: operating at the nanoscale, exploiting “novel” properties associated with that scale, and the ability to control or manipulate matter at that scale [64]. The new NNI definition reads “Nanotechnology is the understanding and control of matter at dimensions of roughly 1 to 100 nm, where unique phenomena enable novel applications” which one commentator has likened to “the usual ‘anything 1–100 nm’ nanotech definition used by so many” [74].

9 Others viewed the (old) NNI definition as similar to the “grab bag” approach in that it includes “any technology with nanoscale features” [21, p. 21].
body of scholarship explores broader issues. Lewenstein takes up the question explicitly in an article simply entitled, “What counts as a ‘social and ethical issue’ in nanotechnology?” [53].

Delineating the relationship between specific ethical and societal issues and nanoscale scientific and engineering activities represents a perhaps equally important task. Given the ethical and societal uncertainties associated with emerging nanotechnologies, science and engineering ethics can be said to entail what Mitcham has described as the injunction of plus respicere, namely, to “take more into account” [59] in order to minimize undesirable socio-technical consequences. Or, as Stephan puts it,

> The technical side of engineering is often about *isolating* a system from the rest of the world and *simplifying* it enough so it can be modeled and predicted with sufficient accuracy. But in an important sense, engineering ethics is about *context* and *complexity*: the context of how hardware or software will be used, by whom, and for what, and the complexity of how technology’s effects can spread in ways that its designers never imagined [89, p. 3].

In relation to nanotechnology, Bernes’ call for “an ethics for nanoscale science and technology” [6, p. 10] seeks to include the dimension of ultimate ends or purposes within the plus respicere injunction. Berne states that “conventional rule-based, prescriptive engineering codes and guidelines are necessary but insufficient for the ethical development of nanotechnology” [6, p. 10]. For while necessary, “professional codes cannot address the deeper philosophical questions” of purpose and beliefs behind the purposes [6, p. 13]. Elsewhere, Berne distinguishes three dimensions of nanoethics [8, pp. 79–92]. The first dimension includes explicit rules, standards, protocols, and the like. Second dimension nanoethics include “moral claims that are negotiable and subject to change under the vagaries of will, power, and perception” [8, p. 82]; while third dimension nanoethics are about metaethical claims about purpose, belief and meaning [8, p. 88].

While deep philosophical reflection may be an end in itself, R&D contexts will perhaps rarely admit of opportunities to set or change course as a result of sustained inquiry and deep reflection. In light of this, others have proposed an incremental approach that scales to the routine decision-making of actual R&D activities [34,83]. These authors suggest that opportunities for reflection ought to be cultivated, but that they need to respect operative conditions and dynamics if openings for integration throughout R&D processes are to be identified and capitalized upon. Sweeney articulates this pragmatic view as well by focusing on the relation between the “tangible technological applications” of the work of scientists and engineers in relation

---

10 Mitcham explicitly discusses engineering design research in this article. In the case of the anticipatory uncertainties that an emerging technology like nanotechnology entails, Mitcham’s observations can arguably be extended to include science and engineering activities more generally.
to “influences that development of nanotechnology research have on society and vice versa” [93, p. 440]. Thus, explicitly conceptualizing the scope of scientific and engineering research activities that would be affected by integration efforts represents another essential task.

Along these lines, recent developments in the philosophy of technology recommend an “empirical turn” [51] that takes in sociological studies of engineering design activities (e.g., [10,12]) into account in order to be effective. Van de Poel and Verbeek provide a similar approach in combining science and technology studies and engineering ethics [98]. These developments emphasize the practical and empirical context of R&D dynamics—a context that must be taken into account by any attempts to integrate ethical and technical research.

3.1.3 Scope of Responsibility

Closely related is question of how the notion of responsibility is to be conceptualized. As Rip’s description of the Dutch approach to assessing the societal implications of nanotechnology states,

‘responsible innovation’ is not just a matter of an innovation actor changing his (or her?) ways. Actors are part of larger patterns which drive (i.e. enable and constrain) them, in a sense govern them. The analysis of such patterns and articulation of such understanding with innovation actors is important to make actions more reflective, even if this may not be sufficient to change the patterns [77].

While traditional ethical frameworks focus on individual agency and responsibility, in R&D contexts both of these terms (agency and responsibility) take on a distributed sense and therefore cannot easily and comprehensively be located at the individual level. For instance, while intentionality can be located within individuals and groups, intentions often fail to lead to intended results, both because of the trial and error nature of much of R&D and also because different techno-scientific actors may have different objects of conception in mind for the same artifact or project (e.g., [12]). The challenge of defining responsibility is compounded by the generic nature of many nanotechnologies insofar as they are intended as enabling technologies to broadly support the design and function of numerous and more developed materials, systems, processes, devices, and applications that are “downstream” of R&D activities themselves.

The uncertainty regarding the locus of responsibility is thus partly a consequence of the limits of predictive certainty and partly a consequence of the complex and hierarchical nature of technological development. Responsibility for socio-technical integration then, in theory, may need to be located at multiple levels and throughout distributed networks of interacting agents and patterns. This notion simultaneously would increase the role of institutions and technology administration processes in
“macroethical” deliberations [42] and the role of individuals and small groups in building more “reflexive awareness” of the ongoing processes of interaction and decision-making within which they are embedded [33]. Preliminary results have indicated that “reflexive awareness” of engineering research decisions at the nanoscale can lead to positive impacts for addressing EHS considerations and also for enhancing engineering research effectiveness [33].

3.1.4 Scope of Participation

One of the explanations for the Congressional emphasis on integration is that previous research programs lacked mechanisms for linking social and ethical intelligence to decisions making in a timely manner and in order to have a practical effect. Among ethics policy precedents, the Human Genome Project’s (HGP) Ethical, Legal, and Societal Implications (ELSI) program is most visible. This was the first “self-critical federal science” program [47], and it has received some attention in connection with nanotechnology policy. Commenced at the beginning of the HGP and funded at a level of 3–5% of the Program’s budgets, although the ELSI program’s stated purpose was to “forestall adverse effects” [84], the program had been widely criticized for failing to accomplish its mission, both because it lacked mechanisms for feeding into R&D and policy, and also because it placed little emphasis on overseeing the practical impact of its outputs [31].

One of the witnesses during the April, 2003 House Science Committee’s hearing on the societal implications of nanotechnology alluded to the ELSI program, criticizing it as a potential model for nanoethics:

I would not advise you to pass a Nanoethicist Full Employment Act, sponsoring the creation of a new profession. Although the new academic research in this area would be of some value, there is also a tendency for those who conduct research about the ethical dimensions of emerging technology to gravitate toward the more comfortable, even trivial questions involved, avoiding issues that might become a focus of conflict. The professional field of bioethics, for example (which might become, alas, a model for nanoethics), has a great deal to say about many fascinating things, but people in this profession rarely say “no” [102].

As a remedy, Winner suggested tempering expert views and research with real world contextualization—a sort of stakeholder review intended to incorporate wider feedback into research, design, and policy processes—in order to keep meaningful public policy issues on the research agenda:

One way to avoid the drift toward moral and political triviality is to encourage social scientists and philosophers to present their findings in forums in which people from business, the laboratories, environmental organizations, churches, and other groups can join the discussion.
In addition, however, to tempering the ethics research program in such a way, Winner also recommended directly involving average citizens as a different body of intelligence altogether drawn from the “ultimate stakeholders—the general public”:

I believe Congress should seek to create ways in which small panels of ordinary, disinterested citizens, selected in much the way that we now choose juries in cases of law, be assembled to examine important societal issues about nanotechnology. The panels would study relevant documents, hear expert testimony from those doing the research, listen to arguments about technical applications and consequences presented by various sides, deliberate on their findings, and write reports offering policy advice.

Winner’s suggestion is in line with a growing movement to “democratize” science and in particular, nanotechnology [39]. Nanotechnology is already characterized by a high degree of interdisciplinarity. Training a new cadre of ethicists and social scientists who are competent to work “hand in hand” [88] or “side by side” [3] with scientists and engineers will introduce notable changes in science education, training, and conduct. Public participation in nanoscale science and engineering would add yet another twist and would most likely need to occur outside the lab and closer to policy making processes than R&D processes. Nevertheless, projects that promote the public engagement of nanotechnology have emphasized a role for scientists in face to face deliberations (e.g., [48]).

3.1.5 Scope of Regulation

The problem of effectively anticipating and addressing in “real-time” [40] social and ethical implications of nanotechnology is compounded by the institutionalized practice of regulating technologies either at the beginning or ending stages of their development cycles, while largely ignoring the development process itself. Early regulation often takes the form of “yes or no” funding decisions and priorities, and “end of the pipe” regulation is often too late to be fully effective. Efforts to assess emerging technology during R&D make sense in terms of cost and effectiveness. For instance, “technology assessment is most effective when applied in the early stages of R&D, when changes are easier” [95]. This type of approach, however, is orthogonal to the “linear model” of innovation that is used to explain and justify many science policy decisions [75]. The linear model is based on the assumptions that techno-scientific activities are ethically neutral and that research funding inevitably lead to development, diffusion, and social benefits (largely in the form of economic performance). While the linear model has been rhetorically denounced by scholars and policy makers alike, it continues to act as “America’s [and other nations] de facto science policy” [57]. For instance, a Canadian report on science and technology indicators observes that the linear model
is so highly abstract that it does not really explain the complexities of innovation in real world. Yet it still informs many policy discussion. This fact alone has led economic historian Nathan Rosenberg (1991)\(^{11}\) to say [the linear model] “is dead, but it won’t lie down” [87, p. 29].

On the other hand, more closely integrating the work of social and natural scientists appears to be in line with a visible policy trend of “enforcing linearity” that consists of more and more social intervention and governance strategies [61].

The tasks presented above stem from the overall focus of this chapter on nanotechnology ethics policy questions of how social and ethical issues might be considered and addressed by social scientists and ethicists working with scientists and engineers. The tasks presented above are hardly meant to be all inclusive. Instead, they represent an attempt to characterize some of the uncertainty that arises when the notion of socio-technical integration emerges in interdisciplinary discussions.

4. Conclusion

From the standpoint of its promoters, there is much at stake in the heavy investments being made in nanotechnology R&D, including the state of science and technology funding and education, economic growth, medical benefits, and international competition. Yet, as several sources discussed have stated or implied, without effective strategies to address societal considerations of nanotechnology, it is likely that public support of nanotechnology could go the way of GM crops in Europe and nuclear power in the US. The large public investments and intended rapid pace of nanotechnology development may in fact condense problems, whether real or perceived, at an accordingly rapid rate. Thus, with much at stake—including goals touching on special, regional, national, and collective interests—policy makers can be expected to continue to take steps intended to assure that nanotechnology objectives not miscarry or come up short. The anticipation of negative public perceptions has led policy makers to prescribe an ethics policy for nanotechnology that implies new capabilities for ethical researchers and research programs and that holds implications as well as for nanoscale science and engineering researchers.

While it is impossible to know in advance which ethical or societal issues may indeed warrant concern, US nanotechnology legislation signals an intention to engage such possibilities at early and ongoing stages. This intention appears strengthened by an explicit focus on “responsible nanotechnology development” both within the US and internationally. However, given several challenges, including the generic quality

of nanotechnologies as enabling technologies and the capacity of ethics and societal research to interface practically with R&D contexts and dynamics, it is clear that new capacities will be needed if implementation of the ethics policy described herein is to be effective.

REFERENCES


[40] Giles J., “What is there to fear from something so small?”, *Nature* 426 (December 18/25, 2003) 750.


Author Index

Numbers in *italics* indicate the pages on which complete references are given.

A

Abdallah, S.D., 284, 294
Achtert, V., 51, 95
Adamson, D.H., 50, 94
Adleman, L., 8, 35
Aguirre, A.H., 19, 36
Ah, C.S., 82, 100
Agrawal, B., 188, 245
Aime, J., 56, 96
Akturk, A., 213, 214, 247
Alam, M.A., 178, 183, 244, 245
Albalak, R.J., 50, 95
Albert, K.J., 107, 127, 140, 153, 160, 162, 163
Alder, B.J., 206, 246
Alekso, A., 83, 100
Allara, D.L., 48, 94
Allen, G., 274, 291
Allen, T.W., 237, 249
Allison, D.P., 77, 99
Aluru, N.R., 219, 220, 228, 230, 248
Alvarez, M., 80, 99
Amato, I., 254, 270
Anantram, M.P., 210, 215–217, 247
Andersen, O.K., 207, 246
Anderson, E., 194, 245
Appenzeller, J., 172, 208, 214, 216, 243, 247
Åqvist, J., 233, 249
Aranawa, H., 77, 99
Arenadas, R.C., 106, 122, 160
Arenda, J.H., 176, 243
Arnall, A.H., 279, 291
Arnold, C., 130, 143, 152, 154, 162
Aron, S., 15, 36
Artaki, M., 187, 245
Asano, K., 194, 245
Ashcroft, F.M., 221, 248
Ashley, S., 9, 35
Ashwell, G.J., 197, 245
Astumian, R.D., 9, 35
Austin, M.D., 61, 97
Aviram, A., 197, 198, 245, 246
Avouris, P., 172, 208, 214, 216, 243, 247
Awschalom, D.D., 172, 243

B

Baccarani, G., 174, 178, 213, 243, 244
Bachilo, S.M., 209, 247
Backhouse, C.J., 283, 295
Bacquet, R.J., 233, 249
Baecker, A., 118, 120, 121, 161
Baert, K., 83, 100
Bailey, R.E., 65, 97
Bailey, T.C., 61, 97
Bainbridge, W.S., 277, 279, 295
Baird, D., 283, 289, 291
Baker, N.A., 233, 249
Bakken, G.A., 127, 162
Bakoglu, H.B., 173, 243
Baldwin, J.W., 197, 245
Ball, M.K., 77, 78, 99
Balzano, L., 209, 247
Bamberger, E., 77, 99
Banerjee, S.K., 194, 217, 245, 247

297
<table>
<thead>
<tr>
<th>Author</th>
<th>Page Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Banoo, K.</td>
<td>189, 245</td>
</tr>
<tr>
<td>Bao, Y.P.</td>
<td>85, 101</td>
</tr>
<tr>
<td>Baragiola, R.A.</td>
<td>49, 94</td>
</tr>
<tr>
<td>Barcilon, V.J.</td>
<td>222, 248</td>
</tr>
<tr>
<td>Bartlett, P.N.</td>
<td>79, 99</td>
</tr>
<tr>
<td>Bath, J.</td>
<td>91, 101</td>
</tr>
<tr>
<td>Battiston, F.M.</td>
<td>77, 99</td>
</tr>
<tr>
<td>Bau, H.H.</td>
<td>67, 98</td>
</tr>
<tr>
<td>Baum, R.</td>
<td>274, 284, 291</td>
</tr>
<tr>
<td>Becke, A.D.</td>
<td>206, 246</td>
</tr>
<tr>
<td>Becker, H.</td>
<td>63, 97</td>
</tr>
<tr>
<td>Beckers, R.</td>
<td>15, 36</td>
</tr>
<tr>
<td>Bek, S.</td>
<td>237, 249</td>
</tr>
<tr>
<td>Bell, D.J.</td>
<td>69, 98</td>
</tr>
<tr>
<td>Belluscio, L.</td>
<td>106, 160</td>
</tr>
<tr>
<td>Beni, G.</td>
<td>16, 36</td>
</tr>
<tr>
<td>Bennett, I.</td>
<td>279, 291</td>
</tr>
<tr>
<td>Bentley, P.J.</td>
<td>11, 13, 19, 23, 32, 33, 35–37</td>
</tr>
<tr>
<td>Berendsen, H.J.C.</td>
<td>228, 248</td>
</tr>
<tr>
<td>Bergman, A.A.</td>
<td>49, 94</td>
</tr>
<tr>
<td>Berne, R.W.</td>
<td>276, 278, 284, 286, 291</td>
</tr>
<tr>
<td>Bernèche, S.</td>
<td>228, 248</td>
</tr>
<tr>
<td>Berry, L.L.</td>
<td>45, 93</td>
</tr>
<tr>
<td>Besenbacher, F.</td>
<td>6, 34</td>
</tr>
<tr>
<td>Besteman, K.</td>
<td>81, 100</td>
</tr>
<tr>
<td>Bethune, D.S.</td>
<td>58, 96</td>
</tr>
<tr>
<td>Bhadriraju, K.</td>
<td>80, 100</td>
</tr>
<tr>
<td>Bhalla, N.</td>
<td>32, 33, 35</td>
</tr>
<tr>
<td>BIA—Berufsgenossenschaftliches Institut für Arbeitsschutz, 279, 291</td>
<td></td>
</tr>
<tr>
<td>Biegel, B.</td>
<td>215, 216, 247</td>
</tr>
<tr>
<td>Biel, L.</td>
<td>127, 162</td>
</tr>
<tr>
<td>Biggin, P.C.</td>
<td>226, 230, 248</td>
</tr>
<tr>
<td>Bijkers, W.</td>
<td>287, 291</td>
</tr>
<tr>
<td>Bilings, A.</td>
<td>46, 94</td>
</tr>
<tr>
<td>Blackwell, T.</td>
<td>19, 36</td>
</tr>
<tr>
<td>Blanco, A.</td>
<td>64, 97</td>
</tr>
<tr>
<td>Biotekjaer, K.</td>
<td>182, 183, 244</td>
</tr>
<tr>
<td>Boardman, A.D.</td>
<td>174, 180, 243</td>
</tr>
<tr>
<td>Bocek, D.</td>
<td>77, 99</td>
</tr>
<tr>
<td>Bodenhoeler, K.</td>
<td>137, 163</td>
</tr>
<tr>
<td>Boilot, P.</td>
<td>126, 133, 162</td>
</tr>
<tr>
<td>Boisen, A.</td>
<td>80, 99</td>
</tr>
<tr>
<td>Boit, C.</td>
<td>45, 93</td>
</tr>
<tr>
<td>Böker, A.</td>
<td>49, 51, 94, 95</td>
</tr>
<tr>
<td>Bokor, J.</td>
<td>194, 245</td>
</tr>
<tr>
<td>Bonabeau, B.</td>
<td>12, 14–17, 23, 35</td>
</tr>
<tr>
<td>Bonabeau, E.</td>
<td>15, 36</td>
</tr>
<tr>
<td>Bond, P.J.</td>
<td>277, 291</td>
</tr>
<tr>
<td>Bonsma, E.</td>
<td>12, 35</td>
</tr>
<tr>
<td>Bossus, A.</td>
<td>82, 100</td>
</tr>
<tr>
<td>Bouhacina, T.</td>
<td>56, 96</td>
</tr>
<tr>
<td>Bourillot, E.</td>
<td>83, 101</td>
</tr>
<tr>
<td>Bowen, R.C.</td>
<td>177, 243, 244</td>
</tr>
<tr>
<td>Boykin, T.B.</td>
<td>177, 243, 244</td>
</tr>
<tr>
<td>Bradley, J.-C.</td>
<td>67, 98</td>
</tr>
<tr>
<td>Bradley, K.</td>
<td>218, 248</td>
</tr>
<tr>
<td>Bras, W.</td>
<td>50, 95</td>
</tr>
<tr>
<td>Brault, R.G.</td>
<td>45, 93</td>
</tr>
<tr>
<td>Braunstein, P.</td>
<td>83, 101</td>
</tr>
<tr>
<td>Breed, J.</td>
<td>230, 249</td>
</tr>
<tr>
<td>Brehmer, L.</td>
<td>197, 245</td>
</tr>
<tr>
<td>Brews, J.R.</td>
<td>168, 242</td>
</tr>
<tr>
<td>Brezger, B.</td>
<td>54, 55, 95, 96</td>
</tr>
<tr>
<td>Briggs, S.</td>
<td>212, 247</td>
</tr>
<tr>
<td>Brink, M.</td>
<td>218, 247</td>
</tr>
<tr>
<td>Brinke, G.</td>
<td>50, 95</td>
</tr>
<tr>
<td>Brooks, F.P.</td>
<td>3, 5, 34</td>
</tr>
<tr>
<td>Bruecka, S.R.J.</td>
<td>64, 65, 97</td>
</tr>
<tr>
<td>Bryson, S.T.</td>
<td>3, 26, 34</td>
</tr>
<tr>
<td>Bucciarelli, L.L.</td>
<td>287, 291</td>
</tr>
<tr>
<td>Buiks, J.</td>
<td>49, 94</td>
</tr>
<tr>
<td>Buot, F.A.</td>
<td>176, 177, 243</td>
</tr>
<tr>
<td>Burgin, T.P.</td>
<td>197–199, 245</td>
</tr>
<tr>
<td>Burke, P.J.</td>
<td>213, 247</td>
</tr>
<tr>
<td>Bush, G.W.</td>
<td>274, 291</td>
</tr>
<tr>
<td>Butt, H.-J.</td>
<td>77, 99</td>
</tr>
<tr>
<td>Byerly Jr., R.</td>
<td>289, 295</td>
</tr>
</tbody>
</table>

C

Cahay, M., 176, 243
Cai, W., 66, 98
Cajigas, J.C.e.a., 139, 163
Calleja, M., 80, 99
Camazine, S., 15, 36
Campbell, P.M., 6, 34
Cantello, M., 46, 94
Cappella, B., 77, 99
Carceller, J.E., 176, 243
Cartujo, P., 176, 243
Castillo, P., 46, 94
Catacuzzeno, L., 235, 249
Catalano, R.K., 283, 293
Cava, M.P., 197, 245
Cehvary, J.A., 206, 246  
Celotta, R.J., 53, 55, 95, 96  
Ceperley, D.M., 206, 246  
Chacham, H., 206, 246  
Ceperley, D.M., 206, 246  
Chang, S., 66, 98  
Chapline, M.G., 144–146, 164, 218, 247  
Chau, L., 66, 98  
Chauhan, S., 82, 100  
Chemical Industry, S.o., 139, 163  
Chen, B., 197, 245  
Chen, C., 66, 98  
Chen, C.S., 80, 100  
Chen, D., 182, 184, 187, 233, 245, 249  
Chen, G.Y., 77, 99  
Chen, H.Y., 50, 81, 95, 100  
Chen, J., 198, 246  
Chen, P., 50, 94  
Chen, W., 82, 100  
Chen, W.R., 118, 161  
Chen, Y., 155, 166  
Chen, Z.R., 50, 95  
Cheng, G., 81, 100  
Cheng, S., 66, 98  
Cherkaoui-Malki, M., 83, 101  
Chettiar, U.K., 66, 98  
Chi, V.L., 3, 5, 34  
Chien, R.D., 63, 97  
Childs, P.A., 178, 244  
Chiu, S.W., 237, 249  
Cho, K., 218, 219, 247, 248  
Chomski, E., 64, 97  
Chopra, S., 144, 165  
Chou, S.Y., 61, 63, 97  
Chou, T.W., 144, 145, 164  
Chretien, L., 15, 36  
Chris Wang, C.R., 66, 98  
Chrisey, L.A., 77, 99  
Christensen, T.A., 118, 119, 161  
Chu, S., 53, 95  
Chu, S.D., 45, 93  
Chun, K., 43, 93  
Chung, K.Y., 53, 95  
Chung, S.-H., 233, 235–237, 249  
Ciarlo, D.R., 127, 162  
Classen, M., 180, 244  
Clark, W.M., 45, 93  
Clarke, J., 6, 34  
Clary, G.J., 3, 5, 34  
Cleveland, J.P., 77, 99  
Coalson, R.D., 235, 249  
Cobb, M.D., 282, 284, 291  
Coe, C.A., 19, 36  
Coglan, A., 135, 163  
Cohen, R.E., 207, 246  
Colbert, D.T., 58, 96  
Cole, J., 222, 248  
Collins, P.G., 144, 145, 164, 218, 248  
Colomi, A., 17, 36  
Colton, R.J., 77, 99  
Colvin, V.L., 256, 270, 279, 291  
Comini, E., 144, 148, 150, 165  
Committee for the Review of the National Nanotechnology Initiative, 280, 284, 294  
Committee on Technology, Nanoscale Science, Engineering and Technology Subcommittee, 275, 281, 283, 285, 294  
Connell, J.O., 224, 248  
Cook, R., 182, 245  
Cooper, K.E., 237, 249  
Corbisier, P., 82, 100  
Cordova, L.E., 206, 207, 246  
Corelli, J.C., 45, 93  
Corry, B., 233, 235–237, 249  
Court, E.B., 274, 295  
Crane, C., 58, 96  
Crespo-Biel, O., 69, 98  
Crichton, M., 279, 291  
Crichton, W., 256, 270  
Crone, W.C., 268, 271  
Cui, Y., 144, 148, 151–153, 164, 165  
Cwik, T.A., 177, 244  
Cyrano Sciences, Inc., 133, 162  

D
Daar, A.S., 274, 295
Daggett, V., 230, 248
Dai, H., 144–146, 164, 165, 217, 218, 247, 248
Dai, Z.R., 144, 148, 150, 165
Datta, S., 177, 203, 205–207, 213, 214, 216, 217, 244, 246, 247
Dauksher, W.J., 61, 97
Davenport, R.J., 140, 163
Davies, J.C., 274, 292
Davis, M.E., 233, 249
Davis, R.C., 218, 247
De, V.K., 188, 245
De Beeck, M., 83, 100
De La Mothe, J., 290, 292
de Pablo, J.J., 50, 94
decher, G., 69, 98
Dehmelt, H., 53, 95
Dekker, C., 81, 100
DeKoven, B.M., 56, 96
Delaurens, F., 178, 213, 244
Demarest, J.J., 49, 94
dique, J.-L., 15, 36
dequesnes, M., 219, 248
dereux, A., 83, 101
derosa, P.A., 206, 207, 246
DeRouchey, J., 51, 95
derycke, V., 172, 208, 214, 216, 243, 247
deshpande, M.R., 197, 245
Deshpande, V., 67, 98
detrain, C., 15, 36
devries, M.S., 58, 96
di Carlo, A., 179, 208, 217, 218, 244, 246–248
di Caro, G., 17, 36
di Francesco, D., 222, 248
di Ventra, M., 199, 246
dickey, M.D., 61, 97
dickinson, T.A., 139, 140, 143, 163
diLello, N., 67, 98
division of Engineering and Physical Sciences, 280, 284, 294
dolan, G.J., 8, 34
dolbya, L, 50, 95
doelman, B.J., 133, 162
dong, L.X., 69, 98
dordi, B., 69, 98
dorigo, M., 12, 14–17, 23, 26, 27, 35, 36
doubleday, R., 283, 292
drachev, V.P., 66, 98
drees, R.C., 72, 98
dresselhaus, G., 145, 165, 209, 247
dresselhaus, M.S., 145, 165, 209, 247
drewsen, M., 54, 95
drexler, K.E., 2, 4, 5, 9, 27, 34, 35, 253, 270, 278, 279, 285, 292
driankov, D., 127, 162
driel, H.M., 64, 97
droodsky, U., 54, 95
dubey, R.S., 87, 101
ducker, W.A., 77, 99
duff, D.G., 27, 35
dufour, P.R., 290, 292
duncan, A., 180, 244
dupree, S.A., 238, 249
dutton, R.W., 173, 182, 184, 243, 245
E
eastwood, J., 180, 238, 244
eberhart, R., 12, 14–16, 18, 20, 35, 36
ebersberger, B., 45, 93
ebert, W., 83, 100
egashira, K., 46, 94
ehrlich, E.E., 51, 95
egler, D.M., 5, 34, 255, 270
Eisenberg, R.S., 222, 228, 233, 235, 237, 240, 248, 249
Ekerdt, J.G., 61, 97
ekinci, K.L., 43, 93
eklund, P.C., 145, 165
elbs, H., 50, 51, 94, 95
eilibol, O.H., 144, 164
elkaakour, Z., 56, 96
ellenbogen, J.C., 152, 165, 198, 246
ellison, K.D., 268, 271
elsner, J., 207, 246
elsner, M., 207, 246
emlen Jr., J.T., 18, 36
endy, D., 285, 292
engel, A., 77, 99
engelbrecht, A.P., 19, 36
engelhardt, H., 237, 249
erie, D.A., 77, 99
etc, 263, 270
etc group, 279, 292
European commission, 275, 282, 292
Evoy, S., 67, 98, 144, 165
F
facello, M., 72, 98
Fahmy, T.M., 91, 102
falvo, M.R., 3, 5, 34
fan, X., 58, 96
fan, X.F., 194, 245
fang, Y., 81, 100
AUTHOR INDEX

Fawcett, W., 174, 180, 243
FD.A, 262, 270
Feder, B.J., 279, 292
Feller, S.E., 228, 248
Fendler, K., 77, 99
Ferry, D.K., 176, 177, 190, 191, 243–245
Feynman, R.P., 8, 35, 252, 270, 278, 292
Finot, E., 83, 101
Fiolhais, C., 206, 246
Firestein, S., 106, 110, 111, 114–116, 122, 160, 161
Fischer, S., 46, 94
Fischetti, M.V., 180, 181, 190, 194, 244, 245
Fisher, E., 277, 281, 283, 286, 288, 292
Flory, P.J., 49, 94
Fodor, S.P., 40, 93
Fong, P.M., 91, 102
Food and Drug Administration, 281, 292
Forrest, L.R., 228, 248
Foss, C.A., 66, 97
Frauenheim, Th., 207, 246
Friedrich, R.W., 106, 118–120, 160
Fritz, J., 78, 99, 137, 138, 163
Frohman, R., 118, 290, 292, 294
Fujita, J., 55, 96
Fukuyama, Y., 19, 36
Fulton, T.A., 8, 34

G
Gaberlein, S., 82, 100
Gain, C., 46, 94
Galick, A.T., 176, 216, 243, 247
Galizia, C.G., 118, 161
Gambardella, L.M., 17, 36
Gámiz, F., 176, 194, 243, 245
Gao, G.T., 219, 220, 230, 248
Gao, Y.F., 50, 95
Gardner, C.L., 177, 184, 244, 245
Gardner, J., 79, 99
Garg, A., 218, 248
Gary, D.S., 80, 100
Ge, H., 61, 97
Geng, H., 66, 98
Geohagan, D.B., 58, 96
Gerber, Ch., 77, 78, 99
Gershenson, N., 8, 35
Gestri, G., 106, 127, 160
Gheorghe, M., 208, 217, 218, 246, 248
Giles, J., 283, 292
Gillespie, D., 235, 237, 249
Ginzewski, J.K., 77, 78, 99, 198, 246
Gingeras, T.R., 40, 93
Glass, R.S., 127, 162
Glasman, G., 111, 143, 161
Gnudi, A., 178, 213, 244
Goasguen, S., 177, 244
Godoy, A., 176, 243
Gogotsi, Y., 67, 98
Goh, M.C., 56, 96
Goldberger, J., 144, 164
Goldenetz, A., 45, 93
Goldie, K., 77, 99
Goldman, S., 233, 249
Goldman, N., 178, 184, 211–214, 244, 245, 247
Goldwater, D., 26, 27, 35
Golubinelli, P., 183, 245
Goodnick, S., 179, 244
Gore, J., 213, 218, 247
Gorman, G., 58, 96
Gorman, J.F., 283, 293
Gornik, E., 179, 244
Goschnick, J., 130, 143, 152, 154, 162
Goss, S., 15, 17, 36
Gourdon, A., 6, 34
Govindan, T.R., 176, 215, 216, 243, 247
Govindanraj, A., 64, 97
Goyal, A., 91, 102
Gozdz, A.S., 43, 93
Grabchak, S., 64, 97
Gray, C.G., 233, 249
Green, J.E., 155, 166
Green, S.J., 91, 101
Griffiths, S., 26, 27, 35
Grimes, C.A., 143, 144, 163
Gross, R., 26, 27, 35
Grothaus, J.T., 50, 95
Groves, M.E., 283, 293
Grubin, H.L., 176, 243
Grzybowski, B., 26, 32, 35
Gu, J., 63, 97
Gunn, J.B., 170, 174, 242
Güntherodt, H.-J., 77, 78, 99
Guo, J., 63, 97
Guo, T., 58, 96
Gupta, R., 55, 96
Guston, D., 282, 283, 289, 293

H

Ha, D.H., 82, 100
Hagleitner, C., 142, 163
Hailer, B., 67, 98
Haines, D.E., 106, 110, 111, 114–117, 125, 143, 159
Hajnal, Z., 207, 246
Hall, J.S., 3, 12, 34
Hall, L.E., 205, 246
Hallstrom, P., 72, 98
Ham, Y.M., 43, 93
Hamadeh, H., 45, 93
Hamaguchi, C., 181, 182, 188, 190, 244, 245
Han, C.C., 50, 95
Han, Y., 69, 98
Han, Y.W., 91, 101
Hänni, W., 45, 93
Hänisch, T.W., 53, 95
Hänisch, W., 194, 245
Hänßel, H., 51, 95
Hansma, P.K., 77, 99
Hara, M., 144, 165
Harada, Y., 91, 101
Harmey, A., 224, 248
Harms, M., 130, 143, 152, 154, 162
Harison, C., 50, 94
Hartl, W., 180, 244
Hashimoto, H., 71, 98
Hasler, P., 173, 180, 243
Haugk, M., 207, 246
Hauptmann, J.R., 208, 247
Hayashi, M., 91, 101
He, P., 81, 100
He, X., 50, 94
Heath, J.R., 155, 166
Hedge, R., 171, 242
Heeger, A.J., 172, 242
Heering, H.A., 81, 100
Heim, U., 63, 97
Heinlein, R.A., 278, 293
Heinze, S., 216, 247
Henderson, J.I., 203, 205, 246
Henrickson, L., 178, 244
Henze, C.E., 3, 26, 34
Heppner, F., 18, 36
Herbig, J., 49, 94
Herkert, J., 288, 293
Hermans, L., 83, 100
Herrera, J.E., 209, 247
Hertel, T., 208, 247
Hess, K., 168, 173, 174, 176, 180, 238, 242–244
Highes, T.V., 197, 245
Higman, J.M., 181, 244
Hille, B., 172, 243
Hinchliffe, A., 205, 246
Hines, C., 277, 293
Hirao, K., 64, 97
Hirschberg, M., 230, 248
Hirvensalo, M., 172, 243
Hisamatsu, D., 194, 245
Hishida, T., 91, 101
Hla, S.W., 7, 34
Hockney, R., 180, 238, 244
Hodgkin, A.L., 222, 248
Hoffman, R., 205, 246
Hoffmann, H., 45, 93
Holland, O.E., 13, 15, 35, 36
Holliday, C., 283, 293
Holst, M.J., 233, 249
Hong, J., 40, 92
Hong, S., 49, 94
Hong, W., 77, 99
Honig, B., 233, 249
Hood, E., 256, 270
Höpfner, U., 197, 245
Horiguchi, S., 71, 98
Hornyak, G.L., 66, 97
Hosch, C., 197, 245
House Committee on Science, 276, 281, 282, 293
<table>
<thead>
<tr>
<th>Name</th>
<th>Page Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hoyles, M.</td>
<td>237, 249</td>
</tr>
<tr>
<td>Hu, C.</td>
<td>173, 194, 243, 245</td>
</tr>
<tr>
<td>Hu, Z.</td>
<td>81, 100</td>
</tr>
<tr>
<td>Huang, A.</td>
<td>77, 99</td>
</tr>
<tr>
<td>Huang, C.C.</td>
<td>43, 93</td>
</tr>
<tr>
<td>Huang, W.</td>
<td>69, 98</td>
</tr>
<tr>
<td>Huber, M.</td>
<td>85, 101</td>
</tr>
<tr>
<td>Häckel, E.P.</td>
<td>204, 246</td>
</tr>
<tr>
<td>Hueschen, M.R.</td>
<td>174, 243</td>
</tr>
<tr>
<td>Hull, R.</td>
<td>49, 94</td>
</tr>
<tr>
<td>Hummer, G.</td>
<td>219, 220, 230, 248</td>
</tr>
<tr>
<td>Hunt, W.D.</td>
<td>137, 138, 163</td>
</tr>
<tr>
<td>Hush, N.S.</td>
<td>205, 246</td>
</tr>
<tr>
<td>Huskens, J.</td>
<td>69, 98</td>
</tr>
<tr>
<td>Huseini, G.A.</td>
<td>218, 247</td>
</tr>
<tr>
<td>Huw Arnall, A.</td>
<td>262, 270</td>
</tr>
<tr>
<td>Huxley, A.F.</td>
<td>222, 248</td>
</tr>
<tr>
<td>Huyberechts, G.</td>
<td>83, 100</td>
</tr>
<tr>
<td>Hyonchol, K.</td>
<td>77, 99</td>
</tr>
<tr>
<td>Ichihashi, T.</td>
<td>58, 96</td>
</tr>
<tr>
<td>ICGRG—International Risk Governance Council</td>
<td>275, 293</td>
</tr>
<tr>
<td>Iijima, S.</td>
<td>7, 34, 58, 96, 209, 247</td>
</tr>
<tr>
<td>Ikai, A.</td>
<td>77, 99</td>
</tr>
<tr>
<td>Ikkala, O.</td>
<td>50, 95</td>
</tr>
<tr>
<td>Illumina, I.</td>
<td>141, 163</td>
</tr>
<tr>
<td>Im, W.</td>
<td>228, 237, 248, 249</td>
</tr>
<tr>
<td>Interagency Working Group on Nanoscience, Engineering and Technology</td>
<td>278, 294</td>
</tr>
<tr>
<td>Ishigami, M.</td>
<td>218, 248</td>
</tr>
<tr>
<td>Ito, A.</td>
<td>86, 101</td>
</tr>
<tr>
<td>Iwasaki, H.</td>
<td>91, 101</td>
</tr>
<tr>
<td>Jabbour, Z.J.</td>
<td>55, 96</td>
</tr>
<tr>
<td>Jack, J.J.B.</td>
<td>222, 248</td>
</tr>
<tr>
<td>Jackson, K.A.</td>
<td>206, 246</td>
</tr>
<tr>
<td>Jacob, T.</td>
<td>108, 112, 124, 160</td>
</tr>
<tr>
<td>Jacoboni, C.</td>
<td>180, 238, 244</td>
</tr>
<tr>
<td>Jacobs, P.</td>
<td>83, 100</td>
</tr>
<tr>
<td>Jacobson, J.</td>
<td>26, 27, 35</td>
</tr>
<tr>
<td>Jaeger, H.M.</td>
<td>51, 95</td>
</tr>
<tr>
<td>Jakobsson, E.</td>
<td>219, 220, 230, 237, 248, 249</td>
</tr>
<tr>
<td>Jakumeit, J.</td>
<td>179, 180, 244</td>
</tr>
<tr>
<td>Jamieson, A.T.</td>
<td>61, 97</td>
</tr>
<tr>
<td>Janata, J.</td>
<td>127, 129, 132, 134, 135, 162</td>
</tr>
<tr>
<td>Jang, Y.T.</td>
<td>144, 145, 164</td>
</tr>
<tr>
<td>Javey, A.</td>
<td>217, 218, 247</td>
</tr>
<tr>
<td>Jeffersis, G.S.X.E.</td>
<td>106, 117, 160</td>
</tr>
<tr>
<td>Jensen, A.</td>
<td>208, 247</td>
</tr>
<tr>
<td>Jensen, K.L.</td>
<td>176, 177, 243</td>
</tr>
<tr>
<td>Jensen, M.O.</td>
<td>224, 248</td>
</tr>
<tr>
<td>Jeong, U.</td>
<td>50, 94</td>
</tr>
<tr>
<td>Jepsen, O.</td>
<td>207, 246</td>
</tr>
<tr>
<td>Jerome, J.J.</td>
<td>184, 245</td>
</tr>
<tr>
<td>Jiang, D.</td>
<td>66, 98</td>
</tr>
<tr>
<td>Jiang, P.</td>
<td>6, 34</td>
</tr>
<tr>
<td>Jin, S.</td>
<td>143, 163</td>
</tr>
<tr>
<td>Joachim, C.</td>
<td>6, 34, 254, 263, 270</td>
</tr>
<tr>
<td>Joachim, C.</td>
<td>198, 246</td>
</tr>
<tr>
<td>John, S.</td>
<td>64, 97</td>
</tr>
<tr>
<td>Johnson, R.C.</td>
<td>148, 165</td>
</tr>
<tr>
<td>Johnson, S.C.</td>
<td>61, 97</td>
</tr>
<tr>
<td>Jones II, L.</td>
<td>197, 245</td>
</tr>
<tr>
<td>Jordan, P.</td>
<td>236, 249</td>
</tr>
<tr>
<td>Jordan, P.C.</td>
<td>233, 249</td>
</tr>
<tr>
<td>Joseph, S.</td>
<td>219, 220, 230, 248</td>
</tr>
<tr>
<td>Jovanovic, D.</td>
<td>216, 247</td>
</tr>
<tr>
<td>Joy, B.</td>
<td>256, 271, 279, 293</td>
</tr>
<tr>
<td>Juengst, E.T.</td>
<td>288, 293</td>
</tr>
<tr>
<td>Jullens, R.A.</td>
<td>45, 93</td>
</tr>
<tr>
<td>Jung, G.-Y.</td>
<td>155, 157, 166</td>
</tr>
<tr>
<td>Jung, S.K.</td>
<td>82, 100</td>
</tr>
<tr>
<td>Jungnickel, G.</td>
<td>207, 246</td>
</tr>
<tr>
<td>Jurs, P.C.</td>
<td>127, 162</td>
</tr>
<tr>
<td>Kaewkamnerpong, B.</td>
<td>13, 23, 35</td>
</tr>
<tr>
<td>Kaler, K.V.I.S.</td>
<td>283, 295</td>
</tr>
<tr>
<td>Kamakura, Y.</td>
<td>181, 182, 244</td>
</tr>
<tr>
<td>Kamimura, A.</td>
<td>27, 35</td>
</tr>
<tr>
<td>Kan, E.C.</td>
<td>182, 184, 187, 245</td>
</tr>
<tr>
<td>Kaneva, I.</td>
<td>82, 100</td>
</tr>
<tr>
<td>Kannangara, K.</td>
<td>58, 96</td>
</tr>
<tr>
<td>Kappl, M.</td>
<td>77, 99</td>
</tr>
<tr>
<td>Karan Caulfield, T.</td>
<td>283, 295</td>
</tr>
<tr>
<td>Karim, A.</td>
<td>50, 95</td>
</tr>
<tr>
<td>Karplus, M.</td>
<td>230, 231, 248</td>
</tr>
<tr>
<td>Karshikoff, A.</td>
<td>237, 249</td>
</tr>
<tr>
<td>Karunatillake, N.C.</td>
<td>12, 35</td>
</tr>
<tr>
<td>Kaschner, R.</td>
<td>207, 246</td>
</tr>
<tr>
<td>Name</td>
<td>Pages</td>
</tr>
<tr>
<td>-----------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Kastalsky, A.</td>
<td>174, 243</td>
</tr>
<tr>
<td>Kathawala, G.</td>
<td>240, 241, 249</td>
</tr>
<tr>
<td>Kauer, J.S.</td>
<td>118, 122, 161</td>
</tr>
<tr>
<td>Kawai, T.</td>
<td>197, 245</td>
</tr>
<tr>
<td>Kaxiras, E.</td>
<td>207, 246</td>
</tr>
<tr>
<td>Kay, L.M.</td>
<td>118, 120, 161</td>
</tr>
<tr>
<td>Kearnes, M.</td>
<td>283, 289, 293</td>
</tr>
<tr>
<td>Kedzierski, J.</td>
<td>194, 245</td>
</tr>
<tr>
<td>Keimel, C.</td>
<td>63, 97</td>
</tr>
<tr>
<td>Keiper, A.</td>
<td>274, 278, 284, 293</td>
</tr>
<tr>
<td>Keller, A.</td>
<td>50, 95</td>
</tr>
<tr>
<td>Keller, P.E.</td>
<td>126, 161</td>
</tr>
<tr>
<td>Kennedy, J.</td>
<td>12, 14–16, 18, 20, 35, 36</td>
</tr>
<tr>
<td>Kerkhoven, T.</td>
<td>176, 187, 243, 245</td>
</tr>
<tr>
<td>Kermitt, M.</td>
<td>127, 162</td>
</tr>
<tr>
<td>Kerr, I.D.</td>
<td>230, 249</td>
</tr>
<tr>
<td>Kho, D.H.</td>
<td>50, 94</td>
</tr>
<tr>
<td>Khushf, G.</td>
<td>279, 283–285, 293</td>
</tr>
<tr>
<td>Kiang, C.H.</td>
<td>58, 96</td>
</tr>
<tr>
<td>Kildishev, A.V.</td>
<td>66, 98</td>
</tr>
<tr>
<td>Kim, D.</td>
<td>50, 95</td>
</tr>
<tr>
<td>Kim, E.</td>
<td>80, 100</td>
</tr>
<tr>
<td>Kim, E.K.</td>
<td>61, 97</td>
</tr>
<tr>
<td>Kim, F.</td>
<td>66, 98</td>
</tr>
<tr>
<td>Kim, H.</td>
<td>50, 95</td>
</tr>
<tr>
<td>Kim, J.K.</td>
<td>50, 94</td>
</tr>
<tr>
<td>Kim, S.H.</td>
<td>43, 93</td>
</tr>
<tr>
<td>Kim, W.</td>
<td>82, 100</td>
</tr>
<tr>
<td>King, T.-J.</td>
<td>194, 245</td>
</tr>
<tr>
<td>Kini, A.D.</td>
<td>106, 122, 160</td>
</tr>
<tr>
<td>Kinloch, I.A.</td>
<td>58, 96</td>
</tr>
<tr>
<td>Kiricsi, I.</td>
<td>60, 96</td>
</tr>
<tr>
<td>Kisimoto, T.</td>
<td>55, 96</td>
</tr>
<tr>
<td>Kitade, S.</td>
<td>50, 95</td>
</tr>
<tr>
<td>Kizilyali, I.C.</td>
<td>187, 245</td>
</tr>
<tr>
<td>Klein, D.L.</td>
<td>6, 34</td>
</tr>
<tr>
<td>Klimeck, G.</td>
<td>177, 243, 244</td>
</tr>
<tr>
<td>Kluksdahl, N.C.</td>
<td>176, 177, 243</td>
</tr>
<tr>
<td>Knebel, D.</td>
<td>77, 99</td>
</tr>
<tr>
<td>Knoch, J.</td>
<td>172, 243</td>
</tr>
<tr>
<td>Knoll, A.</td>
<td>51, 95</td>
</tr>
<tr>
<td>Koerber, R.</td>
<td>130, 154, 162</td>
</tr>
<tr>
<td>Koga, K.</td>
<td>219, 220, 230, 248</td>
</tr>
<tr>
<td>Köhler, Th.</td>
<td>207, 246</td>
</tr>
<tr>
<td>Kohn, E.</td>
<td>83, 100</td>
</tr>
<tr>
<td>Kohn, W.</td>
<td>206, 246</td>
</tr>
<tr>
<td>Kokaji, S.</td>
<td>27, 35</td>
</tr>
<tr>
<td>Kolmakov, A.</td>
<td>144, 148, 152–154, 164</td>
</tr>
<tr>
<td>Kolmogorov, A.N.</td>
<td>152, 165</td>
</tr>
<tr>
<td>Kometer, K.</td>
<td>179, 244</td>
</tr>
<tr>
<td>Kondi, K.</td>
<td>86, 101</td>
</tr>
<tr>
<td>Kong, J.</td>
<td>144–146, 164, 218, 247, 248</td>
</tr>
<tr>
<td>Konya, Z.</td>
<td>60, 96</td>
</tr>
<tr>
<td>Kornfield, J.A.</td>
<td>50, 95</td>
</tr>
<tr>
<td>Kosaki, M.</td>
<td>198, 199, 246</td>
</tr>
<tr>
<td>Koster, G.F.</td>
<td>207, 246</td>
</tr>
<tr>
<td>Koumanov, A.</td>
<td>237, 249</td>
</tr>
<tr>
<td>Kraus, G.</td>
<td>50, 51, 94, 95</td>
</tr>
<tr>
<td>Krauss, P.R.</td>
<td>61, 97</td>
</tr>
<tr>
<td>Kreskovsky, J.P.</td>
<td>176, 243</td>
</tr>
<tr>
<td>Krizman, A.M.</td>
<td>176, 177, 243</td>
</tr>
<tr>
<td>Kroes, P.</td>
<td>287, 293</td>
</tr>
<tr>
<td>Krupke, R.</td>
<td>152, 165</td>
</tr>
<tr>
<td>Krupp, F.</td>
<td>283, 293</td>
</tr>
<tr>
<td>Kubena, R.L.</td>
<td>45, 93</td>
</tr>
<tr>
<td>Kubiak, C.P.</td>
<td>203, 205, 246</td>
</tr>
<tr>
<td>Kuekes, P.J.</td>
<td>155, 166</td>
</tr>
<tr>
<td>Kumar, S.</td>
<td>11, 35</td>
</tr>
<tr>
<td>Kunikiyo, T.</td>
<td>181, 182, 244</td>
</tr>
<tr>
<td>Kuo, C.</td>
<td>194, 245</td>
</tr>
<tr>
<td>Kurnikova, M.G.</td>
<td>235, 249</td>
</tr>
<tr>
<td>Kurokawa, H.</td>
<td>27, 35</td>
</tr>
<tr>
<td>Kusterer, J.</td>
<td>83, 100</td>
</tr>
<tr>
<td>Kuyucak, S.</td>
<td>233, 235–237, 249</td>
</tr>
<tr>
<td>Kwok, K.S.</td>
<td>198, 246</td>
</tr>
</tbody>
</table>

**L**

<table>
<thead>
<tr>
<th>Name</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lachaud, J.-P.</td>
<td>17, 36</td>
</tr>
<tr>
<td>Lacroute, Y.</td>
<td>83, 101</td>
</tr>
<tr>
<td>Laegsgaard, E.</td>
<td>6, 34</td>
</tr>
<tr>
<td>Laidig, K.E.</td>
<td>230, 248</td>
</tr>
<tr>
<td>Lakshmikantham, M.V.</td>
<td>197, 245</td>
</tr>
<tr>
<td>Lang, H.P.</td>
<td>77, 78, 99, 137, 138, 163</td>
</tr>
<tr>
<td>Lang, N.D.</td>
<td>199, 246</td>
</tr>
<tr>
<td>Lansdorp, B.M.</td>
<td>80, 100</td>
</tr>
<tr>
<td>Latessa, L.</td>
<td>208, 247</td>
</tr>
<tr>
<td>Latruffe, N.</td>
<td>83, 101</td>
</tr>
<tr>
<td>Laurent, G.</td>
<td>106, 116–121, 124, 160, 161</td>
</tr>
<tr>
<td>Laureyn, W.</td>
<td>83, 100</td>
</tr>
<tr>
<td>Laureys, W.</td>
<td>83, 100</td>
</tr>
<tr>
<td>Laux, S.E.</td>
<td>176, 180, 190, 238, 243–245, 249</td>
</tr>
<tr>
<td>Law, M.</td>
<td>144, 164</td>
</tr>
<tr>
<td>Lear, J.</td>
<td>233, 249</td>
</tr>
<tr>
<td>Leburton, J.P.</td>
<td>212, 247</td>
</tr>
<tr>
<td>Lechuga, L.M.</td>
<td>80, 99</td>
</tr>
</tbody>
</table>
Leckband, D., 77, 99
Lee, C., 43, 93
Lee, C.H., 173, 180, 243
Lee, C.L., 66, 97
Lee, D.-S., 131, 162
Lee, E.-C., 194, 245
Lee, G.U., 77, 99
Lee, J., 81, 100
Lee, J.Y., 45, 93
Lee, J.Y.M., 45, 93
Lee, K.B., 49, 94
Lee, S.-H., 137, 138, 163
Lee, S.C., 183, 245
Lee, T., 6, 34
Lefebvre, P.A., 85, 101
Leffingwell, J.C., 114, 115, 161
Legay, G., 83, 101
Lehmann-Horn, F., 83, 100
Lenard, S.W., 64, 97
Léonard, F., 217, 247
Letsinger, R.L., 85, 101
Leung, C.C., 178, 244
Leung, O.M., 56, 96
Leung, R., 268, 271
Levine, D., 72, 98
Levit, C., 3, 26, 34
Levitt, M., 230, 248
Lewenstein, B.V., 284, 286, 293
Lewis, G., 46, 94
Lewis, N.S., 133, 162
Li, C.Y., 144, 145, 164
Li, J., 69, 98, 144, 164
Li, M., 61, 97
Li, S.C., 237, 249
Li, Y., 58, 96, 220, 248
Liang, H., 50, 94
Lieber, C.M., 8, 34, 143, 144, 148, 163, 165
Liebig, D., 180, 244
Lin, H., 178, 244
Lin, P.S., 43, 93
Lin, Q., 184, 245
Lina, X., 64, 97
Lindelof, P.E., 208, 247
Linford, M.R., 218, 247
Lipshutz, R.J., 40, 93
Liu, H., 67, 98
Liu, H.Q., 144, 164
Liu, X., 80, 100
Liz-Marzán, L.M., 66, 97
Lopez, C., 64, 97
López-Villanueva, J.A., 176, 243
Loss, D., 172, 243
Lou, K.-L., 224, 248
Love, J.C., 2, 3, 6, 7, 18, 34, 198, 246
Lu, D., 220, 248
Lu, M., 51, 95
Lu, W., 50, 95
Lu, Y.J., 144, 145, 164
Luan, S., 69, 98
Luckham, P.F., 77, 99
Ludwigs, S., 51, 95
Lugli, P., 179, 180, 208, 217, 218, 238, 244, 246–248
Luna, E.H., 19, 36
Lundstrom, M., 213, 214, 216, 217, 247
Lundstrom, M.S., 177, 178, 183, 189, 216, 244, 245, 247
Luo, C., 69, 98
Luo, L., 66, 98
Luryi, S., 174, 243
Lynden-Bell, R., 230, 249
Lyon, S.A., 61, 97

M

Mack, A., 87, 101
MacLeod, K., 118, 120, 121, 161
Macnaghten, P., 283, 289, 293
Macoubrie, J., 282, 284, 291
Macucci, M., 176, 216, 243, 247
Mahajan, R.L., 277, 281, 283, 286, 288, 292
Mahalingam, V., 69, 98
Maiti, A., 210, 247
Mallouk, T.E., 67, 98
Mamonov, A.B., 235, 249
Mancini, D.P., 61, 97
Maniezzo, V., 17, 36
Manne, S., 77, 99
Mansky, P., 51, 95
Mao, Z., 218, 248
Mariotti, A., 46, 94
Marla, S.S., 85, 101
Marlo, A., 66, 98
Marrink, S.J., 228, 248
Marte, P., 55, 96
Martel, R., 172, 208, 214, 216, 243, 247
Martin, A.S., 197, 245
Martin, B.R., 67, 98
Martin, C.R., 66, 97
Martin, D.K., 274, 295
Martin, R.M., 181, 182, 244
Marty, A., 277, 293
Masaki, K., 190, 245
Mascioli, T., 256, 271
Masaki, K., 190, 245
Mathes, D.T., 49, 94
Mathur, S., 154, 166
Matsui, S., 45, 55, 93, 96
Matsushita, Y., 50, 95
Matuoka, G., 44, 93
Maury, P., 69, 98
Mayer, T.S., 67, 98
Mayergoyz, I.D., 178, 244
Mays, J., 51, 95
Mazziar, C.M., 181, 244
Mazzoni, M.S.C., 219, 248
McCammon, J.A., 233, 249
McEuen, P.L., 6, 34, 213, 247
McEuen, P.L., 6, 34, 213, 247
Mead, C., 173, 180, 243
Mehl, M.J., 207, 246
Mehn, D., 60, 96
Mehta, M.D., 283, 295
Meijers, A., 287, 293
Meinders, J.D., 188, 245
Meiners, J.C., 50, 94
Melchior, L., 176, 243
Melhuish, C., 15, 35
Melhuish, C.R., 13, 35
Melosh, N.A., 152, 155, 165
Mench, M., 82, 100
Menon, M., 219, 248
Menzel, R., 118, 161
Mergeay, M., 82, 100
Mertens, R., 55, 83, 96, 100
Meschede, D., 53, 95
Meseguer, F., 64, 97
Metcalfe, H., 53, 95
Metzger, R.M., 197, 245
Meunier-Prest, R., 83, 101
Meyer, E., 77, 78, 99
Meyers, G.F., 56, 96
Middleton, P.L., 45, 93
Miercke, L.W., 224, 248
Miguez, H., 64, 97
Milburn, C., 278, 284, 294
Miller, C.A., 268, 271
Mirkin, C.A., 49, 85, 86, 94, 101
Mitcham, C., 283, 286, 290, 292, 294
Miura, K., 64, 97
Miyatsuji, K., 190, 245
Mizuno, H., 181, 182, 244
Mlynk, J., 50, 54, 55, 94–96
Mnyusiwalla, A., 284, 294
Mock, J.J., 85, 101
Modi, A., 144, 145, 147, 164
Mondia, J.P., 64, 97
Mori, K., 43, 93
Montemagno, C.D., 91, 101
Moore, A., 284, 294
Moore, G., 168, 242
Mor, G.K., 144, 165
Morgan, D., 45, 93
Mori, K., 45, 93, 106, 111, 115–118, 160
Morifuji, M., 181, 182, 244
Morinaga, M., 55, 96
Moriyama, S., 44, 93
Morkoç, H., 174, 243
Morkved, T., 51, 95
Morkved, T.L., 51, 95
Mortimore, D., 12, 35
Moskovits, M., 144, 164
Mowbray, D.J., 65, 97
Moy, G., 233, 236, 249
Mrksich, M., 49, 94
Mucic, R.C., 85, 101
Mueller, R., 83, 100
Mulchandani, A., 82, 100
Mulchandani, P., 82, 100
Muller, A.H.E., 51, 95
Muller, C.J., 197–199, 245
Muller, D.J., 77, 99
Muller, U.R., 85, 101
Muller, W.T., 6, 34
Mulliken, R.S., 208, 246
Mulvaney, P., 66, 97
Murai, F., 44, 93
Murata, S., 27, 35
Musin, R.N., 154, 166
Mustieles, F.J., 178, 213, 244

N
Nagao, H., 106, 111, 115–118, 160
Nakatani, A.I., 50, 95
Nam, J., 192, 245
Nam, J.M., 86, 101
Namba, S., 45, 93
Narayanan, A., 67, 98
National Nanotechnology Initiative, 285, 294
National Research Council, 259, 271, 280, 284, 294
National Science Foundation, 252, 271, 275, 294
Natoli, V.D., 181, 182, 244
Nature, 280, 294
Nealey, P.F., 50, 94
Neihaus, Th., 208, 247
Nelson, B.J., 69, 98
Neumeister, J.M., 77, 99
New Scientist, 148, 165
Nicholls, A., 233, 249
Nicollian, E.H., 168, 242
Nie, S., 65, 97
Niedermann, Ph., 45, 93
Niesz, K., 60, 96
Nigård, J., 208, 247
NIH Nanomedicine Roadmap, 91, 102
Niklewski, D.J., 207, 246
Nikolaev, P., 58, 96
Ning, T.H., 173, 243
Nitzan, A., 235, 249
NN, 262, 271
Noble, D., 222, 248
Noda, I., 50, 95
Nollert, P., 224, 248
Nolmer, W., 235, 237, 249
Nordan, M.M., 284, 285, 294
Nordmann, A., 284, 294
Nordquist, K.J., 61, 97
Nordstrom, M., 80, 99
Novak, J.P., 144, 165
Noworyta, J.P., 219, 220, 230, 248
NSTC—National Science and Technology Council, 275, 278, 281, 283, 285, 294
Nusair, M., 206, 246
Nuzzo, R.G., 48, 94

O
Oberthaler, M.K., 55, 96
Ochiai, N., 50, 95
Odeh, F., 178, 184, 213, 244, 245
Odin, C., 50, 96
Ohibitu, K., 44, 93
Okazaki, S., 43, 44, 93
Okumura, M., 44, 93
Olbrich, A., 45, 93
Omaran, M., 19, 36
Osada, T., 77, 99
Oscarsson, S.O., 49, 94
Osman, M.A., 176, 177, 243
Oyafuso, F., 177, 243
Ozin, G.A., 64, 97

P
Pacelli, A., 180, 244
Pan, Z.W., 144, 148, 150, 165
Pantelides, S.T., 199, 246
Papaconstantopoulos, D.A., 207, 246
Park, H.J., 82, 100
Park, J., 213, 218, 247
Park, M., 50, 94
Park, S.I., 49, 94
Parr, R.G., 206, 246
Pasquarelli, A., 83, 100
Pasteels, J.M., 17, 36
Pastoriza-Santos, I., 66, 97
Patil, M.B., 174, 243
Patolsky, F., 84, 101
Paun, G., 172, 243
Pearce, T., 127, 131, 132, 135, 152, 162, 163
Pecchia, A., 208, 218, 247, 248
Pedemonte, E., 50, 95
Pederson, M.R., 206, 246
<table>
<thead>
<tr>
<th>Author</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peng, S.</td>
<td>144, 145, 164, 218, 219, 247, 248</td>
</tr>
<tr>
<td>Pengfei, Q.F.</td>
<td>143, 144, 163</td>
</tr>
<tr>
<td>Pennington, G.</td>
<td>211–214, 247</td>
</tr>
<tr>
<td>Pennycook, S.J.</td>
<td>58, 96</td>
</tr>
<tr>
<td>Penrose, L.S.</td>
<td>26, 35</td>
</tr>
<tr>
<td>Penrose, R.</td>
<td>26, 35</td>
</tr>
<tr>
<td>Perdew, J.P.</td>
<td>206, 246</td>
</tr>
<tr>
<td>Pereira, G.G.</td>
<td>50, 94</td>
</tr>
<tr>
<td>Pereira, R.S.</td>
<td>78, 99</td>
</tr>
<tr>
<td>Perez-Juste, J.</td>
<td>66, 97</td>
</tr>
<tr>
<td>Perez-Orive, J.</td>
<td>118, 120, 121, 161</td>
</tr>
<tr>
<td>Pergamit, G.</td>
<td>2, 5, 34, 127, 162</td>
</tr>
<tr>
<td>Persad, D.L.</td>
<td>207, 246</td>
</tr>
<tr>
<td>Peters, A.</td>
<td>53, 95</td>
</tr>
<tr>
<td>Peterson, C.</td>
<td>2, 5, 34, 277–279, 284, 285, 294</td>
</tr>
<tr>
<td>Pfaul, T.</td>
<td>54, 95, 96</td>
</tr>
<tr>
<td>Phillips Jr., A.</td>
<td>180, 244</td>
</tr>
<tr>
<td>Pickering, J.P.</td>
<td>56, 96</td>
</tr>
<tr>
<td>Pielke Jr., R.A.</td>
<td>289, 295</td>
</tr>
<tr>
<td>Pilarski, L.M.</td>
<td>283, 295</td>
</tr>
<tr>
<td>Piner, R.D.</td>
<td>49, 94</td>
</tr>
<tr>
<td>Pirone, D.M.</td>
<td>80, 100</td>
</tr>
<tr>
<td>Pitsikalis, M.</td>
<td>51, 95</td>
</tr>
<tr>
<td>Pleiss, J.</td>
<td>72, 98</td>
</tr>
<tr>
<td>Polushkin, E.</td>
<td>50, 95</td>
</tr>
<tr>
<td>Pompeo, F.</td>
<td>209, 247</td>
</tr>
<tr>
<td>Poole, K.</td>
<td>77, 99</td>
</tr>
<tr>
<td>Popov, V.N.</td>
<td>59, 96</td>
</tr>
<tr>
<td>Porezag, D.</td>
<td>207, 246</td>
</tr>
<tr>
<td>Pötz, W.</td>
<td>176, 177, 243, 244</td>
</tr>
<tr>
<td>Prentiss, M.</td>
<td>80, 100</td>
</tr>
<tr>
<td>Price, P.J.</td>
<td>180, 244</td>
</tr>
<tr>
<td>Priebe, C.E.</td>
<td>106, 127, 160</td>
</tr>
<tr>
<td>Primilov, A.</td>
<td>224, 248</td>
</tr>
<tr>
<td>Puech, P.H.</td>
<td>77, 99</td>
</tr>
<tr>
<td>Puretzky, A.A.</td>
<td>58, 96</td>
</tr>
<tr>
<td>Q</td>
<td></td>
</tr>
<tr>
<td>Qiu, J.</td>
<td>64, 97</td>
</tr>
<tr>
<td>Quake, S.</td>
<td>40, 92</td>
</tr>
<tr>
<td>R</td>
<td></td>
</tr>
<tr>
<td>Rabe, J.</td>
<td>137, 163</td>
</tr>
<tr>
<td>Rahman, A.</td>
<td>177, 244</td>
</tr>
<tr>
<td>Rai-Choudhury, P.</td>
<td>6, 34</td>
</tr>
<tr>
<td>Raju, A.R.</td>
<td>64, 97</td>
</tr>
<tr>
<td>Rakow, N.A.</td>
<td>142, 163</td>
</tr>
<tr>
<td>Raman, S.</td>
<td>67, 98</td>
</tr>
<tr>
<td>Ramseyer, J.-P.</td>
<td>77, 99</td>
</tr>
<tr>
<td>Rao, C.N.R.</td>
<td>64, 97</td>
</tr>
<tr>
<td>Rasaiah, J.</td>
<td>219, 220, 230, 248</td>
</tr>
<tr>
<td>Rasaiah, J.C.</td>
<td>230, 249</td>
</tr>
<tr>
<td>Ratner, D.</td>
<td>106, 160</td>
</tr>
<tr>
<td>Rawlett, A.M.</td>
<td>198, 246</td>
</tr>
<tr>
<td>Raychaudhuri, A.K.</td>
<td>66, 97</td>
</tr>
<tr>
<td>Reed, M.A.</td>
<td>8, 14, 35, 197–199, 245, 246</td>
</tr>
<tr>
<td>Reeder, G.</td>
<td>72, 98</td>
</tr>
<tr>
<td>Reggiani, L.</td>
<td>180, 183, 244, 245</td>
</tr>
<tr>
<td>Register, L.F.</td>
<td>176, 194, 217, 243, 245, 247</td>
</tr>
<tr>
<td>Register, R.A.</td>
<td>50, 94</td>
</tr>
<tr>
<td>Reimann, C.T.</td>
<td>49, 94</td>
</tr>
<tr>
<td>Reimers, J.R.</td>
<td>205, 246</td>
</tr>
<tr>
<td>Reinhoudt, D.N.</td>
<td>69, 98</td>
</tr>
<tr>
<td>Reininger-Mack, A.</td>
<td>87, 101</td>
</tr>
<tr>
<td>Ren, Z.</td>
<td>177, 216, 244, 247</td>
</tr>
<tr>
<td>Renstrom, P.J.</td>
<td>61, 97</td>
</tr>
<tr>
<td>Resasco, D.E.</td>
<td>209, 247</td>
</tr>
<tr>
<td>Resnick, D.J.</td>
<td>61, 97</td>
</tr>
<tr>
<td>Resnick, M.</td>
<td>18, 36</td>
</tr>
<tr>
<td>Ressler, K.J.</td>
<td>111, 161</td>
</tr>
<tr>
<td>Reusch, D.B.</td>
<td>45, 93</td>
</tr>
<tr>
<td>Reuss, R.H.</td>
<td>45, 93</td>
</tr>
<tr>
<td>Reynaeris, D.</td>
<td>46, 93</td>
</tr>
<tr>
<td>Reynolds, C.W.</td>
<td>18, 36</td>
</tr>
<tr>
<td>Richard, G.</td>
<td>46, 94</td>
</tr>
<tr>
<td>Rieder, K.H.</td>
<td>7, 34</td>
</tr>
<tr>
<td>Riegelman, M.</td>
<td>67, 98</td>
</tr>
<tr>
<td>Ringhofer, C.</td>
<td>176, 177, 191, 243–245</td>
</tr>
<tr>
<td>Rip, A.</td>
<td>275, 287, 295</td>
</tr>
<tr>
<td>Ritz, A.</td>
<td>50, 94</td>
</tr>
<tr>
<td>Riu, J.</td>
<td>66, 98</td>
</tr>
<tr>
<td>Rius, F.X.</td>
<td>66, 98</td>
</tr>
<tr>
<td>Robinett, W.</td>
<td>5, 34</td>
</tr>
<tr>
<td>Robitizki, A.</td>
<td>87, 101</td>
</tr>
<tr>
<td>Robitizki, A.A.</td>
<td>87, 101</td>
</tr>
<tr>
<td>Roco, M.</td>
<td>259, 261, 271</td>
</tr>
<tr>
<td>Roco, M.C.</td>
<td>252, 259, 271, 277, 295</td>
</tr>
</tbody>
</table>
Roco, M.S., 275, 277, 279, 295
Rogers, J.A., 80, 100
Roller, D., 72, 98
Rosei, F., 6, 34
Rosenblatt, S., 213, 224, 247
Rosenbusch, J.P., 224, 248
Rothuizen, H., 78, 99
Rotkin, S.V., 219, 220, 248
Roufas, K., 27, 35
Roukes, M.L., 6, 34
Roux, B., 228, 230, 231, 236, 237, 248, 249
Royal Academy of Engineering, 282, 284, 295
Royal Society, 282, 284, 295
Royal Society, London, 260, 271
Rozenberg, G., 172, 243
Russell, T.P., 51, 95
Ryu, D.Y., 50, 94
S
Saad, Y., 176, 243
Sader, J.E., 78, 99
Sadowski, J., 208, 247
Saint, N., 224, 248
Saito, R., 209, 247
Saitou, N., 44, 93
Sakita, Y., 44, 93
Sakurai, H., 197, 245
Salamanca-Buentello, F., 274, 295
Salerno, J., 19, 37
Salman, A., 19, 36
Salomaa, A., 172, 243
Saltzman, W.M., 91, 102
Samanta, M.P., 203, 205, 246
Samarth, N., 172, 243
Samuel, J.R., 197, 245
Samia, A.C.S., 64, 97
Sankararamakrishnan, R., 230, 249
Sankey, O.F., 207, 246
Sansen, W., 83, 100
Sansom, M.S.P., 226, 228, 230, 248, 249
Saraniti, M., 179, 244
Sarewitz, D., 279, 282, 283, 286, 289, 291, 293, 295
Sarychev, A.K., 66, 98
Satkowski, M.M., 50, 95
Schawlow, A.L., 53, 95
Schawlow, A.L., 53, 95
Schifman, S., 107, 126, 127, 129–132, 134–137, 139, 160
Schirmer, T., 224, 248
Schleier-Smith, M.H., 152, 165
Schmid, R.D., 72, 98
Schmidt, J.J., 91, 101
Schmidt, P.O., 55, 96
Schmidt, T., 208, 247
Schneider, T., 130, 154, 162
Schreiner, J., 237, 249
Schulz, K., 207, 246
Schuler, E., 284, 295
Schulten, K., 220, 224, 241, 248, 249
Schultz, D.A., 85, 101
Schultz, P.G., 6, 34
Schultz, S., 85, 101
Schulze, Th., 54, 55, 95, 96
Schunack, M., 6, 34
Schutz, C.N., 236, 249
Schweitzer, E.K., 5, 34, 255, 270
Science, 280, 295
Scientific American, 254, 259, 271
Seefeld, S., 237, 249
Seeger, T., 77, 99
Seeman, N.C., 90, 101
Seifert, G., 207, 208, 217, 246
Seifert, K., 77, 99
Seitz, J.T., 56, 96
Selberherr, S., 186, 245
Seminario, J.M., 198, 199, 206, 207, 246
Senate Committee on Commerce, Science, and Transportation, 288, 295
Sendova-Franks, A., 15, 36
Sept., D., 233, 249
Shackleton, M., 12, 35
Shalaev, V.M., 66, 98
Sham, L.J., 206, 246
Shanbhag, N., 171, 242
Shankar, K.S., 66, 97
Sharon, A., 46, 94
Sharon, R., 230, 248
Shea, H.R., 208, 247
Shen, T.C., 177, 244
Shen, Z., 90, 101
Shepherd, G.M., 118, 161
Sheriff, B.A., 155, 166
Shi, Y., 12, 14–16, 20, 35
Shibata, Y., 44, 93
Shichijo, H., 174, 243
Shim, M., 217, 218, 247
Shimizu, F., 55, 96
Shimotsuma, Y., 64, 97
Shinagawa, H., 91, 101
Shiokawa, T., 45, 93
Shipman, R., 12, 35
Shockley, W., 168, 242
Shodjai, F., 290, 295
Shu, C.-W., 182, 184, 245
Shu, C.W., 184, 245
Siedle, P., 77, 99
Silverbrook, K., 205, 246
Simmons, M., 58, 96
Singer, P., 274, 295
Singer, P.A., 284, 294
Singh, D.R., 85, 101
Siri, O., 83, 101
Sitti, M., 69, 71, 98
Skolnick, M.S., 65, 97
Slater, J.C., 207, 246
Smalley, R.E., 58, 96
Smith, A.M., 65, 97
Smith, B.J., 61, 97
Smith, G., 58, 96
Smith, G.R., 226, 230, 248
Smith, J.C., 49, 94
Smith, K., 77, 99
Smith, S.D., 50, 95
Snow, S.J., 6, 34
Snyder, E.J., 5, 34
Snyder, M., 40, 93
Sobel, N., 106, 160
Solda, P., 82, 100
Song, J.H., 66, 98
Sonoda, K., 188, 245
Sotomayor, M., 241, 249
Sotzing, G.A., 133, 162
SpaceDaily, 135, 163
Spener, F., 82, 100
Srivastava, D., 219, 248
Stacey, N.A., 61, 97
Steckl, A.J., 45, 93
Steelman, J.R., 282, 289, 295
Stensgaard, L., 6, 34
Stephan, K.D., 286, 295
Stephenson, N., 258, 271
Stern, F., 176, 243
Sternberg, M., 208, 217, 246
Stettler, M.A., 178, 183, 244, 245
Stevens, E.H., 45, 93
Stevens, F., 72, 98
Sbix, G., 5, 7, 27, 34, 285, 295
Stockert, J.A., 66, 97
Stopfer, M., 118, 120, 161
Stratton, R., 182, 184, 185, 244
Streetman, B.G., 168, 174, 242, 243
Stroscio, M.A., 176, 243
Stroud, R.M., 224, 248
Stuart, C., 282, 295
Stubbs, D.D., 137, 138, 163
Stuhler, J., 54, 95
Subramanian, A., 69, 98
Suehiro, J., 144, 165
Suhai, S., 207, 246
Sullivan, S.L., 111, 161
Suls, J., 83, 100
Summers, C.J., 144, 145, 154, 164, 166
Sun, Y., 80, 100
Sunstein, C.R., 284, 295
Suo, Z., 50, 95
Superfine, R., 3, 5, 34
Suslick, K.S., 142, 163
Svizhenko, A., 210, 215, 216, 247
Swain, S., 174, 180, 243
Swainey, A.E., 283, 284, 287, 295
Swiss-Re, 282, 295
Sysoev, V.V., 153, 165
Sze, S.M., 168, 170, 171, 242

T

Tachibana, H., 197, 245
Tahan, C., 268, 271
Tai, G.C., 184, 245
Tajkhoshid, E., 224, 248
<table>
<thead>
<tr>
<th>Author</th>
<th>Pages</th>
<th>Pages</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Takahashi, Y.</td>
<td>50, 95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Takenaka, M.</td>
<td>181, 182, 244</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Takeuchi, H.</td>
<td>194, 245</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tamayo, J.</td>
<td>80, 99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tan, J.L.</td>
<td>80, 100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tanaka, H.</td>
<td>219, 220, 230, 248</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tanaka, K.</td>
<td>86, 101</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tang, C.C.</td>
<td>154, 166</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tang, J.M.</td>
<td>228, 248</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tang, J.Y.</td>
<td>173, 243</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tang, T.W.</td>
<td>183, 192, 245</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tang, Y.</td>
<td>237, 249</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tani, T.</td>
<td>91, 101</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taniguchi, K.</td>
<td>181, 182, 188, 190, 244, 245</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taniguchi, N.</td>
<td>252, 271</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tasch, A.F.</td>
<td>181, 244</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taubes, G.</td>
<td>132, 162</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taylor, R.M.</td>
<td>3, 5, 34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taylor II, R.M.</td>
<td>5, 34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tenigiyo, N.</td>
<td>45, 93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tepper, A.</td>
<td>289, 296</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tersoff, J.</td>
<td>216, 217, 247</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thaxton, C.S.</td>
<td>86, 101</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The Economist</td>
<td></td>
<td>274, 292</td>
<td></td>
</tr>
<tr>
<td>Theis, T.N.</td>
<td>284, 296</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Theraulaz, G.</td>
<td>12, 14–17, 23, 35, 36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thess, A.</td>
<td>58, 96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thielecke, H.</td>
<td>87, 101</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thomas, E.L.</td>
<td>50, 95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thornber, K.K.</td>
<td>187, 245</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thundat, T.</td>
<td>77, 99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tian, W.</td>
<td>203, 205, 246</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tibazarwa, C.</td>
<td>82, 100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tielman, D.P.</td>
<td>226, 228, 230, 248</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tien, J.</td>
<td>80, 100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Toader, O.</td>
<td>64, 97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Todokoro, H.</td>
<td>44, 93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tombler, T.W.</td>
<td>217, 218, 247</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tomic, O.</td>
<td>127, 162</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tomita, K.</td>
<td>27, 35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tour, J.M.</td>
<td>8, 14, 35, 197–199, 207, 245, 246</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Toyoda, K.</td>
<td>45, 93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tran, P.</td>
<td>233, 249</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treilakis, A.</td>
<td>176, 240, 243, 249</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tsai, J.-H.</td>
<td>192, 245</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tseng, Y.C.</td>
<td>144, 165</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tsien, R.W.</td>
<td>222, 248</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tsuchiya, H.</td>
<td>191, 245</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tu, Y.</td>
<td>81, 100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tucker, J.R.</td>
<td>177, 244</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turberfield, A.J.</td>
<td>91, 101</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ueno, H.</td>
<td>190, 245</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uijin, S.</td>
<td>19, 37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ulman, A.</td>
<td>48, 94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Upadhyay, S.N.</td>
<td>87, 101</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urban, V.</td>
<td>51, 95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urbaś, A.M.</td>
<td>51, 95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>US Congress</td>
<td>280–282, 296</td>
<td></td>
<td></td>
</tr>
<tr>
<td>USPTO</td>
<td>261, 271</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Utalut, M.</td>
<td>45, 93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vab Brussel, H.</td>
<td>46, 93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valentini, L.</td>
<td>144, 164, 165</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Van de Poel, I.</td>
<td>287, 296</td>
<td></td>
<td></td>
</tr>
<tr>
<td>van der Lelie, D.</td>
<td>82, 100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>van der Straaten, T.</td>
<td>240, 241, 249</td>
<td></td>
<td></td>
</tr>
<tr>
<td>van der Straaten, T.A.</td>
<td>228, 248</td>
<td></td>
<td></td>
</tr>
<tr>
<td>van Ekenstein, G.A.</td>
<td>50, 95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Van Gerwen, P.</td>
<td>83, 100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Van Osenbruggen, C., 46, 93</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vancea, J.</td>
<td>45, 93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vancso, G.J.</td>
<td>56, 96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Varani, L.</td>
<td>183, 245</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vasileska, D.</td>
<td>191, 245</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vassar, R.</td>
<td>111, 161</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vazquez, J.</td>
<td>58, 96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Venter, J.C.</td>
<td>110, 161</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ventura, D.</td>
<td>178, 213, 244</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Venugopal, R.</td>
<td>177, 216, 244, 247</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbeek, P.-P.</td>
<td>287, 296</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vettiger, P.</td>
<td>78, 99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vickers, N.J.</td>
<td>118, 161</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vogl, P.</td>
<td>179, 180, 244</td>
<td></td>
<td></td>
</tr>
<tr>
<td>von Allmen, P.</td>
<td>177, 243</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vos, M.</td>
<td>46, 94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vosko, S.H.</td>
<td>206, 246</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vuillaume, D.</td>
<td>197, 245</td>
<td></td>
<td></td>
</tr>
<tr>
<td>W</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wachter, E.A.</td>
<td>77, 99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Walenz, B.</td>
<td>72, 98</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Author Index

Walker, M., 237, 249
Walt, D.R., 127, 139, 162, 163
Walter, T., 130, 154, 162
Wan, Q., 144, 164
Wang, C., 177, 244
Wang, C.R.C., 66, 97
Wang, D.W., 155, 166
Wang, F., 106, 116–118, 160
Wang, J., 16, 36
Wang, J.W., 106, 160
Wang, Q., 50, 94, 217, 218, 247
Wang, W., 80, 100
Wang, X., 194, 245
Wang, X.D., 144, 145, 154, 164, 166
Wang, X.L., 181, 244
Wang, X.Q., 154, 166
Wei, T.F., 85, 101
Weil, V., 284, 296
Weißenhorn, A.L., 77, 99
Weisman, R.B., 209, 247
Weiss, W., 67, 98
Whang, D., 143, 163
Whitesides, G.M., 2, 3, 6, 7, 18, 26, 32, 34, 35, 80, 100
Whittaker, J.D., 218, 247
Wid, P., 127, 162
Wiertz, F.G.M., 81, 100
Wilk, L., 206, 246
Williams, R.S., 5, 34, 155, 166
Willis, R., 282, 296
Willmout, F.M., 50, 95
Willner, I., 66, 98
Wilson, C.G., 61, 97
Wilson, J., 275, 282, 283, 289, 293, 296
Wilson, C.D., 49, 94
Wilson, M., 58, 96
Wind, S.J., 172, 208, 214, 243, 247
Windle, A.H., 58, 96
Wineland, D., 53, 95
Winner, L., 257, 271, 274, 284, 288, 296
Winstead, B., 177, 192, 194, 244, 245
Wolynes, P., 237, 249
Wong, A.M., 106, 160
Wordelman, C.J., 238, 249
Wordeman, M.R., 174, 243
Wright Jr., W.V., 5, 34
Wu, B., 192, 245
Wu, Q., 81, 100
Wu, W., 61, 97
Wu, X., 197, 245
WWC, 258, 271
Wynne, B., 283, 284, 293, 296
Wyns, L., 82, 100
X
Xi, J., 91, 101
Xia, D., 64, 65, 97
Xia, T.S., 217, 247
Xia, Y., 80, 100
Xu, F., 49, 94
Xu, J.Z., 81, 100
Xue, Y., 206, 207, 217, 246, 247
Y
Yaish, Y., 213, 218, 247
Yamaguchi, H., 43, 93
Yamaji, M., 181, 182, 244
Yamakawa, S., 190, 245
Yamakita, M., 26, 27, 35
Yamazaki, T., 44, 93
Yan, H., 90, 101
Yang, P., 66, 98
Yang, P.D., 144, 164
Yang, W., 206, 246
Yang, Z., 176, 243
Yim, M., 27, 35
Yoda, H., 44, 93
Yoder, P.D., 181, 182, 244
Yong, Y., 77, 99
Yoshida, E., 27, 35
Yoshida, H., 19, 36
Yoshihara, Y., 106, 111, 115–118, 160
Yu, H.N., 173, 243
Yu, Y.Y., 66, 97
Yu, Z., 61, 97, 173, 182, 184, 243, 245
Yua, K., 66, 98
Yuan, H.K., 66, 98
Yun, M.H., 144, 164
Yun, W.S., 82, 100
Yun, Y.J., 82, 100

Z

Zaborosch, C., 82, 100
Zachariae, U., 237, 249
Zacharias, A.G., 207, 246
Zandler, G., 179, 244
Zenner, G.M., 268, 271
Zettl, A., 218, 248
Zettl, H., 51, 95
Zhang, J., 69, 98
Zhang, W., 256, 271
Zhang, X., 46, 80, 90, 94, 100, 101, 106, 110, 111, 160
Zhang, Y.F., 154, 166
Zhang, Y., 66, 98
Zhanga, Y., 66, 98
Zhao, H., 106, 111, 160
Zhao, J., 81, 100
Zhao, X.M., 80, 100
Zhao, Y., 80, 100
Zheng, M., 152, 165
Zheng, X.C., 219, 220, 230, 248
Zhong, Z., 155, 166
Zhou, C., 197–199, 218, 245, 247
Zhou, G.B., 144, 165
Zhou, J.R., 177, 244
Zhu, H., 40, 93
Zhu, J., 49, 60, 94, 96
Zhu, J.J., 81, 100
Zhu, Z., 66, 98
Zimmer, C., 117, 161
Zou, Z., 106, 117–119, 121, 160
This page intentionally left blank
Subject Index

A
Acetone, detection, 140
ACO Meta-Heuristic, 17
Adenylyl cyclase (AC), 114
Adhesive, nanostructured, 41
AFM see Atomic force microscopy
Agency, 287
American Nanotechnology Preparedness
Center, 280
Amperometric biosensors, 81
Ant colony optimisation (ACO) algorithm, 12,
16, 17–18
Antennal lobe, 121
Ants, 14, 15, 16, 17
Arc-discharge, 58, 59
Array Matrix, 141
Artificial cells, 89
Artificial intelligence (AI), 11
Artificial neural networks (ANNs), 127
Artificial noses see Electronic noses
Artificial respirocytes, 89
Assemblers, 253
Assembly protocol, 28
component shape and, 28–31
Atom holography, 55–6
Atom lithography, 52–6
complex patterns, 53–6
laser cooling and trapping, 52–3
one-dimensional standing-wave focusing, 53
Atom optics, 53
Atomic force microscopy (AFM), 5, 6, 49, 255
and CNT deformation, 210
direct writing by, 56–7
in nanomechanical sensors, 77–9
in nanoscale perception, 71–2
Atomic nanofabrication (ANF), 53
ATP, 114
Attenuation coefficient, 202
Auditory system, 122
Autonomous Nanotechnology Swarm
(ANTS), 13

B
Ball milling, 60
Ballistic regime, 170, 214–15
Barley lectin (BL), 119–20
Basis functions, 205
BD, 237, 238
BeadArray, 141
BeadChip, 141
Beck-3 Perdew–Wang 91 (B3PW91) model,
206
Benzene, aromatic structures based on, 198
Bias, 253
bio-MOCA, 240
Bioactuators, 74
see also Nanobiomotors
Biobarcode assay, 86
Bioethics, 257, 288
Biological ionic channels/nanochannels see
Ionic channels
Biological machine, 220
Biological processes
space scales, 226
time scales, 225–6
Biomarkers, 266
Biomimetic channels, 172, 221, 225
see also Ionic channels, simulation
Biomolecular systems, 172
Biosensors, 10, 74, 75
in vitro, 75
in vivo, 75
see also Nanobiosensors
Bird flocking, 12, 18
Block copolymers, 49–50
Blood-extraction chip, 41
Blue-detuning, 53
Bohm representation, 192
Boids, 18
Boltzmann equation, 175, 177, 191
moments of, 182
higher-order, 183
simulation approaches using, 182–6
solution approaches, 177–80
see also Monte Carlo particle methods
Boltzmann statistics, 233
Bottom–up technology, 2–3, 4
Brownian Dynamics (BD), 237, 238
Buckyballs, 7, 265
Bulk technology, 2

C
Calcium channels, 235
Calmodulin, 148
Calorimetric sensors, 142
cAMP, 114
Cantilever sensor arrays, 78–9, 137–8
Cantilever sensors, 76–9, 137–8
polymer-based, 80–1
Carbon black, 132–4
Carbon fullerenes, 7, 265
definition, 270
Carbon nanotube sensors, 82, 145–7, 218, 219
consideration for electronic noses, 151–2
sensitivity, 151
ammonia, 145
nitrous oxide, 145
oxygen, 145
Carbon nanotubes (CNT), 7, 145, 208–20, 265
applications, 8, 172, 208–9, 213–20
field-effect transistors, 213–17
high-frequency device, 213
microwave, 213
nanoelectromechanical systems, 208, 217–20
see also Carbon nanotube sensors
armchair, 210, 217
classification, 209
in conductometric sensors see Carbon nanotube sensors
contacts, 216–17
definition, 270
discovery, 209
functionalization, 152, 218
growth, 57–60, 209, 217–18
multi-walled (MWNT), 58, 145, 147, 209
in nanopore structures, 225
nanorobotic manipulation, 69, 70
properties, 208
examination, 5
simulation, 210–13
atomistic approach, 214–16
continuum approach, 219
mode-space approach, 216
molecular dynamics, 218–20
Monte Carlo methods, 211–12, 214
multiscale, 219
quantum approach, 210–11, 214–15
semi-classical particle, 211–13
single-walled (SWNT), 58, 145–7, 209
bulk separation, 152
uses, 58
in wet environments, 218, 219–20
zig-zag, 210, 211, 216, 217
Carrier continuity equation, 183, 185
Cell-based robots, 91
Cellular Automata approach, 179
Centers for Nanotechnology in Society (CNS), 281
Chambers Path Integrals, 178
Channel doping, 193, 227
ChemFETs, 135
Chemical agent detectors, 139
Chemical vapor deposition (CVD), 59–60
Chiral indices, 209
Chirality, 209–10
Cloud in Cell (CiC) assignment, 238
C-MOS architecture, 171
CMOS transistors, 257, 258
CNG ion channels, 114
CNT see Carbon nanotubes
Codes of conduct, 282
Colloid synthesis, 64
Colorimetric sensors, 142
Combinatorial optimisation problems, 17
Compact approaches, 188
Compensation, 234
Complexity, as inherent problem, 255
Components
in self-assembling systems, 28
assembly protocol and shape of, 28–31
Computer aided design (CAD), 173
Condensed matter (CM) physics, 264
Conducting polymer sensors, 131–4
Conductometric biosensors, 82
Continuum models
carbon nanotubes, 219
ionic channels, 231–3, 235–7
Contrast enhancement, 86
Coulomb interaction, 238
Cribiform plate, 115
Crystal-98 codes, 207
Current density, 185, 186
Current drive improvement, 193–4
Cyanose 320, 133

D
Defense Advanced Research Projects Agency (DARPA), 107, 158, 259
Deformation potentials, 181
Demos, 283
Density Functional Theory (DFT), 206–7, 219
Density-functional tight-binding (DFTB) method, 207–8
Density matrix, 176
Deprotonation, 198
Design space, encapsulation, 33
Deterministic particle methods, 178
Device simulation hierarchy for semiconductor devices, 173–88
Boltzmann equation approaches, 177–80
compact approaches, 188
drift-diffusion approaches, 186–7, 232, 236
moments of Boltzmann equation approaches, 182–6
Monte Carlo methods, 180–2
quantum approaches, 176–7
issues in nanoscale silicon devices, 188–97
augmented Monte Carlo simulation, 190
current drive improvement, 193–4
leakage current control, 192–3
narrow conduction channel realization, 195–7
quantum correction approaches, 190–2
source-drain isolation, 194–5
see also Carbon nanotubes (CNT), simulation; Molecular conduction simulation
DFT, 206–7, 219
DFTB, 207–8
Dichloromethane, detection, 140
Dielectric permittivity, 232, 236
Diffusion coefficient, 187
field dependent, 187
Diffusivity models, 236
Dip-pen nanolithography, 49
Direct tunneling, 201
Discriminant function analysis (DFA), 127
DNA, 8, 41
base pair difference detection, 138
hybridization detection, 82–3
interaction forces, 77
in pollution screening, 87
DNA computing, 8
DNA motors, 90–1
DNT, 140
Doping, channel, 193, 227
Drift-diffusion approaches, 186–7, 232, 236
empirically extended, 187, 188
Drift velocities, 169
Drug delivery, nanobiotechnology for, 91–2
DuPont, 283

E
E-beam resists, 43
Ecosystem models, 12
Elastic field, in nanophase separation, 50
Electric field, in nanophase separation, 50–1
Electric organs, 223
Electrochemical machining (ECM), 46
Electrochemical sensors, 129–36
conductance-based, 129–34
conducting polymer sensors, 131–4
metal oxide sensors, 129–31
potentiometric, 134–6
Electrodischarge machining (EDM), 46
Electron beam lithography (EBL), 6, 43–4, 49
Electron density, 182
Electron gasses, 267
Electron–phonon scattering model, 181
Electronic noses, 79, 126–43
applications, 107, 126
calorimetric sensors, 142
categories, 127
colorimetric nose, 142
designing on nanometer-scale, 151–6
device choice considerations, 151–3
nanomemory sensing system, 155–6
nanowire sensing array, 154–5
next steps, 143
R&D overview, 128
sample handling, 126
sensing, 126–7
signal processing, 127
see also
Electrochemical sensors;
Mass-change sensors; Nanosensors;
Optical sensors
Electronic and Telecommunications Research
Institute (ETRI), 130–1
Electropolymerization, 132
Energy, 28
Energy balance equation, 183, 185
Energy density, 182
Energy flow, 185, 186
Energy flow factor, 186
Energy Transport model, 182, 184–6
deficiencies, 189
“Engines of destruction”, 278
Environment, 28
Environmental biosensors, 87
Environmental Defense, 283
Environmental Protection Agency (EPA), 275, 281
Enzyme electrode, 81
EOS, 12
EPA, 275, 281
Ergodic principle, 239
Essentially Non-Oscillatory (ENO) scheme, 184
ETC Group, 262–3, 268, 279
Ethanol, detection, 82, 142
Ethics
nanotechnology and, 277–80
see also
Nanoethics; Socio-technical integration
Ethics policies
definition, 275
for nanotechnology, 275–6, 280–2
see also
Nanoethics; Socio-technical integration
Evolutionary Algorithms (EAs), 12, 179–80
Evolutionary computation model, 32–3
Evolvable Neural Interface (ENI), 13
Explosives detection, 140
FDA, 261–2, 275, 281
Fear, 256
Feedback
negative, 16
positive, 15–16
Fermi–Dirac statistics, 233
Fermi golden rule, 212
Fermi levels, 202–3, 214, 215, 233
FETs see
Field-effect transistors
Fiber optics see
Photonics
Field dependent mobility, 187
Field-effect transistors (FETs), 134–6, 213–17
nanowire-based, 84–5
see also
FinFETs; MOSFETs
Filtration, 224
Finely focused ion beam (FFIB) lithography, 48
FinFETs, 194–5, 196, 227
Flat-panel displays, 208–9
Fluctuations, amplification of, 16
Fluid pumping, 224
Focused ion beam (FIB) lithography, 44–5
Food and Drug Administration (FDA), 261–2, 275, 281
Foresight Institute, 4
Form aggregation space, 23
Fowler–Nordheim tunneling, 201
Fracture-peeling mechanism, 56–7
Franz–Wiedemann law, 183, 186
Function optimisation space, 23
Golf proteins, 114
GaAs
carrier transport behavior, 169, 170
silicon vs, 170
Gaussian-98 code, 207
Gene therapy, 10
Generalized Gradient Approximation (GGA), 206
Genetic algorithms, 11–12
Genetically modified organisms (GMO) see GM agriculture
GGA, 206
Giant magneto-resistance (GMR), 267
Glomeruli, 115–16
Glucose molecule absorption, 78, 79
GM agriculture, 257, 277, 279, 280
Gold contacts in molecular devices, 197, 198, 199
spectral functions, 204
GPCRs, 112
Gramicidin A, 228, 231, 240–1
Granule cells, 116
Gray goo scenario, 278, 279
Greenpeace, 262, 263, 279
Green’s functions, 203–4, 206–7
non-equilibrium (NEGF), 208, 215, 216, 217
retarded, 215
Gunn effect, 174
H
Haptics, in nanobiotechnology, 70–3
Hartree–Fock theory, 205
HAZMATCAD Plus, 139
HD DVD, 65, 66
Heat conduction equation, 183
Heterojunction transistors, 177
High electron mobility transistors (HEMT), 170
Higher-level neural system (HLNS), 13
Highest occupied molecular orbital (HOMO), 202–3
Highly Immersive Molecular Modeling (HIMM), 72
HOMO, 202–3
Hopping conduction, 201, 220
Hot carrier transport, 173–4, 235
Hot embossing nanoimprint lithography (HE-NIL), 63
House Science Committee, 280, 281, 288
Hückel model, 204–5
Human Genome Project (HGP), 275, 280
ELSI program, 288
Hybrid impedimetric biosensors, 82–5
Hydrodynamic model, 182–4
deficiencies, 189
Hydrogen detection, 64
HyperChem, 204
I
I7 receptor, 122
Illumina, 141
Imagination, study of, 278–9
Immunosolation therapies, 10
Inflammation, 75
Innovation, linear model, 289–90
Intel, 7, 257–8
Intelligent systems, 11
Intentionality, 287
Interactions, multiple, 16
Interglomerular neurons, 116–17
International Council for Risk Governance, 275
International Dialogue for Responsible Nanotechnology R&D, 275
Inversion layers, 266–7
Ionic channels, 172, 220–41
biological functions, 223–4
differences from solid-state devices, 234
equivalent circuit models, 222
experimental methodologies, 222–3
hybrid structures incorporating, 224
identification, 222
properties, 223
protein structure, 227, 231
schematic depiction, 222
simulation, 224–41
atomistic, 228–31
Brownian dynamics (BD) approach, 237, 238
coarse grained, 228, 231–41
continuum models, 231–3, 235–7
molecular dynamics (MD) approach, 228–31, 236
Monte Carlo methods, 238–41
solid-state numerical approaches in, 234–5
Ionization sensors, 147
K
KAMINA, 130, 131, 143, 152, 154–5
KcsA, 228
Kidney, 224
Kinesin, 90
Kohn energy functional, 207
Kohn–Sham equation, 206, 208
Kohn–Sham total energy, 207

L
Landauer formula, 203
Landmines, 140
Laser-ablation, 58
Laser-assisted direct imprint (LADI), 63
Laser beam machining (LBM), 46
Laser cooling, 52–3
lbest position, 23
LCAO, 207
LDA, 206, 219
Leakage, definition, 269
Leakage current control, 192–3
Lennard–Jones potential, 229, 237, 238
Linear combination of atomic orbitals (LCAO), 207
Liouville equation, 176
Lithographically induced self-assembly (LISA), 51–2
Lithography
atom see Atom lithography
electron beam (EBL), 6, 43–4, 49
finely focused ion beam (FFIB), 48
focused ion beam (FIB), 44–5
optical, 43
soft, 6
Local Density Approximation (LDA), 206, 219
“Lock-and-key” sensing elements, 153
Long-range interaction, 238
Lorentz–Berthelot mixing rule, 230
Low field mobility, 174, 187, 232
Lower-level neural system (LLNS), 13
Lowest unoccupied molecular orbital (LUMO), 202–3

M
Magnetic immunoassay nanobiosensors, 86
Mass-change sensors, 136–9
Matthew’s Law, 252–3
Mechanical field, in nanophase separation, 50
Mechanosensitivity, 224
MEMS, 268, 285
definition, 270
technology, 137
Mesoscopics, 254–5
definition, 268
Meta-materials, 266
definition, 269
Metal-oxide-semiconductor field-effect transistors see MOSFETs
Metal oxide sensors, 129–31
Metalloporphyrins, 142
Methanol, detection, 140
“Microbivore”, 89
Microcantilevers, 76–9, 137–8
polymer-based, 80–1
Microelectromechanical systems see MEMS
Microfabrication, 42
Microsensor Systems, 139
Microtechnology, 258, 268
Microwave applications, 213
Milling, 42
MNT, 278–9, 284–5
MoleApps, 158
Molecular alligator clip, 198
Molecular conduction simulation, 200–8
Molecular manufacturing, 278
Molecular nanotechnology (MNT), 278–9, 284–5
Molecular systems see Biomolecular systems;
Organic molecular devices
Molecular technology see Bottom–up technology
MoleSensing Program, 107
Molybdenum, 218
Momentum balance equation, 183
Monte Carlo particle methods, 180–2
augmented for advanced MOSFETs, 190
for carbon nanotubes, 211–12, 214
for FinFETs, 195, 196
for ionic channels, 238–41
Moore’s law, 7, 168, 171
MOSFETs, 134, 189–90, 213–14, 216, 217
CNT, 213–14
high energy tail determination, 180
silicon/oxide interface, 226–7
structure, 168
MTRAN, 27
Mulliken charge analysis, 208
Mushroom body, 121
Myosin, 90

N
Nanobelts, 150–1, 152
Nanobiology, 264
Nanobiomotors, 88–91
cell-based robots, 91
nanorobots for medicine application, 88–90
protein nanomotors, 90–1
Nanobiosensors, 75–87
environmental, 87
hybrid, 87
magnetic immunoassay, 86
see also Nanoelectrical sensors;
Nanomechanical sensors;
Nanoptical sensors
Nanobiotechnology, 40–1, 74–92, 264
definition, 74
for drug delivery, 91–2
future, 92
for sensing and actuating, 74–91
significance, 74
see also Nanobiomotors; Nanobiosensors
for therapeutics, 92
see also Nanofabrication
Nanocomputers, 7–9
Nanocrystals, 265, 269
NanoDesign, 11
Nanodevices, functional, 73–4
Nanoelectrical sensors, 81–7
amperometric biosensors, 81
conductometric biosensors, 82
hybrid impedimetric biosensors, 82–5
potentiometric biosensors, 82
Nanoelectromechanical systems (NEMS),
CNTs in, 208, 217–20
Nanoenabling technologies, 260
Nanoethics
dimensions, 286
see also Ethics policies; Socio-technical integration
Nanofabrication, 6–7, 42–74
atomic (ANF), 53
bottom-up methods, 6, 7, 47–60
counterparts in conventional scale, 47–8
nanotube/nanowire growth, 57–60, 209,
217–18
see also Atom lithography; Self-assembly
emerging technologies, 67–70
functional nanomaterials, 64–7
future direction, 70–4
replication methods, 60–3, 69
hot embossing, 63
nanoimprint, 60–3, 67–9
top–down methods, 6–7, 42–7
counterparts in conventional scale, 42–3
electron beam lithography (EBL), 6,
43–4, 49
focused ion beam (FIB) lithography, 44–5
precision machining, 45–7
Nanohorns, 7
nano-HUB, 240
Nanoimprint lithography (NIL), 60–3, 67–9
Nanomanipulators, 4–5, 11
Nanomechanical sensors, 76–81
AFM-based, 77–9
design, 70–4
polymer, 80–1
working principle, 76–7
Nanomedicine, 4, 9–10
Nanomembranes, 157
Nanomemory sensing system, 155–6
Nanometer, 252
Nanomotion planning, 72
Nanooptical sensors, 85–6
biobarcode assay, 86
contrast enhancement, 86
plasmon-resonant nanoparticles, 85–6
Nanoparticles, 64–5, 258, 265–6
definition, 270
plasmon-resonant, 85–6
risk, 266
see also Quantum dots
Nanoparticle structures, 225
Nanoscale structures, 49–51
Nanomechanical sensors, 76–81
Nanorobotic manipulation, 69, 70
Nanorobots, 9
for medicine application, 88–90
Nanorods, 66–7
Nanosensors, 143–51
  biological entities as, 87
  nanobelts as, 150–1, 152
  nanowires as, 82, 147–9, 152–3
R&D overview, 144
  see also Carbon nanotube sensors
Nanostructures
  active, 10
  passive, 10
Nanosurgery, 71–2
Nanosystems
  molecular, 10
  see also Organic molecular devices
  systems of, 10
Nanotechnology
  computer science benefits for, 3, 10–14
  definitions, 257, 258–63, 267–8, 284–5
  development, 4–10
  and ethics, 277–80
  see also Ethics policies; Nanoethics; Socio-technical integration
  funding race, 277
  future, 256–7
  legislation, 280–1, 282
  “long term”, 284
  machine-phase, 253, 255
  mainstream, 284, 285
  molecular (MNT), 278–9, 284–5
  “near term”, 284
  politics, 256–7, 261, 262–3
  public perceptions, 258
  R&D conduct, 282–3
  reality, 252, 254, 255, 260
  research scope, 284–5
  science perspectives, 263–7
  scope, 254–6
  self-assembling, 26–33
  see also Self-assembly
  societal implications, 256–7, 277, 279
  see also Nanoethics; Socio-technical integration
  status, 274
  threats, 256–7, 266, 267
  vision, 252, 253, 254, 255, 260
  see also Perceptive particle swarm optimisation; Swarm intelligence
Nanotubes, carbon see Carbon nanotubes
  Nanowires, 8, 66–7, 269
  as conductometric sensors, 82, 147–9, 152–3
  consideration for electronic noses, 152–3
  crossbar arrays of, 155
  in field-effect transistors, 84–5
  as nanocrystals, 265
  in nanomemory sensing system, 155–6
  sensing array of, 154–5
  silicon (SiNWs), 148, 149
  tin dioxide (SnO2), 148–50, 152, 153
  National Institute for Occupational Safety and Health, 275
  National Institute for Standards and Technology, 275
  National Nanotechnology Initiative (NNI), 10, 254, 259
  definitions of nanotechnology, 285
  foundation, 259, 274, 277
  goals, 275
  NRC review, 280
  research sponsored by, 107
  strategic plan, 281
  National Research Council (NRC), 280, 282
  National Science Foundation (NSF)
  nanotechnology economic impact estimate, 277
  nanotechnology funding, 259, 275, 281
  and nanotechnology societal implications, 279
  social sciences role, 283
  Nature (London), 261, 263
  Nature Nanotechnology, 261, 262
  Near Grid (NG) assignment, 238
  NEGF, 208, 215, 216, 217
  Nernst–Planck equation, 232–3, 234
  Netted systems, 26, 27
  Network for Computational Nanotechnology, 240
  NG assignment, 238
  NNI see National Nanotechnology Initiative
  Nose, artificial see Electronic noses
  Nose-on-a-chip, 135–6
  NoseChips, 133
  NRC, 280, 282
  NSF see National Science Foundation
O

Observation directions, 23
Odor identification, 119–22
- combinatorial coding scheme for, 118, 119
Odor space, 121
- characterizing, 122
Olfaction
- operational principles, 125
- physiology, 108–25
- see also Odor identification; Odor space;
  Olfactory bulb; Olfactory cortex;
  Olfactory epithelium

Olfactory bulb (OB), 110, 115–19
- signal pre-processing, 115–17
- temporal signaling, 118–19
- zones, 117–18
- flow chart of process, 123
Olfactory cortex, 119–22
- areas, 121
- sensory map, 120
- temporal signaling, 120
Olfactory epithelium, 109–14
- receptor neurons, 109–12, 116, 122
- stem cells, 111
- supporting cells, 111
- transduction, 112–14
- zones, 111–12
Olfactory genes, 110–11
Olfactory receptor (OR) neurons, 109–12, 116, 122
Olfactory systems
- Drosophila (fruit fly), 117
- human, 105–6
- locust, 121
- moth, 119
- mouse, 116, 120
- sample handling, 124
- sensing, 124
- signal processing, 14–15
- zebrafish, 118–19

ompF porin, 228
oNose, 141
Optical excitation response, 265
Optical fibers, 139
Optical sensors, 139–41
- see also Nanoptical sensors
Optimisation problems, combinatorial, 17

Optoelectronics see Photonics
Orbitals, carbon atom, 214–15
Organic cells, 9
Organic molecular devices, 171–2, 197–200
- see also Molecular conduction simulation
Overlap matrix, 205
Overshoot regime, 170, 187

P

PACE, 11
Particle–particle–particle–mesh (p^3M) scheme, 238
Particle steering model, 195–7
Particle swarm optimisation (PSO), 12, 16, 18–19
- PPSO vs, 20, 21
Patch clamp, planar, 83–4
Patent and Trademark Office, 261
Patenting, 263–4
Pauli exclusion principle, 229, 233
pbest position, 23
PDMS, 80
Perception radius, 23
Perceptive Particle Swarm Optimisation (PPSO), 13, 19–23
- conventional PSO vs, 20, 21
  for nanotechnology, 23–6
Periglomerular cells, 116
pH electrode, 82
Phase-locked loop (PLL) unit, 138
Pheromone, 14
Photocurable nanoimprint lithography (P-NIL), 61, 62
Photolithography, 6
Photonic band-gap materials, 266
- definition, 269
Photonic information technology, 64
Photonics, definition, 270
Physical mutator operator, 179
π-bonding, 204
Piezoelectric devices, 136–9
Plasmon-resonant nanoparticles, 85–6
PMMA, 43, 61, 69
Poisson equation, 233, 237, 238, 240, 241
Poisson–Boltzmann model, 233–4, 235
Pollution screening, 87
Polyaniline, 132, 134
### PolyBot, 27
Polymer materials, fabrication, 69
Polyphlylene, 198
Polyyprrole, 132
Potential, effective, 191
Potentiometric biosensors, 82
Power dissipation, 171
**PPSO** see Perceptive particle swarm optimisation
Precision machining, 45–7
Principal component analysis (PCA), 127
Programmable artificial cell evolution (PACE), 11
Protein nanomotors, 90–1
Proton channels, 223–4
Proton conduction, 220
Protonation states, 231
Pseudopotentials
   - *ab initio*, 181–2
   - empirical, 181–2
**PSO** see Particle swarm optimisation

### R
Random mutation operators, 179–80
Reactive ion etching (RIE), 45, 61
Real space transfer, 174
Receptor neurons, 109–12, 116, 122
Recombination process, 179–80
Rectification, molecular, 197, 199
Red-detuning, 53
Reflective awareness, 288
Relaxation times, 182
   - for energy, 184
   - microscopic, 182, 184–5
   - for momentum, 184
Resistance, of molecule, 203
Resonant tunneling diodes, 177
Responsibility, scope of, 287–8
Reverse engineering, 233
Royal Academy of Engineering, 282
Royal Society (London), 260, 282

### S
s-bots, 27
SAMs, 48–9
SAW sensors, 137, 138–9
Scanning probe microscopes (SPM), 5, 6–7
Scanning tunneling microscopy (STM), 5, 6, 200–1
   - in atom lithography, 56
Scattering Matrix approach, 178–9
Schottky barrier, 177, 198, 202, 215–16
Science, democratization, 289
Science Media Centre, 4
Security, national, 104, 158
Self-assembly, 26–33, 48–52
   - biological, 26, 27
   - closed forms, 31
   - evolutionary computation model, 32–3
   - framework, 28
   - illustration, 28–32
   - results, 30–2
   - in integrated nanofabrication, 67–9
   - lithographically induced (LISA), 51–2
   - nanophase separation, 49–51
   - open forms, 31
   - principles, 28
   - self-assembled monolayers (SAMs), 48–9
Self-diffusion coefficient, 230
Self-energy matrices, 215
Self-forces, 238
Self-organisation, 12, 15–16 components, 15–16
Self-reconfiguration, 26, 27
Semiconductor devices
conduction channels, 226–7
see also Device simulation; Field-effect transistors; FinFETs; MOSFETs
Semiconductor nanowires see Nanowires
Senate Committee on Commerce, Science and Transportation, 283
SFIL, 61–3
σ-bonding, 204
Silicon, 264–5
carrier mobility, 169
C-MOS architecture, 171
GaAs vs, 170
properties, 265
as semiconductor of choice, 168
technology maturity, 171
Single-electron tunneling (SET) transistors, 8
Size quantization, 176
Smart drugs, 10
Smell, sense of see Olfaction
Smiths Detection, 133
Socio-technical integration, 281, 284–90
scope of inquiry and interactions, 285–7
of nanotechnology research, 284–5
of participation, 288–9
of regulation, 289–90
of responsibility, 287–8
Solid-state physics, 264
Source-drain isolation, 194–5
Source-drain transmission, 215
Spectral functions, 204
Spin, definition, 268
Spintronics, 208, 253
definition, 268
STALK, 73
Step and flash imprint lithography (SFIL), 61–3
Stigmergy, 12, 15–16
STM see Scanning tunneling microscopy
Streptavidin, 82, 148
SU-8, 80
Superconductivity, 267
definition, 269
Superfluid, 269
Surface acoustic-wave (SAW) devices, 137, 138–9
Surface field, in nanophase separation, 50
Surface stain field, in nanophase separation, 50
Swarm, 12
Swarm intelligence, 12, 14–19
techniques, 16–19
Swarm-bot, 27
SwissRe, 282

“Technological backfire”, 274
Telecommunications networks, autonomous, 12
Telenanrobotics, 71
Temporal signaling, in olfaction, 118–19, 120
Termites, 14, 15, 16, 19–20
Therapeutics, nanobiotechnology for, 92
Thermal nanoimprint lithography (T-NIL), 61
Thermionic emission, 201
Thiol, 198, 199
Tight-binding (TB) schemes, 207
Tin dioxide sensor arrays, 130–1, 133–4
TNT, 140
Toluene, detection, 140, 142
Tongue, 224
Top–down technology, 2
Tour wires, 198
Toxic agent detection, 106, 107
see also Electronic noses
Toxicity, 256
Toxicology, 258
Transconductance, 213
Transduction, 112–14
Transferred-electron effect, 212–13
Transistor, definition, 269
Transmission function, 203–4
Transport Monte Carlo (TMC), 238–41
Travelling salesman problem (TSP), 17
Tumor cells, 41
Tunneling current, 5
U
Ultrasonic machining, 46
Unilever, 283

V
van der Waals interaction, 229
Velocity, average, 182
Velocity overshoot, 174, 187, 194
Virtual reality (VR), 70–3

W
Virus detection, 84–5
Visual system, 122
Volatile organic compounds (VOCs), 132

Water, properties within constriction, 230
White blood cell, artificial, 89, 90
Wigner distribution function, 176, 191
Woodrow Wilson Center, 258
Contents of Volumes in This Series

Volume 42
Nonfunctional Requirements of Real-Time Systems
   TEREZA G. KIRNER AND ALAN M. DAVIS
A Review of Software Inspections
   ADAM PORTER, HARVEY SIY, AND LAWRENCE VOTTA
Advances in Software Reliability Engineering
   JOHN D. MUSA AND WILLA EHRLICH
Network Interconnection and Protocol Conversion
   MING T. LIU
A Universal Model of Legged Locomotion Gaits
   S. T. VENKATARAMAN

Volume 43
Program Slicing
   DAVID W. BINKLEY AND KEITH BRIAN GALLAGHER
Language Features for the Interconnection of Software Components
   RENATE MOTSCHEIN-PITRIK AND ROLAND T. MITTERMEIR
Using Model Checking to Analyze Requirements and Designs
   JOANNE ATLEE, MARSHA CHECHIK, AND JOHN GANNON
Information Technology and Productivity: A Review of the Literature
   ERIK BRYNJOLFSSON AND SHINKYU YANG
The Complexity of Problems
   WILLIAM GASARCH
3-D Computer Vision Using Structured Light: Design, Calibration, and Implementation Issues
   FRED W. DEPIERO AND MOHAN M. TRIVEDI

Volume 44
Managing the Risks in Information Systems and Technology (IT)
   ROBERT N. CHARETTE
Software Cost Estimation: A Review of Models, Process and Practice
   FIONA WALKERDEN AND ROSS JEFFERY
Experimentation in Software Engineering
   SHARI LAWRENCE PFLEEGER
Parallel Computer Construction Outside the United States
   RALPH DUNCAN
Control of Information Distribution and Access
   RALF HAUSER
Asynchronous Transfer Mode: An Engineering Network Standard for High Speed Communications
   RONALD J. VETTER
Communication Complexity
Eyal Kushilevitz

Volume 45
Control in Multi-threaded Information Systems
Pablo A. Straub and Carlos A. Hurtado
Parallelization of DOALL and DOACROSS Loops—a Survey
A. R. Hurson, Joford T. Lim, Krishna M. Kavi, and Ben Lee
Programming Irregular Applications: Runtime Support, Compilation and Tools
Joel Saltz, Gagan Agrawal, Chialin Chang, Raja Das, Guy Edjlali, Paul Havlak, Yuan-Shin Hwang, Bongki Moon, Ravi Ponnusamy, Shamik Sharma, Alan Susman, and Mustafa Uysal
Optimization Via Evolutionary Processes
Srilata Raman and L. M. Patnaik
Software Reliability and Readiness Assessment Based on the Non-homogeneous Poisson Process
Amrit L. Goel and Kune-Zang Yang
Computer-Supported Cooperative Work and Groupware
Jonathan Grudin and Steven E. Poltrock
Technology and Schools
Glen L. Bull

Volume 46
Software Process Appraisal and Improvement: Models and Standards
Mark C. Paulk
A Software Process Engineering Framework
Jyrki Kontio
Gaining Business Value from IT Investments
Pamela Simmons
Reliability Measurement, Analysis, and Improvement for Large Software Systems
Jeff Tian
Role-Based Access Control
Ravi Sandhu
Multithreaded Systems
Krishna M. Kavi, Ben Lee, and Alli R. Hurson
Coordination Models and Language
George A. Papadopoulos and Farhad Arbab
Multidisciplinary Problem Solving Environments for Computational Science
Elias N. Houstis, John R. Rice, and Naren Ramakrishnan

Volume 47
Natural Language Processing: A Human–Computer Interaction Perspective
Bill Manaris
Cognitive Adaptive Computer Help (COACH): A Case Study
Edwin J. Selker
Cellular Automata Models of Self-replicating Systems
James A. Reggia, Hui-Hsien Chou, and Jason D. Lohn
Ultrasound Visualization
Thomas R. Nelson
CONTENTS OF VOLUMES IN THIS SERIES

Patterns and System Development
Brandon Goldfedder

High Performance Digital Video Servers: Storage and Retrieval of Compressed Scalable Video
Seungyup Paek and Shih-Fu Chang

Software Acquisition: The Custom/Package and Insourse/Outsource Dimensions
Paul Nelson, Abraham Seidmann, and William Richmond

Volume 48

Architectures and Patterns for Developing High-Performance, Real-Time ORB Endsystems
Douglas C. Schmidt, David L. Levine, and Chris Cleeland

Heterogeneous Data Access in a Mobile Environment – Issues and Solutions
J. B. Lim and A. R. Hurson

The World Wide Web
Hal Bergel and Douglas Blank

Progress in Internet Security
Randall J. Atkinson and J. Eric Klinker

Digital Libraries: Social Issues and Technological Advances
Hsinchun Chen and Andrea L. Houston

Architectures for Mobile Robot Control
Julio K. Rosenblatt and James A. Hendler

Volume 49

A Survey of Current Paradigms in Machine Translation
Bonnie J. Dorr, Pamela W. Jordan, and John W. Benoit

Formality in Specification and Modeling: Developments in Software Engineering Practice
J. S. Fitzgerald

3-D Visualization of Software Structure
Mathew L. Staples and James M. Bieman

Using Domain Models for System Testing
A. von Mayrhauser and R. Mraz

Exception-Handling Design Patterns
William G. Bail

Managing Control Asynchrony on SIMD Machines—a Survey
Nael B. Abu-Ghazaleh and Philip A. Wilsey

A Taxonomy of Distributed Real-time Control Systems
J. R. Acre, L. P. Clare, and S. Sastry

Volume 50

Index Part I
Subject Index, Volumes 1–49

Volume 51

Index Part II
Author Index
Cumulative list of Titles
Table of Contents, Volumes 1–49
Volume 52
Eras of Business Computing
ALAN R. HEVNER AND DONALD J. BERNDT
Numerical Weather Prediction
FERDINAND BAER
Machine Translation
SERGEI NIRENBURG AND YORICK WILKS
The Games Computers (and People) Play
JONATHAN SCHAEFFER
From Single Word to Natural Dialogue
NEILS OLE BENSON AND LAILA DYBKJAER
Embedded Microprocessors: Evolution, Trends and Challenges
MANFRED SCHLETT

Volume 53
Shared-Memory Multiprocessing: Current State and Future Directions
PER STEUSTRÖM, ERIK HAGERSTEU, DAVID I. LITA, MARGARET MARTONOSI, AND MADAN VERNGOPAL
Shared Memory and Distributed Shared Memory Systems: A Survey
KRISHNA KAUI, HYONG-SHIK KIM, BEU LEE, AND A. R. HURSON
Resource-Aware Meta Computing
JEFFREY K. HOLLINGSWORTH, PETER J. KELCHER, AND KYUNG D. RYU
Knowledge Management
WILLIAM W. AGRESTI
A Methodology for Evaluating Predictive Metrics
JASRETT ROSENBERG
An Empirical Review of Software Process Assessments
KHALED EL EMAM AND DENNIS R. GOLDENSON
State of the Art in Electronic Payment Systems
N. ASOKAN, P. JANSON, M. STEIVES, AND M. WAIDNES
Defective Software: An Overview of Legal Remedies and Technical Measures Available to Consumers
COLLEEN KOTYK VOSSLER AND JEFFREY VOAS

Volume 54
An Overview of Components and Component-Based Development
ALAN W. BROWN
Working with UML: A Software Design Process Based on Inspections for the Unified Modeling Language
GUILHERME H. TRAVASSOS, FORREST SHULL, AND JEFFREY CARVER
Enterprise JavaBeans and Microsoft Transaction Server: Frameworks for Distributed Enterprise Components
AVRAHAM LEFF, JOHN PROKOPEK, JAMES T. RAYFIELD, AND IGNACIO SILVA-LEPE
NORMAN F. SCHNEIDEWIND
Computer Technology Changes and Purchasing Strategies
GERALD V. POST
Secure Outsourcing of Scientific Computations
MIKHAIL J. ATALLAH, K. N. PANTAZOPOULOS, JOHN R. RICE, AND EUGENE SPAFFORD
Volume 55
The Virtual University: A State of the Art
LINDA HARASIM
The Net, the Web and the Children
W. NEVILLE HOLMES
Source Selection and Ranking in the WebSemantics Architecture Using Quality of Data Metadata
GEORGE A. MIHAILA, LOUIQA RASCHID, AND MARIA-ESTER VIDAL
Mining Scientific Data
NAREN RAMAKRISHNAN AND ANANTH Y. GRAMA
History and Contributions of Theoretical Computer Science
JOHN E. SAVAGE, ALAN L. SALEM, AND CARL SMITH
Security Policies
ROSS ANDERSON, FRANK STAJANO, AND JONG-HYEON LEE
Transistors and IC Design
YUAN TAUR

Volume 56
Software Evolution and the Staged Model of the Software Lifecycle
KEITH H. BENNETT, VACLAV T. RAJLICH, AND NORMAN WILDE
Embedded Software
EDWARD A. LEE
Empirical Studies of Quality Models in Object-Oriented Systems
LIONEL C. BRIAND AND JÜRGEN WÜST
RICHARD J. FATEMAN
Quantum Computing and Communication
PAUL E. BLACK, D. RICHARD KUHN, AND CARL J. WILLIAMS
Exception Handling
PETER A. BUHR, ASHIF HARJI, AND W. Y. RUSSELL MOK
Breaking the Robustness Barrier: Recent Progress on the Design of the Robust Multimodal System
SHARON OVIATT
Using Data Mining to Discover the Preferences of Computer Criminals
DONALD E. BROWN AND LOUISE F. GUNDERSON

Volume 57
On the Nature and Importance of Archiving in the Digital Age
HELEN R. TIBBO
Preserving Digital Records and the Life Cycle of Information
SU-SHING CHEN
Managing Historical XML Data
SUDARSHAN S. CHAWATHE
Adding Compression to Next-Generation Text Retrieval Systems
NIVIO ZIVIANI AND EDLENO SILVA DE MOURA
Are Scripting Languages Any Good? A Validation of Perl, Python, Rexx, and Tcl against C, C++, and Java
LUTZ PRECHELT
Issues and Approaches for Developing Learner-Centered Technology  
**CHRIS QUINTANA, JOSEPH KRAJCIK, AND ELLIOT SOLOWAY**  
Personalizing Interactions with Information Systems  
**SAVERIO PERUGINI AND NAREN RAMAKRISHNAN**  

**Volume 58**  
Software Development Productivity  
**KATRINA D. MAXWELL**  
Transformation-Oriented Programming: A Development Methodology for High Assurance Software  
**VICTOR L. WINTER, STEVE ROACH, AND GREG WICKSTROM**  
Bounded Model Checking  
**ARMIN BIERE, ALESSANDRO CIMATTI, EDMUND M. CLARKE, OFER STRICHMAN, AND YUNSHAN ZHU**  
Advances in GUI Testing  
**ATIF M. MEMON**  
Software Inspections  
**MARC ROPER, ALASTAIR DUNSMORE, AND MURRAY WOOD**  
Software Fault Tolerance Forestalls Crashes: To Err Is Human; To Forgive Is Fault Tolerant  
**LAWRENCE BERNESTHEIN**  
Advances in the Provisions of System and Software Security—Thirty Years of Progress  
**RAYFORD B. VAUGHN**  

**Volume 59**  
Collaborative Development Environments  
**GRADY BOOCH AND ALAN W. BROWN**  
Tool Support for Experience-Based Software Development Methodologies  
**SCOTT HENNINGER**  
Why New Software Processes Are Not Adopted  
**STAN RIFKIN**  
Impact Analysis in Software Evolution  
**MIKAEL LINDVALL**  
Coherence Protocols for Bus-Based and Scalable Multiprocessors, Internet, and Wireless Distributed Computing Environments: A Survey  
**JOHN SUSTERSIC AND ALI HURSON**  

**Volume 60**  
Licensing and Certification of Software Professionals  
**DONALD J. BAGERT**  
Cognitive Hacking  
**GEORGE CYBENKO, ANNARITA GIANI, AND PAUL THOMPSON**  
The Digital Detective: An Introduction to Digital Forensics  
**WARREN HARRISON**  
Survivability: Synergizing Security and Reliability  
**CRISPIN COWAN**  
Smart Cards  
**KATHERINE M. SHELFER, CHRIS CORUM, J. DREW PROCACCINO, AND JOSEPH DIDIER**
CONTENTS OF VOLUMES IN THIS SERIES

Shotgun Sequence Assembly
Mihai Pop

Advances in Large Vocabulary Continuous Speech Recognition
Geoffrey Zweig and Michael Picheny

Volume 61
Evaluating Software Architectures
Roseanne Tesoriero Tvedt, Patricia Costa, and Mikael Lindvall

Efficient Architectural Design of High Performance Microprocessors
Lieven Eeckhout and Koen De Bosschere

Security Issues and Solutions in Distributed Heterogeneous Mobile Database Systems
A. R. Hurson, J. Ploskonka, Y. Jiao, and H. Haridas

Disruptive Technologies and Their Affect on Global Telecommunications
Stan McClellan, Stephen Low, and Wai-Tian Tan

Ions, Atoms, and Bits: An Architectural Approach to Quantum Computing
Dean Copsey, Mark Oskin, and Frederic T. Chong

Volume 62
An Introduction to Agile Methods
David Cohen, Mikael Lindvall, and Patricia Costa

The Timeboxing Process Model for Iterative Software Development
Pankaj Jalote, Aweejeet Palit, and Priya Kurien

A Survey of Empirical Results on Program Slicing
David Binkley and Mark Harman

Challenges in Design and Software Infrastructure for Ubiquitous Computing Applications
Guruduth Banavar and Abraham Bernstein

Introduction to MBASE (Model-Based (System) Architecting and Software Engineering)
David Klapholz and Daniel Port

Software Quality Estimation with Case-Based Reasoning
Taghi M. Khoshgoftaar and Naem Seliya

Data Management Technology for Decision Support Systems
Surajit Chaudhuri, Umeshwar Dayal, and Venkatesh Ganti

Volume 63
Techniques to Improve Performance Beyond Pipelining: Superpipelining, Superscalar, and VLIW
Jean-Luc Gaudiot, Jung-Yup Kang, and Won Woo Ro

Networks on Chip (NoC): Interconnects of Next Generation Systems on Chip
Theocharis Theocharides, Gregory M. Link, Narayanan Vijaykrishnan, and Mary Jane Irwin

Characterizing Resource Allocation Heuristics for Heterogeneous Computing Systems

Power Analysis and Optimization Techniques for Energy Efficient Computer Systems
Wissam Chedid, Chansu Yu, and Ben Lee

Flexible and Adaptive Services in Pervasive Computing
Byung Y. Sung, Mohan Kumar, and Behrooz Shirazi
Search and Retrieval of Compressed Text  
AMAR MUKHERJEE, NAN ZHANG, TAO TAO, RAVI VIJAYA SATYA, AND WEIFENG SUN

Volume 64
Automatic Evaluation of Web Search Services  
ABDUR CHOWDHURY
Web Services  
SANG SHIN
A Protocol Layer Survey of Network Security  
JOHN V. HARRISON AND HAL BERGHEL
E-Service: The Revenue Expansion Path to E-Commerce Profitability  
ROLAND T. RUST, P.K. KANNAAN, AND ANUPAMA D. RAMACHANDRAN
Pervasive Computing: A Vision to Realize  
DEBASHIS SAHA
Open Source Software Development: Structural Tension in the American Experiment  
COSKUN BAYRAK AND CHAD DAVIS
Disability and Technology: Building Barriers or Creating Opportunities?  
PETER GREGOR, DAVID SLOAN, AND ALAN F. NEWELL

Volume 65
The State of Artificial Intelligence  
ADRIAN A. HOPGOOD
Software Model Checking with SPIN  
GERARD J. HOLZMANN
Early Cognitive Computer Vision  
JAN-MARK GEUSEBROEK
Verification and Validation and Artificial Intelligence  
TIM MENZIES AND CHARLES PECHEUR
Indexing, Learning and Content-Based Retrieval for Special Purpose Image Databases  
MARK J. HUISKES AND ERIC J. PAUWELS
Defect Analysis: Basic Techniques for Management and Learning  
DAVID N. CARD
Function Points  
CHRISTOPHER J. LOKAN
The Role of Mathematics in Computer Science and Software Engineering Education  
PETER B. HENDERSON

Volume 66
Calculating Software Process Improvement’s Return on Investment  
RINI VAN SOLINGEN AND DAVID F. RICO
Quality Problem in Software Measurement Data  
PIERRE REBOURS AND TAGHI M. KHOSHGOFTAAR
Requirements Management for Dependable Software Systems  
WILLIAM G. BAIL
Mechanics of Managing Software Risk  
WILLIAM G. BAIL
The PERFECT Approach to Experience-Based Process Evolution
   BRIAN A. NEIMEH AND WILLIAM E. RIDDLE
The Opportunities, Challenges, and Risks of High Performance Computing in Computational Science and Engineering
   DOUGLASS E. POST, RICHARD P. KENDALL, AND ROBERT F. LUCAS

Volume 67
Broadcasting a Means to Disseminate Public Data in a Wireless Environment—Issues and Solutions
   A.R. HURSON, Y. JIAO, AND B.A. SHIRAZI
Programming Models and Synchronization Techniques for Disconnected Business Applications
   AVRAHAM LEFF AND JAMES T. RAYFIELD
Academic Electronic Journals: Past, Present, and Future
   ANAT HOVAV AND PAUL GRAY
Web Testing for Reliability Improvement
   JEFF TIAN AND LI MA
Wireless Insecurities
   MICHAEL STHULTZ, JACOB UECKER, AND HAL BERGHEL
The State of the Art in Digital Forensics
   DARIO FORTE

Volume 68
Exposing Phylogenetic Relationships by Genome Rearrangement
   YING CHIH LIN AND CHUAN YI TANG
Models and Methods in Comparative Genomics
   GUILLAUME BOURQUE AND LOUXIN ZHANG
Translocation Distance: Algorithms and Complexity
   LUSHENG WANG
Computational Grand Challenges in Assembling the Tree of Life: Problems and Solutions
   DAVID A. BADER, USMAN ROSHAN, AND ALEXANDROS STAMATAKIS
Local Structure Comparison of Proteins
   JUN HUAN, JAN PRINS, AND WEI WANG
Peptide Identification via Tandem Mass Spectrometry
   XUE WU, NATHAN EDWARDS, AND CHAU-WEN TSENG

Volume 69
The Architecture of Efficient Multi-Core Processors: A Holistic Approach
   RAKESH KUMAR AND DEAN M. TULLSEN
Designing Computational Clusters for Performance and Power
   KIRK W. CAMERON, RONG GE, AND XIZHOU FENG
Compiler-Assisted Leakage Energy Reduction for Cache Memories
   WEI ZHANG
Mobile Games: Challenges and Opportunities
   PAUL COULTON, WILL BAMFORD, FADI CHEHIMI, REUBEN EDWARDS, PAUL GILBERTSON, AND OMER RASHID
Free/Open Source Software Development: Recent Research Results and Methods
   WALT SCACCHI
Volume 70
Designing Networked Handheld Devices to Enhance School Learning
   JEREMY ROSCHELLE, CHARLES PATTON, AND DEBORAH TATAR
Interactive Explanatory and Descriptive Natural-Language Based Dialogue for Intelligent Information Filtering
   JOHN ATKINSON AND ANITA FERREIRA
A Tour of Language Customization Concepts
   COLIN ATKINSON AND THOMAS KÜHNE
Advances in Business Transformation Technologies
   JUHYOUNG LEE
Phish Phactors: Offensive and Defensive Strategies
   HAL BERGHEL, JAMES CARPENTER, AND JU-YEON JO
Reflections on System Trustworthiness
   PETER G. NEUMANN