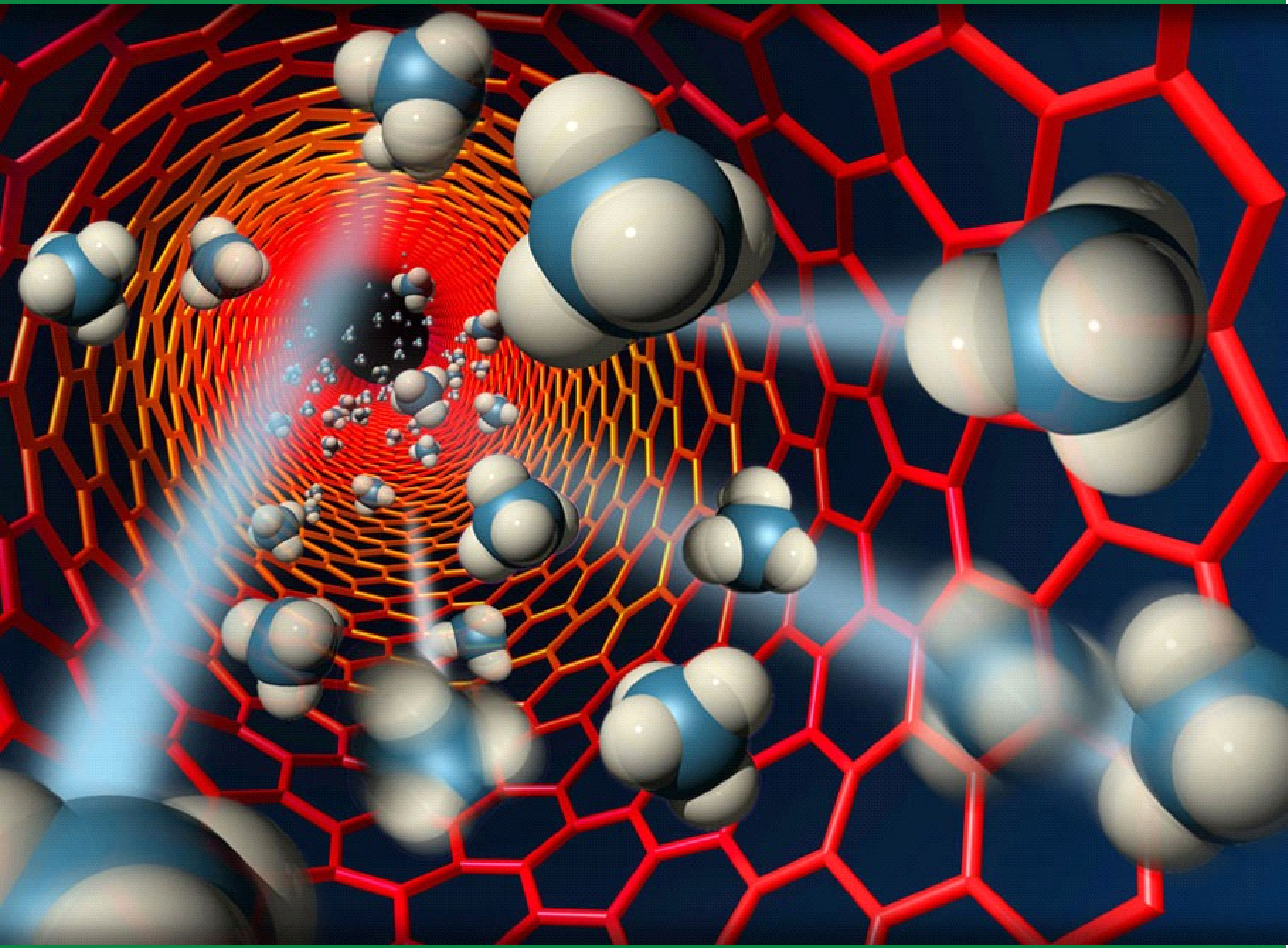
NANOTECHNOLOGY

JEREMY RAMSDEN



Jeremy Ramsden

Nanotechnology

Nanotechnology

ISBN 978-87-7681-418-2

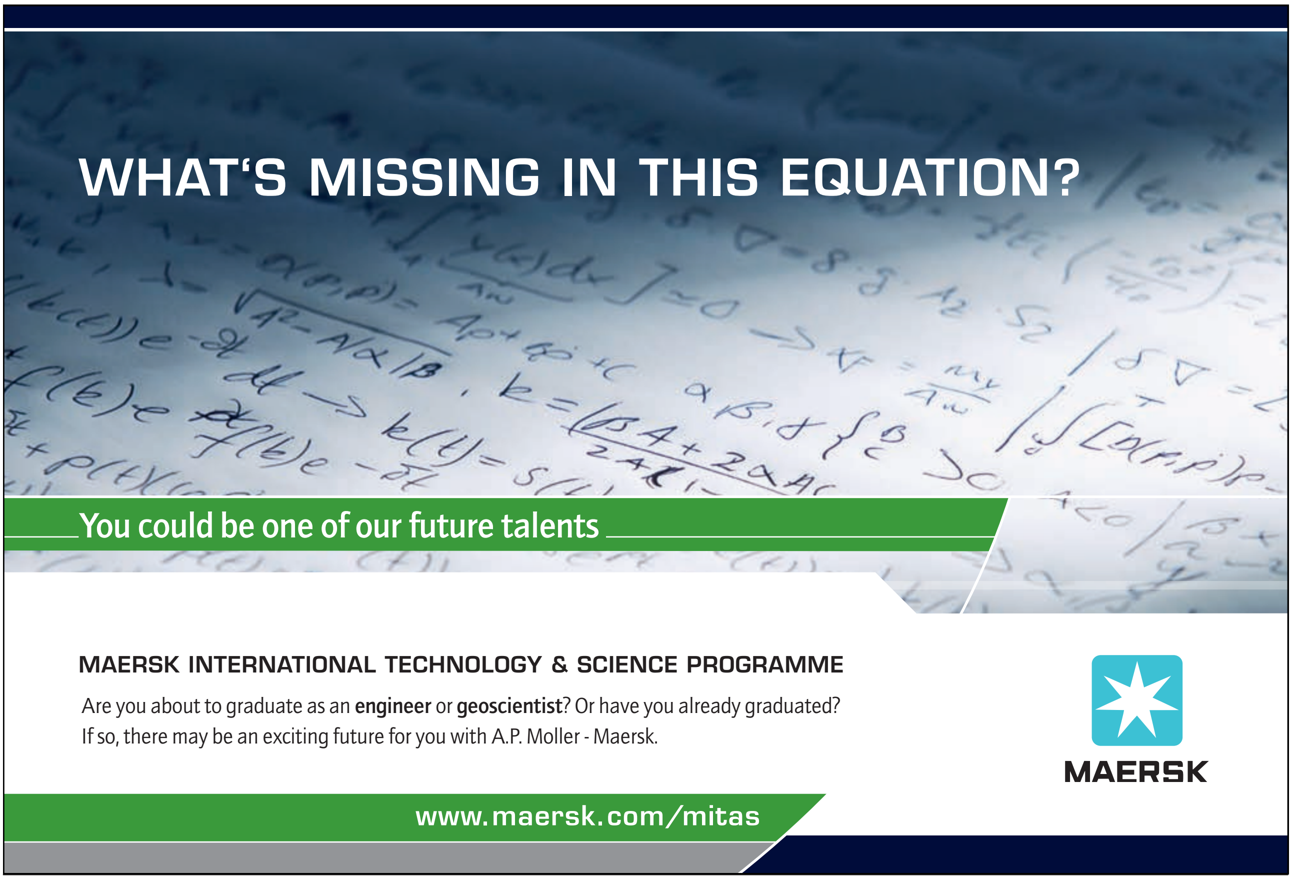
Contents

Guide to the reader 8

1. What is nanotechnology? 10
1.1 Definitions 10
1.2 History of nanotechnology 13
1.3 Context of nanotechnology 15
1.4 Further reading 19

2. Motivation for nanotechnology 20
2.1 Materials 20
2.2 Devices 24
2.3 Systems 25
2.4 Issues in miniaturization 26
2.5 Other motivations 26

3. Scaling laws 28
3.1 Materials 28
3.2 Forces 34
3.3 Device performance 35
3.4 Design 35
3.5 Further reading 37



4. Nanometrology 38
4.1 Imaging nanostructures 38
4.2 Nonimaging approaches 41
4.3 Other approaches 42
4.4 Metrology of self-assembly 45
4.5 Further reading 46

5. Raw materials of nanotechnology 47
5.1 Nanoparticles 47
5.2 Nanofibres 51
5.3 Nanoplates 51
5.4 Graphene-based materials. 53
5.5 Biological effects of nanoparticles 54
5.6 Further reading 57

6. Nanodevices 58
6.1 Electronic devices 62
6.2 Magnetic devices 64
6.3 Photonic devices 65
6.4 Mechanical devices 67
6.5 Fluidic devices 67
6.6 Biomedical devices 68
6.7 Further reading 69

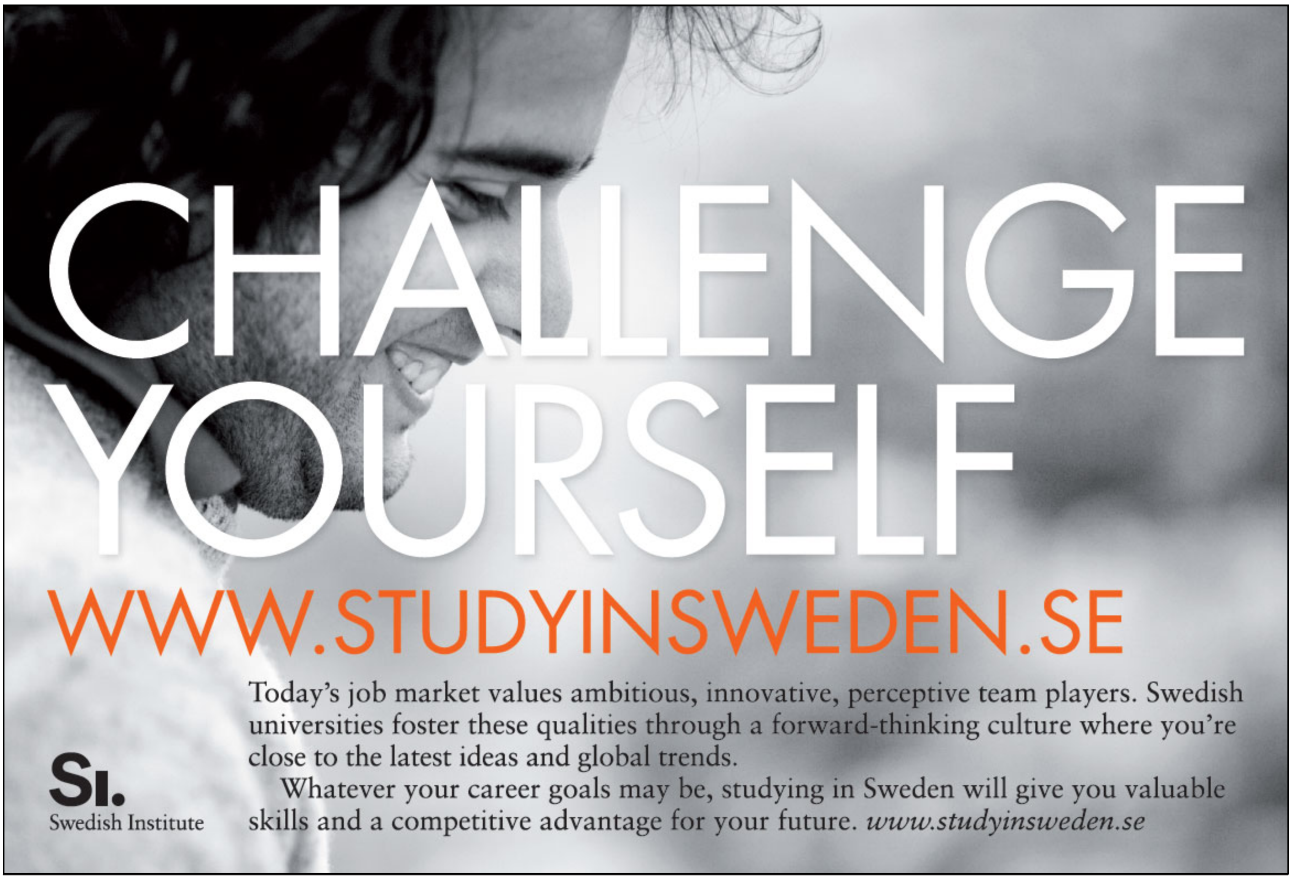
7. Nanofacture 71
7.1 Top-down methods 71
7.2 Molecular manufacturing 73
7.3 Bottom-up methods 74
7.4 Intermolecular interactions 83
7.5 Further reading 91

8. Bionanotechnology 93
8.1 Biomolecules 95
8.2 Characteristics of biological molecules 96
8.3 Mechanism of biological machines 97
8.4 Biological motors 99
8.5 The cost of control 101
8.6 Biophotonic devices 103
8.7 DNA as construction material 104
8.8 Further reading 105

9. New fields of nanotechnology 106
9.1 Quantum computing and spintronics 106
9.2 Nanomedicine 109
9.3 Energy 113
9.4 Three concepts 114
9.5 Further reading 116

10. Implications of nanotechnology **117**
10.1 Enthusiasm 117
10.2 Neutrality 119
10.3 Opposition and scepticism 120
10.4 A sober view of the future 122
10.5 Further reading 123

Index **124**



Guide to the reader

Welcome to this Study Guide to nanotechnology.

Nanotechnology is widely considered to constitute the basis of the next technological revolution, following on from the first Industrial Revolution, which began around 1750 with the introduction of the steam engine and steelmaking (and which parallelled, or perhaps caused, upheavals in land ownership and agriculture practice). The Industrial Revolution constituted as profound a change in society and civilization as the earlier Stone, Bronze and Iron revolutions, each of which ushered in a distinctly new age in the history of human civilization. A second Industrial Revolution began around the end of the 19th century with the introduction of electricity on an industrial scale (and which paved the way for other innovations such as wireless communication), and most recently we have had the Information Revolution, characterized by the widespread introduction of computing devices and the internet.

Insofar as the further development of very large-scale integrated circuits used for information processing depends on reducing the sizes of the individual circuit components down to the nanoscale (i.e., a few tens of nanometres), the Information Revolution has now become the Nano Revolution—just as steam engines powered dynamos for the industrial generation of electricity. But, nanotechnology brings its own distinctive challenges, notably: (i) handling matter at the atomic scale (which is what nanotechnology is all about—a synonym is "atomically precise engineering") means that *qualitatively different behaviour* needs to be taken into account; and (ii) in order for atomically precisely engineered objects to be useful for humans, they need to be somehow multiplied, which introduces the problem of *handling vast numbers of entities*.

One should not underestimate the multidisciplinary nature of nanotechnology. This forces researchers to adopt a manner of working more familiar to scientists in the 19th century than in the 21st. Many active fields in nanotechnology research demand an understanding of diverse areas of science. Sometimes this problem is solved by assembling teams of researchers but members of the team still need to be able to effectively communicate with one another. An inevitable consequence of this multidisciplinarity is that the range of material that needs to be covered is rather large. As a result, some topics have had to be dealt with rather sketchily in order to keep the size of this book within reasonable bounds, but I hope I may be at least partly excused for this by the continuing rapid evolution of nanotechnology, which in many cases would make additional details superfluous since their relevance is likely to be soon superseded. Fundamental discoveries will doubtless continue to be made in the realm of a very small—and given the closeness of discoveries to technology in this field, in many cases they will doubtless be rapidly developed into useful products.

References to the original literature are only given (as footnotes to the main text) when I consider the original article to be seminal, or that reading it will bring some additional illumination. At the end of each chapter I list some (mostly relatively short) authoritative review articles (and a few books) that could be usefully read by anyone wishing to go into more detail. These lists do not include standard texts on topics such as the general properties of matter, electricity and magnetism, optics, quantum mechanics, and so forth.

Chapter 1

What is nanotechnology?

1.1 Definitions

Let us briefly recall the bald definition of nanotechnology: "the design, characterization, production and application of materials, devices and systems by controlling shape and size of the nanoscale".[1] The nanoscale itself is at present consensually considered to cover the range from 1 to 100 nm.[2] A slightly different nuance is given by "the deliberate and controlled manipulation, precision placement, measurement, modelling and production of matter at the nanoscale in order to create materials, devices, and systems with *fundamentally new* properties and functions" (my emphasis). Another formulation floating around is "the design, synthesis, characterization and application of materials, devices and systems that have a *functional organization* in at least one dimension on the nanometre scale" (my emphasis). The US Foresight Institute gives: "nanotechnology is a group of emerging technologies in which the structure of matter is controlled at the nanometer scale to produce *novel* materials and devices that have useful and *unique* properties" (my emphases). The emphasis on control is particularly important: it is this that distinguishes nanotechnology from chemistry, with which it is often compared: in the latter, motion is essentially uncontrolled and random, within the constraint that it takes place on the potential energy surface of the atoms and molecules under consideration. In order to achieve the desired control, a special, nonrandom *eutactic* environment needs to be a available. How eutactic environments can be practically achieved is still being vigorously discussed.[3]

---

[1] E. Abad et al., *NanoDictionary.* Basel: Collegium Basilea (2005).

[2] This scale (and indeed the definitions) are currently the subject of discussions within the International Standards Organization (ISO) aimed at establishing a universal terminology.

[3] E.g., F. Scott et al., NanoDebate. *Nanotechnology Perceptions* 1 (2005) 119–146.

A very succinct definition of nanotechnology is simply "engineering with atomic precision".

However, we should bear in mind the "fundamentally new properties" and "novel" and "unique" aspects that some nanotechnologists insist upon, wishing to exclude existing artefacts had happen to be small.

Another debated issue is whether one should refer to "nanotechnology" or "nanotechnologies". The argument in favour of the latter is that nanotechnology encompasses many distinctly different kinds of technology. But there seems to be no reason not to use "nanotechnology" in a collective sense, since the different kinds are nevertheless all united by (striving for) control at the atomic scale.

Elaborating somewhat on the definitions, one can expand nanotechnology along at least three imaginary axes:

1. The axis of tangible objects, in order of increasing complexity: *materials, devices and systems*. Note that the boundaries between these three can be crossed by such things as "smart" materials.
2. The axis starts with passive, static objects (such as nanoparticles) whose new (i.e., different from those of bulk objects having the same chemical composition) properties arise from their small size. It continues with active devices (e.g., able to transduce energy, or store information, or change their state)—that is, their dynamical properties are explicitly considered. Further along the axis are devices of ever more sophistication and complexity, able to carry out advanced information processing, for example. Finally, we come to manufacture (nanomanufacturing, usually abbreviated to *nanofacture*), also called atomically precise manufacturing (APM), i.e. *processes*, and *nanometrology*, which of course comprises a very varied collection of instruments and procedures. Sometimes these are considered under the umbrella of "productive nanosystems", which implies a complete paradigm of sustainable nanofacture.
3. The axis starts with *direct* nanotechnology: materials structured at the nanoscale (including nanoparticles), devices with nanoscale components, etc.; continues with *indirect* nanotechnology, which encompasses things like hugely powerful information processors based on very large scale integrated chips with individual circuit components within the nanoscale; and ends with *conceptual* nanotechnology, which means the scrutiny of engineering (and other, including biological) processes at the nanoscale in order to understand them better.

Within the context of active devices, it is often useful to classify them according to the media on which they operate—electrons, photons or liquid materials, for example. Thus, we have molecular *electronics*, and single electron devices made from scaled-down bulk materials such as silicon; nano*photonics*, which is nowadays often used as an umbrella term to cover planar optical waveguides and fibre optics, especially when some kind of information processing is involved; and nano*fluidics*, smaller versions of the already well established micromixers used to accomplish chemical reactions. This classification is, however, of only limited utility, because many devices involve more than one medium: for example, nanoelectro*mechanical* devices are being intensively researched as a way of achieving electronic switching, opto*electronic* control is a popular way of achieving photonic switching, and photochemistry in miniaturized reactors involves both nanophotonics and nanofluidics.

1.2 History of nanotechnology

Reference is often made to a lecture given by Richard Feynman in 1959 at Caltech (where he was working at the time). Entitled "There's Plenty of Room at the Bottom", he envisaged machines making the components for smaller machines (a familiar enough operation at the macroscale), themselves capable of making the components for yet smaller machines, and simply continuing the operation until the atomic realm was reached. Offering a prize of $1000 for the first person to build a working electric motor with an overall size not exceeding 1/64th of an inch, he was dismayed when a student presented him not long afterwards with a laboriously hand-assembled (i.e., using the technique of the watchmaker) electric motor of conventional design that met the specified criteria.

In Feynman we find the germ of the idea of the assembler, a concept later elaborated by Eric Drexler.[4] The assembler is a universal nanoscale assembling machine, capable not only of making nanostructured materials, but also copies of itself as well as other machines. The first assembler would be laboriously built atom by atom, but once it was working numbers would evidently grow exponentially, and when a large number became available, universal manufacturing capability, and the nano-era, would have truly arrived.

A quite different approach to the nanoscale starts from the microscopic world of precision engineering, progressively scaling down to ultraprecision engineering (Figure 1.1). The word "nanotechnology" was coined by Norio Taniguchi in 1983 to describe the lower limit of this process.[5] Current ultrahigh-precision engineering is able to achieve surface finishes with a roughness of a few nanometres. This trend is mirrored by relentless miniaturization in the semiconductor processing industry. Ten years ago the focus was in the micrometre domain. Smaller features were described as decimal fractions of a micrometre. Now the description, and the realization, is in terms of tens of nanometres.

A third approach to nanotechnology is based on self-assembly. Interest in this arose because, on the one hand, of the many difficulties in making Drexlerian assemblers, which would appear to preclude their realization in the near future, and on the other hand, of the great expense of the ultrahigh precision approach. The inspiration for self-assembly seems to have come from the work of virologists who noticed that pre-assembled components (head, neck, legs) of bacteriophage viruses would further assemble spontaneously into a functional virus merely

---

[4] K.E. Drexler, Molecular engineering: an approach to the development of general capabilities for molecular manipulation. *Proc. Natl Acad. Sci. USA* 78 (1981) 5275–5278.

[5] N. Taniguchi, On the basic concept of nano-technology. *Proc. Intl Conf. Prod. Engng Tokyo, Part II* (Jap. Soc. Precision Engng).

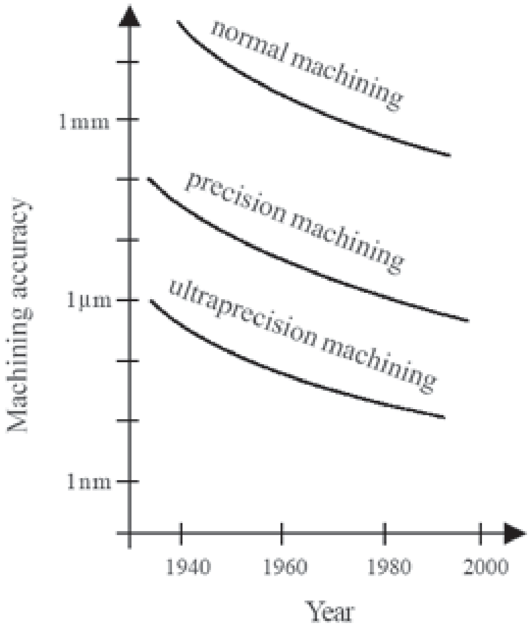


Figure 1.1: The evolution of machining accuracy (after Norio Taniguchi).

upon mixing and shaking in a test-tube.

Nanoparticles mostly rank as passive nanostructures. At present, they represent almost the only part of nanotechnology with commercial significance. However, it is sometimes questioned whether they can truly represent nanotechnology because they are not new. For example, the Flemish glassmaker John Utynam was granted a patent in 1449 in England for making stained glass incorporating nanoparticulate gold; the Swiss medical doctor and chemist von Hohenheim (Paracelsus) prepared and administered gold nanoparticles to patients suffering from certain ailments in the early 16th century. The fabrication of nanoparticles by chemical means seems to have been well established by the middle of the 19th century (e.g., Thomas Graham's method for making ferric hydroxide nanoparticles). Wolfgang Ostwald lectured extensively in the USA, and wrote up the lectures in what became a hugely successful book on the subject, "Die Welt der vernachlässigten Dimensionen" (published in 1914). Many universities had departments of colloid chemistry, at least up to the middle of the 20th century, then slowly the subject seemed to fall out of fashion, until its recent revival as part of nanotechnology.

1.3 Context of nanotechnology

Scientific revolutions. The development of man is marked by technological breakthroughs. So important are they that the technologies (rather than, for example, modes of life) give their names to the successive epochs: the Stone Age, the Bronze Age, the Iron Age, rather than the age of hunting, pastoralism, agriculture, urbanization etc. The most significant change in our way of life during the last two or three millennia was probably that brought about by the Industrial Revolution that began in Britain around the middle of the 18th century; by the middle of the 19th century it was in full swing in Britain and, at first patchily, but later rapidly, elsewhere in Europe and North America. This in turn was replaced by the Information Revolution, marked by unprecedented capabilities in the gathering, storage, retrieval and analysis of information, and heavily dependent upon the high-speed electronic digital computer. We are still within that epoch, but the next revolution already appears to be on the horizon, and it is thought that it will be the Nano Revolution.

There are a couple of things worth noting about these revolutions. Firstly, the exponential growth in capabilities. This is sometimes quite difficult to accept, because an exponential function is linear if examined over a sufficiently small interval, and if the technology (or a technological revolution) unfolds over several generations, individual perceptions tend to be strongly biased towards linearity. Nevertheless, empirical examination of available data shows that exponential development is the rule (Ray Kurzweil has collected many examples, and in our present epoch the best demonstration is probably Moore's law), although it does not continue indefinitely, but eventually levels off. Secondly, very often a preceding technological breakthrough provided the key to a successive one. For example, increasing skill and knowledge in working iron was crucial to the success of the steam power and steel that were the hallmarks of the Industrial Revolution, which ultimately developed the capability for mass production of the very large-scale integrated electronic circuits needed for realizing the Information Revolution.

Why do people think that the next technological revolution will be that of nanotechnology? Because once we have mastered the technology, the advantages of making things "at the bottom" will be so overwhelming it will rapidly dominate all existing ways of doing things. Once iron-making and working had been mastered, no one would have considered making large, strong objects out of bronze; no one uses a slide rule now that electronic calculators are available; and so forth.

What are the advantages of nanotechnology? They follow from miniaturization, novel combinations of properties, and a universal fabrication technology. Typical of the benefits of miniaturization is the cell (mobile) phone. The concept was developed in the 1950s, but the necessary circuitry would have occupied a large multistorey building using the technology of the day (thermionic valves). Materials made with carbon nanotubes can be light and very strong, and transparent and electrically conducting. Universal fabrication, based on assemblers (personal nanofactories) would enable most artefacts required by humans to be made out of acetylene and a source of energy.

How close are we to realizing the Nano Revolution? Miniaturization of circuitry is already far advanced. Components and chips can now be made with features in the size range of tens of nanometres. The World Wide Web would be scarcely conceivable without the widespread dissemination of powerful personal computers enabled by mass-produced integrated circuits. Materials based on carbon nanotubes are still very much at the experimental stage. Nevertheless, prototypes have been made and the difficulties look to be surmountable. Assembly-based nanofacture seems still to be some way in the future. To demonstrate feasibility, computer simulations are generally adduced, together with biological systems (e.g., the rotary motor, a few nanometres in diameter, which is at the heart of the ubiquitous enzyme ATPase, found in abundance in practically all forms of life). Nevertheless, actual experiments demonstrating assembly with atomic precision are still in a primitive state.

What might the benefits be? Reports published during the last few years are typically euphoric about nanotechnology and all the benefits it will bring. Many of the examples are, however, of a relatively trivial nature and do not seem to represent sufficient breakthrough novelty to constitute a revolution. Thus, we already have nanostructured textiles that resist staining, self-cleaning glass incorporating nanoparticulate photocatalysts capable of decomposing dirt (Figure 9.3); nanoparticle-based sun creams that effectively filter out ultraviolet light without scattering it and are therefore transparent; even lighter and stronger tennis racquets made with carbon fibre or even carbon nanotube composites; and so forth. None of these developments can be said to be truly revolutionary in terms of impact on civilization. The Industrial Revolution was very visible because of the colossal size of its products: gigantic bridges (e.g., the Forth bridge), gigantic steamships (e.g., the *Great Eastern*), and, most gigantic of all if the entire network is considered as a single machine, the railway. And the steel for these constructions was produced in gigantic works; a modern chemical plant or motor-car factory may cover the area of a medium-sized town. In sharp contrast, the products of nanotechnology are, by definition, very small. Individual assemblers would be invisible to the naked eye. But of course the *products* of the assemblers would be highly visible and pervasive—such as ultralight strong materials from which our built environment would be constructed.

Microprocessors grading into nanoprocessors are a manifestation of indirect nanotechnology, responsible for the ubiquity of internet servers (and hence the World Wide Web) and cellular telephones. The impact of these information processors is above all due to their very high-speed operation, rather than any particular sophistication of the algorithms governing them. Most tasks, ranging from the diagnosis of disease to surveillance, involve pattern recognition, something that our brains can accomplish swiftly and seemingly effortlessly for a while, but which require huge numbers of logical steps when reduced to a form suitable for a digital processor. Sanguine observers predict that despite the clumsiness of this “automated reasoning”, ultimately artificial thinking will surpass that of humans—this is Kurzweil’s “singularity”. Others predict that it will never happen. To be sure, the singularity is truly revolutionary, but is as much a product of the Information Revolution as of the Nano Revolution, even though the latter provides the essential enabling technology.

Conceptual nanotechnology implies scrutinizing the world from the viewpoint of the atom or molecule. In medicine this amounts to finding the molecular basis of disease, which has been underway ever since biochemistry became established, and which now encompasses all aspects of disease connected with the DNA molecule and its relatives. There can be little doubt about the tremendous advance of knowledge that it represents. It, however, is part of the more general scientific revolution that began in the European universities founded from the 11th century onwards—and which is so gradual and ongoing that it never really constitutes a perceptible revolution. Furthermore, it is always necessary to counterbalance the reductionism implicit in the essentially analytical atomic (or nano) viewpoint by insisting on a synthetic systems approach at the same time. Nanotechnology carried through to Productive Nanosystems could achieve this, because the tiny artefacts produced by an individual assembler have somehow to be transformed into something macroscopic enough to be serviceable for mankind.

Can nanotechnology help to solve the great and pressing problems of contemporary humanity? Although, if ranked, there might be some debate about the order, most people would include rapid climate change, environmental degradation, depletion of energy, unfavourable demographic trends, insufficiency of food, and nuclear proliferation among the biggest challenges. Seen from this perspective, nanotechnology is the continuation of technological progress, which might ultimately be revolutionary if the quantitative change becomes big enough to rank as qualitative. For example, atom-by-atom assembly of artefacts implies discarded ones can be disassembled according to a similar principle, hence the problem of waste (and concomitant environmental pollution) vanishes. More advanced understanding at the nanoscale should finally allow us to create artificial energy harvesting systems, hence the potential penury of energy disappears. If the manufacture of almost everything becomes localized, the transport of goods (another major contributor to environmental degradation) should dwindle to practically nothing. Localized energy production would have a similar effect. However, the achievement of this state of affairs depends on the advent of the personal nanofactory, or something resembling it, which is by no means inevitable. Perhaps the nanobot is somewhat closer to realization. Would indefatigably circulating nanobots inside our bodies enable our lives to be extended almost indefinitely? And what would be the consequences?

Nanoscience. Is there a need for this term? Sometimes it is defined as “the science underlying nanotechnology”. But this really is biology, chemistry and physics—or “molecular sciences”. It is the *technology* of designing and making functional objects at the nanoscale that is new; *science* has long been working at this scale, and below. No one is arguing that fundamentally new physics emerges at the nanoscale; rather, it is the new combinations of phenomena manifesting themselves at that scale that constitute the new technology. The term “nanoscience” therefore appears to be wholly superfluous if it is used in this sense. As a synonym of conceptual nanotechnology, however, it does have a useful meaning: the science of mesoscale approximation. The description of a protein as a string of amino acids is a good example. At the mesoscale, one does not need to inquire into details of the internal structure (at the atomic and subatomic levels) of the amino acids.

1.4 Further reading

K.E. Drexler, *Engines of Creation*. New York: Anchor Books/Doubleday (1986).

R. Feynman, There's plenty of room at the bottom. In: *Miniaturization* (ed. H.D. Gilbert), pp. 282–296. New York: Reinhold (1961).

R. Kurzweil, *The Singularity is Near*. New York: Viking Press (2005).

J.J. Ramsden, What is nanotechnology? *Nanotechnology Perceptions* 1 (2005) 3–17.

Chapter 2

Motivation for nanotechnology

In this chapter, we look at some of the reasons why one might want to make things very small, viewing nanotechnology along the "materials, devices, and systems" axis introduced in Chapter 1.

2.1 Materials

Most of the materials around us are composites. Natural materials such as wood are highly structured and built upon very sophisticated principles. The basic structural unit is cellulose, which is a polymer of the sugar glucose, but cellulose on its own makes a floppy fabric (think of cotton or rayon), hence to give it strength and rigidity it must be glued together into a rigid matrix. This is accomplished by the complex multiring aromatic molecule lignin. The design principle is therefore akin to that of reinforced concrete: Steel rods strengthen what is itself a composite of gravel and cement.

The principle of combining two or more pure substances with distinctly different properties (which might be mechanical, electrical, magnetic, optical, thermal, chemical, and so forth) in order to create a composite material that combines the desirable properties of each to create a multifunctional substance has been refined over millennia, presumably mostly by trial and error. Typically, the results are, to a first approximation, additive. Thus we might write a sum of materials and their properties like

	cellulose	high tensile strength	self-repellent
+	lignin	weak	sticky
=	wood	strong	cohesive

Empirical knowledge is used to choose useful combinations, in which the desirable properties dominate—one might have ended up with a weak and repellent material. The vast and growing accumulation of empirical knowledge, now backed up and extended by fundamental knowledge of the molecular-scale forces involved, usually allow appropriate combinations to be chosen. The *motif* of strong fibres embedded in a sticky matrix is very widely exploited, other examples being glass fibre- and carbon fibre-reinforced polymers.

Essentially, the contribution of nanotechnology to this effort is simply to take it to the ultimate level, in the spirit of "shaping the world atom-by-atom".[1]

Rather like the chemist trying to synthesize an elaborate multifunctional molecule, the materials nanotechnologist aims to juxtapose different atoms to achieve multifunctionality. This approach is known as mechanosynthetic chemistry or, in its large-scale industrial realization, as molecular manufacturing. The famous experiment of Schweizer and Eigler, in which they rearranged xenon atoms on a nickel surface to form the logo "IBM",[2] represented a first step in this direction. Since then, there has been intensive activity in the area, but it still remains uncertain to what extent arbitrary combinations of atoms can be assembled disregarding chemical concepts, and whether the process can ever be scaled up to provide macroscopic quantities of materials.

Most of the recognized successes in nanomaterials so far have been not in the creation of totally new materials through mechanosynthesis, (which is still an unrealized goal) but in the more prosaic world of blending. For example, one adds hard particles to a soft polymer matrix to create a hard, abrasion-resistant coating. As with atomically-based mechanosynthesis, the results are, to a first approximation, additive. Thus we might again write a sum like

	polypropylene	flexible	transparent
+	titanium dioxide	rigid	opaque
=	thin film coating (paint)	flexible	opaque

This is not actually very new. Paint, a blend of pigment particles in a matrix (the binder), has

---

[1] The subtitle of a report on nanotechnology prepared under the guidance of the US National Science and Technology Council Committee on Technology in 1999.

[2] E.K. Schweizer and D.M. Eigler, Positioning single atoms with a scanning tunneling microscope. *Nature (Lond.)* 344 (1990) 524–526.

been manufactured for millennia. What is new is the detailed attention paid to the nanoparticulate additive. Its properties can now be carefully tailored for the desired application. If one of components is a recognized nanosubstance—a nanoparticle or nanofibre, for example—it seems to be acceptable to describe the blend as a nanomaterial.

Terminology. According to Publicly Available Specification (PAS) 136:2007,[a] a nanomaterial is defined as a material having one or more external dimensions in the nanoscale *or* (my emphasis) which is nanostructured. It seems to be more logical to reserve the word "nano-object" (which, according to PAS 136:2007, is a synonym of nanomaterial) for the first possible meaning. This covers nanoparticles, nanorods, nanotubes, nanowires, and so forth. In principle, ultrathin paper would also be included in this definition. The term "nanostructured" is defined as "possessing a structure comprising contiguous elements with one or more dimensions in the nanoscale but excluding any primary atomic or molecular structure." This definition should probably be strengthened by including the notion of *deliberate* in it. Its use would then be properly confined to materials engineered "atom by atom". Nanoparticles in a heap are contiguous to one another, but the heap is not structured in an engineering sense, hence a collection of nanoparticles is not a nanomaterial. Substances made simply by blending nano-objects with a matrix should be called nanocomposites. "Nanosubstance" is not defined in PAS 136:2007.

---

[a] Published by the British Standards Institute.

The biggest range of applications for such nanocomposites is in thin film coatings—in other words paint. Traditional pigments may comprise granules in the micrometre size range; grinding them a little bit more finely turns them into nano-objects. Compared with transparent varnish, paint then combines the attribute of protection from the environment with the attribute of colour. The principle can obviously be (and has been) extended practically *ad libitum*: by adding very hard particles to confer abrasion resistance; metallic particles to confer electrical conductivity; tabular particles to confer low gas permeability, and so on. Two relatively old products even today constitute the bulk of the so-called nanotechnology industry: carbon black (carbon particles ranging in size from a few to several hundred nanometres) added to the rubber tyres for road vehicles as reinforcing filler; and crystals of silver chloride, silver bromide and silver iodide ranging in size from tens of nanometres to micrometres, which form the basis of conventional silver halide-based photography.

Why nanoadditives? Since it is usually more expensive to create nanosized rather than microsized matter, one needs to justify the expense of downscaling. As matter is divided ever more finely, certain properties become qualitatively different (see Chapter 3). For example, the optical absorption spectrum of silicon vapour is quite different from that of a silicon crystal, even though the vapour and crystal are chemically identical. When a crystal becomes very small, the melting point falls, there may be a lattice contraction (that is, the atoms move closer together)—these are well understood consequences of Laplace's law, and may be very useful for facilitating a sintering process. If the radius of the crystal is smaller than the Bohr radius of the electron in the bulk solid, the electron is confined and has a higher energy than its bulk counterpart. The optical absorption and fluorescent emission spectra shift to higher energies. Hence, by varying the crystal radius, the optical absorption and emission wavelengths can be tuned.

Chemists have long known that heterogeneous catalysts are more active if they are more finely divided. This is a simple consequence of the fact that the reaction takes place at the interface between the solid catalyst and the rest of the reaction medium. For a given mass, the finer the division, the greater the surface area. This is not in itself a qualitative change, although in an industrial application there may be a qualitative transition from an uneconomic to an economic process.

Our planet has an oxidizing atmosphere, and has had one probably for at least 2000 million years. This implies that most metals, other than gold, platinum and so forth (the noble metals), will be oxidized. Hence, many kinds of metallic nanoparticles will not be stable in nature.

Carbon-based materials, especially fullerenes in carbon nanotubes, are often considered to be the epitome of a nanomaterial. Carbon has long been an intriguing element because of the enormous differences between its allotropes of graphite and diamond. The carbon nanomaterials are based on another new form, graphene (see §5.4).

2.2 Devices

A device turns something into something else. Synonyms are machine, automaton, transducer, encoder, and so forth. Possible motivations for miniaturizing a device are:

1. Economizing on material. If one can accomplish the same function with less material, the device should be cheaper, which is often a desirable goal—provided that it is not more expensive to make. In the macroscopic world of mechanical engineering, if the material costs are disregarded, it is typically more expensive to make something very small; for example, a watch is more expensive than a clock, for equivalent timekeeping precision. On the other hand, when things become very large, as in the case of the clock familiarly known as Big Ben for example, costs again start to rise, because special machinery may be needed to assemble the components, and so on. We shall return to the issue of fabrication in Chapter 7.
2. Performance (expressed in terms of straightforward input-output relations) may be enhanced by reducing the size. This is actually quite rare. For most micro electromechanical systems (MEMS) devices, such as accelerometers, performance is degraded by downscaling, and the actual size of the devices currently mass-produced for actuating automotive airbags represents a compromise between economy of material, not taking up too much space nor weighing two much, the and still-acceptable performance.

Downscaling. An accelerometer (which transduces force into electricity) depends on the inertia of a lump of matter for its function, and if the lump becomes too small, the output becomes unreliable. Similarly with photodetectors (that transduce photons into electrons): due to the statistical and quantum nature of light, the smallest difference between two levels of irradiance that can be detected increases with diminishing size. On the other hand, there is no intrinsic lower limit to the physical embodiment of one bit of information. One bit could be embodied by the presence of a neutron, for example. Information processing and storage is the ideal field of application for nanotechnology. The lower limit of miniaturization is only dependent on practical considerations of "writing" and "reading" the information. Hence nanotechnology is particularly suited to information processors.

3. Functionality may be enhanced by reducing the size. Using the same example as in the previous item, it would not be practicable to equip mass-produced automobiles with macroscopic accelerometers with a volume of about 1 litre and weighing several kilograms. Another example is cellular telephony, already mentioned. A similar consideration applies to implantable biosensors for monitoring clinical parameters in a patient. In other words, miniaturization increases accessibility.

2.3 Systems

The essence of a system is that it cannot be usefully decomposed into its constituent parts. Two or more objects constitute a system if the following conditions are satisfied:

- One can talk meaningfully of the behaviour of the whole of which they are the only parts
- The behaviour of each part can affect the behaviour of the whole
- The way each part behaves and the way its behaviour affects the whole depends on the behaviour of at least one other part
- No matter how one subgroups the parts, the behaviour of each subgroup will affect the whole and depends on the behaviour of at least one other subgroup.

Typically, a single nanodevice is complex enough to be considered a system, hence a "nanosystem" generally signifies a system whose components are nanoscale devices. An example of a system that can truly be called "nano" is the foot of the gecko, many species of which can run up vertical walls and across ceilings. Their feet are hierarchically divided into tens of thousands of minute pads that allow a large area of conformal contact with irregular surfaces. The adhesive force is provided by the Lifshitz-van der Waals interaction (see §7.4), normally considered to be weak and short range, but additive and hence sufficiently strong in this embodiment if there are enough points of contact. Attempts to mimic the foot with a synthetic nanostructure have only had very limited success, because the real foot is living and constantly adjusted to maintain the close range conformal contact needed for the interaction to be sufficiently strong to bear the weight of the creature.

2.4 Issues in miniaturization

Considering the motor-car as a transducer of human desire into translational motion, it is obvious that the nanoautomobile would be useless for transporting anything other than nano-objects. The main contribution of nanotechnology to the automotive industry is in providing miniature sensors for process monitoring in various parts of the engine and air quality monitoring in the saloon; additives in paint giving good abrasion resistance, possibly self-cleaning functionality, and perhaps novel aesthetic effects; new ultrastrong and ultralightweight composites incorporating carbon nanotubes for structural parts; sensors embedded in the chassis and bodywork to monitor structural health; and so forth.

Scaling up. In other cases, scaling performance up to the level of human utility is simply a matter of massive parallelization. Nanoreactors synthesizing a medicinal drug simply need to work in parallel for a reasonably short time to generate enough of the compound for a therapeutically useful dose. With information processors, the problem is the user interface: a visual display screen must be large enough to display a useful amount of information, a keyboard for entering instructions and data must be large enough for human fingers, and so forth.

2.5 Other motivations

The burgeoning worldwide activity in nanotechnology cannot be explained purely as a rational attempt to exploit "room at the bottom". Two other important human motivations are doubtless also playing a role. One it is simply "it hasn't been done before"—the motivation of the mountaineer ascending a peak previously untrodden. The other is the perennial desire to "conquer nature." Opportunities for doing so at the familiar macroscopic scale have become very limited, partly because so much has already been done—in Europe, for example, there are hardly any marshes left to drain or rivers left to dam, two of the most typical arenas for "conquering nature"—and partly because the deleterious effects of such "conquest" are now far more widely recognized, and the few remaining undrained marshes and undammed rivers are likely nowadays to be legally protected nature reserves. But the world at the bottom, as Feynman picturesquely called it, is uncontrolled and largely unexplored.

Finally, the space industry has a constant and heavily pressing requirement for making payloads as small and lightweight as possible. Nanotechnology is ideally suited to this end user—provided nanomaterials, devices and systems can be made sufficiently reliable.

Chapter 3

Scaling laws applied to nanotechnology

The main point to be discussed in this chapter is how properties and behaviour change as the characteristic dimension is reduced. Of particular interest are discontinuous changes occurring at the nanoscale. Some very device-specific aspects of this topic are discussed in Chapter 6.

3.1 Materials

An object is delineated by its boundary. Dividing matter into small particles has an effect on purely physical processes. Suppose a spherical object of radius r is heated by internal processes, and the amount of heat is proportional to the volume $V = 4\pi r^3/3$. The loss of heat to the environment will be proportional to the surface area, $A = 4\pi r^2$. Now let the object be divided into n small particles. The total surface area is now $n^{1/3} 4\pi r^2$. This is the basic reason why small mammals have a higher metabolic rate than larger ones—they need to produce more heat to compensate for its relatively greater loss through the skin in order to keep their bodies at the same steady temperature. This also explains why so few small mammals are found in the cold regions of the earth.

Chemical reactivity. Consider a heterogeneous reaction A + B → C, where A is a gas or a substance dissolved in a liquid and B is a solid. Only the surface atoms are able to come into contact with the environment, hence for a given mass of material B the more finely it is divided the more reactive it will be, in terms of numbers of C produced per unit time.

The above considerations do not imply any discontinuous change upon reaching the nanoscale.

Granted, however, that matter is made up of atoms, the atoms situated at the boundary of an object are qualitatively different from those in the bulk (Figure 3.1). A cluster of six atoms (in two-dimensional Flatland) has only one bulk atom, and any smaller cluster is “all surface”. This may have a direct impact on chemical reactivity (considering here, of course, heterogeneous reactions). It is to be expected that the surface atoms are individually more reactive than their bulk neighbours, since they have some free valences (i.e., bonding possibilities). Consideration of chemical reactivity (its enhancement for a given mass, by dividing matter into nanoscale-sized pieces) suggests a discontinuous change when matter becomes “all surface”.

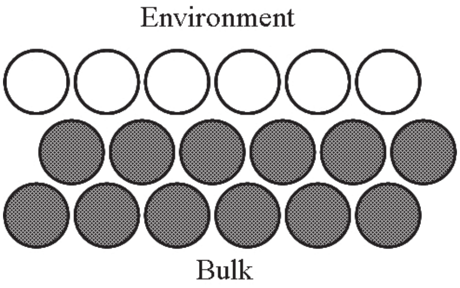


Figure 3.1: The boundary of an object shown as a cross-section in two dimensions. The surface atoms (white) are qualitatively different from the bulk atoms (grey), since the latter have six nearest neighbours (in the two-dimensional cross-section) of their own kind, whereas the former only have four.

In practice, however, the surface atoms may have already satisfied their bonding requirements by picking up reaction partners from the environment. For example, many metals become spontaneously coated with a film of their oxide when left standing in air, and as a result are chemically more inert than pure material. These films are typically thicker than one atomic layer. On silicon, for example, the native oxide layer is about 4 nm thick. This implies that a piece of freshly cleaved silicon undergoes some lattice disruption enabling oxygen atoms to effectively penetrate deeper than the topmost layer. If the object is placed in the “wrong” environment, the surface compound may be so stable that the nanoparticles coated with it are actually less reactive than the same mass of bulk matter. A one centimetre cube of sodium taken from its protective fluid (naphtha) and thrown into a pool of water will act in a lively fashion for some time, but if the sodium is first cut up into one micrometre cubes, most of the metallic sodium will have already reacted with moist air before it reaches the water.

Solubility. The vapour pressure P of a droplet, and by extension the solubility of a nanoparticle, increases with diminishing radius r according to the Kelvin equation

$$k_\mathrm{B}T\ln(P/P_0) = 2\gamma v/r \tag{3.1}$$

where k_B is Boltzmann's constant, T the absolute temperature, P_0 the vapour pressure of the material terminated by an infinite planar surface, γ the surface tension (which may itself be curvature-dependent), and v the molecular volume.

Electronic energy levels. Individual atoms have discrete energy levels and their absorption spectra correspondingly feature sharp individual lines. It is a well known feature of condensed matter that these discrete levels merge into bands, and the possible emergence of a forbidden zone (band gap) determines whether we have a metal or a dielectric.

Stacking objects with nanoscale sizes in one, two or three dimensions (yielding nanoplates, nanofibres and nanoparticles, with, respectively, confinement of carriers in two, one or zero dimensions) constitute a new class of superlattices or superatoms. These are exploited in a variety of nanodevices (Chapter 6). The superlattice gives rise to sub-bands with energies

$$E_n(k) = E_n^{(0)} + \hbar^2k^2/(2m^*) \tag{3.2}$$

where $E_n^{(0)}$ is the nth energy level, k the wavenumber, and m^* the effective mass of the electron, which depends on the band structure of the material.

Similar phenomena occur in optics, but since the characteristic size of photonic band crystals are in the micrometre range, they are, strictly speaking, beyond the scope of nanotechnology.

Electrical conductivity. Localized states with Coulomb interactions cannot have a finite density of states at the Fermi level, which has significant implications for electron transport within nanoscale material. By definition, at zero Kelvin all electronic states of a material below the Fermi level are occupied and all states above it are empty. If an additional electron is introduced, it must settle in the lowest unoccupied state, which is above the Fermi level and has a higher energy than all the other occupied states. If, on the other hand, an electron is moved from below the Fermi level to the lowest unoccupied state above it, it leaves behind a positively charged hole, and there will be an attractive potential between the hole and the electron. This lowers the energy of the electron by the Coulomb term $-e^2/(\epsilon r)$ where e is the electron charge, ϵ the dielectric permittivity, and r the distance between the two sites. If the density of states at the Fermi level is finite, two states separated by but very close to the

Fermi level could be chosen, such that the energy difference was less than $e^2/(\epsilon r)$, which would mean—nonsensically—that the electron in the upper state (above the Fermi level) has a lower energy than the electron located below the Fermi level. The gap in states that must therefore result is called the Coulomb gap, and materials with a Coulomb gap are called Coulomb glasses.

If the size of the conductor is significantly smaller than the mean free path of the electron between collisions, it can traverse the conductor ballistically, and the resistance is $h/(2e^2)$ per sub-band, independent of material parameters.

Ferromagnetism. In certain elements, exchange interactions between the electrons of adjacent ions lead to a very large coupling between their spins, such that, above a certain temperature, the spins spontaneously align with each other. The proliferation of routes to synthesizing nanoparticles of ferromagnetic substances has led to the discovery that when the particles are below a certain size, typically a few tens of nanometres, the substance still has a large magnetic susceptibility in the presence of an external field, but lacks the remanent magnetism characteristic of ferromagnetism. This phenomenon is known as superparamagnetism. There is thus a lower limit to the size of the magnetic elements in nanostructured magnetic materials for data storage, typically about 20 nm, below which room temperature thermal energy overcomes the magnetostatic energy of the element, resulting in zero hysteresis and the consequent inability to store magetization orientation information.

Electron confinement. The Bohr radius r_B of an electron moving in a condensed phase is given by

$$r_\mathrm{B} = \frac{h^2\epsilon}{e^2 m^* \pi} \tag{3.3}$$

where h is Planck's constant. Typical values range from a few to a few hundred nanometres. Therefore, it is practically possible to create particles whose radius r is smaller than the Bohr radius. In this case the energy levels of the electrons (a similar argument applies to defect electrons, positive holes) increase, and the greater the degree of confinement, the greater the increase. Hence the band edge of optical adsorption (and band-edge luminescent emission) blue shifts with decreasing r for $r < r_\mathrm{B}$. This is sometimes called a quantum size effect in the scientific literature, and nanoparticles with this property are called quantum dots.

Integrated optics. Light can be confined in a channel or plate made from a transparent material having a higher refractive index than that of its environment. Effectively, light propagates in such a structure by successive total internal reflexions at the boundaries. The channel (of fibre) can have a diameter, or the plate and thickness, less than the wavelength of the light. Below a certain minimum diameter or thickness (the cut-off), however, typically around one third of the wavelength of the light, propagation is no longer possible. The science and technology of light guided in thin structures is called integrated optics and fibre optics, and sometimes nanophotonics. However, the cut-off length is several hundred nanometres, and does not therefore truly fall into the nano realm as it is currently defined.

Chemical reactivity. Consider the prototypical homogeneous reaction A + B → C. Supposing that the reaction rate coefficient k_f is much less than the diffusion-limited rate, that is, $k_f \ll 4\pi(d_\mathrm{A} + d_\mathrm{B})(D_\mathrm{A} + D_\mathrm{B})$, where d and D are the molecular radii and diffusivities respectively. Then[1]

$$\frac{\mathrm{d}c}{\mathrm{d}t} = k_f[\langle a\rangle\langle b\rangle + \Delta^2(\gamma_t)] = k_f\langle ab\rangle \tag{3.4}$$

where a and b are the numbers (concentrations) of A and B, and the angular brackets denote expected numbers, and γ_t is the number of C molecules created up to time t. The term $\Delta^2(\gamma_t)$ expresses the fluctuations in γ_t: $\langle\gamma_t^2\rangle = \langle\gamma_t\rangle^2 + \Delta^2(\gamma_t)$: supposing that γ_t approximates to a Poisson distribution, then $\Delta^2(\gamma_t)$ will be of the same order of magnitude as $\langle\gamma_t\rangle$. The kinetic mass action law (KMAL) putting $\langle a\rangle = a_0 - c(t)$ etc., the subscript 0 denoting initial concentration at $t = 0$, is a first approximation in which $\Delta^2(\gamma_t)$ is supposed negligibly small

---

[1] See A. Rényi, Kémiai reakciók tárgyalása a sztochasztikus folyamatok elmélete segítségével. *Magy. Tud. Akad. Mat. Kut. Int. Közl.* 2 (1953) 83–101.

compared to $\langle a\rangle$ and $\langle b\rangle$, implying that $\langle a\rangle\langle b\rangle = \langle ab\rangle$, whereas strictly speaking it is not since a and b are not independent. The neglect of $\Delta^2(\gamma_t)$ is justified for molar quantities of starting reagents (except near the end of the process, when $\langle a\rangle$ and $\langle b\rangle$ become very small), but not for reactions in nanomixers.

These number fluctuations, i.e. the $\Delta^2(\gamma_t)$ term, will constantly tend to be eliminated by diffusion. On the other hand, because of the correlation between a and b, initial inhomogeneities in their spatial densities lead to the development of zones enriched in either one or the other faster than the enrichment can be eliminated by diffusion. Hence instead of A disappearing as t^{-1} (when $a_0 = b_0$), it is consumed as $t^{-3/4}$, and in the case of a reversible reaction, equilibrium is approached as $t^{-3/2}$. Deviations from perfect mixing are more pronounced in dimensions lower than three.

Occurrence of impurities. If p is the probability that an atom is substituted by an impurity, then the probability of exactly k impurities among n atoms is

$$b(k; n, p) = \binom{n}{k} p^k q^{n-k} \tag{3.5}$$

where $q = 1 - p$. If the product $np = \lambda$ is of moderate size (~ 1), the distribution can be simplified to:

$$b(k; n, p) \approx \frac{\lambda^k}{k!} e^{-\lambda} = p(k; \lambda) \tag{3.6}$$

the Poisson approximation to the binomial distribution. Hence the smaller the device, the higher the probability that it will be defect-free. The relative advantage of replacing one large device by m devices each $1/m$th of the original size is $m^{1-k} e^{np(1-1/m)}$, assuming that the nanification does not itself introduce new impurities.

Mechanical properties. The spring constant (stiffness) k of a nanocantilever varies with its characteristic linear dimension as l, and its mass m as l^3. Hence the resonant frequency of its vibration $\omega_0 = \sqrt{k/m}$ varies as $1/l$. This ensures a fast response—in effect, nanomechanical devices are extremely stiff. Since the figure of merit (quality factor) Q equals ω_0 divided by the drag (friction coefficient), Q, especially for devices operating in a high vacuum, can be many orders of magnitude greater than the values encountered in conventional devices. On the other hand, under typical terrestrial operating conditions water vapour and other impurities may condense onto moving parts, increasing drag due to capillary effects, and generally degrading performance.

3.2 Forces

The magnitudes of the forces (gravitational, electrostatic, etc.) between objects depend on their sizes and the distance z between them. At the nanoscale, gravitational forces are so weak that they can be neglected. Conversely, the range of the strong nuclear force is much smaller, and can also be neglected. Of particular importance are several forces (e.g., the van der Waals force) that are electrostatic in origin. They are discussed in Chapter 7, since they are especially important for self-assembly.

A cavity consisting of two mirrors facing each other disturbs the pervasive zero-point electromagnetic field, because only certain wavelengths can fit exactly into the space between the mirrors. This lowers the zero-point energy density in the region between the mirrors, resulting in an attractive Casimir force. The force falls off rapidly with the distance z between the mirrors (as z^{-4}), and hence is negligible at the microscale and above, but at a separation of 10 nm it is comparable with atmospheric pressure (10^5 N/m), and therefore can be expected to affect the operation of nanoscale mechanical devices.

3.3 Device performance

Analysis of device performance begins by noting how key parameters scale with device length: area (and power and thermal losses) as length squared, volume and mass as length cubed, electromagnetic force as length to the fourth power, natural frequency as inverse length, and so on. Relationships such as these are used to derive the way a device's performance scales as it is made smaller.

Natural phenomena comprising discrete noninteracting entities (such as photons) can be approximated by the Poisson distribution (eqn 3.6). A fundamental property of this distribution is that its variance equals its mean. The uncertainty (e.g., of the magnitude of a certain exposure of an object to light) expressed as a standard deviation therefore equals the square root (of exposure).

When objects become very small, the number of entities conveying information necessarily also becomes small. Small signals are more vulnerable to noise. Repetition of a message is the simplest way of overcoming noise. A nanoscale, device using only one entity (e.g., an electron) to convey one bit of information would, in most circumstances, be associated with an unacceptably high equivocation in the transmission of information.

3.4 Design

Although the most obvious consequence of nanotechnology is the creation of very small objects, an immediate corollary is that there must be a great many of these objects. If r is the relative device size, and R the number of devices, then usefulness may require that $rR \sim 1$, implying the need for 10^9 devices. This corresponds to the number of components (with a minimum feature length of about 100 nm) on a very large-scale integrated electronic chip, for example. At present, all these components are explicitly designed and fabricated. But will this still be practicable if the number of components increases by a further two and more orders of magnitude?

Because it is not possible to give a clear affirmative answer to this question, alternative routes to the design and fabrication of such vast numbers are being explored. The human brain serves as an inspiration here. Its scale is far vaster: it has $\sim 10^{11}$ neurons, and each neuron has hundreds or thousands of connexions to other neurons. There is insufficient information contained in our genes to specify the all these interconnexions. Rather, our genes specify an algorithm for generating them.

In this spirit, evolutionary design principles may become essential for designing nanodevices. An example of an evolutionary design algorithm is shown in Figure 3.2. It might be initialized by a collection of existing designs, or guesses at possible new designs. Since new variety within the design population is generated randomly, the algorithm effectively expands the imagination of the human designer.

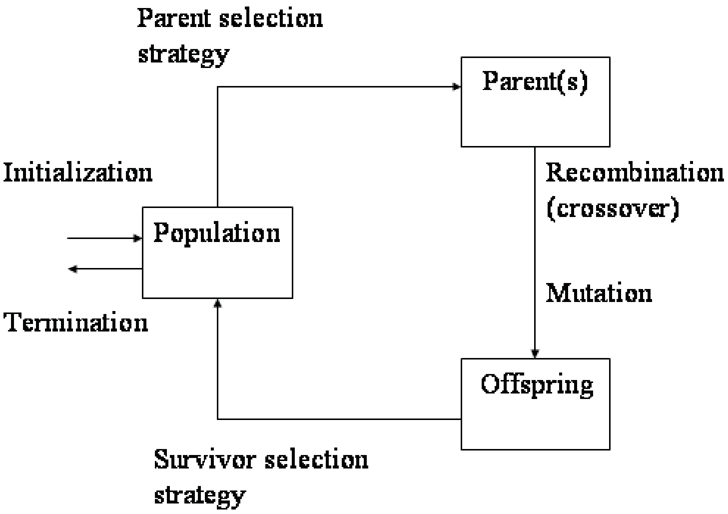


Figure 3.2: An evolutionary design algorithm. All relevant design features are encoded in the genome (a very simple genome is for each gene to be a single digit binary value indicating absence (0) or presence (1) of a feature). The genomes are evaluated ("survivor selection strategy")—this stage could include human (interactive) as well as automated evaluation—and only genomes fulfilling the evaluation criteria are retained. The diminished population is then expanded in numbers and in variety—typically the successful genomes are used as the basis for generating new ones via biologically-inspired processes such as recombination and mutation.

Although this strategy enables the design size (i.e., the number of individual features that must be explicitly specified) to be expanded practically without limit, one sacrifices knowledge of the exact internal workings of the device, introducing a level of unpredictability into device performance that may require a new engineering paradigm to be made acceptable.

Component failure and redundancy. As the number of components on a “chip” is increased, it may become more cost-effective to build in functional redundancy, such that failures of some of the components will not affect the performance of the whole (more explicitly, their failure would be detected by their congeners, who would switch in substitutes). Eqn (3.5) can be used to estimate likely numbers of failures, as a first approximation, considering them to all occur independently of each other.

3.5 Further reading

W. Banzhaf et al., From artificial evolution to computational evolution. *Nature Reviews Genetics* 7 (2006) 729–735. A research agenda.

C. Hierold, From micro- to nanosystems: mechanical sensors go nano. *J. Micromech. Microengng* 14 (2004) S1–S11. Quantitative analysis of performance scaling with device size.

Please click the advert

FOSS

Sharp Minds - Bright Ideas!

Employees at FOSS Analytical A/S are living proof of the company value - First - using new inventions to make dedicated solutions for our customers. With sharp minds and cross functional teamwork, we constantly strive to develop new unique products - Would you like to join our team?

FOSS works diligently with innovation and development as basis for its growth. It is reflected in the fact that more than 200 of the 1200 employees in FOSS work with Research & Development in Scandinavia and USA. Engineers at FOSS work in production, development and marketing, within a wide range of different fields, i.e. Chemistry, Electronics, Mechanics, Software, Optics, Microbiology, Chemometrics.

The Family owned FOSS group is the world leader as supplier of dedicated, high-tech analytical solutions which measure and control the quality and production of agricultural, food, pharmaceutical and chemical products. Main activities are initiated from Denmark, Sweden and USA with headquarters domiciled in Hillerød, DK. The products are marketed globally by 23 sales companies and an extensive net of distributors. In line with the corevalue to be ‘First’, the company intends to expand its market position.

We offer
A challenging job in an international and innovative company that is leading in its field. You will get the opportunity to work with the most advanced technology together with highly skilled colleagues.

Read more about FOSS at www.foss.dk - or go directly to our student site www.foss.dk/sharpminds where you can learn more about your possibilities of working together with us on projects, your thesis etc.

Dedicated Analytical Solutions

FOSS
Slangerupgade 69
3400 Hillerød
Tel. +45 70103370
www.foss.dk

Chapter 4

Nanometrology

It is all very well to manufacture things with atomic precision, but how do we know we have done it? Measuring the results of fabrication, whether top-down or bottom-up, is especially important during method development.

4.1 Imaging nanostructures

Ever since the invention of the microscope in the 17th century, science has been confronted with the challenge of exploring phenomena that are not directly visible to the human eye. The same extension of the senses applies to “colours” only visible using infrared or ultraviolet radiation, sounds of a pitch too low or too high to be audible, and forces too slight to be sensed by the nerves in our fingers. Although artists sometimes maintain that there is a qualitative distinction between the visible and the invisible, scientists have not found this distinction to be particularly useful. Therefore, for them the problem of “visualizing” atoms is only technical, not conceptual.

Among the senses, it is probably fair to say that sight is pre-eminent. Therefore, we shall pay most attention to how nano-objects can be seen and located with nanometre precision.

Improvements in lenses, and other developments in microscope design, eventually enabled magnifications of about 2000-fold to be reached. With that, objects around 100 nm in size could just be visualized by a human observer peering through the eyepiece of the microscope. The classical microscope runs into the fundamental limitation of spatial resolving power Δx, due to the wavelike nature of light (Abbe):

$$\Delta x = \lambda/2(\text{N.A.}) \tag{4.1}$$

where λ is the wavelength of the illuminating light and N.A. is the numerical aperture of the microscope condenser. To address this problem, one can

- reduce the wavelength of the light
- operate in the near field, rather than the far field
- renounce direct imaging
- use a totally different approach (profiles).

Reduce the wavelength. Although shorter-wavelength varieties of radiation (ultraviolet, X-rays) are well known, as the wavelength diminishes, it becomes very hard to construct the lenses needed for the microscope. However, one of the most important results emerging from quantum mechanics is the de Broglie relationship linking wave and particle properties:

$$\lambda = h/p \tag{4.2}$$

where λ is the wavelength associated with a particle of momentum $p = mv$, where m and v are the mass and velocity, respectively, and h is Planck's constant, with a numerical value of 6.63×10^{-34} J s. Knowing the mass and velocity of a particle, we can immediately calculate the wavelength!

The electron had been discovered not long before the formulation of the de Broglie relationship, and was known to be a particle of a certain rest mass ($m_e = 9.11 \times 10^{-31}$ kg) and electrostatic charge e. We know that opposite charges attract, hence the electron can be accelerated to a desired velocity simply by application of an electric field. In other words, the wavelength can be tuned as required! Furthermore, ingenious arrangements of magnetic fields can be used to focus electron beams. The transmission electron microscope was invented by Ernst Ruska in the 1930s. Nowadays, high-resolution electron microscopy can indeed image matter down to atomic resolution. The space through which the electrons pass, including around the sample, must be evacuated, because gas molecules would themselves scatter, and be ionized by, fast-moving electrons, completely distorting the image of the sample. If the sample is very thin, the modulation (according to electron density) of electrons transmitted through the sample can be used to create an electron density map (transmission electron microscopy, TEM). Otherwise, a finely focused beam can be raster-scanned over the sample and the reflected electrons used to create a topographical image (scanning electron microscopy, SEM, first developed by Manfred von Ardenne, also in the 1930s). In this case, if the sample is not electrically conducting, a thin layer of a metal, typically palladium, must be evaporated over its surface to prevent the accumulation of those electrons that are not reflected. Alternatively, if the sample is a semiconductor with a not-too-large band gap, it might be practicable to heat it in order to make it sufficiently conducting. Continuous incremental improvements in the technology of scanning electron microscopy now makes it possible to obtain images in the presence of air at a pressure of a few thousandths of an atmosphere. This is called environmental scanning electron microscopy (ESEM). Some resolution is thereby sacrificed, but on the other hand it is not necessary to dehydrate the sample, nor is it necessary to coat it with a metal if it is nonconducting—the remaining air suffices to conduct excess electrons away.

Near field microscopy. The principle is shown in Figure 4.1. The obtainable resolution is below the diffraction limit applicable to far-field optics (eqn 4.1). The resolution depends on the fineness of the construction, especially the diameter of the optical fibre-based dielectric probe illuminating the sample. The relative motion, with subnanometre control, between sample and dielectric probe is accomplished using piezoelectric crystals (as in scanning probe microscopies, see below).

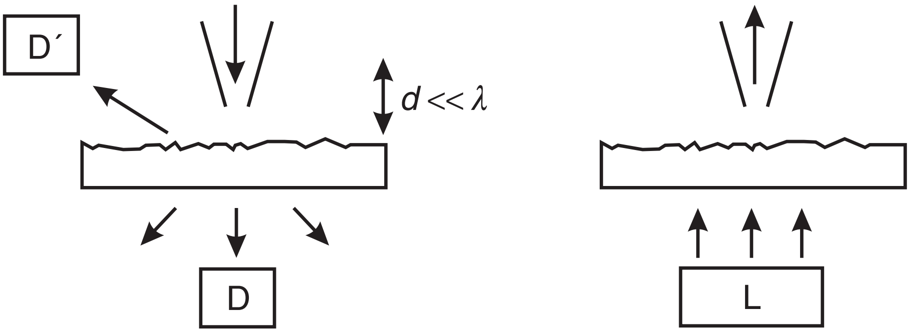


Figure 4.1: Scanning near field optical microscopy (SNOM), also known as near field scanning optical microscopy (NSOM). On the left, SNOM in illumination mode: a dielectric probe (e.g., a tapered and surface-metallized optical fibre) positioned at a distance $d \ll \lambda$ from the surface illuminates the sample from above. Either the transmitted or the reflected light is collected in the far field (detectors D or D′, respectively). On the right, SNOM in collection mode: the sample is illuminated from far below (source L). A dielectric probe in the near field collects the light transmitted through the sample.

4.2 Nonimaging approaches

The best-known approach of this type is probably X-ray diffraction. A beam of X-rays is made to impinge on the sample making an angle θ with a plane of atoms within it, and the spatial distribution of the scattered X-rays is measured. Because the wavelength λ of X-rays is of the order of interatomic-plane distance d (tenths of a nanometre), crystalline material, or at least material with some order in its atomic arrangement, diffracts the beam. The key condition for constructive interference of the reflected beam is Bragg's law:

$$d \sin\theta = n\lambda \,, n = 1, 2, \ldots \tag{4.3}$$

This metrology technique was developed soon after the discovery of X-rays by Röntgen in 1895, in other words long before the era of nanotechnology.

The distribution of collimated, typically partially coherent light scattered from a diffusely reflecting surface is suitable for determining its statistical roughness up to about a quarter of a wavelength (i.e., about 150 nm for typical visible light sources). If the surface is specularly reflecting, the illuminating light should itself be a speckle pattern, whose phase distribution is modulated by the asperity.

4.3 Other approaches

Stylus-based profilers have long been used by engineers to determine the surface roughness of objects. A sharp-pointed stylus equipped with some means of determining its vertical displacement is simply dragged over the surface perpendicular to the surface plane. This device was being progressively miniaturized, and the ability to determine subnanometre vertical displacement was achieved for the first time by the Topografiner, invented by scientists at the US National Standards Institute.[1] This nanoscale vertical resolution was achieved using the electron tunnelling effect, another quantum phenomenon (given the existence of two levels having the same energy, there is a finite probability for an electron occupying one of the energy levels to pass to the other one (if unoccupied), depending exponentially on the spatial distance separating the levels). The current measured between an electrically conducting stylus and an electrically conducting sample can therefore be converted into the sample topography. Another indispensable technological advance was the perfection of piezoelectric motion controllers in the (x, y) plane (i.e., that of the surface) and in the z direction (perpendicular to the surface). The stylus could now be raster-scanned very close to the surface. A feedback circuit can be arranged to appropriately adjust the z displacement in order to keep the tunnelling current constant. The perfected instrument was called the scanning tunnelling microscope (STM).[2]

The principle, of miniature styli moving over the sample surface and at each position returning some information about topography, or friction, or chemical nature etc. has meanwhile been vastly extended to cover dozens of different scanning probe microscopies, as the family is called. The most important (in the sense of being the most widely used) is called the atomic force microscope (Figure 4.2).[3]

It is a considerable advantage over electron microscopy that many probes can operate in air at atmospheric pressure, and even in liquids. On the other hand, the forces applied to the sample features are relatively large, and unless extreme care is taken may distort or even destroy the sample during imaging. Furthermore, the presence of liquid (even thin films of water coating the sample surface and scanning tip) gives rise to capillary forces, which may, for example, pull the tip towards the sample surface. The technique is, however, being continuously improved. An important innovation has been the introduction of "tapping mode", in which the cantilever oscillates vertically, thereby minimizing contact of the tip with the sample, and permitting the use of lock-in amplification to reduce noise.

---

[1] R. Young et al., The Topografiner: an instrument for measuring surface microtopography. *Rev. Sci. Instrum.* 43 (1972) 999–1011.

[2] G. Binnig et al., Surface studies by scanning tunneling microscopy. *Phys. Rev. Lett.* 49 (1982) 57–61.

[3] G. Binnig et al., Atomic force microscope. *Phys. Rev. Lett.* 56 (1986) 930–933.

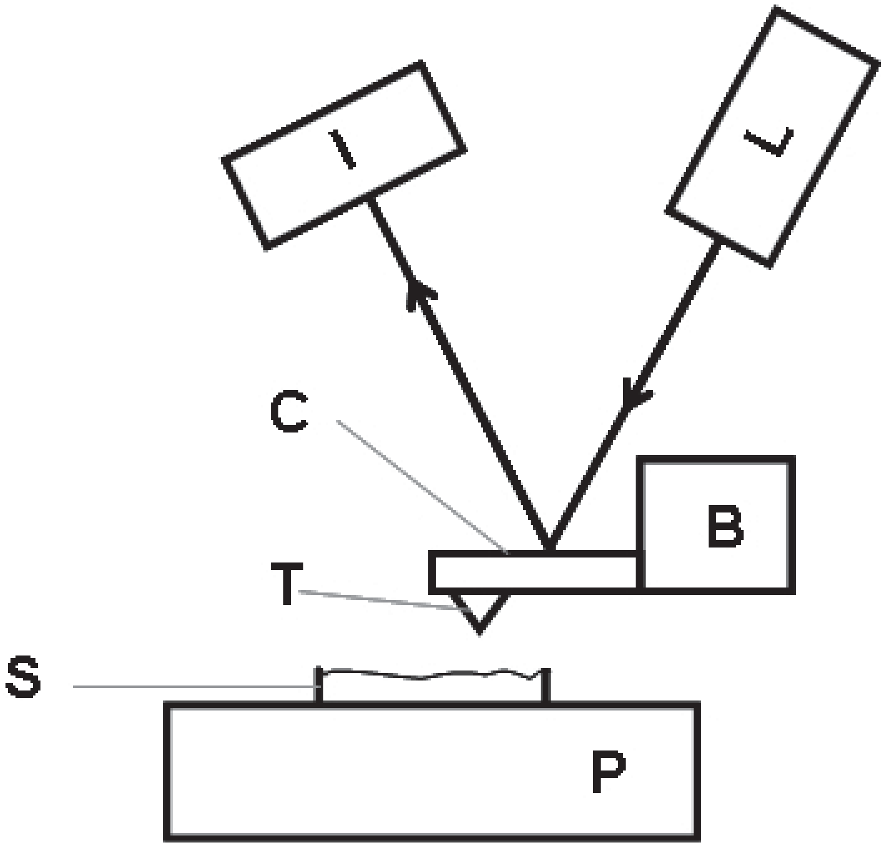


Figure 4.2: The atomic force microscope. The sample S is mounted on the platform P, in relation to which the block B can be moved in the x, y plane (parallel to the surface of S) and z direction (perpendicular to the surface of S). A flexible cantilever is mounted on the block, and on the end of the cantilever is a sharp tip. In order to record the vertical displacement of the tip as a result of the asperity of S, the beam from a diode laser L is reflected off the cantilever onto a split photodiode D. The tip is scanned across the sample (i.e., in the x, y plane) while remaining almost in contact with it; sample asperity displaces the tip vertically, and the corresponding deflexions of the cantilever are faithfully recorded as the ratio of the signals from the two halves of the photodiode, from which sophisticated signal processing allows the three-dimensional topography of the surface to be extracted. A quadruply split photodiode enables sideways deflexions of the cantilever due to friction to be recorded as well.

AFM resolution. The vertical resolution is limited only by the piezoelectric crystal that moves the sample relative to the tip, and the arrangement for detecting cantilever deflexion. Subnanometre resolution is easily obtainable. Lateral resolution is typically limited by the radius R of the tip. Current technology is able to routinely mass produce silicon or silicon nitride tips with R equal to a few tens of nanometres. The apparent lateral dimension L of a feature of radius r is

$$L = 4\sqrt{Rr} \tag{4.4}$$

For features only a few nanometres in size, ultrasharp tips have to be used, but not only are they troublesome to make (e.g., by carefully etching a standard tip), but they are also extremely fragile, easily broken, and if imaging in liquid, impurities dissolved in the liquid may rapidly deposit on the tip, increasing its radius.

An innovative solution to the problem of excessive lateral force being applied to the sample is scanning ion current microscopy, SICM (Figure 4.3).

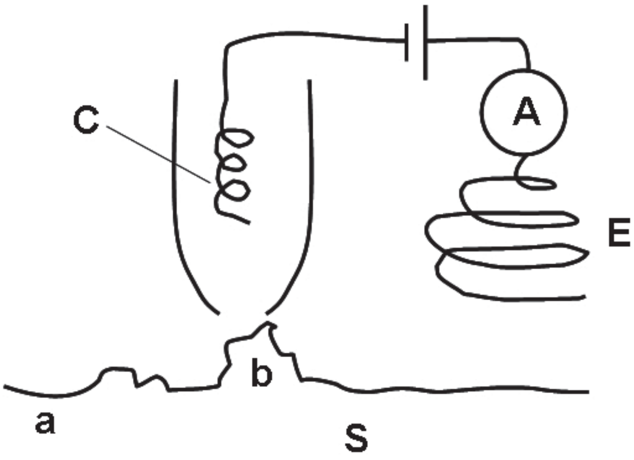


Figure 4.3: The scanning ion current microscope (SICM). A capillary C containing the working electrode is moved relative to the sample S. The magnitude of the current (measured by the ammeter A) between the working electrode and the large counterelectrode E depends on the gap between the tip of the capillary and the sample surface. Sample and electrodes are bathed in an electrolyte. When C is above feature b, the gap is small, the resistance is high on the current is low. Above feature a, the gap is relatively large, the resistance low, and the current high.

4.4 Metrology of self-assembly

A typical scenario (see §7.3.3) is to disperse nano-objects randomly in a liquid and allow them to adsorb and assemble on a solid substratum (i.e., at the solid/liquid interface). In order to establish basic relations between the self-assembled structure and the characteristics of the nano-objects, the liquid, and the substratum, some non-perturbing, *in situ* means of (a) counting the added particles, (b) determining their rate of addition, and (c) determining the structure of the self-assembled film is required.

Optical methods (scanning angle reflectometry (SAR), ellipsometry) rely on monitoring changes in the reflectance of the solid/liquid interface due to the accumulation of the nano-objects. Optical waveguide lightmode spectroscopy (OWLS) relies on perturbation of the evanescent field generated by waves guided along the substratum by the nano-objects, provided their polarizability is different from that of the liquid (Figure 4.4).[4]

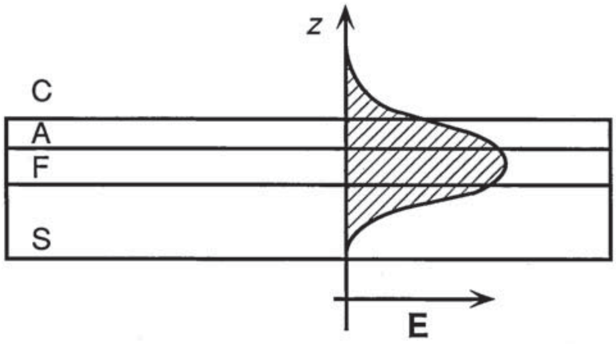


Figure 4.4: The electromagnetic field distribution of a zeroth-order guided wave in a four-layer slab waveguide (S, support; F, high refractive index film; A, self-assembled adlayer, C, cover medium (liquid in which the nano-objects would be dissolved). Note the exponentially decaying evanescent fields in the zones S and A,C. The highest sensitivity of the phase velocity of the guided modes to adlayer structure is obtainable with thin waveguides whose thickness (i.e., that of the F-layer) is close to the cut-off limit, but in this case only two orthogonal modes can be excited in the waveguide.

---

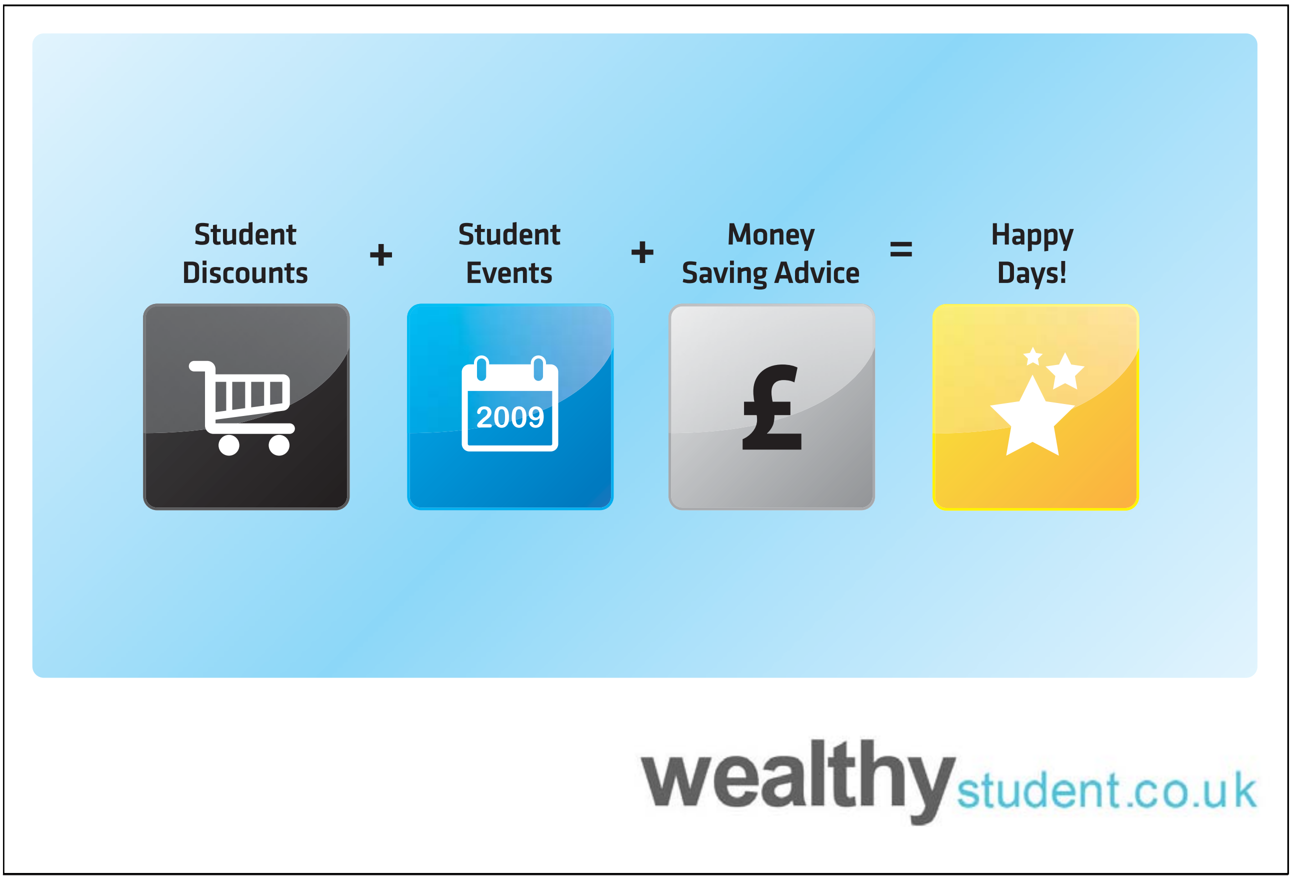
[4] See, e.g., J.J. Ramsden, High resolution molecular microscopy. In: *Proteins at Solid-Liquid Interfaces* (ed. Ph. Dejardin), pp. 23–49. Heidelberg: Springer-Verlag (2006).

If the substratum is a thin metal film, optically-excited surface plasmon resonance (SPR) can also be used to monitor the presence of the nano-objects, although it is less sensitive and less informative than OWLS. If an electrode-coated piezoelectric crystal can be used as the substratum, changes in the resonant vibration frequency of the crystal, and the dissipation of its oscillation, can also provide useful information (the quartz crystal microbalance, QCM).

4.5 Further reading

J.J. Ramsden, Experimental methods for investigating protein adsorption kinetics at surfaces. *Q. Rev. Biophys.* 27 (1994) 41–105. Comprehensive survey of different techniques useful for monitoring self-assembly processes.

M. Wieczorowski, A.G. Mamalis, M. Rucki and S.N. Lavrynenko. Interferometry and scanning microscopy in asperity measurement of biomedical surfaces. *Nanotechnology Perceptions* 4 (2008) 265–288. Excellent account of nanometrology applied to quantifying surface roughness.



Chapter 5

Raw materials of nanotechnology

This chapter is mostly about how nano-objects are manufactured. Carbon-based materials are dealt with in a separate section (§5.4) because of their unique importance.

5.1 Nanoparticles

One can use either a top-down (comminution and dispersion) or bottom-up (nucleation and growth) approach. The decision which to adopt depends on which can deliver the specified properties, and then on cost.

Comminution and dispersion means taking bulk material and fragmenting it. Crushing and grinding have typically been treated as low-technology operations. Theoretical scientists seeking to formalize phenomenological mechanistic rules (e.g., random sequential fragmentation) have found they have had little impact on the industry!

The main advantages are universality and low cost. Even soft organic matter (e.g., grass) can be ground by first freezing it in liquid nitrogen.

The main disadvantages are polydispersity of the final particles, and the introduction of many defects. Furthermore, the product may become contaminated by the material used to make the grinding machinery. The smaller the particles, the worse the contamination.

Crushing and grinding are venerable industrial processes, but the advent of nanotechnology has given rise to novel, very well controlled methods of achieving monodisperse nanoparticle generation by comminution and dispersion. One such process is electroerosion dispersion (EED);[1] in which granulated metal is ground into a fine powder by electrical discharges—typically a few hundred volts are discharged in a microsecond. The plasma temperature in the discharge filament is 10 000 to 15 000 K, sufficient to melt any metal.

Nucleation and growth describes the first-order phase tradition from an atomically dispersed phase to a solid condensed phase. During the first stage of the transition fluctuations in the homogeneous, metastable parent phase result in the appearance of small quantities of the new phase. The unfavourable process of creating an interface opposes the gain in energy through the reduction in supersaturation of the parent phase, leading to a critical size of nucleus, n^*, above which the nucleic develop rapidly and irreversibly into the new phase.

When atoms cluster together to form the new phase, they begin to create an interface between themselves and their surrounding medium, which costs energy. Denoting the interfacial tension by γ, and using subscripts 1 and 2 to denote the new phase and surrounding medium, respectively (see §7.4, the energy cost is $A\gamma_{12}$, where A is the area of the cluster's surface, equal to $(4\pi)^{1/3}(3nv)^{2/3}$, where n is the number of atoms in the cluster, and v the volume of one atom. At the same time each atom contributes to the cohesive energy of the new phase. Summing these two contributions, at first the energy will increase with increasing n, but ultimately the (negative) cohesive energy of the bulk will win (Figure 5.1).

In order to synthesize nanoparticles via nucleation and growth, firstly the atoms are dispersed (dissolved) in a medium under conditions such that the dispersion is stable. Then, one or more of the external parameters is changed such that the bulk phase of the material now dispersed is stable. This could be accomplished, for example, by cooling the vapour of the material. The formation of the new bulk phase is a first order phase transition involving nucleation. Chance fluctuations will generate critical nuclei (see Figure 5.1).

Compound particles can be synthesized by chemical reaction. Suppose the formula of the desired substance is MX, where M represents a metal such as silver or cadmium, and X a metalloid such as sulfur or selenium. One then prepares two solutions of soluble compounds of M and X (for example, silver nitrate and sodium sulfide), which are then mixed together.

Two key challenges in this process are (i) to obtain particles that are as uniform (monodisperse) as possible, and (ii) to be able to control the mean size. In the case of synthesis by chemical reaction, the key parameter is the rate of mixing. Two extreme situations yield the desired

---

[1] M.K. Monastyrov et al., Electroerosion dispersion-prepared nano- and submicrometre-sized aluminium and alumina powders as power-accumulating substances. *Nanotechnology Perceptions* 4 (2008) 179–187.

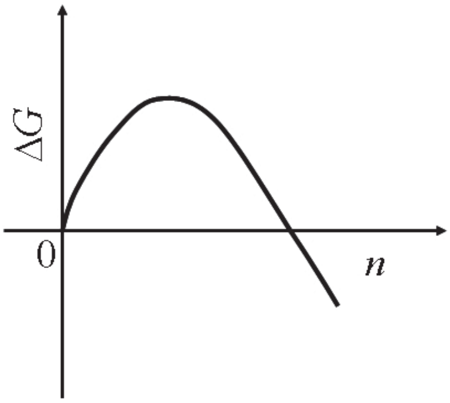


Figure 5.1: Sketch of the variation of free energy of a cluster containing n atoms. The maximum corresponds to the critical nucleus size. Clusters that have managed through fluctuations to climb up the free energy slope to reach the critical nucleus size have an equal probability to shrink back and vanish, or to grow up to microscopic size.

monodispersity: ultrarapid mixing of very concentrated solutions, and ultraslow mixing of very dilute solutions.

In the former case, a very large number of critical nuclei are formed almost simultaneously; (the rate of creation of critical nuclei is proportional to the supersaturation; that is, the ratio of the actual concentration to the solubility product of MX) growth of material onto the initially formed nuclei is too slow to compete with fresh nucleation in sinking the added mass. Conditions should be chosen such that the nuclei are just able to grow sufficiently large to be effectively irreversibly stable before all the M and X ions have been scavenged by the formation of nuclei. Further growth to any desired size can then be achieved in a separate, subsequent, stage by adding fresh material at a rate just sufficient to allow all the nuclei to grow without creating any new ones.

In the latter case, nuclei are formed extremely rarely and are unable to grow beyond the size of minimum stability because of the lack of material; diffusion of fresh material to the few nuclei formed initially is too slow to prevent new nuclei been formed in order to sink the added reagents. Once a sufficient number of nuclei has been synthesized, they can be grown up to the desired size as in the previous case. This approach is very effective for synthesizing monodisperse noble metal particles (e.g., gold) by very slowly reducing the solution of a salt of the metal.

Because of the Kelvin relation (eqn 3.1), larger particles will have a slightly lower solubility than smaller ones. Therefore, there will be a slight tendency for the smaller ones to dissolve, and for their material to be deposited onto the bigger ones. This process is known as Ostwald ripening and under certain conditions may permit the size distribution of a collection of particles to be narrowed, albeit at the price of increasing the mean size.

Once a collection of nuclei has been synthesized, it is very easy to grow shells of different materials around them; one simply needs to ensure that the new material is added at a sufficient rate to allow all the particles to grow uniformly, and not so rapidly that fresh nuclei are formed.

The interfacial free energy for aggregation of particles made from material 1 in the presence of medium 2 is given by (see §7.4):

$$\Delta G_{121} = \Delta G_{11} + \Delta G_{22} - 2\Delta G_{12} \tag{5.1}$$

where ΔG_{11} and ΔG_{22} are the cohesive energies of materials 1 and 2, and ΔG_{12} is the solvation energy. Note that water has a very large cohesive energy. Therefore, particles of almost any insoluble material synthesized in water are likely to aggregate, unless appropriate measures to ensure their hydration are taken. A useful strategy is to synthesize the particles in the presence of a very hydrophilic material such as polyethylene glycol or a polyion such as hexametaphosphate, which is able to adsorb on the surface of the particles and effectively hydrate them.

Michael Faraday's famous synthesis of gold nanoparticles used citrate ions to hydrate their surface.

5.2 Nanofibres

Terminology. "Nanofibre" is the generic term describing nano-objects with two external dimensions in the nanoscale. A nanorod is a rigid nanofibre, a nanotube is a hollow nanofibre, and a nanowire is an electrically conducting nanofibre.

Three approaches can be used to synthesize nanofibres. For some substances, under certain conditions, the natural growth habit is acicular. Therefore, the nucleation methods described in the previous section can be used to generate nuclei, followed by a growth stage to elongate them.

Heterogeneous nucleation can be induced at the solid/gas interface by predepositing small catalytic clusters. Upon addition of vapour, condensation on the clusters and growth perpendicular to the solid substratum takes place. This is used as an efficient way of synthesizing carbon nanotubes. A drawback of the method is that the preparation is almost inevitably contaminated with the catalyst.

If uniform nanopores can be formed in a membrane (e.g., by laser drilling or by self-assembly) they can be used as templates for nanofibre formation. The material for the fibre should be deposited as a shell on the inner surface of the pores (if the goal is to make nanotubes), or else should completely fill the pores (for nanorods). Nanofibres, especially nanorods, formed by either of the two previous methods can also be used as templates for making nanotubes of a different material.

5.3 Nanoplates

Until now, thin coatings on a substratum have not been considered as nano-objects, but simply as thin films, because typically they have been more than 100 nm thick. Exceptions are Langmuir films, transferred to solid substrata using the Langmuir-Blodgett and Langmuir-Schaefer techniques; these films might only be a few nanometres thick. Exceptionally laterally cohesive Langmuir films can be manipulated as free-standing objects. Nevertheless, the trend is to develop thinner functional surfaces by coating or otherwise modifying bulk material, and insofar as the coating or modification is engineered with atomic precision, it belongs to nanotechnology.

Langmuir films and the Langmuir-Blodgett and Langmuir-Schaefer techniques. The precursors are molecules of general formula XP, where X is (typically) an apolar chain (e.g. an alkyl chain), called the "tail", and P is a polar "head" group such as oligoethylene oxide, or phosphatidyl choline. When spread on water they mostly remain at the water/air interface, where they can be compressed to form two-dimensional liquid-like and solid-like arrays. The Langmuir-Blodgett technique refers to the transfer of the floating monomolecular films to solid substrata by vertically dipping them into and out of the bath. In the Langmuir-Schaefer technique, the substratum is pushed horizontally through the floating monolayer. Very stable multilayer films can be assembled by making P a chelator for multivalent metal ions, which bridge lateral neighbours and/or successive layers (assembled head-head and tail-tail). Lateral stability can be increased by UV-irradiation of films with an unsaturated alkyl chain (photopolymerization).

5.4 Graphene-based materials

Inspired by learning about naphthalene and anthracene, countless school children have doubtless doodled endless fused polyaromatic rings. It has long been known that graphite is composed of stacks of such polyaromatic sheets, which are called graphene. Due to convincing theoretical work, it was however long believed that two-dimensional crystals cannot exist. The ostensive demonstration of their existence (graphene sheets) have, *post hoc*, led to the explanation that their stability is due to undulations of the sheet.

Graphene. The graphene lamellae stacked to make bulk graphite were from the ease of their detachment (e.g., writing with graphite on paper) known to be only weakly bound to each other. Individual sheets of graphene can actually be peeled off graphite using adhesive tape. Alternatively, a crystal of silicon carbide can be heated under vacuum to 1300 °C; the silicon evaporates and the remaining carbon slowly reorganizes to form some graphene.

Carbon nanotubes. The carbon nanotube is a seamless tube made by rolling up graphene. It was long known that carbon filaments are formed by passing hydrocarbons over hot metal surfaces, especially iron and nickel. The actual nature of carbon nanotubes was however only established relatively recently (by Iijima in 1991). Multiwall carbon nanotubes consists of several concentric tubes of graphene nested inside each other. The three methods for producing carbon nanotubes are the laser furnace, the carbon arc (i.e., vaporizing graphitic electrodes), and (plasma enhanced) chemical vapour deposition (Figure 5.2). Carbon nanotubes are often closed at one or both ends by a hemisphere of fullerene.

Major problems currently remain with the large-scale utilization of carbon nanotubes. The most severe one appears to be: dispersing them in a liquid (they tend to be strongly aggregated into bundles); reducing their length (a 20 nm diameter tube may be 20 μm long as fabricated, unnecessary for many applications); and manipulating them into a desired position. Hence *in situ* growth on a flat substrate is preferred, especially for field emitting applications (see Chapter 6).

Carbon nanoparticles. Fullerenes (also known as soluble carbon or buckyballs) can be thought of as graphene curled up to form an enclosed spherical shell. They exist as C_{60}, C_{70}, etc. They can be made in a carbon arc, but burning a hydrocarbon feedstock with strict control of the oxygen supply is a more controllable method. The fullerenes can be separated

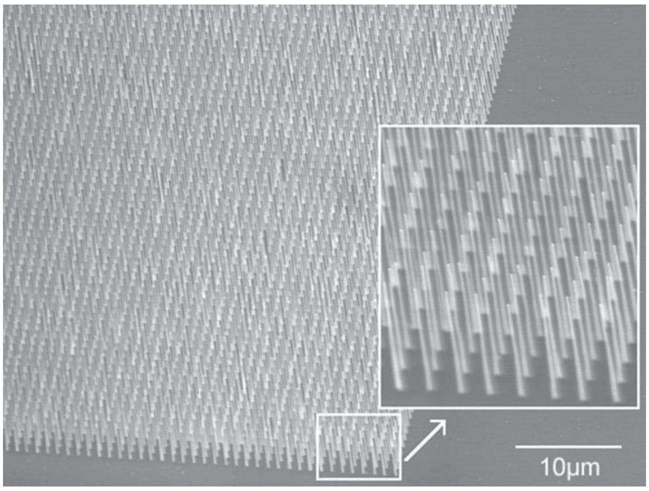


Figure 5.2: A forest of carbon nanotubes produced by plasma-enhanced chemical vapour deposition. The substratum must first be covered with metal (e.g., Fe or Ni) catalyst islands. Hydrocarbon feedstock (acetylene) is then passed over the substratum heated to several hundred °C. The acetylene decomposes at the surface of the catalyst and the carbon nanotubes grow up from the catalyst particle, or grow up beneath it(pushing it up). Courtesy of Dr Ken Teo, AIXTRON.

from coproduced soot by dissolving them out.

5.5 Biological effects of nanoparticles

The toxicity of chemicals and materials can arise in two ways:

- triggering an adverse immune response
- acting as a poison.

The immune response engendered by an artificial material in contact with the blood or tissues typically arises because proteins dissolved in the blood or other biofluids adsorb onto the surface of the material and change their conformation (generally because of an entropic driving force). The native protein is thereby transformed into a foreign protein, recognized as such by circulating immune cells, which trigger the usual apparatus for eliminating foreign invaders into action. Any immovable artificial material will become a permanent site of inflammation.

Poisoning usually has a specific biochemical mechanism. Typically, a poison binds to the active site on enzyme, preventing it from binding its customary substrate. The classic example is carbon monoxide, which binds the haem group of haemoglobin, very effectively outcompeting oxygen binding.

Most substances are not toxic in elemental form. In order to eliminate the threat from ingesting mercury ions, bacteria merely reduce them (with an enzyme called mercury reductase) to metallic mercury. Mercury-containing clinical thermometers that happen to break in the mouth of the patient are dangerous because of jagged pieces of glass, not because of the toxicity of the matter.

Materials in nano form will exacerbate these two potential sources of toxicity in three ways (this field of investigation is often called "nanotoxicity"): firstly, if a block of material is divided up and dispersed as nanoparticles, its surface area is vastly increased (by many orders of magnitude). Hence what might have been a negligible immune reaction becomes a severe one. Secondly, nanoparticles are usually far more chemically reactive than the same substance in bulk form, because of their high curvature (cf. eqn 3.1). It follows that metallic nanoparticles are more readily ionized than bulk metal. The toxicity of (for example) lead or silver arises not from the metal in elemental form, but from atoms of the metal that are detached and ionized from the block of substance. Thirdly, nanoparticles, being small, can penetrate through structures that would prevent larger particles from traversing them. In the human body, barriers are generally of two forms: the lipid bilayer enclosing individual cells, and tightly packed layers of cells. There is evidence that nanoparticles can pass between cells arranged in such tightly packed layers, such as those constituting the blood-brain barrier, and that they can pass through the lipid bilayer into the cytoplasm of individual cells much as some macromolecules of comparable size are able to do.

Relatively little is presently known about what nanoparticles do once they are actually inside the living cell. They may denature any soluble proteins with which they come into contact, but may also catalyse reactions. This will depend on their chemical nature, shape, crystal structure, and size. Artificial joints implanted inside the body are a significant source of nanoparticles, which are produced by abrasion of rubbing surfaces (Figure 5.3).

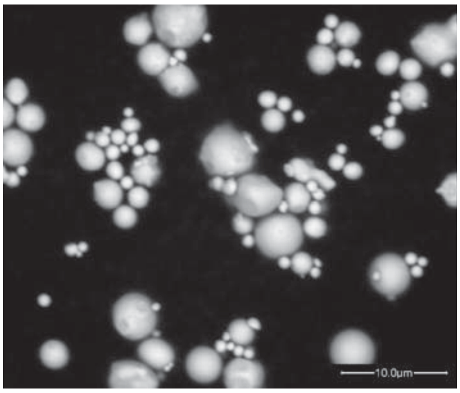


Figure 5.3: Scanning electron micrograph of Co-Cr particles retrieved from tissues adjacent to a human implant at revision surgery. The scale bar at the lower right is 10 μm long. Courtesy of Prof. Peter Revell.

The body is equipped with a number of mechanisms that are able to eliminate molecular threats. The mercury reductase enzyme has already been mentioned as an example. Such enzymes are however quite specialized, and are more likely to be found in bacteria than in higher organisms such as *H. sapiens*. We rely a great deal on our sophisticated immune system, which is, however, geared to deal with macromolecules (biopolymers, viruses, bacteria). Nanoparticles may lack the features needed to trigger an effective immune response. The shape of a nanoparticle is of particular importance. Acicular particles, such as asbestos and certain zeolites, are very problematical because they are much longer than the macrophages that attempt to engulf and destroy foreign invaders. Macrophages are simply unable to completely engulf something like an asbestos fibre, and even if they were to succeed, they would not be able to degrade it. Therefore, the presence of asbestos fibres inside the body permanently activates the macrophages, which seem to continue, indefinitely and futilely, to try to ingest and destroy the fibres.

It is sometimes pointed out that human beings have long been exposed to natural nanoparticles (e.g., smoke) and therefore (good) have developed appropriate defence mechanisms for rendering them harmless. Caution is, however, in order, because many unusually shaped particles made from exotic materials can be prepared nowadays, many of them for the first time, and wholly different from what we have already encountered.

5.6 Further reading

Arikawa Mineyuki, Fullerenes—an attractive nano carbon material and its production technology. *Nanotechnology Perceptions* 2 (2006) 121–114. Excellent account of the industrial preparation of fullerenes.

B.O. Boscovic, Carbon nanotubes and nanofibres. *Nanotechnology Perceptions* 3 (2007) 141–158. Comprehensive survey of carbon nanotubes.

G.A. Ozin and A.C. Arsenault, *Nanochemistry: A Chemical Approach to Nanotechnology.* London: RSC Publishing (2005). Encyclopaedic collection of chemical methods of nanomaterial generation.

P.A. Revell, The biological effects of nanoparticles. *Nanotechnology Perceptions* 2 (2006) 283–298. Highly informative review of the topic.

Chapter 6

Nanodevices

A device is fundamentally an information processor, in other words a transducer that encodes (or decodes) information. Binary or Boolean logic, based on zero or one, true or false, presence or absence, and so forth, has very modest physical requirements. There is essentially no intrinsic lower limit to the size of the physical embodiment of "1". Of all technologies, information technology is the one most suited to miniaturizing down to the nanoscale.

The fundamental component of a digital information processor is the switch, or relay (Figure 6.1). Several relays can be connected together to create logic gates, for example a not-and (NAND) gate, a fundamental component of a binary logic processor. Its characteristics can be summarized in the following truth table:

input 1	input 2	output
0	0	1
1	0	1
0	1	1
1	1	0

Table 6.1: Truth table for a NAND gate.

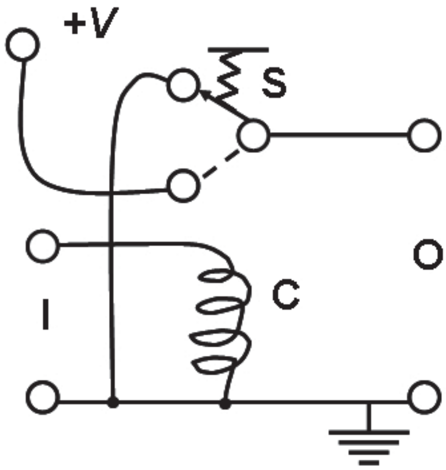
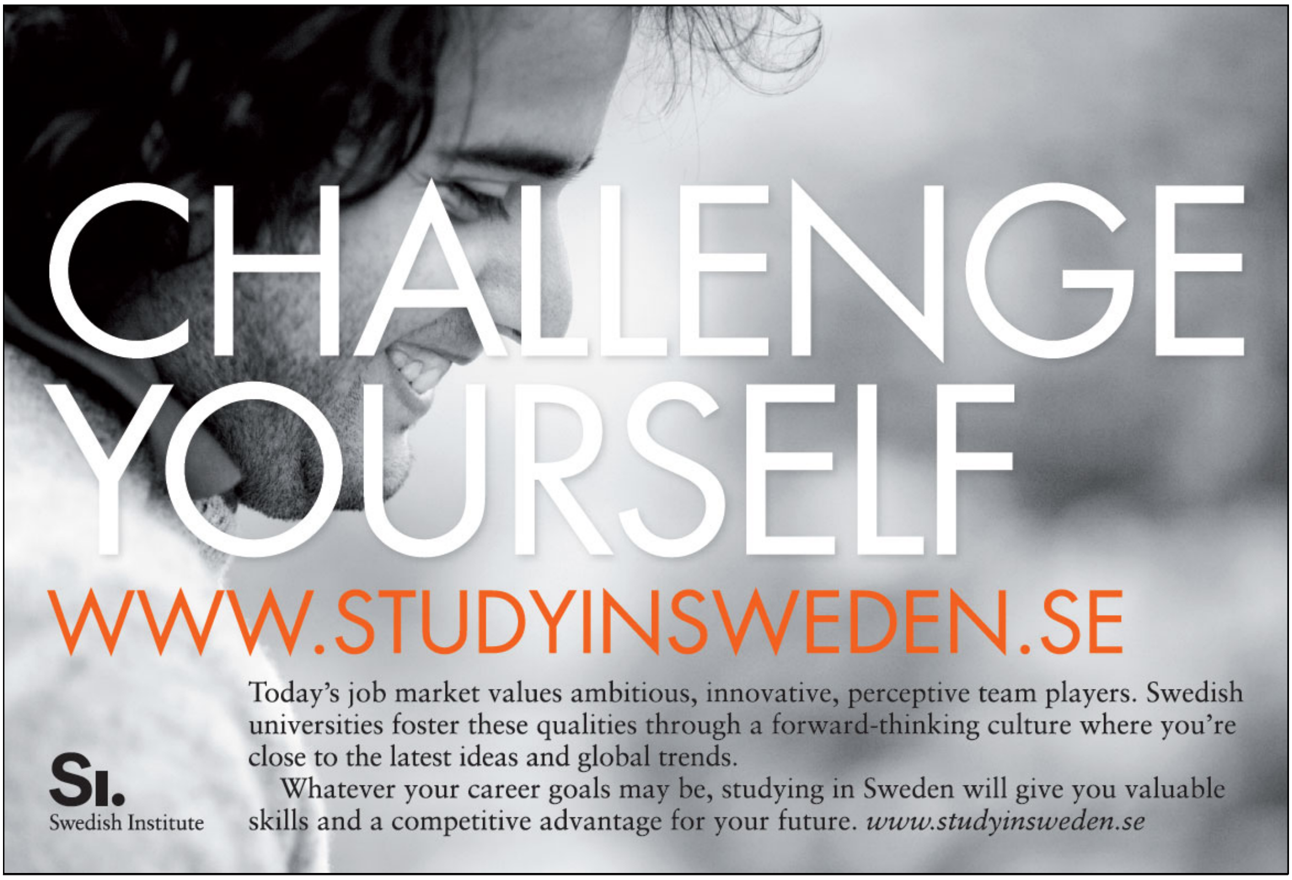


Figure 6.1: A relay or switch. When the coil C is energized by applying a voltage across the input terminals I, it pulls the movable contact arm above it to link voltage $+V$ to the output terminals O. If the restoring spring S is present, setting the input I zero will also cause the output to return to zero. Alternatively, a second appropriately placed coil could be used to move the contact arm in the opposite direction. The device is then bistable and would be suitable for use as a memory element.



The relay has the input-output relationship shown in Figure 6.2.

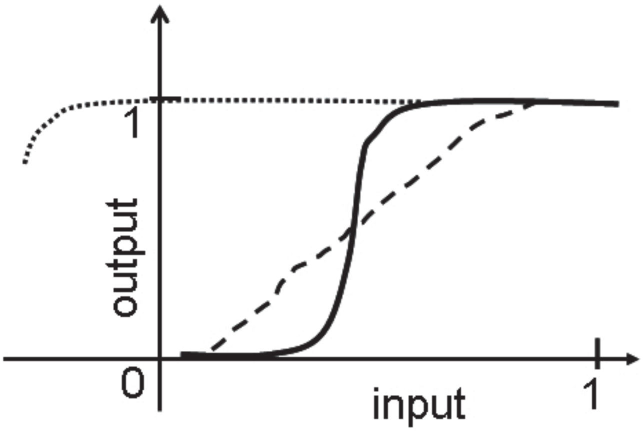


Figure 6.2: Input-output relationships approaching the ideal of a step (Heaviside) function. The input might be a voltage (e.g., applied to the coil of a relay, Figure 6.1, or the gate of a transistor, Figure 6.3) and the output might be the current flowing through the rest of the circuit. Note that in this example the response characterized by a thick solid line will give an output of one for any input (voltage) exceeding about 0.6. For the response characterized by the dashed line, the input would have to exceed about 0.8; i.e., it is less tolerant to deviations from the ideal (the input of one yielding an output of one). The dotted line marks a possible hysteretic response when decreasing the input from 1 to 0 and beyond to negative values, the existence of which opens up the possibility of using the device as a memory element.

The earliest digital computers used electromechanical relays. They are large, slow, expensive, energy-hungry (and hence expensive to run), and unreliable. Frequent operational errors during the execution of programs run with such devices provided Hamming with the inspirational motivation for developing error-correcting codes. Thermionic valves (vacuum tubes) are faster and more reliable, but even more expensive and energy-hungry. The first significant step towards miniaturization was taken with the replacement of relays and valves by solid-state transistors (Figure 6.3). Provided the fabrication does not pose new difficulties (remember that a clock is usually cheaper to make than a watch), miniaturization uses less material in fabrication and less energy in operation (see Chapter 3). At a stroke, the devices became smaller, faster (the electrons carrying the information had less distance to travel), cheaper (not only because the volume of material required was lower, but also because efficient massively parallel fabrication procedures were devised), used less energy, and like all solid-state devices were more reliable (the thermionic valve was more reliable than the relay because it had no mechanical moving parts, but the vacuum could leak and incandescent electron-emitting filaments could break).

A major step in fabrication technology was the introduction of integration. Miniaturization and, concomitantly, parallel fabrication now permits millions of integrated transistors to be fabricated on a single "chip" of silicon, with additional gains in operation speed because the electrons have less far to travel, both within and between components.

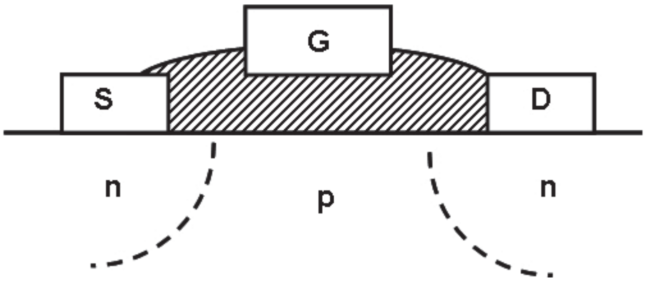


Figure 6.3: A field-effect transistor (FET). Regions marked "n" and "p" are n-type and p-type semiconductors (e.g., appropriately doped silicon). The hatched region is an insulator (e.g., silica). Conductors S, G and D are, respectively, the source, gate and drain. Application of a voltage to G (which plays the role of the coil in the relay) increases the concentration of conduction electrons in the p region and allows current to flow from S to D.

A related device is an information store, or memory. A relay or transistor having the property of bistability could function as an information store (memory), with the disadvantage that it would need to be constantly supplied with electrical power. A more elaborate relay, with two separate coils for switching the current "on" and "off", would be better in this regard, since once its position had been flipped, power could be cut off (see Figure 6.2). Read-only memories do not even require the flipping to be reversible: an early type of read-only memory was paper tape in which holes were punched. Reading was carried out by passing the tape between a pair of electrically conducting rollers. In the absence of a hole, there would be no contact between the rollers. A later development was the use of ferromagnets, which could be poled "up" or "down". Since ferromagnetism cannot exist below a certain volume, this technology is not suitable for ultimate nanoscale miniaturization, but this limit is still far from being reached—the current limitation is the sensitivity of the magnetic field detector (reading head). Memories based on electrical resistance can be fabricated from materials (e.g., NiO) that can be switched from a conducting to an insulating state by applying a voltage pulse. Other materials can have their phase changed from amorphous to crystalline by exposure to light or by passing an electric current, with corresponding changes in reflectance and resistance, but these materials are not especially "nano".

Apart from the logic gates acting as the components of information processors, the other main types of device to be considered are sensors and actuators. A sensor has a clear transduction function. Examples are a magnetic sensor that registers whether the spin in a memory cell is “up” or “down”; a light sensor such as a photodiode that converts light into electricity; a chemical sensor that converts the presence of a certain chemical compound into electricity or light. The main issue in miniaturization is whether the signal exceeds the noise level. An example of an actuator is the coil in the relay (Figure 6.1).

6.1 Electronic devices

For devices in which information is represented as electrostatic charge, a scalar quantity, the lower limit of its magnitude is the charge e of a single electron. Neglecting noise and equivocation issues (see §3.3), single electron devices can be achieved by downscaling the components of a conventional transistor. Developments in fabrication technologies (see Chapter 7) have led to devices with the same architecture as their microscopic counterparts. Truly nanoscale devices using electrons involve single-charge transport in minute tunnel junctions. Several different device configurations designed to exploit the discrete nature of electric charge transport have been or are being investigated.

Another approach to ultraminiaturize electronic components is to use a single molecule as the active medium. Current realizations of such **molecular electronic devices** comprise an organic molecule uniting electron donor (D^+, i.e., a cation) and acceptor (A^-, i.e., an anion) moieties separated by an electron-conducting bridge (π, i.e., a π-conjugated (alkene) chain) between a pair of (usually dissimilar) metal electrodes $M^{(1)}$ and $M^{(2)}$,[1] mimicking a semiconductor p-n junction. When a forward bias is applied across the electrodes, chosen for having suitable work functions, the process $M^{(1)}/D^+–\pi–A^-/M^{(2)} \rightarrow M^{(1)}/D^0–\pi–A^0/M^{(2)}$ occurs, followed by intramolecular tunnelling to regenerate the starting state. Under reverse bias, the energetically unfavourable formation of $D^{2+}–\pi–A^{2-}$ that would be required blocks electron flow, hence we have rectification.

Printed electronic devices. Circuits can be fabricated at extremely low cost by printing onto a suitable substrate. Conventional processes such as screen printing and inkjet are suitable, with inks formulated using “pigments” that are conductive or semiconductive nanoparticles. This technology is especially attractive for radio frequency identification tags (RFID), which are expected to become widely used in packaging, and as security devices on products and even on documents if they can be produced at sufficiently low cost.

Nano field emitters. Due to their extremely high curvature, carbon nanotubes can emit electrons at much lower voltages (a few volts) compared with conventional field emission devices. The principle envisaged application is in flat display screens, competing with liquid crystal technology. They are also attractive as electron guns for scanning electron microscopes, in high-power microwave amplifiers, and for miniature X-ray sources. Carbon nanotube-based electron guns for electron microscopes are undoubtedly the best available (but the global market is insignificant in terms of volume of carbon nanotubes!). Potentially larger volume applications are for constructing electrochemical capacitors, and in electronics (e.g., as connectors between components, especially vertical ones (“vias”) to connect stacked layers). In many of these other applications, however, existing materials (e.g., the much cheaper carbon black) already offer performance close to the theoretical limit.

Nanocomposite transparent electrodes. Display technology is inevitably going to have to change in the near future because of the global dearth of indium, presently used for doping tin oxide (at a fairly high level, of the order of 10%) to create electrically conducting transparent “indium tin oxide” (ITO) thin films on glass, which are used as the counterelectrode in display

---

[1] See, e.g., A.S. Martin et al., Molecular rectifier. *Phys. Rev. Lett.* 70 (1993) 218–221.

devices. Annual consumption is around 800 tonnes, yet the total known reserves are less than 3000 tonnes. As much recycling as possible is carried out, but extraction of the indium is becoming increasingly difficult as devices become smaller and more integrated.

Due to the very low percolation threshold of highly elongated objects dispersed in a matrix, carbon nanotube-doped polymers can be made adequately conductive at levels low enough for the material to remain transparent, which should therefore be able to replace current indium tin oxide-based conducting glass technology.

Nanocapacitor arrays. By analogy with ferromagnetic memory, ferroelectric materials are being investigated for nonvolatile storage. Using a nanoporous template, ferroelectric ceramic (e.g., lead zirconate titanate) can be deposited as nanoscale islands on a suitable metal (e.g., platinum).

6.2 Magnetic devices

Electrons have spin as well as charge. This is of course the origin of ferromagnetism, and hence magnetic memories, but their miniaturization has been limited not by the ultimate size of a ferromagnetic domain but by the sensitivity of magnetic sensors. The influence of spin on electron conductivity was invoked by Nevill Mott in 1936, but remained practically uninvestigated and unexploited until the discovery of giant magnetoresistance (GMR) in 1988. Spintronics, sometimes called magnetoelectronics, which may be loosely defined as the technology of devices in which electron spin plays a role, has three main directions now:

- The development of ultrasensitive magnetic sensors for reading magnetic memories
- The development of spin transistors , and in which the barrier height is determined by controlling the nature of the electron spins moving across it
- The development of devices in which logical states are represented by spin (Chapter 9).

Giant magnetoresistance (GMR). This phenomenon is observed in thin (a few nanometres) alternating layers (superlattices) of ferromagnetic and non-magnetic metals (e.g., iron and chromium). Depending on the width of the nonmagnetic spacer layer, there can be a ferromagnetic or antiferromagnetic interaction between the magnetic layers, and the antiferromagnetic state of the magnetic layers can be transformed into the ferromagnetic state by an external magnetic field. The spin-dependent scattering of the conduction electrons in the nonmagnetic layer is minimal, causing a small resistance of the material, when the magnetic moments of the neighbouring layers are aligned in parallel, whereas for the antiparallel alignment the resistance is high. The technology is already used for read-write heads in computer hard drives. It is noteworthy that the discovery of GMR depended on the development of methods for making high-quality ultrathin films (see Chapter 7). The GMR effect has clearly demonstrated that spin-polarized electrons can carry a magnetic moment through nonmagnetic materials while maintaining spin coherence: this is the meaning of the term “spin transport” nowadays.

A second type of magnetic sensor is based on the magnetic tunnel junction (MTJ). In this device, a very thin *dielectric* layer separates ferromagnetic (electrode) layers, and electrons tunnel through the nonconducting barrier under the influence of an applied voltage. The tunnel conductivity depends on the relative orientation of the electrode magnetizations and the tunnel magnetoresistance (TMR): it is low for parallel alignment of electrode magnetization and high in the opposite case. The magnetic field sensitivity is even greater than for GMR. MTJ devices also have high impedance, enabling large signal outputs. In contrast with GMR devices, the electrodes are magnetically independent and can have different critical fields for changing the magnetic moment orientation. The first laboratory samples of MTJ structures (NiFe-Al_2O_3-Co) were demonstrated in 1995.

6.3 Photonic devices

Another kind of superlattice is made from alternating layers of wider and narrower band gap semiconductors (for example, n-AlGaAs and GaAs, respectively)—called a quantum well. Semiconductor lasers, in which a voltage is applied across the semiconductor crystal that in effect constitutes a Fabry-Perot cavity to create a nonequilibrium population distribution of electrons and holes, whose luminescent recombination generates photons stimulating further emission, were already in existence when Dingle and Henry[2] showed that using quantum wells as the active lasing medium would result in more efficient lasers with lower threshold currents, essentially because quantum confinement of the charge carriers and the optical modes enhances carrier-radiation interaction; moreover the lasing wavelength could be tuned by changing the thickness of the layers. Again, real progress was only made with improvements in the technology of ultrathin film fabrication (see Chapter 7). It follows that reduction of the dimensionality from two to one (quantum wires) to zero (quantum dots) would lead to further improvements, although any dispersity of the size of the dots would smear out the density distribution of the

---

[2] US Patent 3,982,207 (1976).

electronic states, negating the advantage of the zero-dimensional confinement. Early attempts to produce quantum dots a few tens of nanometres in diameter using electron beam lithography followed by the usual semiconductor processing (etching, see §7.1) were bedevilled by damage and contamination introduced by the processing. An important advance came through the exploitation of frustrated wetting (Stranski-Krastanov growth): lattice mismatch between the deposited layer and the substratum results in strain, which was found to be relieved by the spontaneous formation of monodisperse islands (quantum dots).[3]

---

[3] D. Bimberg et al., InAs-GaAs quantum pyramid lasers. *Jpn. J. Appl. Phys.* 35 (1996) 1311–1319.

6.4 Mechanical devices

Careful analysis of the size-dependent performance of microelectromechanical systems (MEMS) such as accelerometers revealed that their performance is degraded if they are further miniaturized down to the nanoscale.[4] Nevertheless, continuing advances in the technologies of structuring materials such as silicon, not only to create layered electronic devices but also mechanical devices, as well as the advent of graphene-based materials, have generated renewed interest in mechanical devices that can now be made at the nanoscale (nanoelectromechanical systems, NEMS). As already pointed out in Chapter 3, ultrasmall cantilevers (taking the cantilever as the prototypical mechanical device) have extremely high resonant frequencies, effective stiffnesses and figures of merit Q and, evidently, very fast response times $\sim Q/\omega_0$. It therefore becomes conceivable that a new generation of relays, constructed at the nanoscale, could again contend with their solid-state (transistor-based) rivals that have completely displaced them at the microscale. Relays have, of course, excellent isolation between input and output, which makes them very attractive as the components of logic gates.

NEMS should however also be very valuable as mass sensors (by the same token, manufacturing variability may be problematical). A large fraction of the atoms of a nanocantilever are inevitably at its surface (cf. Chapter 3), and in some cases it has been found that the addition of molecules to the cantilever surface *increases* resonant frequency because they stiffen the surface “skin”, and this effect predominates over the decrease expected from the increase of resonant mass.

A significant NEMS engineering challenge is the detection of displacements in the picometre or even femtometre range at gigahertz frequencies.

6.5 Fluidic devices

Miniaturizing mixers has been very successful at the microscale, as can be deduced from the huge proliferation of lab-on-a-chip devices for analytical and preparative work in chemistry and biochemistry. Particular advantages are the superior control over flow compared with macroscopic mixers, one very important benefit of which is much more predictable selection of reaction products, wherever several are possible (in consequence, yields can be raised to 100%), and (in principle) great ease of scale up, simply by having many micromixers in parallel

---

[4] C. Hierold, From micro- to nanosystems: mechanical sensors go nano. *J. Micromech. Microengng* 14 (2004) S1–S11.

(although this does not yet seem to have been convincingly demonstrated for any industrial-scale production facility).

It is, however, by no means clear that even greater success will attend further miniaturization down to the nanoscale. On the contrary, performance may be degraded. This needs further investigation.

The key feature of microfluidics, greatly exacerbated with nanofluidics, is the very high surface/volume ratio of flow channels. For water-based fluidics, the surfaces of most metal oxide materials typically used to create channels (e.g., silica) in contact with water are hydroxylated, and hence undergo the reaction material–OH $\rightleftharpoons$ material–O^- + $H^+_{(aq)}$. In acidified water, we would have material–OH + $H^+_{(aq)}$ $\rightleftharpoons$ material–OH_2^+. Thermal energy ensures that the counterions, while ensuring overall neutrality, are distributed diffusely within the fluid side of the interface, resulting in a local excess of electrostatic charge at the interface (as has been described by Gouy and Chapman). If the fluid is moving with respect to the interface, some of the (solvated) counterions will be entrained with the fluid. Conversely, if an electric field is applied parallel to the interface, the solvated counterions will move and entrain fluid, generating flow (has been described by Smoluchowski). Since the spatial extent of the diffuse layer perpendicular to the interface is in the nanoscale range, nanofluidic transport is typically dominated by these electrokinetic phenomena. Control of wetting (see §7.4.3) is also essential.

6.6 Biomedical devices

The devices considered in this section fall into the category of nanobiotechnology, also known as nanomedicine, defined as the application of nanotechnology to human health (see also Chapter 9).

One of the most attractive candidate tasks for a radically new approach is the sequencing of the human genome. The growing fund of medical experience concerning individual patients' responses to pharmaceutical drugs is revealing significant differences between individuals, which in many cases might be due to differences in DNA sequence (cf. §9.2.4). Despite the tremendous boost to the technology of DNA sequencing that came from the international project to sequence the (putatively prototypical) human genome, the basic methods applied were the conventional biochemical ones; the vast increase in throughput was achieved through massive parallelization and automation.

The four different DNA "bases" (or nucleotides, symbolized as A,C,G,T) differ not only in their chemical nature, but also in their physical nature, most significantly as regards size and shape. One of the early motivations for developing the atomic force microscope was the hope that these physical differences could be revealed by rapidly scanning a single strand of DNA. Although the resolution, at least in the presence of liquid water, has so far proved to be inadequate, alternative approaches with the same end in view are being intensively investigated. The favoured scheme is to pass the DNA strand through a nanopore while measuring ionic conductance (of the electrolyte solution in which the DNA is dissolved), either along or across the pore, with the resolution of a single base. The different nucleotides can be thus distinguished, but it is difficult to capture the DNA and drive it through the pore.

The flagship nanomedical system (rather than device) is the “nanobot”, an autonomous robot envisaged to be about the size of a bacterium (i.e., about one micrometre in diameter), and containing many nanodevices (an energy source, a means of propulsion, an information processor, environmental sensors, and so forth). When engineering such devices it is important to note the environment in which they must operate: viscous (highly dissipative), dominated by friction and fluctuations (Brownian motion), and in which inertia plays a negligible role. This is in contrast to the familiar macroscopic mechanisms that follow Newton’s laws: for the nanobot, force is not given by the product of mass and acceleration, but by the product of the coefficient of friction and its velocity, together with superimposed random fluctuations. Any self-propelling nanobot is therefore likely to resemble a motile bacterium rather than a device equipped with nanoscale oars or paddles.

6.7 Further reading

J. Berthier, *Micrographs and Digital Microfluidics.* Norwich, NY: William Andrew (2008). Comprehensive treatment of the technology of fluidics at the microscale, with a special emphasis on electrowetting on dielectrics. Provides an excellent introduction to the development of nanofluidic devices.

S. de Haan, NEMS—emerging products and applications of nanoelectromechanical systems. *Nanotechnology Perceptions* 2 (2006) 267–275. Overview of industrial devices being developed.

T. Hogg, Evaluating microscopic robots for medical diagnosis and treatment. *Nanotechnology Perceptions* 3 (2007) 63–73. Sober assessment of the feasibility of nanobots.

K.K. Likharev, Correlated discrete transfer of single electrons in ultrasmall tunnel junctions. *IBM J. Res. Develop.* 32 (1988) 144–158. A masterly account of the field.

S. Martel, The coming invasion of the medical nanorobots. *Nanotechnology Perceptions* 3 (2007) 165–173. An enthusiastic account of active research towards realizing a nanobot.

M. Zwolak and M. Di Ventra, Physical approaches to DNA sequencing and detection. *Rev. Mod. Phys.* 80 (2008) 141–165. Up-to-date survey of current knowledge.

Chapter 7

Nanofacture

Chapter 5 dealt with the production of nano-objects (including particles and fibres) by essentially chemical means, and which are typically incorporated into what are called nanoproducts by blending, which is not atomically precise manufacture. In this chapter, we shall focus on the technology of true nanoscale engineering. Figure 7.1 summarizes the different approaches under development.

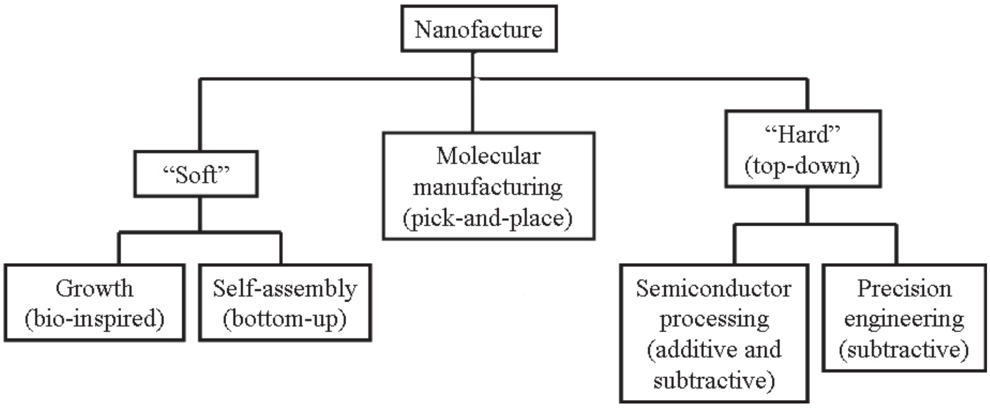


Figure 7.1: Different modes of nanomanufacture (nanofacture) (see text).

7.1 Top-down methods

These share the general feature of requiring large (and also expensive, requiring considerable concentrations of capital) installations. What might be called the traditional route, that of scaling down processes familiar in macro and micro engineering, appears on the extreme right of the diagram, Figure 7.1. Concerted incremental improvement in the entire manufacturing process transforms precision engineering into ultraprecision engineering (Figure 1.1). The stiffness of the parts of mechanical devices used to shape objects is particularly important. These processes are essentially subtractive: material is removed by grinding (etc.).

Semiconductor processing refers to the operations of sequentially modifying (e.g., oxidizing), depositing (additional layers on), and removing (parts of) a substratum (e.g., silicon) over areas selected by exposing photoresist coating the working surface through a mask and then dissolving away the unexposed resist (or the converse). This works well at the micrometre scale (and is used to fabricate very large-scale integrated circuits). Problems of scaling it down to produce features with lateral sizes in the nanorange runs into the diffraction limit of the light used to create the mask (cf. eqn 4.1), partly solved by using light of shorter wavelengths, or high energy electrons.

In contrast to the difficulties of nanoscaling lateral features, very high quality thin films can be deposited with nanometre control perpendicular to the plane of a substratum. These methods are grouped under the heading of physical vapour deposition (PVD). The material to be deposited is evaporated from the reservoir, or sputtered from the target. The most precise control is obtainable with molecular beam epitaxy, developed at AT&T Bell Laboratories in the late 1960s: the evaporated material is beamed onto the substratum under conditions of ultrahigh vacuum. Deposition is typically very slow (several seconds to achieve 1 nm film thickness) and hence can be epitaxial. Ultrathin layers (of the order of the nanometre) with atomically sharp interfaces can be deposited.

Chemical vapour deposition (CVD) is similar to PVD, except that the precursor of the thin layer is a reactive gas or mixture of gases, and the substratum is typically heated to accelerate chemical reaction to form a solid product deposited as a film. The decomposition can be enhanced with a plasma (this typically allows the substratum to be maintained at a lower temperature than otherwise). An example is given in Figure 5.2.

Related technologies are used to *modify* existing surfaces of materials, such as exposure to a plasma, and ion implantation, in which electrostatically charged high-energy (typically 10–100 keV) ions are directed towards the surface, where they arrive with kinetic energies several orders of magnitude higher than the binding energy of the host material, and become implanted in a surface layer that may be tens of nanometres thick.

7.2 Molecular manufacturing

This method literally constructs things atom by atom.[1] A more specific formulation focused on carbon-based, diamondoid structures.[2] An appropriately functionalized molecular tool driven by mechanical forces (such as the tip of a scanning probe microscope) abstracts hydrogen from passivated surfaces to form radicals ("dangling bonds"), where other atoms can be added. Hence this approach is also called "tip-based nanofabrication".

The manipulation of xenon atoms on an ultracold nickel surface to form the letters "IBM" has been demonstrated.[3] Extensive calculations using density functional theory (DFT) and other methods of mechanosynthetic reactions have been published.[4] Atomic extraction has been demonstrated using purely physical forces (albeit not with a specifically functionalized tip).[5] Clearly this approach is still very much in its infancy.

At present, atom-by-atom assembly is very slow and laborious. High throughput can only be achieved by massive parallelization, which in turn is only feasible if the required tools can make themselves. Significant acceleration of the process could take place if "nanoblocks"—pre-assembled (possibly by self-assembly, §7.3.3) units that may comprise dozens or hundreds (or more) atoms—are manipulated.

A related technique, called "dip-pen nanolithography" (DPN), is a way of picking up solutions of molecules ("ink") and allowing it to be transferred to the substratum by capillary action.[6] Although not atomically precise manufacturing, it allows features of the order of 100 nm to be written.

---

[1] K.E. Drexler, Molecular engineering: an approach to the development of general capabilities for molecular manipulation. *Proc. Natl Acad. Sci. USA* 78 (1981) 5275–5278.

[2] K.E. Drexler, *Nanosystems: Molecular Machinery, Manufacturing, and Computation.* Wiley-Interscience (1992).

[3] E.K. Schweizer and D.M. Eigler, Positioning single atoms with a scanning tunneling microscope. *Nature (Lond.)* 344 (1990) 524–526.

[4] E.g., B. Temelso et al., Ab initio thermochemistry of the hydrogenation of hydrocarbon radicals using silicon-, germanium-, tin-, and lead-substituted methane and isobutene. *J. Phys. Chem. A* 111 (2007) 8677–8688.

[5] N. Oyabu, Ó. Custance, I. Yi, Y. Sugawara and S. Morita, Mechanical vertical manipulation of selected single atoms by soft nanoindentation using near contact atomic force microscopy. *Phys. Rev. Lett.* 90 (2003) 176102.

[6] M. Jaschke and H.-J. Butt, Deposition of organic material by the tip of a scanning force microscope. *Langmuir* 11 (1995) 1061–1064.

7.3 Bottom-up methods

It has long been known that many biological systems exhibit remarkable capabilities of assembling themselves starting from a randomly arranged mixture of components. These include the bacteriophage virus (the final stages of assembly), and proteins and ribonucleic acids (RNA), which can be spontaneously transformed from a random coil of the linear polymer to a compact, ordered three-dimensional structure. It is clear that the starting precursors of the final structures have to be very carefully designed.

At the same time, the increasing difficulty of continuing the miniaturization of classical photolithography and its derivatives, and the extreme laboriousness of mechanosynthesis, generated interest in alternative fabrication technologies. The idea of self-assembly ("shake and bake") is to gather precursors in random positions and orientations and supply energy ("shaking") to allow them to sample configuration space. The hugeness of this space suggests that a convergent pathway is inherent in the process in order to allow it to be completed in a reasonable time. Once the precursors are in position, "baking" may be required to strengthen the bonds connecting them and fix the final object permanently.

Although appreciation of self-assembly in biology has played a hugely important inspirational role, the highly specialized chemistry of living systems, the fragility of many of its products, and its inherent variability at many levels have made it unsuitable for mimicking directly and incorporating into our present industrial system. This is particularly so in the case of the food industry. The extreme complexity, both structural and chemical, of its products and the relative ease of letting them grow renders efforts to manufacture food synthetically largely superfluous. More debatable is solar energy conversion. The natural system comprises the photosystems embedded within the chloroplast, whose maintenance requires the rest of the machinery of the cell, and whose effective operation requires a microscopic structure (stem and branches) to support the leaves in which the chloroplasts are embedded. The classical artificial system is the semiconductor photovoltaic cell. There is hope that its efficiency, as well as ease of manufacture, can be enhanced by using nanostructured photoactive components. Most appraisals of the photovoltaic cell as a "renewable" or "sustainable" energy source pay scant regard to the entire manufacturing cycle, and the key question of working lifetime under realistic conditions is scarcely addressed by laboratory trials. There is also a history of considerable efforts to more closely mimic the molecular machinery of the natural photosystems in a nanoconstruction. Nevertheless, except for the ultimate, and still hypothetical, stage of molecularly manufactured nanosystems, none of the proposed solutions come anywhere near the performance (considered as an overall system) of natural photosynthesis, which can simply be grown over vast areas.

7.3.1 Biological growth

Reproducibility is interpreted somewhat differently by living processes. Although the basic building blocks (e.g., proteins) of living organisms are identical, templated from a master specification (see §8.1), organisms are not identical in the way that very large scale integrated circuits (VLSIs) are. What is specified (genetically) is at most an algorithm (subject to local environmental influence) for constructing an organism, or maybe just an algorithm for an algorithm.

This concept of an algorithm specifying how the construction should take place is used for building the nests of social insects, which are constructed stigmergically—each insect is armed with rules specifying what to do in a variety of local circumstances.[7]

There are some relatively unexplored niches for creating nano-objects via biological growth. For example, the magnetic protein ferritin, which is constituted from an iron oxide core surrounded by protein, could in principle be made on a large scale by low-cost biotechnological

---

[7] See, e.g., G. Theraulaz and E. Bonabeau, Coördination in distributed building. *Science* 269 (1995) 686–688.

manufacturing routes for use in magnetic memory devices.

7.3.2 Biopolymer folding

Biopolymer "folding" means the transformation of a linear polymer chain, whose monomers are connected only to their two nearest neighbours, and which adopts a random coil in solution, into a complex three-dimensional structure with additional bonds between distant monomers.

Predicting the final three-dimensional structure is *prima facie* a difficult problem. Energetics are clearly involved, because bonds between distant monomers form spontaneously (if geometric constraints are satisfied), releasing enthalpy and hence lowering the free energy. On the other hand, this raises the entropy because the chain becomes constrained. Finding the free energy minimum by systematically searching configuration space is a practically impossible task for a large molecule with thousands of atoms—it would take longer than the age of the universe. Since the protein molecule can fold within seconds, it seems clear that the solution to the problem lies in determining the pathways. The Principle of Least Action is useful for this purpose: the most expedient path is found by minimizing the action.

Action is the integral of the Lagrangian $\mathcal{L}(= L - F$ for conservative systems, where L and F are respectively the kinetic and potential energies). Minimization of the action is an inerrant principle for finding the correct solution of a dynamical problem; the difficulty lies in the fact that there is no general recipe for constructing $\mathcal{L}$.

A solution leading to a successful algorithm has been recently found for the folding of ribonucleic acid (RNA).[8] Natural RNA polymers are made up from four different "bases", A, C, G and U (see §8.1). As with DNA, multiple hydrogen bonding favours the formation of G–C and A–U pairs, which leads to the appearance of certain characteristic structures. Loop closure is considered to be the most important folding event. F (the potential) is identified with the enthalpy; that is, the number n of base pairings (contacts), and L corresponds to the entropy. At each stage in the folding process, as many as possible new favorable intramolecular interactions are formed, while minimizing the loss of conformational freedom (the principle of sequential minimization of entropy loss, SMEL). The entropy loss associated with loop closure is ΔS_{loop} (and the rate of loop closure $\sim \exp(\Delta S_{\text{loop}})$); the function to be minimized is therefore $\exp(-\Delta S_{\text{loop}}/R)/n$, where R is the universal gas constant. A quantitative expression for ΔS_{loop} can be found by noting that the N monomers in an unstrained loop ($N \geq 4$) have essentially

---

[8] A. Fernández and H. Cendra, In vitro RNA folding: the principle of sequential minimization of entropy loss at work. *Biophys. Chem.* 58 (1996) 335–339.

two possible conformations, pointing either inwards or outwards. For loops smaller than a critical size N_0, the inward ones are in an apolar environment, since the nano-enclosed water no longer has bulk properties, and the outward ones are in polar bulk water. For $N < N_0$, $\Delta S_{\text{loop}} = -RN \ln 2$ (for $N > N_0$, the Jacobson-Stockmayer approximation based on excluded volume yields $\Delta S_{\text{loop}} \sim R \ln N$).

Sequential minimization of entropy loss (SMEL) applied to biopolymer folding is a least-action principle that involves sequentially maximizing the number of contacts while minimizing entropy loss.

7.3.3 Self-assembly

Self-assembly is usually considered to be synonymous with "bottom-up" fabrication. It is particularly valuable for generating nanostructured thin films supported on a substratum. Atomic precision is achievable in the direction perpendicular to the plane of the substratum, and statistical atomic precision (which may be all that is required) is achievable within the plane.

Advantages of "bottom up" fabrication are that it can:

- Assemble onto curved and other nonplanar surfaces
- Be massively parallellized
- Produce structures with features down to a few nm.

Current challenges of "bottom up" nanotechnology include:

- Formulating design rules. Generally we ask: "how to design X to carry out a desired function Y?" *Design* means essentially specifying structure, hence the question can be reworded as: "what structure will give function Y?"[9] Micro (and macro) engineering benefits from vast experience; i.e., a look-up table with structure in the left-hand column and function in the right. There is less experience in the nanoworld, but if a nanostructure is simply a microstructure in miniature, this experience can be transferred. One needs however to ask whether the properties of matter change at the nanometre scale; i.e., do we need a new set of structure-property relations? These relations may also affect the fabrication process.

---

[9] "Function" means properties and performance. Structure is constituted from certain *numbers* of different *types* of entities, *connected* together in a certain way.

- Formulating assembly rules (see below). Current rules are few: either very general (such as the Principle of Least Action, see §7.3.2) or very specific (chemical intuition). Hence general principles need to be adapted to specific cases, and specific heuristic rules or intuition need to be formalized and generalized.

The main disadvantage of bottom-up is that:

- The process is not well understood theoretically. Hence although we need to be able to, at present we cannot design the starting objects (precursors) to achieve a specified final device.

7.3.4 Self-assembled monolayers (SAM)

This is a practical procedure of some importance for modifying the surfaces of objects fabricated by other means. The precursors are molecules of general formula XL, where X is (typically) an apolar chain (e.g. alkyl), and L is a ligand capable of binding to a substratum. Addition of XL to the metal surface results in a closely packed array of XL. The film is stabilized by hydrogen or chemical bonds to the substrate, and lateral LW forces between the X.

Currently the two main types of L are –SH (thiol or mercaptan), which binds strongly to Au, Ag, Pt, Cu, Hg etc.), and organosilanes, which bind strongly (bond covalently) to silica. These chemical requirements are the main constraints limiting the versatility of the technology.

X can be functionalized at the end opposite from L with reactive groups to form molecules RXL. These can profoundly change the wetting properties of the assembled monolayer. For example, whereas octadecanethiol (R = –H) films are both oil and water repellent, if R = –OH then oil and water will spread. More elaborate groups can be incorporated. If they are bulky, the functionalized molecules should be mixed with unfunctionalized ones. Mixtures of L with different chain lengths (e.g. C_{12} and C_{22}) give liquid-like SAMs. SAMs can be patterned using photolithography, or "stamping" (microletterpress), to create patterns on substrata (e.g., gold and/or silica) to which the SAM precursor molecules will bind, leaving other zones free. In this procedure, the required pattern is the first created in relief on a silicon substrate, and which is used as a mould for the elastomeric polymer PDMS (polydimethylsiloxane). The SAM molecules can be used directly as ink to coat the projecting parts of the relief pattern, which is then stamped onto the substratum, or else the ink is some substance that passivates the substratum with respect to the SAM molecules, which then selectively bind as required.

7.3.5 Alternating polyelectrolyte deposition

The method of *alternating polyelectrolyte deposition* (APED) appears to have immense potential as a simple, robust method of surface modification. It requires the substrate to be electrostatically charged when immersed in water. It is then dipped into an aqueous solution of a polyelectrolyte of opposite charge, with which it rapidly decomes coated. Any excess is then washed off, and the coated substrate is dipped into a polyelectrolyte of the opposite charge, with which it now becomes coated, and whose excess is again washed off, and so on.

There are few restrictions on the choices of polyelectrolytes. Much early work was done with polyallylamine as the polycation, and polystyrene sulfonate as the polyanion. The essential feature of the technique is that at each dipping stage the substrate charge is not only neutralized but reversed ("overcharging"), hence allowing the deposition to be repeated indefinitely. This phenomenon contradicts the predictions of the mean-field theories—Gouy-Chapman and Debye-Huckel—of the distribution of ions in the vicinity of charged surfaces ("electrified interfaces"). The discrepancy arises because the charges of the polyions are correlated. Imagine a polyion approaching a surface already covered with its congeners. The new arrival will repel the already-adsorbed ones, creating a correlation hole (i.e., a negative image) permitting attraction and overcharging.

(Monovalent) counterions screen the polyions in the usual Debye-Huckel fashion, diminishing the charging energy of the polyion more than its correlation energy, enhancing the charge inversion. (If the monovalent counterion concentration is very high the correlation disappears and APED is no longer possible.) Multivalent counterions are more difficult to treat theoretically and APED in their presence would appear to be a fruitful area of investigation. The hydrogen ion may play a special role; for example, it has been found that the porosity of built layers can be reversibly controlled by varying the pH.

Instead of polymeric polyions, nanoparticles composed of materials with ionizable surface groups can be used. In this case, although the electrostatic charge of the surface of the coating is always reversed, typically not all the inner charges are compensated because of steric hindrances, and hence electrostatic charges build up, and the long range electrostatic force ultimately prevents further particles from being deposited.

If polymeric polyions are used as the polyelectrolyte of one sign, and ionizable particles as the polyelectrolyte of the opposite sign, the particles act as stress concentrators, thus greatly increasing the toughness of the built material. Large aspect ratio nanoparticles are very useful for diminishing the deleterious effects of defects (pinholes) in multilayer films. In this way sophisticated coatings can be built up. It has been established that the shells of many marine organisms, such as the abalone, are assembled using this principle, producing materials that are both robust and beautiful: anisotropic nanoparticles are dispersed in a biopolymer matrix, which only occupies a few volume percent of the total mass. Natural biopolymers, which are nearly all heteropolymers, primarily based on amino acids as monomers, but also possibly incorporating polysaccharides and nucleic acids, can incorporate enormous functional variety, in ways that we can only dream about at present in synthetic systems.

7.3.6 Random addition of particles to a surface

Many self-assembly processes are based on the addition of nanoparticles to a surface.

Functions for characterizing nanparticle addition: the fraction of occupied surface θ, equal to the number of particles ν per unit area times the area a occupied by one particle; and the fraction of surface ϕ available for adsorption (sometimes called the "available area function").

In general we have for the rate of addition:

$$\mathrm{d}\theta/\mathrm{d}t = k_a c^* \phi(\theta) \tag{7.1}$$

where k_a is the addition rate coefficient ($k_a \sim D \exp(-\Delta G_a)$, where D is the diffusion coefficient of the nanoparticle and ΔG_a is the energy barrier hindering addition).

One of the earliest theories relating ϕ to θ was that of Langmuir: if small particles adsorb to discrete sites larger than the particles,

$$\phi = 1 - \theta \tag{7.2}$$

Substituting this into eqn (7.1) and integrating, we see that in Langmuir adsorption the surface is completely filled up ($\theta \to 1$) exponentially in time (for a uniform rate of arrival of particles at the surface).

In the absence of discrete sites, the particles adsorb wherever they happen to arrive (assumed to be random locations). If, however, a particle arrives such that its centre would fall within the exclusion zone of a previously adsorbed particle (Figure 7.2) its adsorption attempt is rejected. Since the exclusion zone is four times as large as the particle, we should have

$$\phi = 1 - 4\theta \tag{7.3}$$

but as θ increases, exclusion zones will overlap (Figure 7.2), and compensating terms have to be added, proportional to θ^2 for two overlapping particles, and so on:[10]

$$\phi = 1 - b_1\theta + b_2\theta^2 + b_3\theta^3 + \mathcal{O}(\theta)^4 \tag{7.4}$$

with $b_1 = 4$ and the coefficients b_2 and b_3 determined by purely geometrical considerations; $b_3 = 6\sqrt{3}/\pi$ is identical for both irreversible and equilibrium adsorption, whereas the coefficient b_3 varies from about 1.4 for irreversible (random sequential addition, RSA) to about 2.4

---

[10] P. Schaaf and J. Talbot, Surface exclusion effects in adsorption processes. *J. Chem. Phys.* 91 (1989) 4401–4409.

for equilibrium (reversible, whether via desorption and readsorption or via lateral movement) adsorption. In this case $\phi \to 0$ for $\theta < 1$; the "jamming limit" at which $\phi = 0$ is $\theta_J \approx 0.55$ for spheres adsorbing irreversibly.

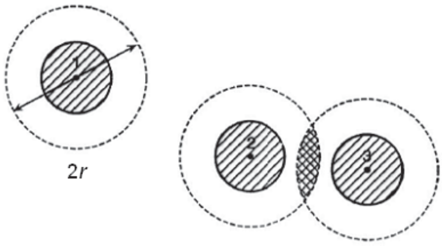


Figure 7.2: The concept of exclusion zone. The particles' projected area is hatched. The area enclosed by the dashed lines is the exclusion zone and has twice the radius of the actual particle. The exclusion zone is defined as that area within which no centre of any particle can be placed without violating the condition of no overlap of hard bodies. The cross-hatched area marks the overlap of the exclusion zones of particles numbered 2 and 3.

The RSA formalism was developed in the context of particles interacting predominantly via hard body repulsion. The particle radius r is implicitly considered to be the hard body radius r. "Soluble" (stably suspended) particles must have repulsive particle-particle interactions and cannot in fact approach each other to a centre-to-centre distance of $2r$, but will behave as particles of an effective radius r', where r' is that value of z at which the total interfacial (IF) interaction energy (see §7.4) $\Delta G^{(\mathrm{IF})}(z) \sim k_B T$.

The ballistic deposition (BD) model was introduced to describe the behaviour of particles falling onto a surface under the influence of gravity. Whereas in RSA if a particle attempts to land with its centre within the exclusion zone around a previously adsorbed particle it is rejected, in BD the particle is not eliminated but rolls along on top of previously adsorbed particles until it finds space to adsorb. The coefficients of eqn (7.4) are then different, namely $b_1 = b_2 = 0$ and $b_3 \approx -9.95$. BD and RSA can be combined linearly in generalized ballistic deposition (GBD), where

$$\phi(\theta) = \phi^{\mathrm{RSA}}(\theta) + j\phi^{\mathrm{BD}}(\theta) \tag{7.5}$$

with the parameter j defined as

$$j = p'/p \tag{7.6}$$

where p' is the probability that a particle arriving via correlated diffusion ("rolling") at a space large enough to accomodate it will remain (i.e., will surmount any energy barrier), and p is the probability that a particle arriving directly at a space large enough to accomodate it will remain. p is clearly related to the lateral interaction ("stickiness") of particles for each other, and as $j \to \infty$ the model describes nanoparticle aggregation at a surface. Essentially, the exclusion zones are thereby annihilated, and ϕ can be simplified to eqn (7.2).

If the adsorbing nanoparticles have a positionally-dependent affinity for each other, the exclusion zones are annihilated, but the particles do not cluster randomly but form a supported two-dimensional crystal. $\phi(\theta)$ can be simplified to eqn (7.2), but the apparent area per particle corresponds to the unit cell size of the crystal.[11]

7.4 Intermolecular interactions

Familiar macro- and micromaterials rely on strong metallic, ionic or covalent bonds to hold the constituent atoms together. On the other hand, the self-assembly of molecules relies on weak (non-covalent) interactions. In other words, self-assembly ultimately depends on intermolecular

---

[11] J.J. Ramsden, G.I. Bachmanova and A.I. Archakov, Kinetic evidence for protein clustering at a surface. *Phys. Rev. E* 50 (1994) 5072–5076.

interactions, which may have to be different on different facets of a nanoparticle in order to enable a low-symmetry assembly.

7.4.1 The concept of surface tension

Surface tension γ is formally defined as the free energy G required to create the extend an interface of area A:

$$\gamma = (\partial G/\partial A)_{T,P} \tag{7.7}$$

where the practically frequently encountered constant temperature and pressure would make the Gibbs free energy the appropriate choice for G. In the Système Internationale, the units of γ are N/m, which is the same as an energy per unit area (J/m^2). It is customary to refer to γ as a surface tension if the increase of area is reversible, and as a surface energy if it is not.

Generally speaking, work needs to be done to create an interface; it has a higher free energy than the bulk. The work of cohesion of a solid is

$$W^{(\mathrm{coh})} = 2\gamma_1 A = -\Delta G^{(\mathrm{coh})} \tag{7.8}$$

(see Figure 7.3), where ΔG is understood to be per unit area. On the other hand, the work of adhesion (needed to separate two dissimilar substances 1 and 2) is given by (see Figure 7.3)

$$W_{12}^{(\mathrm{adh})} = (\gamma_1 + \gamma_2 - \gamma_{12})A = -\Delta G^{(\mathrm{adh})} \tag{7.9}$$

a formalism introduced in the 19th century by Dupré. γ_1 and γ_2 account for the old interfaces lost, and γ_{12} accounts for the new interface gained. Most of the subsequent difficulties experienced by the field of interfacial interactions have concerned the theoretical calculation (prediction) of terms involving two (or more) substances such as γ_{12}.

> The nanoscopic viewpoint is that the microscopic surface tension (or energy) γ_{12} depends on specific chemical interactions between the surfaces of the two substances 1 and 2.

Fowkes, Girifalco and Good introduced the very reasonable assumption that the tension at the interface of substance 1 against substance 2 is lowered by the presence of by an amount equal to the geometric mean of the tensions of the two substances individually,[12] hence equal

---

[12] See, e.g., F.M. Fowkes, Attractive forces at interfaces. *Ind. Engng Chem.* 36 (1964) 40–52.

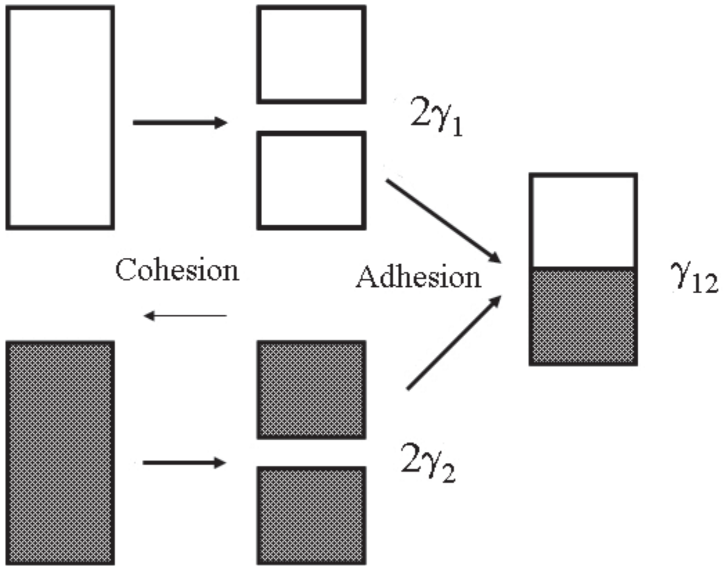


Figure 7.3: Cohesion and adhesion of substances 1 (white) and 2 (grey) (see text).

to $\gamma_1 - (\gamma_1\gamma_2)^{1/2}$, and similarly the tension at the interface of substance 2 against substance 1 is $\gamma_2 - (\gamma_1\gamma_2)^{1/2}$. Summing these two terms, we have

$$\gamma_{12} = \gamma_1 + \gamma_2 - (\gamma_1\gamma_2)^{1/2} = (\sqrt{\gamma_1^{(\mathrm{LW})}} - \sqrt{\gamma_2^{(\mathrm{LW})}})^2 \tag{7.10}$$

called the Girifalco-Good-Fowkes equation. This is equivalent to the work of adhesion being the *geometric mean* of the works of cohesion, i.e. $W_{12} = (W_{11}W_{22})^{1/2}$. The Dupré equation (7.9) then becomes

$$W_{12}^{(\mathrm{adh})} = 2(\gamma_1\gamma_2)^{1/2} \tag{7.11}$$

Fowkes and van Oss developed the idea that the total interfacial energy is linearly separable into the dispersive (London-van der Waals), dipole-induced dipole (Debye), dipole-dipole (Keesom) and electron donor-acceptor terms, and Lifshitz has pointed out that the London-van der Waals, Debye and Keesom interactions are all of the same type (cf. the Hellman-Feynman theorem), with the same dependence of magnitude on separation between the two interacting substances, and hence

$$\gamma^{(\mathrm{total})} = \gamma^{(\mathrm{LW})} + \gamma^{(\mathrm{ab})} \tag{7.12}$$

where LW denotes Lifshitz-van der Waals and ab denotes (Lewis) acid-base, and *a fortiori*

$$\gamma_{12}^{(\mathrm{total})} = \gamma_{12}^{(\mathrm{LW})} + \gamma_{12}^{(\mathrm{ab})} \tag{7.13}$$

Whereas the Lifshitz-van der Waals interaction is always attractive, the sign of the Lewis acid-base interaction depends on the relative proportions of Lewis acids and Lewis bases constituting the two interacting substances. Superscript $\ominus$ will be used to denote electron-donating (Lewis base) and superscript $\oplus$ will be used to denote electron-accepting (Lewis acid) moieties; van Oss has proposed that one might again take the geometric mean, namely

$$\gamma^{\mathrm{ab}} = 2(\gamma^{\ominus}\gamma^{\oplus})^{1/2} \tag{7.14}$$

Two monopolar substances of the same sign will repel each other; attraction depends on the presence of cross-terms. By analogy with eqn 7.11,

$$W_{12}^{(\mathrm{adh})} = 2[(\gamma_2^{\oplus}\gamma_1^{\ominus})^{1/2} + (\gamma_1^{\oplus}\gamma_2^{\ominus})^{1/2}] \tag{7.15}$$

Hence the ab combining law is

$$\gamma_{12}^{(\mathrm{ab})} = 2[(\gamma_1^{\oplus}\gamma_1^{\ominus})^{1/2} + (\gamma_2^{\oplus}\gamma_2^{\ominus})^{1/2} - (\gamma_1^{\oplus}\gamma_2^{\ominus})^{1/2} - (\gamma_1^{\ominus}\gamma_2^{\oplus})^{1/2}] = 2(\sqrt{\gamma_1^{\oplus}} - \sqrt{\gamma_2^{\oplus}})(\sqrt{\gamma_1^{\ominus}} - \sqrt{\gamma_2^{\ominus}}) \tag{7.16}$$

It takes account of the fact that $\ominus$ interacts with $\oplus$, which is why the da interaction can be either attractive or repulsive. In typical biological and related systems, the Lewis acid-base interaction accounts for 80–90% of the total interactions. The most familiar manifestation is hydrogen bonding (e.g., double-stranded DNA (the double helix), globular proteins containing α-helices). The π–π interactions (stacking of alternately electron-rich and electron-deficient aromatic rings) are frequently encountered in self-assembling synthetic organic supermolecules.

Let us now consider two solids 1 and 3 in the presence of a liquid medium 2 (e.g., in which the self-assembly process takes place). ΔG_{123} is the free energy per unit area of materials 1 and 3 interacting in the presence of liquid 2. Using superscript $\|$ to denote the interfacial interaction energies per unit area between infinite parallel planar surfaces,

$$\Delta G_{121}^{\|} = -2\gamma_{12} \tag{7.17}$$

and

$$\Delta G_{123}^{\|} = \gamma_{13}^{-}\gamma_{12} - \gamma_{23} \tag{7.18}$$

From the above equations we can derive:

$$\Delta G_{123}^{(\mathrm{LW,ab})\|} = \Delta G_{22}^{(\mathrm{LW,ab})\|} + \Delta G_{13}^{(\mathrm{LW,ab})\|} - \Delta G_{12}^{(\mathrm{LW,ab})\|} - \Delta G_{23}^{(\mathrm{LW,ab})\|} \; . \tag{7.19}$$

where ΔG_{13} is the free energy per unit area of materials 1 and 3 interacting directly. It follows that:

- LW forces (anyway weak) tend to cancel out;
- the so-called "hydrophobic force" is a consequence of the strong cohesion of water ΔG_{22}. Attraction of suspended solids is only prevented by their hydrophilicity. The sign of ΔG_{12} with 2 = water, provides an unambiguous measure of hydrophobicity: $\Delta G_{12} < 0 \equiv$ hydrophilic; $\Delta G_{12} > 0 \equiv$ hydrophobic.

$\Delta G_{123}^{\|}$ can be used to provide a rapid first estimate of whether adhesion between materials 1 and 3 will take place in the presence of medium 2.

Using the Derjaguin approximation, a sphere of radius r (material 3) interacting with an infinite planar surface (material 1) has the following free energies of interaction, as the function of z, the perpendicular distance between the plane and the nearest point of the sphere:

$$\Delta G^{(\mathrm{LW})}(z) = 2\pi\ell_0^2 \Delta G^{(\mathrm{LW})\|} r/z \; ; \tag{7.20}$$

where ℓ_0 is the equilibrium contact distance (about 0.15 nm);

$$\Delta G^{(\mathrm{da})}(z) = 2\pi\chi\Delta G^{(\mathrm{ab})\|} \exp[(\ell_0 - z)/\chi] r \; , \tag{7.21}$$

where χ is the decay length for the ab interactions; and where electrostatic charges at present,

$$\Delta G^{(\mathrm{el})}(z) = 4\pi\epsilon_0\epsilon\psi_3\psi_1 \ln[1 + \exp(-\kappa z)] r \; ; \tag{7.22}$$

where ψ are the electrostatic surface potentials of materials 1 and 3 and $1/\kappa$ is the Debye length (inversely proportional to the square root of the ionic strength—this is why electrostatic interactions tend not to be very important in salty aqueous systems).

7.4.2 Experimental determination of single-substance surface tensions

The general strategy is to measure the advancing contact angles θ on the material 3 whose surface tension is unknown using three appropriate liquids with different surface tension components (known, e.g., from hanging drop measurements). With these values, three Young-Dupré equations:

$$[\gamma_2^{(\mathrm{LW})}/2 + (\gamma_2^{\oplus}\gamma_2^{\ominus})^{1/2}](1 + \cos\theta) = (\gamma_1^{(\mathrm{LW})}\gamma_2^{(\mathrm{LW})})^{1/2} + (\gamma_1^{\oplus}\gamma_2^{\ominus})^{1/2} + (\gamma_1^{\ominus}\gamma_2^{\oplus})^{1/2} \tag{7.23}$$

can be solved to yield the unknowns $\gamma_1^{(\mathrm{LW})}$, $\gamma_1^{\oplus}$ and $\gamma_1^{\ominus}$. Values of the surface tension parameters for some common materials are given in Tables 7.1 and 7.2. Methods for computing ψ for ionized solids (e.g., polyions, protonated silica surfaces) are based on the Healy-White ionizable surface group model.[13]

7.4.3 Wetting and dewetting

Wetting means the spreading of a liquid over a solid surface; dewetting is its converse, the withdrawal of liquid from a surface. They are basic processes in countless natural and industrial processes. Although pioneering work in characterizing the interfacial tensions upon which wetting depends was reported two hundred years ago by Young, the processes are still relatively poorly understood. Few experimental techniques are available for investigating the important solid/liquid interfaces: the contact angle method is simple and probably still the most important, but only a handful of laboratories in the world have shown themselves capable of usefully exploiting it. The history of dewetting, a phenomenon of no less industrial importance than wetting, is much more recent: quantitative experimental work dates from the early 1990s.

It is essentially intuitive to expect that the spreading of a liquid on a solid depends on γ_{SV} (S = solid, V = vapour, L = liquid). The quantitative relationship was given by Young in 1805 (cf. eqn 7.23):

$$\gamma_{\mathrm{LV}} \cos\theta = \gamma_{\mathrm{SV}} - \gamma_{\mathrm{SL}} \tag{7.24}$$

---

[13] T.W. Healy and L.R. White, Ionizable surface group models of aqueous interfaces. *Adv. Colloid Interface Sci.* 9 (1978) 303–345.

Table 7.1: Surface tension parameters of some solids. Data mostly from C.J. van Oss, *Forces interfaciales en milieux aqueux*, Paris: Masson (1996).

Material	$\gamma^{(\mathrm{LW})}$/mJ m^{-2}	$\gamma^{\oplus}$/mJ m^{-2}	$\gamma^{\ominus}$/mJ m^{-2}
Synthetic polymers			
Nylon 6,6	36	0.02	22
PMMA	41	0	13
Polyethylene	33	0	0
Polyethylene oxide	43	0	64
Polystyrene	42	0	1.1
Polyvinylpyrrolidone	43	0	30
PVC	43	0.04	3.5
Teflon	18	0	0
Carbohydrates			
Cellulose	44	1.6	17
Dextran T-150	42	0	55
Metal oxides			
SiO_2	39	0.8	41
SnO_2	31	2.9	8.5
TiO_2	42	0.6	46
ZrO_2	35	1.3	3.6

The degree of wetting is inversely proportional to the contact angle θ; $\theta = 0$ corresponds to complete wetting. Young's equation (7.24) can be easily derived by noting that the surface tension can be written as a force per unit distance. The interfacial forces acting on the triple line T, where three phases S, L, V (solid, liquid, vapour) meet must sum to zero in a given direction (x, parallel to the interface). More formally, it follows from the condition that (at equilibrium) the energies must be invariant with respect to small shifts dx of the position of T. The structure of T may be very complex. For example, for water containing dissolved electrolyte, the local ion composition may differ from that in the bulk; soft solids may be deformed in the vicinity of T. Although Young's equation ignores these details it provides a remarkably accurate description of contact angles. The region in which deviations from "far field" quantities occur is known as the core, with radius $r_C \sim 10$ nm. Hence for typical drops used in contact angle determinations (with a radius $R \sim 1$ mm), the curvature of T may be neglected; atomic scale heterogeneity on the subnanometre scale may also be neglected.

Table 7.2: Surface tensions of some liquids. Data mostly from C.J. van Oss, *Forces interfaciales en milieux aqueux*, Paris: Masson (1996).

Liquid	$\gamma^{(\mathrm{LW})}$/mJ m^{-2}	$\gamma^{\oplus}$/mJ m^{-2}	$\gamma^{\ominus}$/mJ m^{-2}
Water[a]	22	25.5	25.5
Glycerol	34	3.9	57
Ethanol	19	0	68
Chloroform	27	3.8	0
Octane	22	0	0
n-hexadecane	27.5	0	0
Formamide	39	2.3	40
α-bromonaphthalene	44	0	0
Diiodomethane	51	0	0

[a] Absolute values of $\gamma^{\oplus}$ and $\gamma^{\ominus}$ are not known at present; values are arbitrarily assigned to ensure that the known overall γ is correct.

Please click the advert

FOSS

Sharp Minds - Bright Ideas!

Employees at FOSS Analytical A/S are living proof of the company value - First - using new inventions to make dedicated solutions for our customers. With sharp minds and cross functional teamwork, we constantly strive to develop new unique products - Would you like to join our team?

FOSS works diligently with innovation and development as basis for its growth. It is reflected in the fact that more than 200 of the 1200 employees in FOSS work with Research & Development in Scandinavia and USA. Engineers at FOSS work in production, development and marketing, within a wide range of different fields, i.e. Chemistry, Electronics, Mechanics, Software, Optics, Microbiology, Chemometrics.

The Family owned FOSS group is the world leader as supplier of dedicated, high-tech analytical solutions which measure and control the quality and production of agricultural, food, pharmaceutical and chemical products. Main activities are initiated from Denmark, Sweden and USA with headquarters domiciled in Hillerød, DK. The products are marketed globally by 23 sales companies and an extensive net of distributors. In line with the corevalue to be 'First', the company intends to expand its market position.

We offer
A challenging job in an international and innovative company that is leading in its field. You will get the opportunity to work with the most advanced technology together with highly skilled colleagues.

Read more about FOSS at www.foss.dk - or go directly to our student site www.foss.dk/sharpminds where you can learn more about your possibilities of working together with us on projects, your thesis etc.

Dedicated Analytical Solutions

FOSS
Slangerupgade 69
3400 Hillerød
Tel. +45 70103370

www.foss.dk

Complete wetting is characterized by $\theta = 0$, which implies (from eqn 7.24)

$$\gamma_{\mathrm{LV}} = \gamma_{\mathrm{SV}} - \gamma_{\mathrm{SL}} \tag{7.25}$$

at equilibrium (out of equilibrium, this relation may not hold). The Cooper-Nuttall spreading coefficient S is

$$S = \gamma_{\tilde{\mathrm{S}}\mathrm{V}} - \gamma_{\mathrm{SL}} - \gamma_{\mathrm{LV}} \tag{7.26}$$

where $\gamma_{\tilde{\mathrm{S}}\mathrm{V}}$ is the interfacial tension of a dry solid. Three régimes can thus be defined:

1. $S > 0$. This corresponds to $\gamma_{\tilde{\mathrm{S}}\mathrm{V}} > \gamma_{\mathrm{SV}}$, i.e. the wetted surface has a lower energy than the unwetted one. Hence wetting takes place spontaneously. The thickness h of the film is greater than monomolecular if $S \ll \gamma_{\mathrm{LV}}$. The difference $\gamma_{\tilde{\mathrm{S}}\mathrm{V}} - \gamma_{\mathrm{SV}}$ can be as much as 300 mJ/m^2 for water on metal oxides. Such systems therefore show enormous hysteresis between advancing and receding contact angles. Other sources of hysteresis include chemical and morphological inhomogeneity (contamination and roughness).
2. $S = 0$. Occurs if $\gamma_{\tilde{\mathrm{S}}\mathrm{V}}$ practically equals γ_{SV}, as is typically the case for organic liquids on molecular solids.
3. $S < 0$. Partial wetting. Films thinner than a certain critical value, usually ~ 1 mm, break up spontaneously into droplets.

7.5 Further reading

M.G. Cacace, E.M. Landau and J.J. Ramsden, The Hofmeister series: salt and solvent effects on interfacial phenomena. *Q. Rev. Biophys.* 30 (1997) 241–278. A concise review of ion-specific effects on interfacial interactions

E. Kellenberger, Assembly in biological systems. In: *Polymerization in Biological Systems*, CIBA Foundation Symposium 7 (new series). Amsterdam: Elsevier (1972). Inspirational survey.

J.-M. Lehn, Supramolecular polymer chemistry. In: *Supramolecular Polymers* (ed. A. Ciferri), pp. 615–641. New York: Dekker (2000). Many examples.

P.A. McKeown, J. Corbett, P. Shore and P. Morantz, Ultraprecision machine tools—design principles and developments. *Nanotechnology Perceptions* 4 (2008) 5–14. Good summary of the state-of-the-art in ultraprecision engineering.

A.G. Mamalis, A. Markopoulos and D.E. Manolakos, Micro and nanoprocessing techniques and applications. *Nanotechnology Perceptions* 1 (2005) 63–73. Good summary of the state-of-the-art in semiconductor processing technology.

C.J. van Oss, *Interfacial Forces in Aqueous Media*. New York: Dekker (1994). Useful treatment of the subject.

Chapter 8

Bionanotechnology

Bionanotechnology is defined as the application of biological molecules and systems to nanotechnology (note that biotechnology is the directed use of organisms to make useful products, typically achieved by genetically modifying organisms). If the nanotechnology is then applied to human health (nanomedicine or nanobiotechnology), consistency in terminology would demand that we call it bionanobiotechnology. The discovery of some of the mechanistic details of complicated biological machinery such as the ribosome that encodes the sequence of nucleic acids as a sequence of amino acids (called "translation" in molecular biology) was happening around the time that Eric Drexler was trying to promote his assembler-based view of nanotechnology, and these biological machines provided a kind of living proof of principle that elaborate and functionally sophisticated mechanisms could operate at the nanoscale. Some of these biological machines are listed in Table 8.1. There are many others, such as the mechanism that packs viral DNA ultracompactly in the head of bacteriophage.

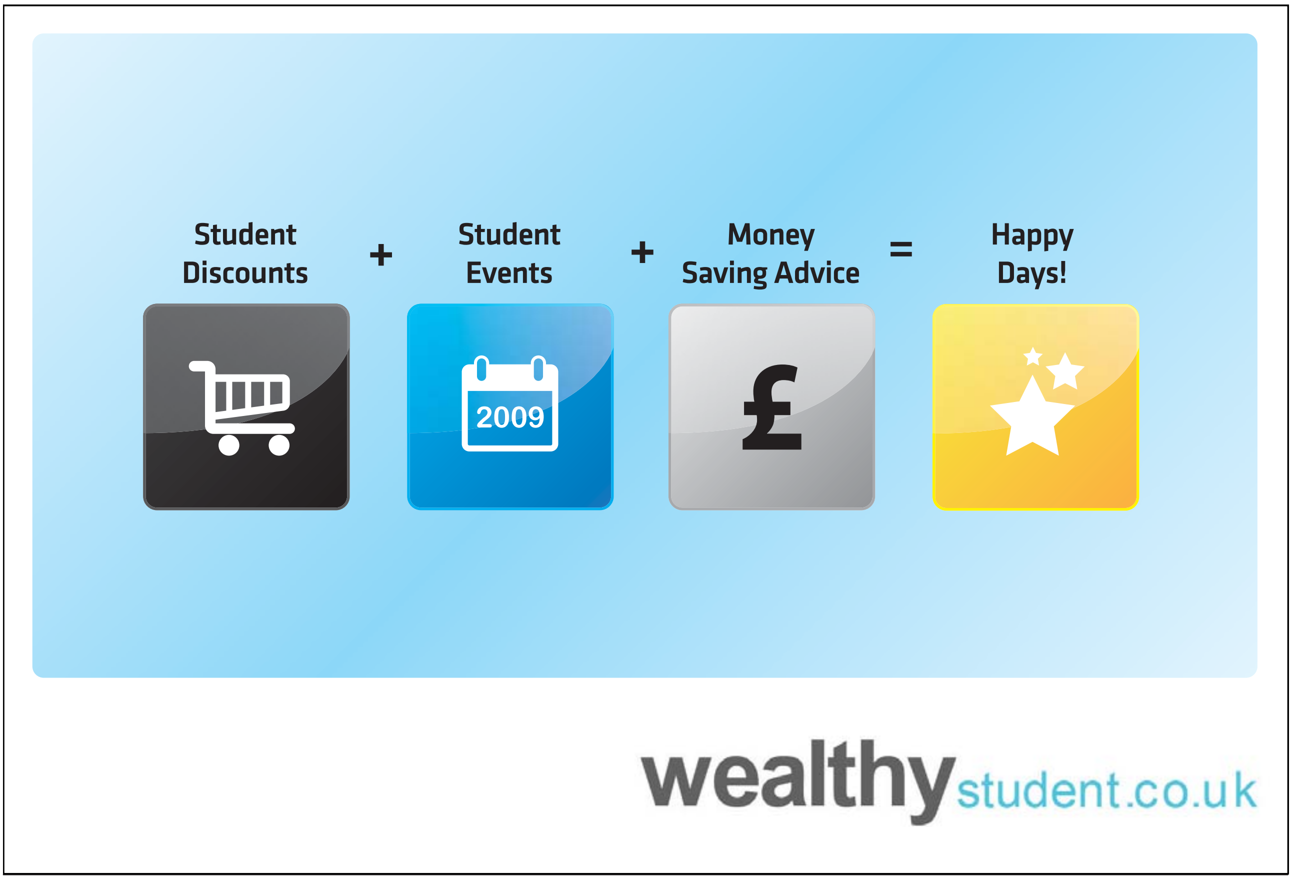
These machines are essentially proteins, which are fabricated by a self-assembly process (§7.3.2). Some of them show consummate scaling up to the macroscopic realm. Muscle is probably the best example: although the actin-myosin pair that is the molecular heart of muscular action develops a force of a few piconewtons, by arranging many "molecular muscles" in parallel, large animals such as elephants can develop kilowatts of power, as humans have known and made use of for millennia.

Table 8.1: Biological nanosized machines.

Name	natural function	state of knowledge[a]
Muscle (myosin)	pulling	C,S,T
Kinesin	linear motion	C,S,T
Nerve	information transmission	T
ATPase	synthesis of ATP from proton e.p.g.[b]	C,S,T
Bacteriorhodopsin	generation of proton e.p.g. from light	C,T
Transmembrane ion pump	moving selected ions against an adverse e.p.g.	C,T
Haemoglobin	oxygen uptake and release	C,T

[a] C, crystal structure determined; S, single-molecule observation of operation; T, theoretical mechanism available.

[b] Electrochemical potential gradient.



8.1 Biomolecules

Many structures and especially devices produced in living systems are constituted from biopolymers designed to fit to congeners with exquisite specificity and precise stoichiometry. One of the challenges of biomimetic nanotechnology is to recreate these attributes with simpler artificial systems—without much success until now. Could one, for example, create an oxygen carrier working like haemoglobin but with a tenth the number of atoms? Possibly, but such a "lean" carrier would be far less resilient to fluctuations in its working environment.

Polypeptides (PP) (proteins) are linear polymers of amino acids (H_2N–CHR–COOH, where R (bonded to the central C) is a variable side chain ("residue")—there are twenty different natural ones. To polymerize them, water is eliminated between –COOH and H_2N– to form the peptide bond), hence there is a common backbone (linked via the "peptide" bond) with variable side chains—short aliphatic groups, small aromatic groups, carboxylate, amine, hydroxyl, etc. functionalities. Template-directed synthesis with a very high yield is used in nature, with the templates being closely related to genes via the genetic code (triplets of nucleotide bases encode each amino acid). After synthesis (polymerization), they fold, often spontaneously, to a compact structure according to a least-action principle (see §7.3.2). Typical natural proteins have $50 \sim 500$ amino acids. Depending on their sequence, they adopt a definite remembered conformation (proteins acting as devices, rather than having a passive structural role, have two or more stable conformations) and can carry out varied functions, ranging from essentially structural or scavenging to enzymes and motors. Some proteins (called glycoproteins) are branched with oligosaccharides attached to certain residues.

Nucleic acids (NA) are polymerized from nucleotides constituted from a sugar, a phosphate group, and a "base" derived from a purine or pyrimidine (aromatic heterocycle). The sugar and phosphate are polymerized by eliminating water to form a linear backbone, with the bases playing the role of the residues in PP. There are 4 natural bases, abbreviated A, C, G, T (in deoxyribonucleic acid, DNA) and A, C, G, U (in ribonucleic acid, RNA). The bases pair preferentially: A with T (or U), via 2 hydrogen bonds, and C with G via 3 hydrogen bonds (complementary base pairing, CBP). Linear polymers are linked via the sugar. Template-directed synthesis with a very high yield is used in nature to create the polymers. The templates are the genes (DNA), and operate according to the principle of CBP. During polymerization RNA spontaneously folds to a definite compact structure according to a least-action principle (see §7.3.2), in which base-pairing via hydrogen bonding is equivalent to the potential energy, and loop and hairpin formation is equivalent to the kinetic energy. DNA forms the famous double helix.

Polysaccharides (PS) and oligosaccharides (OS) are linear or branched polymers of diverse sugar (cyclic oligoalcohol) monomers, linked via water elimination ("condensation") at any of the several hydroxyl groups. The problem of predicting their structure is not yet solved. Polymerization is not templated (i.e., not under direct genetic control) and there is variability (to a degree that is only poorly characterized) in sequence and length of the polysaccharides found fulfilling the same function in comparable organisms.

8.2 Characteristics of biological molecules

The energy contained in a given system can be divided into two categories: (a) the multitude of microscopic or thermal motions sufficiently characterized by the temperature; and (b) the (usually small number of) macroscopic, highly correlated motions, whose existence turns the construction into a machine (a device). The total energy contained in the microscopic degrees of freedom may be far larger than those in the macroscopic ones, but nevertheless the microscopic energy can usually be successfully neglected in the analysis of a construction (in informational terms, the macrostates are remembered, but the microstates are not).

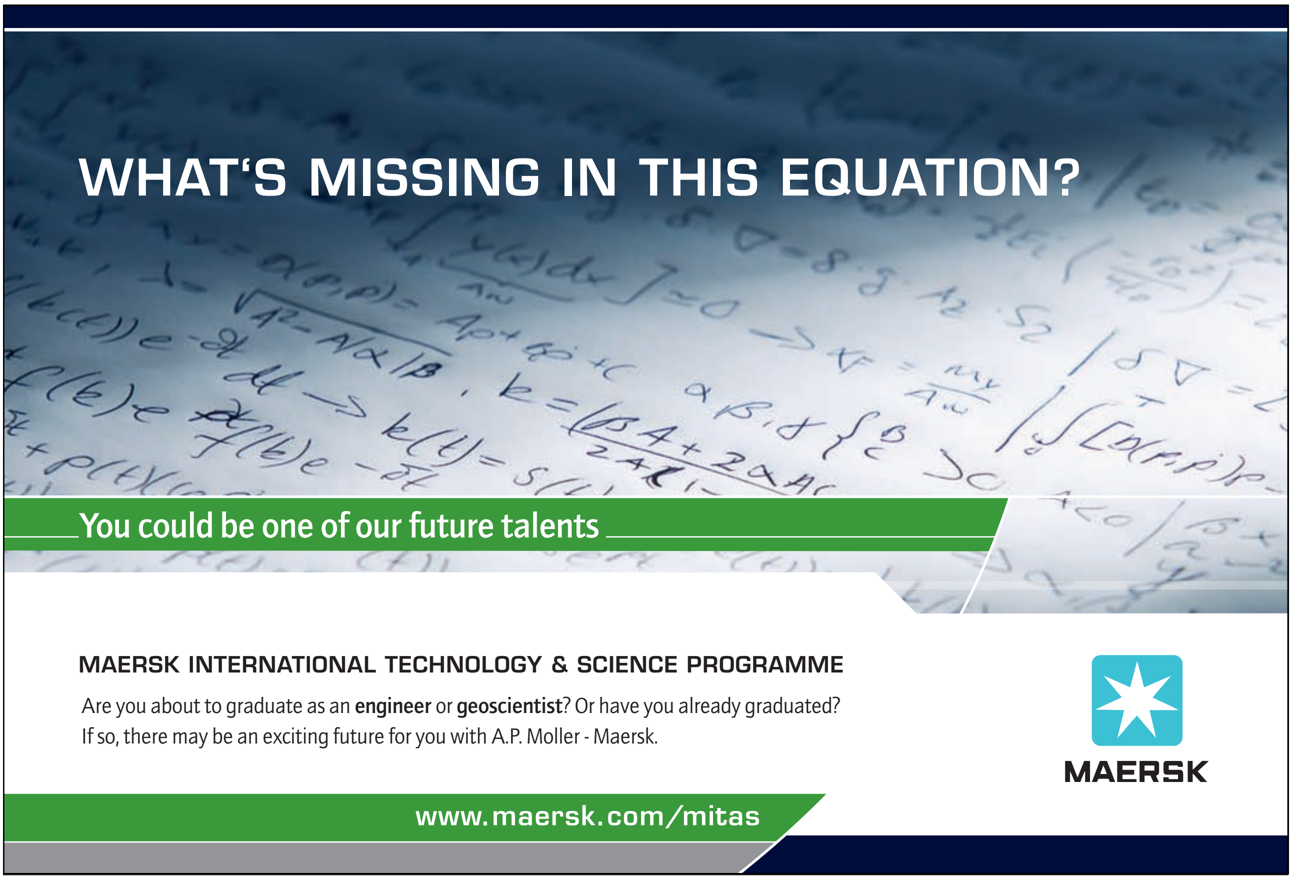
Biological molecules are constructions. Hence a statistical approach, in which the motions of an immense number of individual particles are subsumed into a few macroscopic parameters such as temperature and pressure, is inadequate.

A construction uses only an insignificant fraction of the Gibbs canonical ensemble and hence is essentially out of equilibrium. This is different from thermodynamic nonequilibrium—it arises because the system is being investigated at time scales much shorter than those required for true statistical equilbrium. Such systems exhibit "broken ergodicity", as epitomized by a cup of coffee in a closed room to which cream is added and then stirred. The cream and coffee equilibrate within a few seconds (during which vast amounts of microinformation are generated within the whorled patterns); the cup attains room temperature within tens of minutes; and days may be required for the water in the cup to saturate the air in the room.

Broken ergodicity may be regarded as a generalization of *broken symmetry*, which leads to a new thermodynamic quantity, the *order parameter* ξ whose value is zero in the symmetrical phase. ξ may be thought of as conferring a kind of generalized rigidity on a system, allowing an external force applied at one point to be transferred to another. Some protein molecules demonstrate this very clearly: flash photolysis of oxygenated haemoglobin (causing the oxygen molecule to dissociate from the iron core of the porphyrin (the haem) to which it is bound) causes motion of the iron core of the haem, which results in (much larger) movement at the distant intersubunit contacts, leading ultimately to an overall change in the protein conformation involving hundreds of atoms.

8.3 Mechanism of biological machines

Active proteins (i.e., all proteins apart from those fulfilling a purely passive structural or space-filling role) have two or more stable conformations. Useful as X-ray diffraction is for determining the structure of proteins, it has the disadvantage of usually fixing the protein in just one of these conformations in the crystal that is used to diffract the X-rays. If a fraction of the proteins does happen to be present in one of the other stable conformations, this is usually regarded as "disorder" and any information about it is lost during structural refinement of the diffraction data.



The prototypical example is an enzyme E catalysing (say) the decomposition of a molecule A–B (called the substrate of the enzyme in the biochemical literature) into products A + B. In this case, work has to be done to break the chemical bond A–B; the principle can equally well be applied to any action where work is done, such as pulling on a spring (as in muscle). The binding and release of oxygen to and from haemoglobin also works on this principle.

The enzyme consists of an active site and the rest of the protein, which may be considered to be much bigger than the active site. The "rest" has two stable conformations, E and $\tilde{\mathrm{E}}$. The mechanism proceeds in four stages:

Blumenfeld mechanism of biological machines.

1. A complex is formed between the substrate and the enzyme, A–B + E → (A–B)E*. A–B binds to the active site, releasing free energy and resulting in a local conformational change, which creates a strain between the active site and the rest of the protein. Local fast vibrational relaxation takes place on the picosecond time scale, but the active site is no longer in equilibrium with the rest of the molecule and the resulting strain modifies the energy surface on which the enzymatic reaction takes place. The asterisk denotes that the protein is overall in a strained, nonequilibrium state. Strain creation requires energy, but its magnitude must of course be less than the energy of binding.
2. The complex slowly relaxes to a new conformation $\tilde{\mathrm{E}}$, releasing the energy to drive the energy-requiring breaking of the A–B bond: (A–B)E* → A$\tilde{\mathrm{E}}$B. This is the elementary act of the enzymatic reaction. This conformational relaxation involves making and breaking a multiplicity of weak bonds, but at a slower rate than the reaction being catalysed.
3. The product-enzyme complex is decomposed, i.e. the products are released: A$\tilde{\mathrm{E}}$B → A + B + A$\tilde{\mathrm{E}}^*$. Release of the products from the active site again creates strain between it and the rest of the protein molecule.
4. Finally, the strained enzyme slowly relaxes back to its initial conformation: $\tilde{\mathrm{E}}^*$ → E.

An interesting prediction of this mechanism is that the rate of the overall reaction A–B → A + B should exhibit an inverse Arrhenius temperature dependence, because increasing the temperature accelerates conformational relaxation (step 2), and hence shortens the time during which the strained molecule is able to accelerate the reaction enzymatically.

8.4 Biological motors

Figure 8.1 shows the working cycle of the myosin motor that powers muscle. Key details have been observed via an experimental approach based on single molecule manipulation. For example, myosin is immobilized on the substratum and allowed to interact with actin tethered to beads held in optical traps.[1] It was long assumed that there was a direct correlation between the hydrolysis of ATP and the mechanical work performed by the motor; that is, each hydrolysed molecule resulted in one unit of displacement Δx ("tight coupling model"). Single molecule observations do not, however, support this assumption. Simultaneous monitoring of individual hydrolysis events (by microscopic observation of fluorescently labelled adenosine) and bead displacements due to the mechanical force exerted on the actin have clearly shown that mechanical force might be generated several hundred milliseconds after release of ADP.[2] Apparently, myosin can store the chemical energy from several ATP hydrolysis events, and may then subsequently perform several mechanical work steps ("loose coupling model"). The diagram in Figure 8.1 is therefore somewhat simplistic. Nevertheless, this energy storage is fully within the framework of the Blumenfeld mechanism.

Other mechanical motors such as kinesin moving on microtubules also operate on this principle. The gating of transmembrane ion channels ("ion pumps") presumably operate in a similar fashion, although this still needs to be elucidated. The general principle of conformational strain between a binding site and the rest of the protein (generated by binding or release of a small molecule) driving (electro)chemical or mechanical work can be expected to be universal in living systems.

---

[1] See, e.g., T. Yanagida, Y. Harada and A. Ishijima, Nanomanipulation of actomyosin molecular motors in vitro: a new working principle. *Trends in Biochemical Sciences (TIBS)* 18 (1993) 319–323; T. Funatsu et al., Imaging of single fluorescent molecules and individual ATP turnovers by single myosin molecules in aqueous solution. *Nature* 374 (1995) 555–559; T. Nishizaka et al., Unbinding force of a single motor molecule of muscle measured using optical tweezers. *Nature* 377 (1995) 251–254.

[2] A. Ishijima et al., Simultaneous observation of individual ATPase and mechanical events by a single myosin molcule during interaction with actin. *Cell* 92 (1998) 161–171.

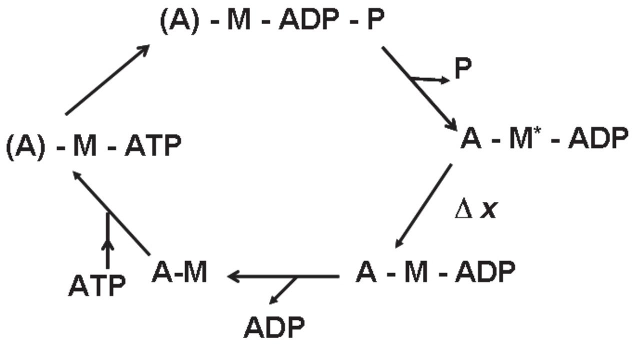


Figure 8.1: The working cycle of muscle. M denotes myosin and A denotes actin; (A) is actin weakly bound to myosin; the asterisk denotes myosin in a strained, nonequilibrium state. Binding of adenosine triphosphate (ATP) to the myosin results in weakened binding to actin, and hydrolysis of the ATP to adenosine diphosphate (ADP) and the subsequent release of phosphate generate strain between the ATP-binding site and the rest of the myosin molecule; the relaxation of this strain drives the movement of the motor (Δx, 8–10 nm) during which mechanical work is done. The hydrolysis of one ATP molecule yields a chemical energy E of 5–20 k_BT; the force E/Δ exerted by one myosin motor can therefore be estimated 2–10 pN.

Another example is the assembly and disassembly of microtubule filaments from the globular protein tubulin. During the normal state of the eucaryotic cell, these filaments pervade the cytoplasm, acting as tracks for kinesin motors transporting molecules and supramolecular complexes. Prior to eucaryotic cell division (mitosis) the duplicated genome is compactified into chromosomes and the nuclear membrane and the microtubule filaments network are degraded. The duplicated genome must be separated and two halves relocated in the two halves of the cell that will become separate cells at the division. How can this be accomplished? Two centrosomes (protein complexes) form asteriated poles at opposite ends of the cell, and microtubules repeatedly elongate at a speed v_g out from them in random directions, followed by catastrophic disassembly leading to abrupt shrinkage with speed $v_s, v_s \gg v_g$. The process continues until a microtubule filament reaches a chromosome, upon which it attaches itself and drags half of it towards the centrosome. The result is each duplicated genome located in separate halves of the cell, after which the rest of the division process takes place. The dynamic instability (assembly-disassembly) is characterized by length fluctuations of the order of the mean microtubule length, hinting at a phase transition. Let f_{gs} denote the frequency of switching from growth to shrinkage and f_{sg} the frequency of switching from shrinkage to growth. When $v_g f_{sg} = v_s f_{gs}$ growth switches from unbounded (corresponding to the assembly of the microtubule filament network) to bounded. At this point the average microtubule length $\bar{\ell} = v_g v_s/(v_s f_{gs} - v_g f_{sg})$ diverges. The molecular origin of growth and shrinkage lies in the fact that tubulin monomers can bind to guanosine triphosphate (GTP), and the complex can spontaneously assemble to form filaments. But the GTP slowly hydrolyses spontaneously to guanosine diphosphate (GDP), thereby somewhat changing the tubulin conformation such that it prefers to be monomeric. However, the monomers can only be released from the end; disassembly can be initiated if the rate of GTP hydrolysis exceeds that of tubulin addition for a while. The overall process is a remarkably effective way of searching a restricted volume for an object when no prior information about the location of the object exists.

8.5 The cost of control

The force F which has to be applied to a molecular lever requires accurate knowledge of its position x if reversible work is to be performed. Specifying the positional accuracy as Δx, the uncertainty principle gives the energy requirement as[3]

$$\Delta E \geq hc/(4\Delta x) \tag{8.1}$$

---

[3] This section is based on B.F. Gray, Reversibility and biological machines. *Nature (Lond.)* 253 (1975) 436–437.

where h is Planck's constant ($= 6.63 \times 10^{-34}$ J s) and c the speed of light in the vacuum ($= 3.00 \times 10^8$ m/s), ΔE is obviously negligible for macroscopic systems millimetres in size. The uncertainty in the force $F(x)$ generated at x is

$$\Delta F = F(x) \pm \Delta x(\mathrm{d}F/\mathrm{d}x) \tag{8.2}$$

To compute the work W done by the system, eqn (8.2) is integrated over the appropriate x interval. The first term on the right hand side yields the reversible work W_{rev}, and the second term yields $-\Delta x \sum_j \mid F_j - F_{j+1} \mid$ for any cycle involving j steps.

The energy conversion factor ϵ is

$$\epsilon = W/(Q + \Delta E) \tag{8.3}$$

where Q is the net energy input during the cycle. With the help of inequality (8.1), the ratio of this to the classical conversion factor $\epsilon_{\mathrm{rev}} = W_{\mathrm{rev}}/Q$ is

$$\epsilon/\epsilon_{\mathrm{rev}} \leq (1 - \alpha/z)/(1 + z) \tag{8.4}$$

where

$$\alpha = hc \sum_j \mid F_j - F_{j+1} \mid /(4QW_{\mathrm{rev}}) \tag{8.5}$$

and the relative energy cost of control is

$$z = \Delta E/Q \tag{8.6}$$

The maximum possible value of the ratio $\epsilon/\epsilon_{\mathrm{rev}}$ is obtained by substituting z by its optimal value z_{opt}, obtained from the turning point of equation (8.4):

$$z_{\mathrm{opt}} = \alpha(1 + \sqrt{1 + 1/\alpha}) \tag{8.7}$$

It is

$$\left(\frac{\epsilon}{\epsilon_{\mathrm{rev}}}\right)_{\mathrm{max}} = \frac{1 - 1/(1 + \sqrt{1 + 1/\alpha})}{1 + \alpha(1 + \sqrt{1 + 1/\alpha})} \tag{8.8}$$

If more energy than z_{opt} is used, then α decreases because of the energy cost of information; if less, then ϵ decreases because of the irreversibility (dissipation etc.).

For a macroscopic system these quantities are insignificant. But consider the myosin motor (Figure 8.1): taking $F_j \approx 2$ pN, the displacement $x \approx 10$ nm, and $Q \approx 0.067$ aJ (the energy released by hydrolysing a single ATP molecule), then the energy cost of optimum control, Qz_{opt}, is equivalent to hydrolysing almost 150 ATP molecules and $(\epsilon/\epsilon_{\mathrm{rev}})_{\mathrm{opt}} = 0.0033$. Reversible operation is evidently far from optimal; chemical to mechanical conversion occurs at a finite rate which may essentially be uncontrolled, i.e. determined intrinsically. This analysis and conclusion allows the loose coupling model for muscle (§8.4) to be rationalized.

8.6 Biophotonic devices

Apart from the marvellous intricacy of the biological machinery that converts light into chemical energy, which at present only serves to inspire nanotechnological mimics, there are other, simpler, photoactive proteins, robust enough to be incorporated into artificial devices. Molecules based on the chromophore rhodopsin (such as the primary optical receptor in the eye) seem to have a special place here.

One of the most remarkable of these photoactive proteins is bacteriorhodopsin, which constitutes about a third of the outer membranes of the archaeon (extremophilic procaryote) *Halobium salinarum*, living in salt lakes. The optically active site of the protein is the conjugated polyene rhodopsin, and when it absorbs a photon of red light, there is a conformational change generating strain between it and the rest of the protein, which translocates a proton across the membrane (according to the mechanism outlined in §8.3). The process is called the bacteriorhodopsin photocycle, and a key intermediate state is called M; the altered interaction between the chromophore and its protein environment gives it an absorption maximum of 410 nm (Figure 8.2).

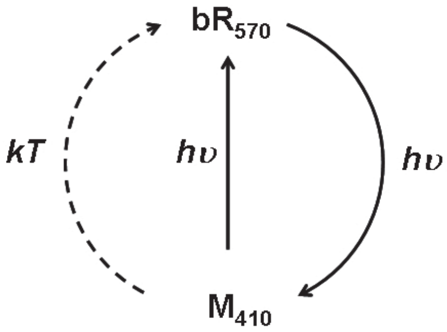


Figure 8.2: The bacteriorhodopsin photocycle. A 570 nm photon absorbed by the ground state bR_{570} (the subscript indicates the wavelength of maximum adsorption of the molecule) rapidly (within a few microseconds) transforms (through a series of intermediate stages) the molecule to the relatively stable intermediate M_{410}. This state slowly relaxes thermally back to the ground state, but it can also be rapidly converted by a 410 nm photon. The thermal stability of the M state can be extended almost indefinitely by genetically modifying the protein.

H. salinarum can be easily grown and the bacteriorhodopsin harvested in the form of "purple membrane fragments"—pieces of outer membrane consisting of an array of bacteriorhodopsin with the membrane lipid filling the interstitial volume. These fragments can be oriented and dried, in which state they can be kept under ambient conditions for 10 years or more without any loss of activity; they have already generated considerable interest as a possible optical storage medium, using a bacteriorhodopsin mutant, the M state of which is almost indefinitely thermally stable. In such an optical memory, the ground state would represent "0" and the M state would represent "1".

Native bacteriorhodopsin can be used to construct an optically switched optical switch (Figure 8.3). Not only can the switch operate extremely rapidly (at megahertz frequencies and above), but only weak light is needed. The remarkable optical nonlinearity of the protein is manifested by exposing it to a single photon!

8.7 DNA as construction material

The specific base-pairing of DNA together with the ease of nucleotide polymerization (it can be accomplished using automated equipment) has engendered interest in the design and construction of artificial nano artefacts of arbitrary shape made from DNA. At present, the design of the required DNA strands is a laborious, empirical process; but in principle both DNA and

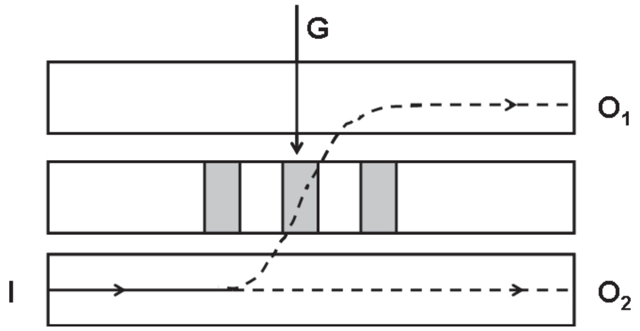


Figure 8.3: An optically switched optical switch. The basic construction is a slab of bacteriorhodopsin-containing purple membrane fragments sandwiched between two optical waveguides. Initially, let us suppose that an optical wave introduced at the input I is guided to the structure to emerge at output O_2. If a grating (indicated by the stripes) is holographically generated in the bacteriorhodopsin slab by illuminating with light of 570 nm (from G, the "gate"), light in the lower waveguide will be coupled out and coupled into the upper waveguide, emerging at output O_1. Destroying the grating by illuminating with light of 410 nm will cause the output to revert to O_2.

RNA could become universal construction materials (provided they are not required to be stable in extreme conditions). The fact that enzymes constructed from RNA are known to exist in nature suggests that ultimately devices could also be made.

8.8 Further reading

L.A. Blumenfeld, D.S. Burbajev and R.M. Davydov, Processes of conformational relaxation in enzyme catalysis. In: *The Fluctuating Enzyme* (ed E.R. Welch), pp. 369–402. New York: Wiley (1986). An account of the Blumenfeld mechanism.

J. Chen, N. Jonoska and G. Rozenberg (eds), *Nanotechnology: Science and Computation.* Berlin: Springer (2006). Mostly about DNA construction and computation.

J.J. Ramsden, Biophysical Chemistry. In: *Encyclopaedia of Chemical Physics and Physical Chemistry*, (eds J.H. Moore and N.D. Spencer), pp. 2509–2544. Philadelphia: IOP (2001). Covers much of the biophysicochemical background.

J. Youell and K. Firman, Biological molecular motors for nanodevices. *Nanotechnology Perceptions* 3 (2007) 75–96. Comprehensive overview of motors in biology.

Chapter 9

New fields of nanotechnology

This chapter deals with three significant areas of research activity aimed at developing novel technology in the medium to long-term. They concern information processing, medicine and energy. The chapter closes with a brief description of three new "nano" terms: nanification, vastification and sensorization.

9.1 Quantum computing and spintronics

Extrapolation of Moore's law to about the year 2020 indicates that component size will be sufficiently small for the behaviour of electrons within them to be perturbed by quantum effects. This implies a profound perturbation of the proper functioning of the technology, and no solution to this problem within the current framework is in view. Quantum computing can be thought of as "making a virtue out of necessity", creating computational devices based on the principles of quantum logic.

The key features of quantum objects of interest for computational purposes are superposition—an object can be in several different states simultaneously—and entanglement (with its environment). Operations can be carried out internally, maintaining superposition, which is only destroyed at the very end of the computation when a single output is required.

The architecture of computers should then be wholly reconceived in order to exploit the peculiarities of quantum mechanics. This means in particular that a particle can exist in two states simultaneously. A cluster of electrons (physically instantiating a bit) in a conventional computer represents either zero or one. The value of a qubit, on the other hand, which might be physically instantiated as a single electron localized on a quantum dot, depends on its position relative to other electrons. For example two electrons can exist in four different states—00, 01, 10, and 11—depending on their relative positions. If the electrons interact (are entangled) with each other, then any operation carried out on one electron will simultaneously be carried out on the other—implying that one operation is carried out on four different states of the same time. Hence a computer with just 32 bits could perform more than a thousand million operations simultaneously.

One of the biggest problems with current supercomputers is energy dissipation. They require tens of kilowatts of energy to run and generate vast amounts of heat. Landauer showed in 1961 that almost all operations required in computation can be performed reversibly, thus dissipating no heat. Reversible computation is possible on a quantum computer.

The physical embodiment of a bit of information—called a qubit in quantum computation—can be any absolutely small object capable of possessing the two logic states 0 and 1 in superposition—e.g. an electron, a photon or an atom. A single photon polarized horizontally (H) could encode the state $|0\rangle$ and polarized vertically (V) could encode the state $|1\rangle$ using the Dirac notation. The photon can exist in an arbitrary superposition of these two states, represented as $a\,|\mathrm{H}\rangle + b\,|\mathrm{V}\rangle$, with $|a|^2 + |b|^2 = 1$. The states can be manipulated using birefringent waveplates, and polarizing beamsplitters are available for converting polarization to spatial location. With such common optical components, logic gates can be constructed. Another possible embodiment of a qubit is electron spin. Figure 9.1 illustrates a not-and gate based on single spin logic. A "true" spintronics device encodes binary information as spin, in contrast to the so-called spin transistor, in which spin merely mediates switching.

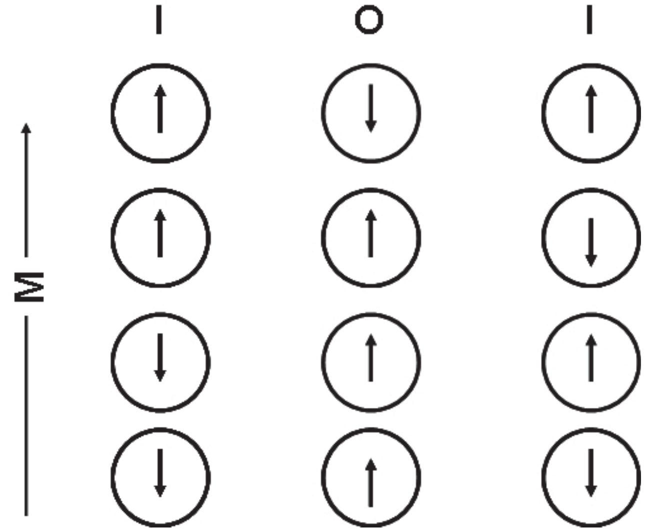


Figure 9.1: A spintronic NAND gate. The two inputs are the left and right columns, and the output is the central column. The physical entities are quantum dots. The upspin state parallel to the global magnetic field **M** represents 1 and the downspin state represents 0. Exchange interaction causes nearest neighbours to prefer antiparallel spins, but if there is a conflict, the global magnetic field biases the output to 1.

One important application already been envisaged is in factorizing large numbers, an operation that is a crucial part of code breaking. In this particular application, the typical performance of prototype quantum computers is already competitive with that of current supercomputers.

9.2 Nanomedicine

Recall that nanomedicine is defined as the application of nanotechnology to health (veterinary practice is usually included), and hence is virtually synonymous with nanobiotechnology—considering health as applied biology.

9.2.1 Drug delivery

Therapy on the basis of externally administered (oral, intravenous, subcutaneous, etc.) pharmaceutical preparations suffers from three more or less severe problems: (i) as a rule, they have had to be administered systemically, and since their therapeutic effect is usually a differential one (i.e., it is slightly more toxic to the offending cells than to the rest of the body) the entire body is poisoned to some degree; (ii) many therapeutically efficacious compounds are destroyed by the regular defence mechanisms of the body against foreign invaders before they arrive at their destination; and (iii) the bilayer lipid membrane, ubiquitous in eucaryotic cells, is increasingly recognized as an important target for medicinal drugs, but the hydrophobicity of the drug molecule required to ensure its high affinity for the membrane is incompatible with transporting it through the predominantly aqueous (hydrophilic) media between the point of administration and the target membrane.

These are of course generalizations with exceptions, but numerous examples show their widespread validity. Nucleic acids (e.g., small interfering RNA) might be very valuable, and certainly they could be targeted very specifically, if they could penetrate to the nuclei of cells, but they are hydrolysed and destroyed by circulating enzymes long before.

The aim of drug delivery is to encapsulate a therapeutic agent to disguise its properties until it reaches its target and is released. The encapsulation must therefore be able to respond to its environment. There is a general requirement that it should not excite an adverse immune response; that is, its surface must be biocompatible, which typically means it should not bind and denature the proteins circulating in the blood. A simple example is to place a drug destined for the stomach in a hollow sphere of calcium carbonate. This material will dissolve in the high concentration of hydrochloric acid found in the stomach, releasing the drug. In all other respects it should be possible to handle the encapsulated drug in the same way as its unencapsulated congener, which is where the nanotechnology comes in: a nanosphere can be treated in the same way as a largish molecule.

9.2.2 Biosensors

The Holy Graal of clinical biosensing is continuous, noninvasive monitoring. Currently, most tests require a sample of the relevant biofluid (e.g., blood) to be drawn from the patient. For most people this is a somewhat unpleasant procedure, and hence the tests are carried out infrequently. It is, however, recognized that much more insight into a patient's pathological state could be obtained by frequent, ideally continuous, monitoring. At present, this is only possible in intensive care stations, where the patient is immobilized, and even then continuous invasive monitoring does not take place (the oxygen content of the blood is monitored noninvasively by analysing the optical reflectance spectrum of the skin covering blood vessels). It seems to be a very difficult technical problem to extend such noninvasive analysis to the plethora of biomarkers currently under intensive study as symptomatic of disease or incipient disease. An alternative approach is to develop sensors so tiny that they can be semipermanently implanted inside the body, where they can continuously monitor their surroundings.

Because of the large and growing number of afflicted people, diabetes has received overwhelmingly the most attention. The sensing requirement is for glucose in the blood. The glucose sensor follows classic biosensing design: a *recognition element* to capture the analyte (glucose) mounted on a *transducer* that converts the presence of captured analyte into an electrical signal (Figure 9.2). The recognition element is typically a biological molecule, the enzyme glucose oxidase, hence (if small enough) this device can be categorized as both nanobiotechnology and bionanotechnology.

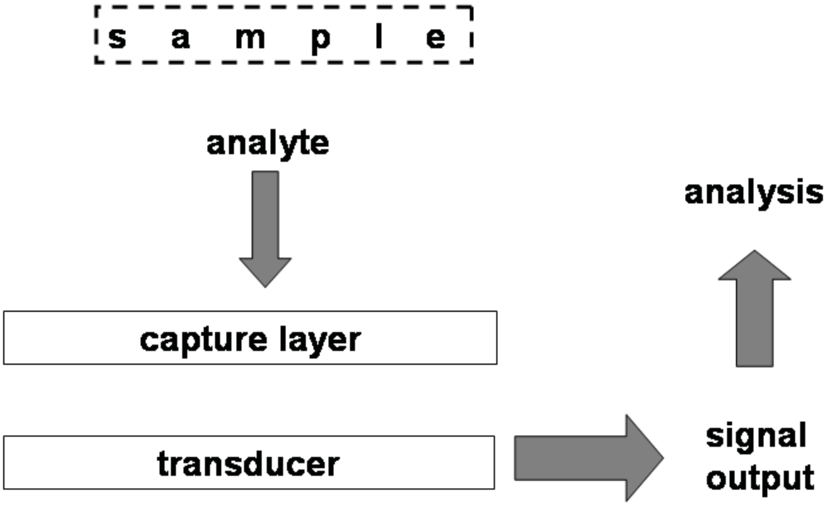


Figure 9.2: A prototypical biosensor. The capture layer concentrates the analyte in the vicinity of the transducer, which reports the concentration of analyte in the capture layer, which is directly related to the concentration in the sample.

Both components of the biosensor are excellent candidates for the application of nanotechnology. Molecular recognition depends on a certain geometrical and chemical arrangement of atoms in some sense complementary to the analyte molecules, together with cooperative motions to enhance affinity. Atom-by-atom assembly therefore represents the perfect way to artificially fabricate recognition elements. The ultimate goal of the transducer is to detect a single captured analyte molecule, hence the smaller it can be made, the better.

9.2.3 Automated diagnosis

Medicine has already become accustomed to depending on heavy computations in the various tomographies that are now routine in large hospitals. Diagnosis is essentially a problem of pattern recognition: an object (in this case, the disease) must be inferred from a collection of features. Although there have already been attempts to ease the work of the physician by encapsulating his or her knowledge in an expert system that makes use of the physician's regular observations, significant progress is anticipated when measurements from numerous implanted biosensors are input to the inference engine. This is an example of indirect nanotechnology: the practical feasibility depends on the availability of extremely powerful processors, based on chips having the very high degree of integration enabled by nanoscale components on the chips.

9.2.4 Custom synthesis

Enough evidence has accumulated for it to be generally recognized that many drugs are typically efficacious against only part of the population. This may partly be due to genetic diversity, and partly to other factors, which have not yet been characterized in molecular detail. In the case of genetic diversity, it is possible that almost all the pharmaceutically relevant DNA sequence variants occur in haplotype blocks, regions of 10,000 to 100,000 nucleotides in which a few sequence variants account for nearly all the variation in the world human population (typically, five or six sequence variants account for nearly all the variation). If a drug that has been found to be efficacious against the majority haplotype variant can be made to be efficacious against the others by small chemical modifications, then the task of the drug developer would not be insuperable. At present, clinical trials do not generally take account of haplotype variation. Advances in sequencing, in which nanotechnology is helping both through the development of nanobiotechnological analytical devices (although it seems that microtechnological "labs on chips" may be adequate to fulfil needs) and through more powerful information processing, should make it possible in the fairly near future for haplotype determination to become routine. It is, however, not known (and perhaps rather improbable) whether small modifications to a drug would adapt its efficacity to other haplotype variants. At any rate, different drugs will certainly be needed to treat different groups of patients suffering from what is clinically considered to be the same disease. The development of nanomixers (if micromixers prove to be inadequate) would represent a key step in making custom synthesis of drugs for groups of patients, or even for individual patients, economically viable.

9.3 Energy

Undoubtedly natural photosynthesis is only possible through an extremely exact arrangement of atoms within the photosystems working within plant cells, and the more precisely artificial light harvesters can be assembled, the more successful they are.[1] An enduring concept in artificial systems is photoelectrochemistry using nanoparticles (see Figure 9.3), although to date no completely successful realization has been achieved.

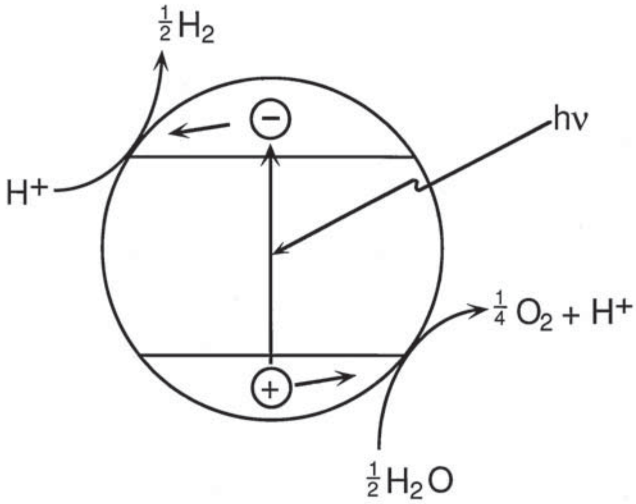


Figure 9.3: Radiant energy harvesting using semiconductor nanoparticles. The absorption of a photon raises an electron from the valence band to the conduction band, leaving a positive hole in the former. The electron ⊖ is a strong reducing agent, and the hole ⊕ a strong oxidizing agent, possibly capable of, respectively, reducing and oxidizing water to hydrogen and oxygen. At present, this concept of nanoparticle photoelectrochemistry can be used to sacrificially destroy persistent organic pollutants allowed to adsorb on the particles. This is the basis of self-cleaning windows.

Nanotechnology, by enabling significant miniaturization of the information processing parts of devices such as computers and mobile phones, introduces a requirement for correspondingly miniature energy sources to power them. This is itself becoming an active field, in which nanomaterials seem likely to play a large part, not least due to the complex, multiple requirements that the material components of the devices have to fulfil.

---

[1] See, e.g., L. Schmidt-Mende et al., Efficiency improvement in solid-state dye-sensitized photovoltaics with an amphiphilic ruthenium dye. *Appl. Phys. Lett.* 86 (2005) 013504.

9.4 Three concepts

Any revolution introduces new terms. A great many of the terms that have been invented for nanotechnology are simply existing words with "nano" prefixing them. Their meaning is therefore self-evident; a formal definition is only needed in some cases to remove ambiguity. In this section, three distinctively new words are introduced.

9.4.1 Nanification

To approach the meaning of the word "nanification", think of miniaturization but in a more all-encompassing fashion. To nanify electronics, for example, is not only to make individual components smaller (right down to the nanoscale) but also to adapt all parts of the industry to that situation, including design aspects. In short, nanification means introducing nanotechnology in an integrated rather than a piecemeal fashion. Hence, to nanify manufacture is, ultimately, to introduce molecular manufacturing, which involves not only the actual assembly devices themselves, but also logistics, indeed the entire supply chain, and the (re)organization of the economic system.

9.4.2 Vastification

It has been already pointed out that vast numbers of objects are a corollary of their being very small. The main qualitatively distinctive consequence of vastification is that explicit specification and control become impracticable. To tackle this problem, an evolutionary approach will be required, notably in design (see §3.2), but very possibly also in operation (e.g., a stigmergic approach). Biomimicry of the human brain may become the favoured approach.

9.4.3 Sensorization

Sensorization means, literally, the incorporation of vast numbers of sensors, something that is only feasible if they are nanoscale sensors, from the viewpoints of both cost and space requirements. Sensorization is likely to lead to a qualitatively different way of handling situations in at least four areas:

- Structural (civil) engineering: bridges, walls, buildings, etc. Sensors—typically optical fibre Bragg gratings, the technology of which already exists—will be incorporated throughout the structure (e.g., embedded in the concrete, or in the wings of an aircraft). The output of these sensors is indicative of strain, the penetration of moisture, and so forth.
- Process (including chemical) engineering: sensors embedded throughout machinery and reaction vessels will monitor physical (e.g., temperature) and chemical (e.g., the concentration of a selected substance) variables.
- Biosensors will be incorporated into the human body, continuously monitoring physiological variables (cf. §9.2.2).
- Sensors will be dispersed throughout the environment (e.g., along rivers and in lakes), reporting the purity of water, and so forth. The concept is in some ways a development of what is already taking place in agriculture ("microfarming"; i.e., intervention guided by high-resolution satellite images of fields, indicating local moisture, etc.).

In most, or perhaps all, cases where sensorization is envisaged, at present we simply do not have data of the spatial and temporal intensity that will be obtainable. Its availability will almost certainly qualitatively change our views. It will perhaps come to seem primitive to base an assessment of health on a single analysis of key physiological biomarkers. Vehicle health—

as appraised by analysing sensor readouts—will become the criterion for the airworthiness of aircraft (etc.).

The two main challenges of this development are (i) how to deal with the vast proliferation of data (an example of vastification) and (ii) what about the reliability of the sensors? The first challenge can presumably be dealt with by automated processing of the data, and human intervention will only be alerted in the event of some unusual pattern occurring. This implies vast data processing capacity, which is, however, anyway an envisaged development of nanotechnology. The second challenge may be dealt with in the same way: the system will come to be able to determine from the pattern of its readouts when sensors are malfunctioning. Responses to these two challenges require a phase in which much new knowledge is gathered in order to enable automated operation, although this cannot render the system immune from unexpected, emergent events.

9.5 Further reading

S. Bandyopadhyay, Single spin devices—perpetuating Moore's law. *Nanotechnology Perceptions* 3 (2007) 159–163.

A. Politi and J.L. O'Brien, Quantum computation with photons. *Nanotechnology Perceptions* 4 (2008) 289–294.

Chapter 10

Implications of nanotechnology

This chapter puts nanotechnology into its social context—inasmuch as that is possible considering that we are still very much at the beginning of any Nano Revolution. But nanotechnology has already evoked a wide spectrum of responses from society. These responses can be roughly classified as emanating from groups of “enthusiasts”, “neutrals”, “sceptics” and “opponents”, whose views are summarized in the following sections.

10.1 Enthusiasm

We may consider Richard Feynman as the original enthusiast for nanotechnology. During the decades following Feynman’s lecture, steady development of miniaturization, especially in ultraprecision engineering and semiconductor processing, ensured strong continuity with existing fabrication technologies rooted in the original Industrial Revolution. A theoretical technologist, K. Eric Drexler, meanwhile strongly advocated directly tackling the challenge posed by Richard Feynman, namely that of literally building things atom by atom. In a series of works he elaborated his ideas, which were based on the “assembler”, a nanoscale device capable of building other devices (including copies of itself). These assemblers are envisaged to form the basis of personal nanofactories, desktop machines containing many assemblers and capable of fabricating almost anything. The personal nanofactory is, in effect, a very sophisticated development of the rapid prototypers that are able to fabricate quite complex shapes using sticky powder as the raw material.

Carbon is the favoured atom for constructing almost anything. Intensive theoretical and experimental work is being undertaken by Robert Freitas, Jr and others in order to demonstrate the feasibility of atom-by-atom assembly of carbon atoms. The envisaged feedstock is acetylene. The personal nanofactory would simply need a supply of acetylene, electricity, and instructions (software) for directing the assembly. The acetylene could be fabricated from carbon dioxide in the atmosphere.

There seems to be a general consensus that as soon as one personal nanofactory has been realized, the technology will spread rapidly throughout the world and will usher in a new age of economics of abundance, rather than of scarcity. Hallowed laws of economics, such as those of supply and demand, will become obsolete under such new conditions.

Most of the corollaries of the age of personal nanofacture are beneficial. Since manufacturing of (almost) everything will be localized at point-of-use, the vast transport industry that represents such a terrible, and indeed environmentally unsustainable, burden on the Earth's ecosystem will become obsolete. It is very likely that, concomitantly, electricity generation will also become localized, possibly based on nanofactured photovoltaic solar cells, eliminating the aesthetic blight (and dangers) of electricity pylons. Some of the gravest current problems threatening the survival of humanity, such as pollution, excessive accumulation of carbon dioxide in the atmosphere, and depletion of conventional energy sources, will be eliminated.

These benefits are quite apart from other nano-enabled technologies that should directly benefit humanity. The main ones are (i) continued exponential growth of computer processing power, which at some point should exceed the power of the human brain (Kurzweil's singularity)—intelligence will then be able to spread rapidly throughout the universe; and (ii) advances in artificial enhancements of the human body, in particular leading to a very significant increase in longevity (in a good state of physical and mental health).

Within this group, most prominently at the Center for Responsible Nanotechnology (CRN), one finds attempts being made to inform and educate society about the implications of the Nano Revolution.

10.2 Neutrality

Members of this group see nanotechnology as steady progress along fairly predictable paths. Their vision of technological development is linear rather than exponential. They will point out the advantages of blending nanoparticulate ultraviolet-absorbing pigments in sunscreen ointment, or varnish, thereby keeping them transparent (which may be desirable for cosmetic or aesthetic reasons). They will be enthusiastic about nanostructured coatings on textile fibres, enabling cravats and other items of clothing to repel spills of food and beverages onto them. They will see the advantages of nanostructured surface coatings for glass windows, or motor-cars, enabling them to be cleaned simply by showering them with water (or letting rainfall do the work). They expect that nanomaterials will lead to low-cost light-emitting diodes suitable for use as general light sources, with a far higher ratio of output light to input power compared with existing lamps based on incandescent filaments, thereby leading to significant energy savings.

For many members of this group, nanotechnology seems to predominantly mean nanoparticles. The recent prominence given to adverse, or possibly adverse, effects of nanoparticles penetrating into the human body have led to loud calls for appropriate regulations on the manufacture, use and disposal of nanoparticles, accompanied by appropriate research—for which there always seems to be a need, since despite the huge volume of work already carried out on the biological effects of nanoparticles, it seems that many of the reported experiments were carried out badly, with inadequate controls and other deficiencies.

Some members of this group downplay the very existence of nanotechnology—as they (correctly) point out, nanoparticles have been made and used for centuries, and chemists have been assembling more and more elaborate constructs at the nanoscale for several decades (supramolecular chemistry). What then is new? One also finds the view that one should not speak of "nanotechnology", but instead of "nanotechnologies", there being not one unified field, but a plethora of different ones with applications as diverse as medicine and power generation; presumably the implication is that the different applications have little to learn from one other and are best pursued separately. Within this group, the view that the very term "nanotechnology" will disappear within less than a decade is quite prevalent.

10.3 Opposition and scepticism

The most prominent members of this group are veritable neo-Luddites, who in earlier epochs would have been equally vociferous in opposing railways, electricity and so forth. Their opposition is not specific to nanotechnology and does not therefore require special consideration here.

Another vociferous section within this group seek to raise awareness of specific dangers that are sufficiently alarming to require tight controls, or even an outright ban, on nanotechnology to be put in place, well in advance of any actual evidence for the dangers (this an enactment of the so-called "precautionary principle"). The best-known of these dangers is the spectre of "grey goo". This picturesque term means assemblers that somehow cease to follow their programmed instructions, but simply replicate themselves, and in the process of such exponential growth consume all the Earth's resources. Even if we only consider assemblers using carbon-based feedstock, this is already quite alarming enough, in view of carbon's central place in life on the planet. Interestingly, it was Drexler himself, one of the enthusiasts, who first discussed the possibility of grey goo in print. Note that these opponents share with the enthusiasts the view that molecular manufacturing will become a reality.

Although grey goo has captured the public imagination, few scientists take the possibility seriously. More solidly-founded opposition is based on the possibility that the fabrication capabilities of personal nanofactories will be used not for benign purposes; that is, for the satisfaction of human needs, both necessities and luxuries; but for aggressive, hostile purposes—acquisitive crime and military domination, for example. Possibly the only reasonable reassurance that can be given on this point is to point out that belligerence has always been present in human society, both individually (criminals) and collectively (warmongers), yet has never ultimately dominated, and humanity has survived. The capability to resist misuse of any new technology will, as before, presumably advance *pari passu* with the capability of misuse.

A significant part of this group is made up of sceptics rather than opponents. They are sceptical about the possibility of ever realizing a personal nanofactory based on molecular manufacturing, hence there is nothing to oppose, except the possible waste of public funds used in attempting to develop the technology. Regarding the benefits proclaimed by the enthusiasts, the sceptics would typically respond as follows. Machine intelligence will never surpass human intelligence, because the latter has noncomputational aspects, which cannot be reproduced mechanically. The arguments can be stated quite straightforwardly,[1] and it seems a pity that the two sides do not engage more consequentially in an effort to converge on a consensus. Therefore, despite the acknowledged exponential growth in artificial processing power, human intelligence will presumably always dominate. The real danger to humanity, as Sydney J. Harris has wittily pointed out, is not that computers will come to think like humans, but that humans will come to think like computers.

---

[1] See, for example, Ramsden, J.J. Computational aspects of consciousness. *Psyche: Problems, Perspectives* 1 (2001) 93–100, for a summary.

Theranostics. This rather ugly word signifies the diagnosis of disease (probably via automated processing of data obtained from implanted biosensors) followed by therapy administered by an implanted device (probably including a drug reservoir, or even a miniature drug factory) conjoined to the sensor. A significant section of the population dislikes the idea that medicine will be placed in the hands of automated systems. This is considered as disempowering. When the same function is undertaken by a human medical practitioner, the proposed therapy can be discussed with the patient who takes the ultimate decision whether to accept it.

The responses to human enhancement are more mixed, and the arguments cannot be stated so straightforwardly in unambiguous terms. Debate on the matter is not eased by the existence of an almost religious attitude (typically called "transhumanism") among some of the enthusiasts for human enhancement. Furthermore, a far larger section of the population (probably the majority) is interested in at least a modest enhancement (in the sense of cosmetic embellishment, "energy drinks", drugs to combat ailments, even those as minor as the common cold, and so forth), compared with those interested in artificial intelligence. It seems to be very difficult to clearly divide unacceptable enhancement (transhumanism?) from acceptable (in the view of the majority of society) such as cosmetic and medical treatment. Insofar as accidents can usually be clearly recognized as such, one could in principle exclude medical treatment to combat the effect of an accident from the category of human enhancement. But what if the accident arose through deliberate participation in a dangerous sport, for example? Clarity is further muddied by the prevalence of avoidable unhealthy lifestyles, especially overeating. Until there is some consistency in all of this, it seems to be futile to attempt to find an ethical standard for nanotechnology-based human enhancement. It is disappointing that there has been relatively little attempt to dig down to some fundamental bases. Given that the technological capability of human enhancement is progressing rapidly, we—humanity—are certainly going to be confronted by these issues in ever-increasingly acute form, including, for example, the demographics of human enhancement. Other issues that will inevitably have to be tackled include such things as the meaning of work.

10.4 A sober view of the future

The future is evolving and unpredictable. Typically, concerted efforts to influence it are only made if the risk of an activity threatens survival. We can consider that risk = hazard × probability of occurrence; the evaluation of risk can only be as good as the estimates of hazard and probability of occurrence. Generally speaking, diversity is a guarantor of survival. Different regions of the world should be strongly encouraged to develop their own varieties of nanotechnology. If one should prove catastrophic, the damage is then more likely to be localized. The general trend towards globalization is inimical to regional variety, and this aspect needs at least as much consideration as the technology itself.

10.5 Further reading

F. Allhoff, P. Lin, J. Moor and J. Weckert (eds), *Nanoethics: The Ethical and Societal Implications of Nanotechnology.* Chichester: Wiley (2007). A collection of thought-provoking essays reviewing present nanotechnology and its possible development over the next 30 to 40 years.

J. Altmann, *Military Nanotechnology.* London: Routledge (2006). A review of the potential military market for nanotechnology and current research activity.

M. Anissimov et al., The Center for Responsible Nanotechnology Scenario Project. *Nanotechnology Perceptions* 4 (2008) 51–64. Presents three scenarios of the world in the not too distant future that has acquired the capability of something close to the personal nanofactory.

R.A. Freitas, Jr, Economic impact of the personal nanofactory. In: N. Bostrom et al., *Nanotechnology Implications: More Essays*, pp. 111–126. Basel: Collegium Basilea (2006). Probably the most detailed attempt hitherto published to assess the economic impact of molecular manufacturing.

R. Kurzweil, *The Singularity Is Near.* New York: Viking Press (2005). A very original work setting out the author's view of the development of high technology.

J.J. Ramsden, The music of the nanospheres. *Nanotechnology Perceptions* 1 (2005) 53–64. Conceived as a critique of Wood et al.'s report.

S. Wood, R. Jones and A. Geldart, *The Social and Economic Challenges of Nanotechnology.* Swindon: Economic and Social Sciences Research Council (2003). The senior author of this report is a psychologist. It gives an eclectic view of how nanotechnology's relation with society might develop.

Index

accessibility, 25
alternating polyelectrolyte deposition, 79
artificial intelligence, 121
atomic force microscope, 42
available area function, 81

bacteriophage assembly, 74
bacteriorhodopsin, 103
ballistic deposition, 83
ballistic transport, 31
biological motors, 99
biomolecule, 95
biopolymer folding, 76
biotechnology, 93
bistability, 61
blending, 21
Blumenfeld mechanism, 97
Bohr radius, 32
broken ergodicity, 96

carbon black, 22
carbon nanotubes, 53, 63
catalysts, 23
cellular telephone, 25
charge correlation, 80
chemical reactivity, 28, 32
chemical vapour deposition, 72
chloroplast, 75
Coulomb gap, 31

design, 35
devices, 24
dip-pen lithography, 73
DNA, 95
DNA sequencing, 68
Drexler, 13, 117

electrokinesis, 68
electron microscopy, 39
energy, 113
energy levels, 30
environment, 115
enzyme, 97
equivocation, 35
eutactic environment, 10
evaporation, 72
exclusion zone, 81

Faraday, 51
ferromagnet, 61
Feynman, 13, 117
fullerene, 53
functional surface, 51

giant magnetoresistance, 64
graphene, 53
grey goo, 120
grinding, 47

Hamming, 60
hysteresis, 60

impurities, 33
inflammation, 54
intermolecular interactions, 83

kinesin, 99

Langmuir film, 51
least action, 76
logic gate, 58

magnetic tunnel junction, 65
materials, 20, 28
mechanosynthesis, 21
memory, 61
microtubule, 101
miniaturization, 26
molecular beam epitaxy, 72
molecular electronics, 63
Moore's law, 106
muscle, 93, 99
myosin, 99

nanification, 114
nanoadditives, 22
nanobiotechnology, 68
nanobot, 69
nanocomposites, 22
nanoelectromechanical systems, 67
nanofibre, 51
nanofluidics, 67
nanomedicine, 68
nanoparticle, 47
nanoparticles, 14
nanoplate, 51
nanopore, 51
nanorod, 51
nanoscience, 18
nanotechnology, definitions, 10
nanotoxicity, 56
nanotube, 51
nanowire, 51
near-field scanning optical microscope, 40
noise, 35
nucleation, 48
nucleic acid, 95
nucleotide, 95

optical waveguide lightmode spectroscopy, 45
Ostwald, 14

paint, 21
percolation, 64
physical vapour deposition, 72
polypeptide, 95
polysaccharide, 96
printed electronics, 63
protein, 95

quantum computing, 106
quantum dot, 32, 66, 107
quantum well, 65
qubit, 108

random sequential addition, 81
relay, 58, 67
revolutions, scientific, 15
RNA, 95

scaling up, 26
scanning ion current microscopy, 44
scanning near field microscopy, 40
scanning probe microscopy, 42
scanning tunnelling microscope, 42
self-assembled monolayers, 79
self-assembly, 45, 77
semiconductor processing, 72
sensorization, 115
single electron devices, 62
solubility, 29
spin transistor, 108
spin transport, 65
spintronics, 64, 106
spreading, 88
sputtering, 72
stamping, 79
stigmergy, 75
superatom, 30
superlattice, 30, 64, 65
surface tension, 30, 84
switch, 58
systems, 25

Taniguchi, 13
terminology, 22
theranostics, 122
transducer, 58, 111
transistor, 60
transport, 118
tubulin, 101

ultraprecision engineering, 72

vastification, 115
vehicle health, 115

weak interactions, 83

X-ray diffraction, 41