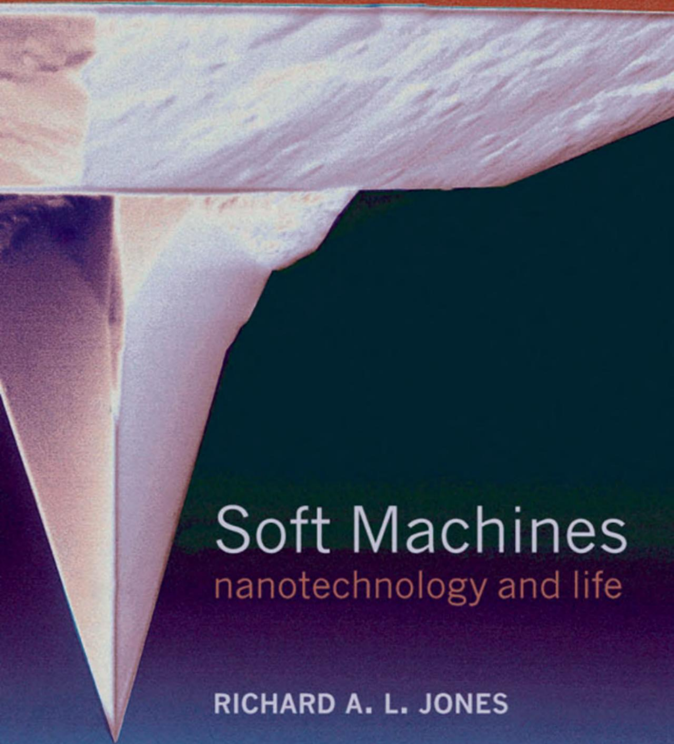

Soft Machines

nanotechnology and life

RICHARD A. L. JONES

Soft Machines

Nanotechnology and Life

Richard A. L. Jones

Department of Physics and Astronomy
University of Sheffield

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford OX2 6DP
Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries

Published in the United States
by Oxford University Press Inc., New York

First published 2004
First published in paperback 2007

British Library Cataloguing in Publication Data
Data available

Library of Congress Cataloging in Publication Data
Data available

Typeset by Newgen Imaging Systems (P) Ltd., Chennai, India
Printed in Great Britain
on acid-free paper by
Ashford Colour Press Ltd, Gosport, Hampshire

ISBN 978–0–19–852855–5 (Hbk.) 978–0–19–922662–7 (Pbk.)

1 3 5 7 9 10 8 6 4 2

Contents

1
Fantastic voyages

A new industrial revolution?

Some people think that nanotechnology will transform the world. Nanotechnology, to these people, is a new technology which is not with us yet, but whose arrival within the next fifty years is absolutely inevitable. Once the technology is mastered, we will learn to make tiny machines that will be able to assemble anything, atom by atom, from any kind of raw material. The consequences, they believe, will be transforming. Material things of any kind will become virtually free, as well as being immeasurably superior in all respects to anything we have available to us now. These tiny machines will be able to repair our bodies from the inside, cell by cell. The threat of disease will be eliminated, and the process of ageing will be only a historical memory. In this world, energy will be clean and abundant and the environment will have been repaired to a pristine state. Space travel will be cheap and easy, and death will be abolished.

Some pessimists see an alternative future—one transformed by nanotechnology, but infinitely for the worse. They predict that we will learn to make these immensely powerful but tiny robots, but that we will not have the wisdom to control them. To the pessimists, nanotechnology will allow us to make new kinds of living, intelligent organisms, who may not wish to continue being our servants. These tiny machines will be able to reproduce, feed, and adapt to their environment, in just the same way as living organisms do. But unlike natural organisms, they will be made from tough, synthetic materials and they will have been carefully designed rather than having emerged from the blind lottery of evolution. Whether unleashed on the world by a malicious act, or developing out of control from the experiments of naïve scientists, these self-replicating nanoscale robots will certainly break out of our custody, and when this happens our doom is assured. The pessimists think that life itself will have no chance in the struggle for supremacy with these nanobots; they will take over the world, consuming its resources and rendering feebler, carbon-based life-forms such as ourselves at best irrelevant, and at worst extinct. In this scenario, we humans will accidentally, and quite possibly with the best of intentions, use the power of science to destroy humanity.

What is now not in dispute is that scientists have an unprecedented ability to observe and control matter on the tiniest scales. Being able to image atoms and molecules is routine, but we can do more than simply observe; we can pick molecules up and move them around. Scientists are also understanding more about the ways in which the properties of matter change when it is structured on these tiny length scales. Technologists are excited by the prospects of exploiting the special properties of nano-structured matter. What these properties promise are materials that are stronger, computers that are faster, and drugs that are more effective than those we have now. Government research funds are flooding into these areas, and start-up companies are attracting venture capital with a vision of nanotechnology that is, perhaps, incremental rather than revolutionary, but which in the eyes of its champions will drive another burst of economic growth in the developed countries. For this kind of enthusiast, usually to be found in government departments and consultancy organisations, nanotechnology is not necessarily going to transform the world; it is just going to make it somewhat more comfortable, and quite a lot richer.

There are some who are simply suspicious of the whole nanotechnology enterprise. They see this as another chapter in a long saga in which different branches of science are hijacked and misused by corporate and state interests. The results will be new products, certainly, but these will be products that no one really needs. The rich will be persuaded by clever marketing to buy expensive cosmetics and ever more sophisticated consumer gadgets, while the poor people of the world continue to live in poverty and ill health. The environment will be further degraded by new nano-materials, even more toxic and persistent than the worst of the chemicals of the previous industrial age.

Some people doubt whether nanotechnology even exists as a single, identifiable technology. We might well wonder what nanotechnology actually is. Is it simply a cynical rebranding of chemistry and materials science, or can we really map out a path from the mundane but potentially lucrative applications of nanoscale science of today to the grand visions of the nanotechnology enthusiasts? Many distinguished scientists are certainly deeply sceptical that the vision of self-replicating nano-robots is achievable even in principle, and they warn that the dream of radical nanotechnology is simply science fiction.

But the visionaries of radical nanotechnology have one unbeatable argument with which to respond to the scepticism of scientists and others. A radical nanotechnology must be possible in principle, because we are here. Biology itself provides a fully-worked-out example of a functioning nanotechnology, with its molecular machines and precise, molecule by molecule, chemical syntheses. What is a bacteria if not a self-replicating, nanoscale robot? Yet the engineering approach that radical nanotechnologists have proposed to make artificial nanoscale robots is very different to the approach taken by life. Where biology is soft, wet, and floppy, the structures that radical nanotechnology envisions are hard and rigid. Are the soft machines that life is built from the unhappy consequence of the contingencies of evolution? When we build a new, synthetic nanotechnology by design, will our creations be able to overcome the frailties of life's designs? Or does life provide us with a model for nanotechnology that we should try and emulate—are life's soft machines simply the most effective way of engineering in the unfamiliar environment of the very small?

This is the central, recurring question of this book. To engage with it, we need to find out in what way the world on the nanoscale is different to the one in which we live our everyday lives, and the extent to which the engineering solutions that evolution has produced in biology are particularly fitted for this very different environment. Then, perhaps, we will be in a position to find our own solutions to the problems of making machines and devices that work at the nanoscale.

The radical vision of nanotechnology

In *Dorian*, Will Self's modern reworking of Oscar Wilde's fable *The picture of Dorian Gray*, the central character is a dissipated hedonist who magically keeps his youthful appearance despite the excesses of his life. At one point, he explores cryonic suspension as a way of staying alive for ever. In a dingy industrial building on the outskirts of Los Angeles, Dorian Gray and his friends look across rows of Dewar flasks, in which the heads and bodies of the dead are kept frozen, waiting for the day when medical science has advanced far enough to cure their ailments. One of Dorian's friends is sceptical, pointing out that the remaining water will swell and burst each cell when it is frozen, and he doubts that technology will ever advance to the point at which the body can be repaired cell by cell.

> *—'Course they will, the Ferret yawned; Dorian says they'll do it with nannywhatsit, little robot thingies—isn't that it, Dorian?*
> *—Nanotechnology, Fergus—you're quite right; they'll have tiny hyperintelligent robots working in concert to repair our damaged bodies.*

This is the way in which the idea of nanotechnology has entered our general culture. This vision has a single source, K. Eric Drexler's 1986 book *Engines of creation*. Drexler imagined a technology in which factories would be shrunk to the size of cells, and equipped with nanoscale machines. These machines would follow a program stored on a molecular tape, and would be able to build anything by positioning atoms in the right pattern. Drexler calls the machines 'assemblers', and the vision of assembler-based technology 'molecular manufacturing'. Of course, if the assemblers can build anything, then they can build copies of themselves—such machines would be self-replicating.

Drexler's vision of assemblers had two origins. On the one hand, molecular biology and biochemistry shows us astounding examples of sophisticated nano-machines. Consider the ribosome, the machine that synthesises protein molecules according to the specification coded in an organism's DNA—this looks very much like Drexler's picture of an assembler. On the other hand, he drew on a famous lecture given in 1959 by the iconic American physicist Richard Feynman, ‘There’s plenty of room at the bottom’, to stress that there were no fundamental reasons why the trends toward miniaturisation that were driving industries like electronics could not be continued right down toward the level of atoms and molecules. Drexler put these two lines of thought together. What would happen if you could create nano-machines that did the same sorts of things as the machines of biochemistry, but which, instead of using the materials that the chance workings of nature had provided biochemistry, used the strongest and most sophisticated materials that science could provide? Surely you would have a nanotechnology that was as far advanced from the humble workings of a bacteria as a jumbo jet is from a sparrow.

With such a powerful nanotechnology, the possibilities would be endless. Instead of factories building cars and aircraft piece by piece, nanotechnology would make manufacturing more like brewing than conventional engineering. You would simply need to program your assemblers, put them in a vat with some simple feedstocks, and wait for the product to emerge. No matter how intricate the product, with nanotechnology it would be barely more expensive to produce than the cost of the raw materials.

If nano-machines can build things from scratch, then they can also repair them. If you regard the results of disease and ageing as simply being a consequence of misarranged patterns of atoms, then the assembler gives you a universal panacea. Drexler envisaged nano-machines as functioning both as drugs of unparalleled power and as surgeons of unsurpassed delicacy. He did not shrink from the ultimate conclusion—that nanotechnology would allow life to be extended indefinitely. For those who cannot wait for science’s slow advance to bring us to this point, there is always the option of putting your body into cold storage and waiting for science to catch up.

What will a future transformed by nanotechnology look like? Many science fiction writers have made an attempt to describe such a future. *The diamond age* by Neal Stevenson is a rich and quite convincing picture, but for all its nuances he presents a world that is more or less a natural extension of modern technological capitalism. Life would be extremely comfortable for the well born and well educated, but considerably less wonderful for those who drew life’s less lucky lottery tickets. Meanwhile, there is another, much more terminal view of what nanotechnology might do to us—the dystopian vision of a world taken over by *grey goo*.

It must have been a slow news day in Britain on 27 April 2003, because the lead headline in *The Mail on Sunday*, a mass-market newspaper of rather conservative character, was about science. Characteristically, the story had a royal angle too; the heir to the British throne, Charles, Prince of Wales, was reportedly very worried about the threat posed by nanotechnology. Scientists were risking a global catastrophe in which an unstoppable plague of maverick self-replicating nano-machines consumed the entire world. As an apocalyptic vision, it certainly beat *The Mail on Sunday’s* usual fare of collapsing house prices and disappearing pensions, but as a story it was rather older. Drexler’s own book, *Engines of* *creation*, warned of a potential dark side to his otherwise utopian dream. What would happen, if having created intelligent, self-replicating nano-robots, these robots decided that they were not happy with their terms of employment? The result would be the destruction and consumption of all existing forms of life by the nanobots—the world will have been taken over by grey goo.[1]

Although what has come to be known as 'the grey goo problem' was discussed by Drexler, what raised the issue to prominence was the publication, in *Wired* magazine, of an article by Bill Joy, the former chief scientist of Sun Microsystems. At the time, the year 2000, *Wired* was the standard-bearer of West Coast technological triumphalism. The article, however, called 'Why the future doesn't need us', painted a grim picture of a future in which advances in robotics, genetic engineering, and nanotechnology rendered humans at best irrelevant, and at worst extinct. The article is very personal, very thoughtful, very wide ranging, and it carries the conviction of an author who knew at first hand both the rapidity of the progress of technology in recent times and the unpredictability of complex systems.

From the *Wired* article, the dangers of nanotechnology slowly permeated into the public consciousness. The article explicitly linked genetic modification (GM) to nanotechnology as twin technologies with similar risks. So, not unnaturally, those activist groups which had cut their teeth opposing GM started to see nanotechnology as the next natural target. After all, the novelist Michael Crichton, who in the novel *Jurassic Park* had so memorably depicted the downside of our ability to manipulate genetic material, chose nanotechnology as the subject of his novel *Prey*.

What has been the scientists' reaction to the growing fears of grey goo? There has been some fear and anger, I think; many scientists watched the controversy about genetic modification with dismay, as in their eyes a hugely valuable, as well as fascinating, technology was hobbled by inaccurate and irresponsible reporting. But mostly the reaction is blank incomprehension. At least genetic modification was actually a viable technology at the time of the controversy, while for a self-replicating nano-machine there is still a very long way to go from the page of the visionary to the laboratory or factory. To a scientist, struggling maybe to get a single molecule to stick where it is wanted on a surface, the idea of a self-replicating nano-robot is so far-fetched as to be laughable.

How have we got to this state, where we have a backlash to a technology that has not yet arrived? In this, maybe scientists are not entirely without blame. Most scientists working in nanotechnology themselves may refrain from making extreme claims about what the science is going to deliver, but (with some notable exceptions) they have not been very quick to lower expectations. One does not have to be very cynical to link this to the very favourable climate for funding that nanoscale science and technology has been enjoying recently.

---

[1] Why grey? Apart from the appealing alliteration, presumably because the nanobots that make up the goo are made of diamond-like carbon.

I do not think that grey goo represents a serious worry, either, but I do think that it is worth thinking through the reasoning underlying the fears. This is because I believe that this reasoning, deeply flawed as it is, betrays a profound underestimation of the power of life itself and the workings of biology, and a complete misunderstanding of the way that nature works on the nanoscale. Until we clear up these misunderstandings we are not going to be able to harness the power that nanotechnology will give us.

In some of the most extremely optimistic visions of nanotechnology, there is a distrust of the flesh and blood of the biological world that is almost Augustinian in its intensity. This underestimation of biology underlies the thinking that produced the grey goo dystopia too. Surely, the argument goes, as soon as human engineers start to engineer nanobots, the feeble biological versions will not stand a chance. After all, when an evolutionary superior species invades the ecological niche of an inferior one, the inferior one is doomed to extinction. In this cartoon view of Darwinism, dumb dinosaurs were outsmarted by quick-thinking mammals, and hapless Neanderthals were inexorably pushed out by our own *Homo sapiens* ancestors. A similar fate is inevitable when our type of life—basically assembled by chance from all sorts of unsuitable materials patently lacking in robustness—meets something that has been properly designed by a college-trained nanotechnologist. What chance will a primitive bug, little more than a water-filled soap bubble, have when it meets a gleaming diamond nanobot with its molecular gears grinding and its nanotube jaws gnashing?

Of course, there are no primitive bugs (at least on Earth), and we ought to know very well that while individual organisms can seem frail, life itself is spectacularly tough. The insights of molecular cell biology show us more and more clearly how optimised nature's machines are for operation at the nanoscale. If the mechanisms nature uses seem odd and counter-intuitive to us, it is because the physical constraints on design are very different on the nanoscale from the constraints in the world we design for. The other insight that we should take from biology is that evolution is an extremely efficient design principle that works as well—possibly better—at the level of molecules as it does with finches and snails. The biological macromolecules that form the basis of the nano-machines in even the simplest-looking cells have themselves evolved to the point at which they are extremely effective at their jobs.

Surely though, a steam engine is better than a horse, strong and lightweight aluminium alloy is a better material to make a wing out of than feather and bone . . . if we can find materials that are so much better than the ones nature has given us to work at the macroscopic level, then surely the same is true at the nanoscopic level? This is Drexler's argument, but I disagree. Nature has evolved to get nanotechnology right. Most of nature exists at the nano-level, the necessary mechanisms and materials were evolved very early on, work extremely well, and look pretty similar in all kinds of different organisms. Big organisms like us consist of mechanisms and materials that have been developed and optimised for the nanoworld, that evolution has had to do the best it can with to make work in the macroworld. We are soft and wet, because soft and wet works perfectly for bacteria. Because we have evolved from bacteria-like organisms we have had to start with the same nano-machinery and try and build something human-sized out of it. No wonder it seems a bit clunky and inadequate on a human scale. But at the nano-level, it is just right.

Nano everywhere

It is difficult to visit a university anywhere in the world nowadays without falling over a building site where a new institute of nanotechnology or nanoscience is due to open. Taking the lead from the USA, where in 1999 President Clinton announced a National Nanotechnology Initiative, governments and science-funding bodies across the world have been pouring hundreds of millions of dollars into the areas of nanoscience and nanotechnology. Scientists have risen to the challenge, and nanotechnology now forms one of the most active areas of scientific endeavour.

So does this mean that Drexler's vision of molecular manufacturing and nanoscale assemblers will soon be with us? No. It is fair to say that most scientists working in the area of nanoscience and technology regard the Drexlerian program as being somewhere along the continuum between the impractical and the completely misguided. Instead, what we see is a great flowering of chemistry, physics, materials science, and electronic engineering, a range of research programs which sometimes have little in common with each other besides the fact that their operations take place on the nanometre scale. Some of this work, that is now called nanoscience or nanotechnology, is actually no different in character to what has been studied in fields like metallurgy, materials science, and colloid science for the last fifty years. Control of the structure of matter on the nanoscale can often bring big benefits in terms of improvements in properties, and this is the basis of many of the improvements which we have seen in the properties of materials in recent years. One could call this branch of nanotechnology *incremental nanotechnology*.

Perhaps more novel are those areas of science where advances in miniaturisation are being scaled down further into the nanoscale. One might call this area *evolutionary nanotechnology*, and the type example is micro-electronics. Driven by the huge size of the worldwide electronics and computing industries, the technologies for making integrated circuits have matured to the point that feature sizes of less than 100 nm are now routine. Related technologies are used to make tiny mechanical devices—micro-electro-mechanical systems—which already find use in applications like acceleration sensors for airbags. At the moment devices that are in production are characterised by length scales of tens or hundreds of microns rather than nanometres, but very much smaller devices are being made in the laboratory. Other types of evolutionary nanotechnology include molecular electronics—the creation of electronic circuits using single molecules as building blocks, as well as concepts that are being developed for packaging molecules and releasing them on a trigger. These are beginning to find applications for delivering drugs efficiently. In evolutionary nanotechnology we are moving away from simply making nanostructured materials, toward making tiny devices and gadgets that actually do something interesting.

So where does this leave nanotechnology in the radical sense that Drexler suggested? A very small proportion of the scientists and technologists who would claim to be nanotechnologists are working directly toward this goal, and indeed many of the most influential of these nanotechnologists are deeply sceptical of the Drexler vision. Does this mean, then, that *radical nanotechnology* will never be developed? My own view is that radical nanotechnology will be developed, but not necessarily along the path proposed by Drexler. I accept the force of the argument that biology gives us a proof in principle that a radical nanotechnology, in which machines of molecular scale manipulate matter and energy with great precision, can exist. But this argument also shows that there may be more than one way of reaching the goal of radical nanotechnology, and that the path proposed by Drexler may not be the best one to follow.

Into the nanoworld

Nanotechnology gets its name from the prefix of a unit of length, the nanometre (abbreviated as nm), and in its broadest definition it refers to any branch of technology that results from our ability to control and manipulate matter on length scales between a nanometre and 100 nanometres or so. One nanometre is one-thousandth of a micrometre or micron. This in turn is one-thousandth of a millimetre. How can we put these rather frighteningly small numbers into context?

Everyone is familiar with the macroworld, the world of our everyday experience. We can directly touch and interact with objects with sizes from around a millimetre up to a metre. This is our human world, and in it we have an intuitive understanding of how things move and behave.

The microworld is less familiar, but not completely foreign. The tiniest mites and insects have sizes of a few hundred microns (or a few tenths of a millimetre); these are visible to those of us with good eyesight as little specks or motes, but we need a magnifying glass or low-power microscope to see very much of the individuality of these objects. These are the smallest things that we have direct experience of—the thickness of a human hair, the thickness of a leaf of paper; these all represent lengths at the upper end of the microworld, around 100 microns.

The microworld is familiar territory to engineers. Precision measuring instruments, like micrometers and vernier callipers, can easily measure dimensions to an accuracy of tens of microns. The experienced workshop technicians in my university's machine shop still, despite metrication, think in terms of one-thousandth of an inch, 25 microns, as a precision to which they can, without trying very hard, build components for scientific instruments.

Biologists, too, work naturally in the microworld; it is the world that can clearly be seen through a light microscope. The largest single cells, an amoeba, or a human egg cell, are just about visible as specks to the naked eye, around 100 microns in size. But most animal and plant cells fall into the range of sizes between 10 microns and 100 microns. The simplest forms of single-celled life—bacteria—are a little bit smaller. These ubiquitous organisms, a few of which are feared as the agents of disease, are usually around a micron in size. Most bacteria are clearly visible in a light microscope, but they are too small to see very much internal structure within them.

The internal structure of cells belongs to the nanoworld. At these sizes, things are too small to see with a light microscope. But new techniques have, in the last fifty years, revealed that within what a hundred years ago was thought of as an unstructured, jelly-like protoplasm, there is a fantastically complex world of tiny structures and machines. Inside each of the cells in our bodies are structures such as mitochondria, tiny bodies made from convoluted foldings of membranes, like crumpled balls of paper. Inside plant cells are chloroplasts, the structures in which light is collected and turned into useful energy. Smaller still we would see ribosomes, the factories in which protein molecules are made according to the specifications of the genetic code that is stored in DNA.

Now we are down to the level of molecules, albeit rather big ones. Biological nanostructures, such as ribosomes, are made up of very big molecules, such as proteins and DNA itself, each of which is made up of hundreds, thousands, or tens of thousands of individual atoms. A typical protein molecule might be somewhere between 3 and 10 nm in size, and will usually look like a compact but knobbly ball. We can make big molecules synthetically too. Long, chain-like molecules consisting of many atoms linked together in a line are called polymers, and they are familiar to us as plastics. Materials like nylon, polythene, and polystyrene are made up of such long molecules. If we could see a single molecule in a piece of polyethylene, then it would look like a fuzzy ball about 10 nm big. Unlike the protein molecule, this is not a compact lump, it would be more like a loosely-folded piece of string.

Small molecules are made up of a few atoms; from the three that make a water molecule, to the tens of atoms that make up a molecule of soap or sugar. An individual atom is a fraction of a nanometre in size, so these small molecules will be around one nanometre big. It is the size of these small molecules that defines the lower end of the nanoworld.

As we have seen, the nanoworld is now the realm of cell biology, and it is our efforts to make structures and devices on this scale that defines nanotechnology. How far have we come toward achieving this goal?

The technology that has come the furthest by shrinking the most has been the electronics industry. The original electronic computers were very much artefacts of the macroworld. Older readers will remember that, in the 1960s and before, the crucial components of a radio were thermionic valves, devices the size of small light bulbs. Before the introduction of transistors, these were at the heart of both amplifiers and logic circuits. So the first computers consisted of rooms full of racks of electronics, the basic unit of which was the centimetre-sized valve.

It was the invention, firstly of the transistor, but most crucially of the integrated circuit, that allowed electronics to move from the macroworld into the microworld. The transistor meant that electronic components could be made entirely in the solid state, doing away with the vacuum-filled glass bulbs of thermionic valves. The integrated circuit allows us to pack many different electronic components onto a single piece of semiconductor, to produce a complete electronic device in one package—the silicon chip.

In the integrated circuit, single components are not individually hewn from the semiconductors they are made from. Instead, lines etched on the surface of the chip define the transistors that are wired up to make the circuits. How small one can make the components is limited by how fine one can draw lines, and it is a reduction in this minimum line size that has driven the colossal increase in available computer power that we are all familiar with. The minimum line size commercially achievable fell below one micron in the mid 1980s, and is currently well below 100 nm.

We live in the macroworld, we have mature technologies that operate in the microworld, and we are beginning our discovery of the nanoworld. Are there any worlds on even smaller scales that remain to be exploited? There is an old rhyme which captures this sense of worlds within worlds and structures on ever smaller scales: ‘Big fleas have little fleas, upon their backs to bite them. And little fleas have littler ones, and so ad infinitum.’ But how small can you go? Is there another world that is even smaller than the nanoworld? Physics tells us that there is such a world, the world of subatomic structure. Can we look forward even further to yet more powerful technologies, which manipulate matter on even finer scales, the worlds of picotechnology and femtotechnology?

We now know that atoms themselves, far from being the indivisible objects imagined by the Greek originators of the concept, have a substantial degree of internal structure. Take, for example, a carbon atom. To a chemist, this is an indivisible ball with a diameter of 0.14 nm. It was the achievement of nuclear physics in the early part of the twentieth century to show that the atom was not an indivisible entity; it has internal structure. Ernest Rutherford, a physicist from New Zealand, was able to show in experiments carried out in Manchester that most of the mass of an atom is concentrated at its centre, in a tiny, dense object called the nucleus. This is small, very small—the nucleus of a carbon atom is about 3 femtometres in diameter (a femtometre being *one-millionth* of a nanometre).

But the nucleus is not where the story stops; it itself is made up of protons and neutrons, which themselves have some finite size. The proton can exist independently; since the nucleus of a hydrogen atom consists of a single hydrogen ion—a hydrogen atom with its accompanying electron stripped off—is a free-living proton. Neutrons, too, can exist independently, but not indefinitely; after their lifetime of about ten minutes a free-living neutron will decay into a proton, and an electron and an antineutrino.

For a while it was thought that protons and neutrons were truly fundamental particles, but it turns out that they, too, are composites. Experiments in the 1960s showed that, in exactly the same way as an atom is mostly empty space, with its mass concentrated in a tiny nucleus, protons and neutrons are composed of much smaller particles. Protons and neutrons are each made up of three particles called quarks.

Is there further internal structure to be found, at still smaller lengths, within the quark? It is currently believed that there is not; quarks, and electrons, are believed to be fundamental particles that are not further divisible. Inasmuch as it makes any sense to talk about the size of these particles at all, they have no finite size—they are true points, without extension in space. In passing, it is worth noting that one might object to this proposition, noting that the suggestion that particles exist that are true points causes all sorts of philosophical and physical problems. This is indeed the case, and it is the business of quantum field theory to sort these problems out.

We can manipulate matter on scales below the atomic; this is the business of nuclear physics. This is now a relatively old technology, and one that has been, to say the least, a mixed blessing for humanity. It is possible to rearrange the protons and neutrons within the nucleus to obtain new elements, even elements that are unknown in nature. But the characteristic of these transformations, as well as the transformations of nuclear fusion and nuclear fission, is that they involve very high energies.

Energy scales

There is a relationship between the length scale at which one is trying to manipulate matter and the relative size of the energy input that is needed to make transformations of matter on that length scale. Roughly speaking, the smaller the length scale on which one operates, the higher the energies that are involved in these transformations. This is why, to look inside the very smallest subatomic particles, particle physicists need to build huge accelerators many miles in diameter. Nuclear physicists probe and manipulate the interior of atomic nuclei; their smaller accelerators can be fitted into tall buildings. Chemists, on the other hand, rearrange the peripheral electrons on atoms, and this they can do simply with a Bunsen burner.

The brute force way of putting energy into something is to heat it up, and the temperature of a material is a measure of the amount of available thermal energy per molecule. The transmutations that chemistry can achieve involve the absorption and release of amounts of energy that correspond to temperatures of hundreds or, at the most, thousands of degrees.

But chemical transformations—even the most highly-energetic ones, such as the detonation of explosives—only tinker with the structure of the outermost edges of the structure of atoms. Only the outermost, most loosely-attached electrons are affected by these changes. To rearrange the nucleus, very much higher temperatures are required. If a gas of the heavy isotope of hydrogen, deuterium, can be heated up to a few hundred million degrees in temperature, then pairs of deuterium nuclei can combine to create helium. In the process, they release a great deal of energy. For this reason, nuclear fusion, if it could be controlled, would be able to provide all of our energy needs. The problem is those enormously high temperatures, which are so much greater than any solid material can sustain.

But the temperatures at which nuclear transformations take place—the temperatures at the centres of the sun and stars—are still tiny compared to the temperatures one needs to transform the deep components of the protons and neutrons—the quarks. At a temperature of around 10^{12} K (around 100 000 times hotter than the temperature at the centre of the sun), the quarks that make up protons come apart from each other to make an undifferentiated soup—the quark–gluon plasma. These conditions are thought to have existed very early in the universe, shortly after the big bang.

These conditions involve unimaginably high levels of energy. What of nanotechnology—what are the natural energy scales that characterise transformations that take place within the molecular machines and structures of our own cells? The energies have to be adapted to the temperatures at which we live, a room temperature of 300 K. These are the energies, not of the violent fusions and sunderings of nuclear physics, but of the rather gentle stickiness of a post-it note. Biology is low-energy physics.

Different physics at the nanoscale

It is an axiom of science that the fundamental laws of physics are constant and unchanging; we believe them to be the same for all objects at all times and in all places. But in the working lives of most physicists, and all engineers and technologists, what one is using to predict and control the behaviour of material things are not the fundamental laws of physics, but a set of approximations and rules of thumb that happen to operate in one particular domain. If we are architects designing a building made of stone, then we use the classical laws of statics. These are a subset of the laws of classical mechanics, which we can think of as an approximation to what we believe to be the ultimate laws governing the behaviour of matter, quantum mechanics, which is appropriate for macroscopic objects. Together with these laws, we use some rules of thumb—that stone is incompressible, but that it will break under tension, for example—that we know are not strictly correct, but which are close enough to being right that they allow us to build buildings that do not collapse. What mixture of approximate laws and rules of thumb we should use will be very different on the nanoscale than the ones we are familiar with from the macroworld.

One key difference is the importance of quantum mechanics. In fact, it is becoming a received truth that the difference between the macroworld and the nanoworld is that, while the macroworld is governed by the classical mechanics of Newton, the nanoworld is governed by the mysterious and counter-intuitive laws of quantum mechanics. Like much received wisdom, there is a kernel of truth in this, surrounded by much that is misleading. The real situation is much more complicated than that. To start with, some very familiar, everyday properties in the macroworld can only be properly understood in terms of quantum mechanics. Why metals conduct electricity, why magnets attract iron, why leaves are green . . . classical mechanics provides no explanations at all for these questions, which can only really be understood in terms of quantum mechanics. On the other hand, quite a lot of what is special about the nanoworld does not depend on quantum mechanics. This is particularly true when there is water around, and the temperature is closer to the comfortable warmth of everyday life than the chilly environs of absolute zero that physicists often like to do experiments in.

The big difference between the macroworld and the nanoworld, if we are not at an ultra-low temperature and in a vacuum, but in a water-filled beaker at room temperature, arises from the fact that water (and everything else) is made of molecules. These molecules are constantly flying around at high speed in random directions, hitting whatever happens to be in their way. This leads to a distinctive feature of the nanoworld—Brownian motion. Everything is continually being shaken up and jiggled around.

The other unfamiliar feature of the nanoworld is its stickiness—when surfaces get close, they almost always like to stick to each other. It is inevitable that when you make things smaller their surfaces get more important, so working around this stickiness problem is a central part of the technology of finely-divided matter. This is well known in those traditional branches of science and technology that deal with finely-divided matter. People who make paint devote a lot of attention to making sure that the tiny paint particles stay in suspension and do not form a sticky goo at the bottom of the tin. But the importance of the problem is maybe not fully appreciated by those who sketch designs for nanoscale machines.

It is these unfamiliar features of the nanoworld that make engineering in this domain so unfamiliar and non-intuitive. Imagine mending your bicycle in the shed one day, as a simple example of the kind of everyday engineering we are familiar with. The parts are rigid, and if we screw them in place they stay where we put them. Mending a nano-bicycle would be very different. The parts would be floppy, and constantly flexing and jiggling about. Whenever different parts touched there would be a high chance that they would stick to each other. Also, the pile of screws that we had left in a pot would have jumped out by themselves and would be zigzagging their way toward the garage door. Nanoscale engineering is going to be very different from human-scale engineering, but if we need lessons then we know where to look. The more we learn about the nanoscale mechanisms that biology uses at the level of the cell, the more we learn how well adapted they are for this unfamiliar world. This book begins to look for some of these biological lessons for nanotechnologists.

2

Looking at the nanoworld

The nanoworld was not invented by Richard Feynman or K. Eric Drexler. Long before the idea of nanotechnology was devised and the word coined, technologies and processes that humans depended on relied on the manipulation of matter at the nanoscale, even if the way these technologies produced their effects were not fully understood at the time. Take the invention of Indian ink by the ancient Egyptians or the discovery of how to make soap; both of these long-established materials undoubtedly rely on nanotechnology in the broad sense, and if these inventions were being made today then their inventors would no doubt be stressing their nanotechnological credentials as they attempted to raise capital for their start-up companies. What makes it possible to think of the nanoworld as a new realm of matter that we can explore and control is the availability of instruments that allow us to see into that realm. It only became possible to appreciate the vast extent of the universe beyond the Earth after the telescope had been discovered. So it is, that the invention of new microscopes, capable of picking out the details of the world on scales smaller than a micron, has enabled us to appreciate the scope of the nanoworld. Seeing is believing.

If you want to look at something small, then you need a microscope. When we talk of microscopes and telescopes, this suggests ways of enhancing our sense of vision. This is perhaps natural given how most of us depend on our sense of sight in our interaction with the ordinary world, in the realm of our own senses. Ordinary light microscopes and telescopes are essentially enhancements of our own eyes, whose technology is in direct descent from the medieval invention of spectacles.

The development of the optical microscope in the seventeenth century opened up a new world whose existence had not been previously suspected—a world filled with tiny animals and plants of strange designs, and ultimately of microbes. Many of these microbes were revealed to be beneficial to humanity, like the yeasts that convert grape juice to wine and the bacteria that convert milk to yoghurt. Others are harmful or even fatal, like the pathogens that cause smallpox and the plague. But the majority simply make their living in their own world without much impact on humans. This world that the light microscope reveals we can call the microworld—the world defined by dimensions between a micron or so—the size of the smallest object that a light microscope can discern—and the fraction of a millimetre that can be made out by the unaided naked eye.

That there is another world even smaller than the microworld—the nanoworld—was clear even before the tools required to image it became available. The lower size limit on the nanoworld is set by the size of molecules, and long before molecules could be directly imaged there were indirect ways of estimating their size. At the end of the nineteenth century and the beginning of the twentieth it was becoming apparent that a whole class of matter was made up of objects that were bigger than molecules but still well below a micron in size. Glues and gums, milk and blood, it was clear that these were not just simple solutions like a solution of sugar or salt. The evidence was unequivocal that colloids, as these materials were called, consisted of a dispersion in water of objects that were characterised by nanoscale dimensions. It was not clear at the time whether these nanoscale components were aggregates of smaller molecules or very large individual molecules—macromolecules.

But even the best light microscopes do not let us see into the nanoworld proper. Fundamental physical limits that arise from the wave nature of light mean that it will always be impossible to discern objects with dimensions much less than a micron with a light microscope of conventional design. To extend our vision into the nanoworld, it is necessary to use different kinds of radiation.

The size and shape of molecules were being determined by X-ray diffraction in the first half of the twentieth century. In particular, the existence of very large molecules—macromolecules—was confirmed. X-ray diffraction is a method that could determine the size and structure of molecules directly; after the development of the technique at the beginning of the twentieth century by Max von Laue and the Braggs (Laurence and William, the most famous father and son team in science), the technique was applied to bigger and bigger molecules. By 1950, the significance of macromolecules in biology was clear, and the importance of determining the structure of biological macromolecules was obvious. At this time diffraction patterns had already been obtained for proteins, and most famously, in 1953, the structure of the macromolecule DNA was solved by Francis Crick, James Watson, Maurice Wilkins, and Rosalind Franklin. X-ray diffraction unambiguously tells us not only the overall size of molecules, but also their internal structure . . . but for many scientists the complicated mathematical relationships that relate the diffraction patterns you see on the photographic plate and the structure of the molecules themselves makes the technique less satisfying than being able to visualise something directly with microscopy. More seriously, the technique does rely on being able to make a crystal—a regular three-dimensional repeating array—from the molecule.

So, despite having fairly convincing evidence of the existence of the nanoworld, and something of its richness and complexity, without a better microscope than the optical instruments available in the first half of the twentieth century there was a lack of immediacy about people's knowledge of the nanoworld. Seeing is believing, and even for the most rational scientists there is something much less satisfying about knowledge that is inferred than knowledge obtained from direct observation. So why can we not make microscopes that operate at higher magnifications, to see the nanoworld directly?

Such microscopes can be made, but not using light—you need to use electrons. The electron microscope was invented in 1931, and soon it was illuminating the richness of the nanoworld. Even within the cells of plants and animals there was another whole world of complexity; wheels within wheels in the shape of a whole cast of tiny organelles. Even in such a humble object as a plastic bag there are hierarchies of structures, built on sheaves of macromolecules. And yet, despite these advances, electron microscopy is still a long way away from the immediacy of the light microscope. The instruments are complicated to operate, temperamental even; they are expensive, and the images are not always easy to interpret without years of experience. Delicate samples can be damaged or even obliterated by the huge doses of radiation that the samples are subjected to. Perhaps most importantly, elaborate and complicated procedures need to be gone through to prepare the sample for examination in the microscope. Soft samples need to be mounted, frozen or dried, cut into tiny slices, coated with metals, and then put into the hostile environment of an ultra-high vacuum. It is a long way away from the convenience of being able to put a drop on a slide and peer down a light microscope at it. The electron microscope does at least have the end product that is a photograph, a simple magnified image, rather than the abstract pattern of spots on a photographic plate that X-ray diffraction produces. But the process of preparing the sample and making the image is less like taking a quick look at an object, and more like commissioning an artist to make a painting. There is no question of capturing any motion; everything in the sample has to be commanded to stay still, and you have to live with the knowledge that the likeness of your image is mediated by the quirks of the artist.

It was the invention of an entirely new type of microscope that ultimately led to what we might call the democratisation of the nanoworld—the development of a microscope capable of visualising individual molecules, but without the need for difficult and potentially destructive sample preparation that electron microscopy imposes, and available at a low enough cost that most nanoscience laboratories could afford one, just as they would have a robust light microscope. These new instruments were the scanning tunnelling microscope and the scanning force microscope (or atomic force microscope), both invented within a few years of each other in IBM's Zurich laboratory.

Scanning probe microscopes rely on an entirely different principle to both light microscopes and electron microscopes, or indeed our own eyes. Rather than detecting waves that have been scattered from the object we are looking at, one feels the surface of that object with a physical probe. This probe is moved across the surface with high precision. As it tracks the contours of the surface, it is moved up or down in a way that is controlled by some interaction between the tip of the probe and the surface. This interaction could be the flow of electrical current, in the case of a scanning tunnelling microscope, or simply the force between the tip and the surface in the case of an atomic force microscope. The height of the tip above the sample surface is recorded as it is scanned across the surface, allowing a three-dimensional picture of that surface to be built up on the controlling computer. In going from a light or electron microscope to a scanning probe microscope we have moved away from looking to touching.

Light microscopy

By medieval times, ageing monks were able to use magnifying glasses to help them read the small writing of their manuscripts, and it was almost certainly a spectacle maker (most probably the Dutchmen Hans and Zaccharias Janssen in the late sixteenth century) who realised that two or more lenses could be combined to make a microscope of considerable magnifying power. The way a microscope works is that a lens produces an inverted, real image, which is then magnified by an eyepiece—which works in essentially the same way as a magnifying glass. The power of a microscope lies in the fact that it has more than one stage of magnification. A good high-powered objective might have a magnifying power of 100, while the eyepiece might have a magnifying power of 10. In producing the final image, these magnifying powers are multiplied together; in this case an object one micron big appears to the observer to have the easily discernible size of a millimetre. Can we push the magnifications up yet further so that we can see nanometre-scale objects? There is no reason why we should not introduce another stage of magnification still.

There is a lot of subtlety in the classical optics that describes how a microscope works. But at its simplest, we can think of a microscope as a device that takes the light emitted from a single point on the sample, and maps it onto another single point on a detector, whether that is the retina of a human eye, a piece of photographic film, or (most commonly, nowadays) a charge-coupled device—an electronic detector of the sort that you have in a digital camera or video camera. In the ray optics that you use to analyse a microscope, you draw lines representing the rays of light as they travel from the sample to the detector, being bent by lenses and blocked by apertures as they make their journey. If the microscope has been designed correctly, then no matter which path the light takes from the sample to the detector, light that leaves one point on the sample arrives at one point on the detector; light that leaves two adjacent points on the sample, separated by some small distance, will arrive at adjacent points on the detector, separated by a larger distance which is simply the distance separating the points on the sample multiplied by the magnification of the microscope. In a less-well-designed microscope, or one made with components such as lenses of lower quality, light leaving a single point on the sample arrives not on one single point on the detector, but on a little area. If the areas arising from light coming from two closely-spaced features on the sample overlap, then we will not be able to distinguish those features. Now our image is not perfect, but blurred. Obviously, the aim of microscope design must be to reduce this blurring effect to the minimum possible.

The limit to this process of design is that the basic picture that it is based on, of light travelling in rays from one point to another, is incorrect, or at least incomplete. We have known for a couple of centuries that light is a wave, not a ray, and because of this even a beam which has been most carefully prepared to be parallel will slowly spread out. Lasers, for example, can emit a beam of light that can be made very parallel, so the spot the beam makes if it hits a screen stays very small, even if the screen is a long way from the laser. But spread out the beam certainly does, no matter how carefully we make the beam parallel (collimate it, to use the technical term). As we get further and further away from the laser, the edges of the spot get less and less sharp, and eventually the beam broadens and becomes fuzzier. The cause of this effect is diffraction, and its effect is that, even in the best-designed microscopes, there is a fundamental lower limit on the size of features that can be distinguished, or resolved. This lower limit is related to the wavelength of light; as this varies from about 400 nm (for violet light) to 700 nm (for red light) this means that it is fundamentally impossible to use a light microscope to image the nanoworld.

Before considering how more sophisticated microscopes can look down further into the nanoworld, it is worth remarking that a simple light microscope is still a remarkably powerful and useful tool for the nanoscientist. There are some important features of the nanoworld that are discernible at the length scales of a light microscope—I am thinking here particularly of the phenomenon of Brownian motion, which we will discuss in detail in Chapter 4, and which plays such an important role in understanding why the nanoworld is so different to the macroscopic world. This is easily observable at home with a hobbyist's microscope, or with the sort of microscope available in schools. This illustrates one of the advantages of light microscopes over other, apparently more powerful techniques with higher available magnifications and resolutions. A light microscope can observe a system as it is, a living system for example, without any special treatment, and you can see how things move around as well as seeing their frozen structure.

Perhaps the most important advantage of using light is that there is a lot of it about, and it is easy to detect. It is not difficult to obtain light from very high intensity sources—discharge lamps, lasers—and it is possible to detect light at very low levels. Even our own naked eyes are very efficient detectors of very dim light (and mammals like cats, adapted for life at night, are much better at this than we are), and modern semiconductor-based detectors can detect a single unit of light—a single photon. Taken together, the ease of generating very intense beams of light, and the efficiency with which we can detect very dim signals, means that if we are trying to visualise something very small, then we should not be limited by the fact that this small object only reflects a very small amount of the light that is incident on it. As we shall see next, this means that it is possible to use a light microscope to see a single molecule, even though the wavelength-limited resolution of a light microscope means that we do not see it at its true size.

Seeing a single (big) molecule

I do not know who the first person to see a single molecule using a light microscope was, but I remember very clearly the single-molecule experiment that first made an impression on me. The subject it addressed was the question of how a polymer molecule could move around in a melt of other polymer molecules. Think of a pan full of spaghetti, full of long, flexible, slippery strands, and ask how can one of these strands, tangled up as it is with all the other strands, move through the pan? The problem is an important one for understanding how molten polymers—like the molten polyethylene that is extruded or moulded to form plastic bags or washing-up bowls—can flow at all. The theoretical physicists Pierre Gilles de Gennes, from Paris, and Sam Edwards, from Cambridge, had the insight that the only way a single polymer molecule could move through such a tangle was if it wiggled head first, its body following the path its head made, like a snake moving through long grass. De Gennes coined the term reptation to describe this kind of motion, and, on the basis of this rather pictorial insight, de Gennes, Sam Edwards, and the Japanese theorist Masao Doi developed a complete and quantitative theory of polymer motion. But how could the theory be proved? There was a lot of indirect evidence—the theory seemed to correctly explain the way the flow characteristics of polymers depended on how long the chains were, for example—but the direct proof, that would convince the sceptics, was still missing. Then, on the cover of *Science* magazine, was a series of images that showed a long molecule starting out in the shape of a letter R, and then moving in exactly the snake-like way that de Gennes and Edwards had imagined, its tail following the sinuous path that its head had made through the (invisible) surrounding forest of other molecules.

The images came from a paper by Stephen Chu, a Stanford physicist who, soon after this paper, won the Nobel prize for physics. But what was perhaps most remarkable about them was that they had been made, not with some fabulously expensive and sophisticated piece of new scientific equipment, but with a plain old light microscope. What were the secrets that allowed this remarkable feat?

The first point is that Chu made life easy for himself by finding a big molecule to look at. He chose DNA, the polymer that carries the genetic code. We will say a lot more about this molecule later, but for now what is important is that it is very long, and very stiff. DNA molecules can easily be tens or even hundreds of microns long—an extraordinarily large figure for a single molecule—and because the molecule is so stiff it does not coil up on itself. Instead, it tends to form grand, sinuous curves. This means that, even though Chu's microscope was still bound by the diffraction limit, the molecule was long enough to be resolved. Of course, although the molecule is very long, it is still only a few nanometres wide, and the light microscope cannot resolve this dimension. So, instead of seeing a very fine, long line, what the microscope seems to show is a rather fuzzy sausage shape.

The second thing that Chu did to make the experiment possible was that, rather than using light scattered or reflected from the molecule to make the image, he used its fluorescence. Fluorescence is an optical process in which some molecules, if they are illuminated by light of one colour, re-emit the light with a different colour. If you have ever worn a white shirt at a party with ultraviolet lighting then you have seen the effects of fluorescence—the ultraviolet light is invisible to the human eye, but fluorescent dyes that are incorporated in washing powder ('for brighter than bright whites') make your shirt glow with visible light. The advantage of fluorescence for the microscopist is that one can illuminate the sample with a very bright light source—for example, a laser—and then use a coloured filter to block any of this light from being scattered into the eyepiece. Since the light that the fluorescent molecule is emitting is a different colour to the light used to illuminate the sample, it passes through the filter. In this way, because the background is dark, even the very weak signal from a single molecule can still be picked out.

DNA by itself does not naturally fluoresce. Before one can use this technique one needs to attach fluorescent dye molecules to the DNA—to label it, in effect (see figure 2.1). This could potentially cause difficulties—one needs to find a dye which will stick to the DNA, and one can ask valid questions about whether the behaviour of the molecule might be affected by having these

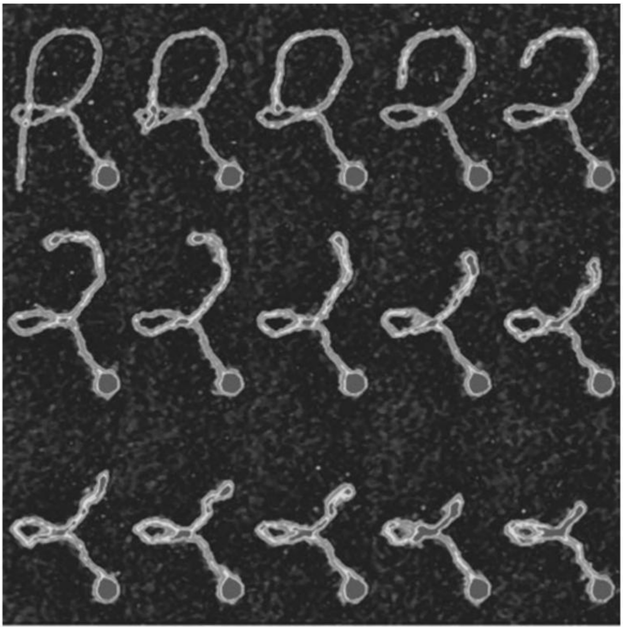

Fig. 2.1 Pulling a single DNA molecule. One DNA molecule has been fluorescently labelled, and attached to a micron-sized polystyrene bead. The bead is moved using 'laser tweezers' through a solution of non-labelled DNA. The images are obtained from optical fluorescence microscopy. See T. T. Perkins, D. E. Smith, and S. Chu. *Science* **264** (1994) 819; this figure by courtesy of Steve Chu.

dye molecules stuck to it. But the need to label can be used to advantage, too, and this is what Chu's experiment relied on. Imagine once again our pan of spaghetti, and suppose that we are looking at the pan with vision that is too blurred to see the width of each piece of spaghetti sharply. If all the pieces of spaghetti are glowing, then all we will see is an undifferentiated block of colour. But if we have taken out one single strand of spaghetti and labelled it with luminous paint before returning it to the pan, then if we look at the pan we will clearly see our labelled strand, glowing against the dark background of the unlabelled strands. This is what Chu did in his experiment; a very low concentration of DNA was fluorescently labelled, leaving most of the DNA in his sample unlabelled, and thus invisible.

If you want to look at single molecules with light microscopy, they need to be highly dilute, so the blurred signals from each molecule do not overlap with each other. So an experiment like Chu's, whose point is to study the way the molecules interact with each other, is only possible by using selective labelling.

Fluorescence microscopy is an important tool in biology, because it allows one to begin to see something of the complicated traffic of molecules within the cell, as well as the static structure of the cell itself. The power of using fluorescent labelling has become even more obvious thanks to the introduction of a new kind of light microscope—the scanning laser confocal microscope. Confocal microscopes suffer from the same fundamental limitations on resolution that an ordinary light microscope has, but have some powerful advantages.

Anyone who has played with a microscope knows the importance of focusing it. If you are looking at a surface, you rotate the focus knob, which physically moves the microscope away from the surface until the blurred features snap into clarity. If you are looking, not at a surface, but at a transparent sample, what you see is just one section of the sample, the thin slice that happens to be at the right distance to be in focus. But the light that comes from the out-of-focus parts of the sample still enters the eyepiece; if you were doing fluorescence microscopy this would give a diffuse glow that would reduce the contrast with which the in-focus parts of the image would stand out from the background.

What a confocal microscope does is remove the glow from the out-of-focus parts of the sample. To achieve this, it has to work in a completely different way to a normal microscope. Instead of illuminating a wide area of the sample and forming an image of this whole area, a laser beam is finely focused down to an intense point, and as this little point of light is scanned across the sample the detector signal is recorded to build up an image. In this way, a very clear image of a slice of sample is built up; if the sample is then successively moved short distances away from the microscope, with each time another slice being imaged, then a complete three-dimensional picture of the sample can be built up, which can be visualised and manipulated using a computer.

The major difficulty in fluorescence microscopy is finding the appropriate fluorescent dye and sticking it to the molecule one is interested in. The most common dyes are smallish organic molecules of the general type that we will discuss in more detail when we come to molecular electronics. At the moment there is some excitement about using tiny inorganic semiconductor particles a few nanometres in size—quantum dots. The advantage of these is that quantum effects mean that by controlling the size of the particle one controls the colour of the fluorescence. Both dyes and quantum dots have the disadvantage from the point of view of biological studies that, if you want to study the processes inside a living cell then you have to get them inside the cell without killing it. Another approach uses the fact that some organisms are naturally fluorescent. The favourite is a deep-sea jellyfish which has a creepy green glow. This comes from a fluorescent protein (imaginatively named green fluorescent protein, or GFP), and genetic engineering has been used to make all kinds of organisms produce variants of GFP attached to their own proteins, essentially making a living label.

Other types of waves

Powerful as light microscopes are, the diffraction limit means that they are fundamentally unable to look properly into the nanoworld. If you are going to use a microscope that depends on waves to form its image, then a proper nanomicroscope is going to need waves whose wavelength is smaller than a nanometre. What kind of waves could we use?

Light, like radio waves, is an electromagnetic wave, which is based on oscillations in the electric and magnetic fields. These oscillations can take place at any wavelength, different wavelengths occupying different positions on the electromagnetic spectrum. Waves with wavelengths in the range of metres are the radio waves, which surround us all the time with a low-energy background of bad pop music. As we have seen, visible light has wavelengths in the range 400–700 nm, according to its colour. Is there a type of electromagnetic radiation, with a wavelength of a nanometre or less, which would allow us to make a microscope whose diffraction limit was low enough to have a clear look at the nanoworld?

There is such a radiation—these are X-rays. The wavelength of X-rays is exactly right, but, frustratingly, it is not easily possible to make a microscope that uses them for the very practical reason that it is very difficult to make a lens that will focus them. As we all know, X-rays easily penetrate most materials, and it is not possible to find a material that will bend X-rays without absorbing them. The X-ray pictures that we are familiar with, if we are unlucky enough to have broken a bone, are not really images in the true sense; they are simply unmagnified shadows. We will see below that X-rays have a vital role in allowing us to make sense of the nanoworld, but we cannot easily use them to make direct images.

So there is no type of electromagnetic radiation that we can use to make a nanoscope. But electromagnetic waves are not the only waves in the world;