Microfluidics, or the art of handling very small liquid samples

The integration of the successive operations of a complex biological analysis protocol into a lab-on-a-chip, involving the handling of very small samples and reagents intended to circulate through nanometre-scale channels, is part and parcel of microfluidics, which offers solutions for moving, splitting, mixing and separating liquids. However, as the dimensions are scaled down, the very properties of liquids change. Basically, edge effects become preponderant, and turbulence, which is needed to mix the liquids, disappears. Novel solutions for handling these kinds of volumes need to be developed.

Generally speaking, nanobiotechnologies are first considered as nanotechnologies, meaning they either present properties at **nanometre**-scale or employ elements of nanometric size or diametre (**particles, wires, tubes**), or they make it possible to observe and handle objects at nanoscale dimensions (for example, individual **molecules**) or study their properties. Following on, they are "bio" in the sense that they either incorporate organic elements from the living world (*e.g.* biological molecules such as **DNA** fragments or **proteins**), or their fields of application are the life sciences (biology or medicine). This field therefore covers part of the instrumentation used in biology, drug delivery systems, and materials development techniques.

Given that virtually all biochemical reactions and biological mechanisms take place in aqueous solution, any technology, nano or otherwise, related to biology necessarily includes a significant amount of "fluidics", *i.e.* liquids handling, actions carried out on these liquids such as mixing, splitting, temperature control, moving, etc. For reasons that will become clear, within nanobiotechnologies, researchers often work with tiny volumes (below a microlitre, *i.e.* below 1 mm^3) or in structures with ultra-small dimensions, such as capillaries of the order of one hundred **microns** or less (around the diametre of a hair). At this point, we are dealing with "microfluidics".

Microfluidics, a key component of integration

This microfluidics aspect is particularly important in the field of **labs-on-chips** or micro total analysis systems (MicroTAS), which aims to develop biological analysis systems presenting two special features in relation to more classical instruments (see *Towards analysis microsystems interfaced with the living world*). The first feature is related to advanced miniaturization and automation. The second is the integration of several step-by-step operations of a complex analysis protocol into a single device. This integration is particularly dependent on controlling movement, distribution, and possibly storage of reagents and sample, as well as the physical conditions of chemical reactions, especially temperature, within the device. This makes microfluidics *one*, if not *the* critical feature in this integration.

These labs-on-chips find applications in a number of fields[1]: *in vitro diagnosis* for "point-of-care" tests that are used in places, such as the doctor's office, the emer-
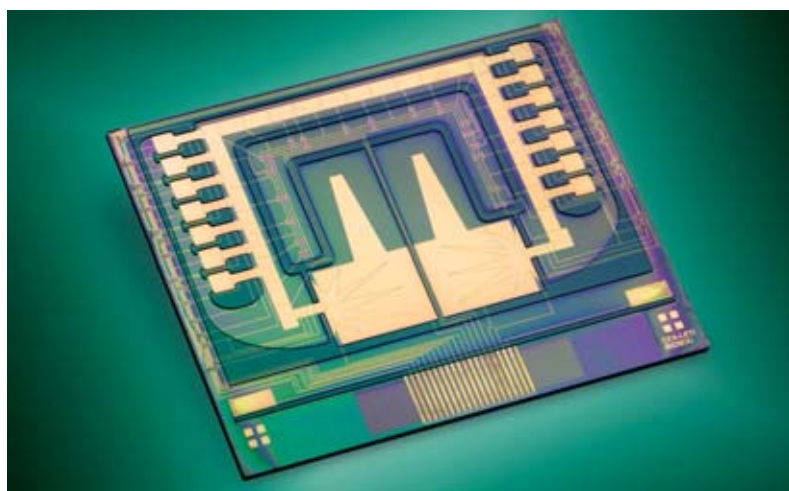
Microfluidics component for carrying out complex biological protocols.

gency room, or at the patient's bedside; *agrifoods control* for detecting chemical or biological contaminants, or the presence of genetically-modified organisms; *environmental control*, for testing water or air quality, such as testing for legionellosis in air-conditioning systems; or *defence and population security*. Since the terrorist attacks of 9-11, the USA has been funding the lion's share of research in the latter applications.

There are several reasons behind the scaling down of the volumes being handled. First, there is increasing need to analyze samples of very small physical size, at the scale of a single cell or cell **compartments**, or including tiny quantities of analytes, such as DNA or **RNA** analysis on the handful of cells in a biopsy. Furthermore, the scaling-down of sample size more often than not favours faster chemical reactions and heat transfer. This is partly due to the fact that when device and sample size are reduced, surface-to-volume ratios increase, thereby speeding up all the phenomena that occur at a surface, such as nucleic acid **hybridization** in **DNA microarrays** (see *Towards analysis microsystems interfaced with the living world*). The speed of a chemical reaction in a volume is proportional to the concentration of the reagents, thus inversely proportional to the volume. In particular, when the operator has few available analytes, it is necessary to work with the smal-

(1) Only applications directly related to nanobiotechnologies will be dealt with here. However, there are implications for other domains, such as fuel cells, chemistry, integrated optics for producing optical switches or variable-focus liquid microlenses for cameras integrated into mobile phones.

Step where a lab-on-a-chip reservoir, designed around the concept of droplet microfluidics, is filled with a fluorescent solution.

P. Stroppa/CEA

lest volumes possible. Lastly, the development of micro-technologies in the broader sense makes it possible to envisage high-volume production of these kinds of devices at relatively low cost.

The specific features of microfluidics

Most microfluidics systems are comprised of capillaries and reaction chambers. To manufacture these, the operator generally starts with a bare sheet of solid material onto which he forms channels and bowls, later adding a cover. To produce these features, the operator uses **lithography** and etching of a solid material such as glass, **silicon** or quartz, or else he moulds or dimples **polymer** materials. Typical channel dimensions are of the order of one hundred microns, but sometimes reach just tens of nanometres. The section is more often than not square or semi-circle in shape. The nature of a liquid's flow regime is determined by its Reynolds number, which depends on the speed of flow, the diametre of the capillary, and the viscosity[2] of the fluid. At these dimensions and in the usual flow regimes (below a centimetre per second), with water or a more viscous aqueous solution, the Reynolds number dictates that there is never **turbulence**. Turbulence cannot therefore be used for mixtures, which can only be formed by diffusion, which is a whole lot slower.

Another characteristic of microfluidics that was hinted at above is the importance of surface phenomena. A particular aspect of this concerns the importance of **capillary forces** that can be exploited to limit or initiate movement of the fluid.

Microfluidics takes advantage of a number of physical principles to move and control liquid flow. Depending on the application needed, the most promising solutions, over and above solutions giving better results in terms of flow rate, speed and accuracy of the various operations, will be those that are most easily integrated (ideally, those where all the elements can be embedded into one lab-on-a-chip), those that are most compatible with the characteristics of the fluid to be handled, and last but not least, those that work out the cheapest.

(2) Viscosity: state of a liquid whose flow properties are slowed by friction between its molecules.

(3) In capillaries with circular cross-section, the pressure difference required to ensure a given flow is inversely proportional to the power 4 of the radius of the capillary.

As we do not have enough room here to cover all the possible solutions that have been researched, interested readers can consult a list of recent publications that offer an excellent grounding on the subject (see *For further information*). We will now take a look at the main approaches.

Moving a fluid by applying a pressure gradient

The most straightforward way of setting a liquid in a capillary into motion is to create a difference in pressure between the two ends of the capillary. This generates a flow that, due to the friction on the capillary walls created by viscosity, gives a parabolic velocity profile. In fact, the liquid moves a lot faster in the middle of the capillary than along the walls. The application of a pressure gradient is widely used as a principle in microfluidics. Despite the obvious advantage of simplicity, this solution nevertheless carries drawbacks. First of all, it is not always easy or even possible to have available a good-quality, generally external source of pressure. In addition, the pressures involved increase very quickly as the size of the conduits decreases[3]. Lastly, certain situations require all the liquid to transit at the same speed in the system, which, here, is not the case. One of the most promising approaches for integrating microfluidics operations at large scale is being developed by Stephen R. Quake's team at **Caltech**. This approach combines, in an elastomer[4] material, a microfluidic network with a network of pressurized air tubes that can be used to create pumps and valves (Figures 1 and 2).

Electro-osmosis to circulate a fluid

In most cases, the moving fluid is a solution of water loaded with various salts. Contact between this solution and certain insulator materials (particularly glass) making up the walls of the capillary gives rise to static charges at the surface, which are locally balanced out in the solution by a layer of opposite charge, which can reach a hundred nanometres in thickness. When a difference in electric potential is applied to the ends of the capillary, this peripheral layer is made to move under an **electrophoretic** effect, which trig-

fluid out

air in/out

fluid in   vertical gap: 30 μm

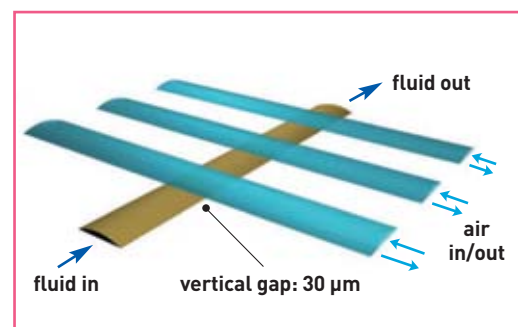Reprinted with permission from *Science* 288: 113-116 (2000), © 2005 AAAS

Figure 1.
Principle of a microfluidic elastomer-based valve.
The capillary network (in blue) can be subjected to air pressurized to locally deform the material and give controlled blocking of fluid circulation in the brown channel (from M.A. UNGER, H.-P. CHOU, T. THORSEN, A. SCHERER and S.R. QUAKE, "Monolithic microfabricated valves and pumps by multilayer soft lithography", *Science*, 288, pp. 113-116, 2000).

gers the movement of all the fluid present inside the capillary. In contrast with the pressure difference, this effect becomes stronger, and therefore more efficient, at smaller capillary diametres. The velocity profile which is thus generated is virtually uniform, in stark contrast to the previous case. Another advantage is that it is easier to integrate into a miniature component. However, this solution does have drawbacks. The voltage sources required are generally relatively high (around a thousand volts), which means **electrolysis** problems at the **electrodes** will have to be solved. Furthermore, for each **ion** species in solution, this effect brings with it an electrophoretic effect that can present complications.

Along the same lines, it is possible to use a magnetohydrodynamic effect. Combining an electric field producing ion movement with an orthogonal magnetic field, which creates a Lorentz force on these ions, makes it possible to induce a fluid movement in the third dimension.

Manipulating fluid as microdroplets

In certain circumstances, particularly when the sample has a volume of the order of a microlitre or less, it can be kept in the form of a microdroplet throughout the protocol sequence. The microdroplet itself can be considered as a chemical reactor. When operating at these scales, capillary forces are high, and when maintained on a **hydrophobic** support, the microdroplet remains in the form of a quasi-sphere attached to a flat surface, or held between two surfaces. The droplet can therefore be moved by disturbance, particularly asymmetric disturbance, of the capillary forces. Once again, several principles apply. The microdroplet can be moved under the action of an **electrostatic** field, a principle called *electrowetting*, and by employing any kind of mechanical stimulation on an asymmetrically-structured support. Finally, the microdroplet can also be subjected to a surface acoustic wave.

The electrowetting approach is particularly promising. Solutions have been developed at the Electronics and Information Technology Laboratory (CEA-Leti: Laboratoire d'électronique et de technologie de l'information) in which the microdroplets move along a network of electrodes that are buried under a layer of hydrophobic **dielectric** material (the EWOD technique, for ElectroWetting On Dielectric, see *Towards analysis microsystems interfaced with the living world*). To actuate the microdroplets, and electric voltage is established between these electrodes and either a conductor wire parallel to the plane of the electrodes or a conductor layer deposited on the surface of the "cover" of the system (Figures 3 and 4). Using this principle, it becomes possible to design electrode configurations to perform basic microfluidics operations, as well as the majority of biological analysis protocols (Figure 5). This provides researchers with a sort of "tool box", or a generic "operator library" that, inspired by how **integrated circuits** are designed in microelectronics, can eventually be used in the design of a lab-on-a-chip based on standard, well-known and well-characterized operations.
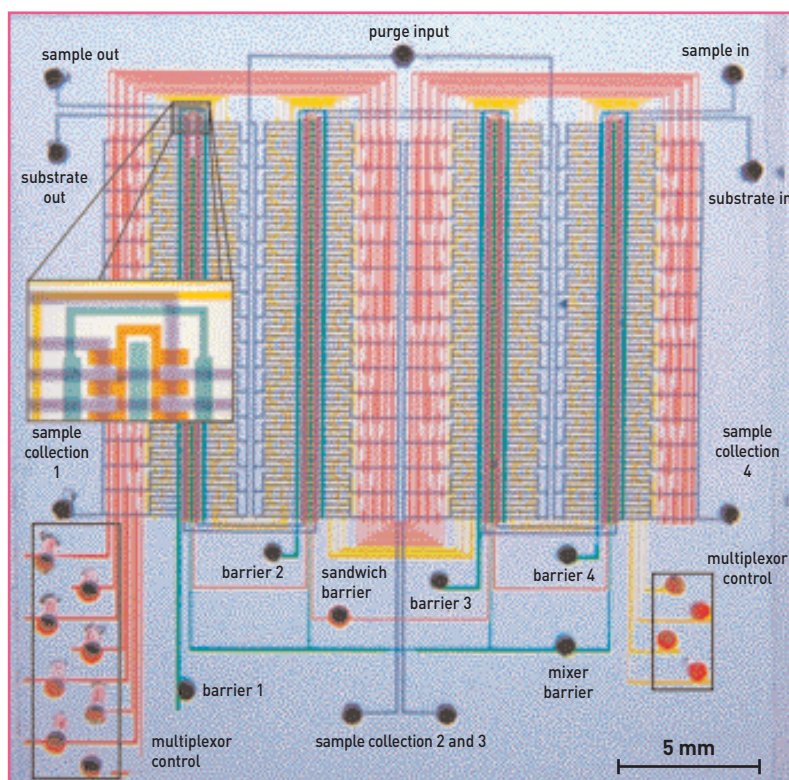
Figure 2.
Schematic diagram of a highly complex microfluidic circuit incorporating the microfluidic valves described in Figure 1. This circuit has 256 reaction chambers, each with a capacity of 750-picolitres. 18 pneumatic channels (in pink, orange and green) actuate 2,056 valves that control access to the reaction chambers. Samples and reagents circulate through the blue channels (from T. Thorsen, S.J. Maerkl and S.R. Quake, "Microfluidic large-scale integration", *Science,* 298, pp. 580-584, 2002).
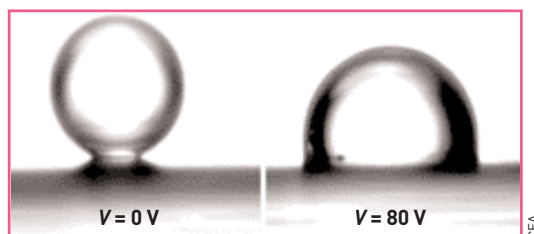

Figure 3.
Electrowetting. A microdroplet is deposited on an electrode coated in a thin layer of hydrophobic dielectric material, such as teflon. When a current is passed through it, the contact angle between the droplet and the surface diminishes, increasing the support's hydrophilicity. Here, the second electrode applying a current (not shown on the photo) is in-plane and concentric to the first electrode.
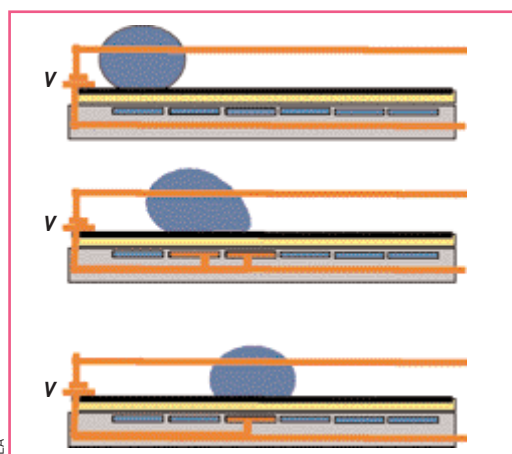

Figure 4.
Microdroplet actuation by electrowetting. By producing a set of sequentially powered electrodes, it becomes possible to move a droplet from electrode to electrode. In this setup, the electrical voltage is established between the electrode located under the dielectric material and a conductor wire acting as a sort of overhead contact line.

Figure 5. Validation of basic microfluidic operations. Left, in-plane movement (from S.-K. Fan, C. Hashi and C.-J. Kim, "Manipulation of multiple droplets on NxM grid by cross-reference EWOD driving scheme and pressure-contact packaging", *IEEE Conf. MEMS, Kyoto, Japan*, Jan. 2003, pp. 694-697); middle, splitting (from S.K. Cho, H. Moon and C.-J. Kim, "Creating, transporting, cutting, and merging liquid droplets by electrowetting-based actuation for digital microfluidic circuits", *Journal of Microelectromechanical Systems*, 12(1), pp. 70-80, 2003); right, forming a droplet from a reservoir.
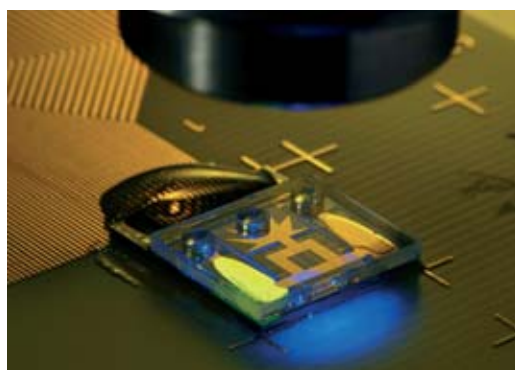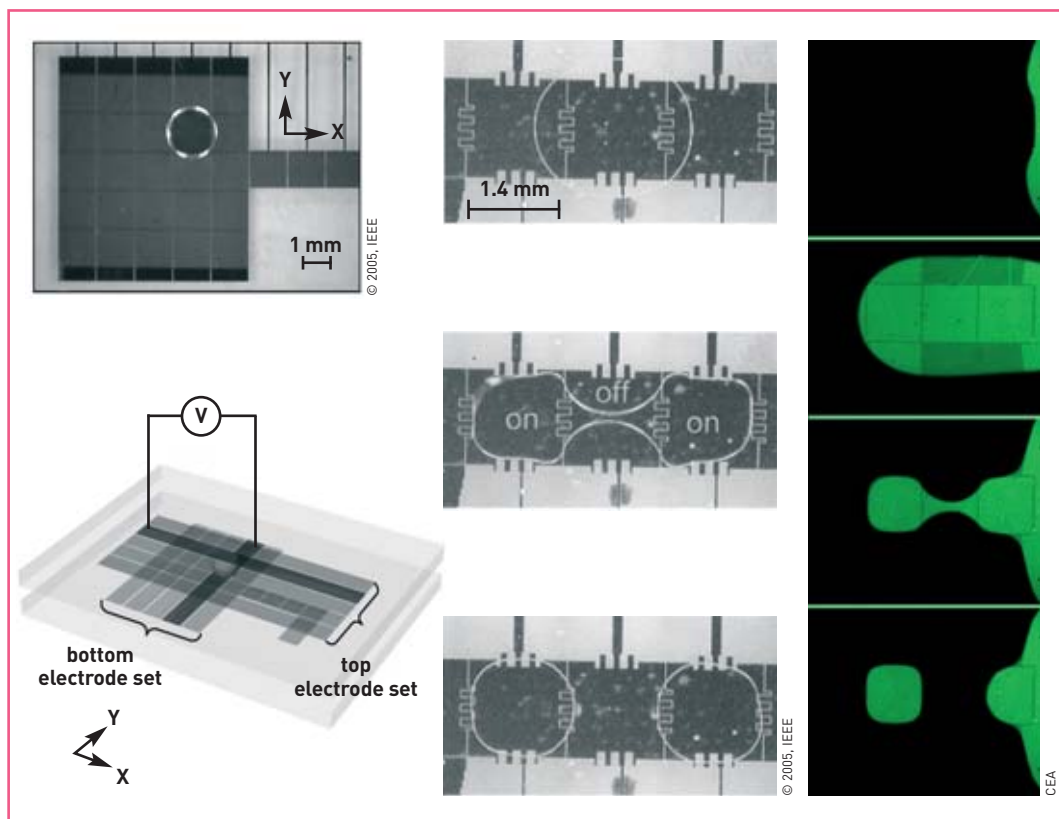
Testing a lab-on-a-chip developed around the concept of droplet microfluidics. Generation of two droplets, with one fluorescent, that move and combine.

Towards the emergence of high-potential solutions

Despite still being in its early days, the field of microfluidics is booming. While the problems it addresses, which are mainly driven by needs in biological analysis, have been clearly identified, there has been an extremely wide and varied range of potential solutions put forward by the scientific community. The most promising of these solutions are expected to emerge over the next few years. We still do not have a comprehensive understanding of the phenomena involved, but progress in this direction will almost certainly improve results and boost the design of computer **models**.

> **Pierre Puget** and **Yves Fouillet**
Technological Research Division
CEA-Leti, Grenoble Centre

FOR FURTHER INFORMATION

Y. Fouillet and J.-L. Achard, "Microfluidique discrète et biotechnologie", *Comptes-rendus. Physique*, 5(5), pp. 577-588, 2004.

D. Erickson and D.Q. Li, "Integrated microfluidic devices", *Analytica Chimica Acta*, 507(1), pp. 11-26, 2004.

D.J. Laser and J.G. Santiago, "A review of micropumps", *Journal of Micromechanics and Microengineering*, 14(6), pp. R35-R64, 2004.

H.A. Stone, A.D. Stroock and A. Adjari, "Engineering flows in small devices: Microfluidics toward a Lab-on-a-Chip", *Annual Review of Fluid Mechanics*, 36, pp. 381-411, 2004.

A   From the macroscopic to the nanoworld, and vice versa...

In order to gain a better idea of the size of microscopic and nanoscopic* objects, it is useful to make comparisons, usually by aligning different scales, *i.e.* matching the natural world, from molecules to man, to engineered or fabricated objects (Figure). Hence, comparing the "artificial" with the "natural" shows that artificially-produced nanoparticles are in fact smaller than red blood cells.

Another advantage of juxtaposing the two is that it provides a good illustration of the two main ways of developing nanoscale systems or objects: *top-down* and *bottom-up*. In fact, there are two ways

* From the Greek *nano meaning* "very small", which is also used as a prefix meaning a billionth (10^{-9}) of a unit. In fact, the **nanometre** (1 nm = 10^{-9} metres , or a billionth of a metre), is the master unit for nanosciences and nanotechnologies.

300-mm silicon wafer produced by the Crolles2 Alliance, an illustration of current capabilities using top-down microelectronics.

into the nanoworld: molecular manufacturing, involving the control of single atoms and the building from the ground up, and extreme miniaturization, generating progressively smaller systems.

Top-down technology is based on the artificial, using macroscopic materials that we chip away using our hands and our tools: for decades now, electronics has been applied using silicon as a substrate and what are called "wafers" as workpieces. In fact, microelectronics is also where the "top-down" synthesis approach gets its name from. However, we have reached a stage where, over and above simply adapting the miniaturization of silicon, we also
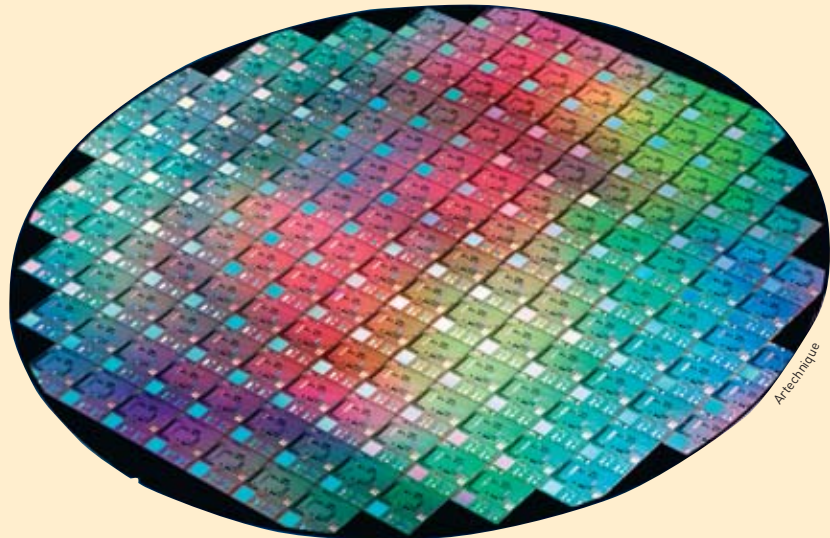
have to take on or use certain physical phenomena, particularly from quantum physics, that operate when working at the nanoscale.

The bottom-up approach can get around these physical limits and also cut manufacturing costs, which it does by using component self-assembly. This is the approach that follows nature by assembling molecules to create proteins, which are a series of amino acids that the super-molecules, *i.e.* nucleic acids (DNA, RNA), are able to produce within cells to form functional structures that can reproduce in more complex patterns. Bottom-up synthesis aims at structuring the material using
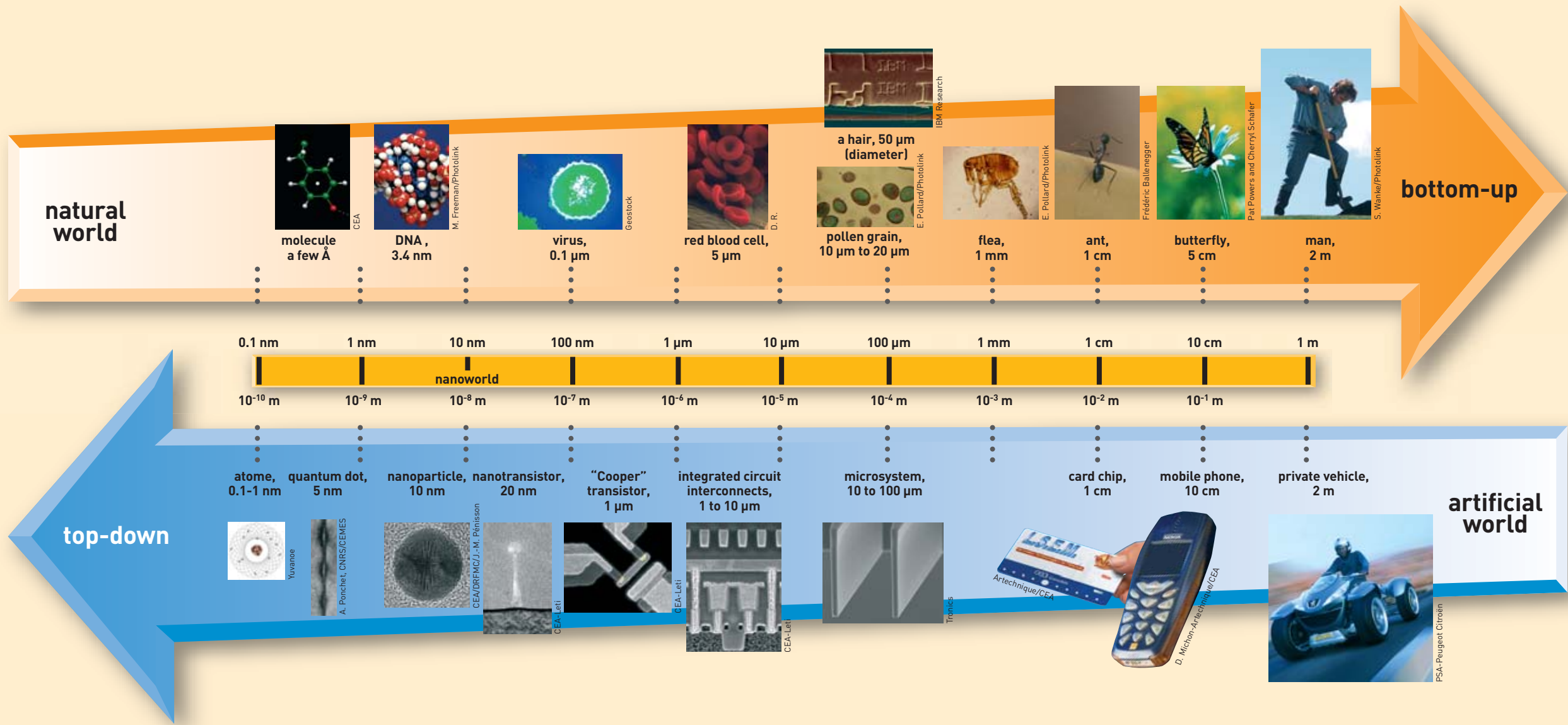
"building blocks", including atoms themselves, as is the case with living objects in nature. Nanoelectronics seeks to follow this assembly approach to make functional structures at lower manufacturing cost.

The nanosciences can be defined as the body of research into the physical, chemical or biological properties of nano-objects, how to manufacture them, and how they self-assemble by auto-organisazation.

Nanotechnologies cover all the methods that can be used to work at molecular scale to reorganize matter into objects and materials, even progressing to the macroscopic scale.

natural world

bottom-up

| molecule a few Å | DNA, 3.4 nm | virus, 0.1 µm | red blood cell, 5 µm | pollen grain, 10 µm to 20 µm | a hair, 50 µm (diameter) | flea, 1 mm | ant, 1 cm | butterfly, 5 cm | man, 2 m |

| 0.1 nm | 1 nm | 10 nm | 100 nm | 1 µm | 10 µm | 100 µm | 1 mm | 1 cm | 10 cm | 1 m |

nanoworld

| 10^{-10} m | 10^{-9} m | 10^{-8} m | 10^{-7} m | 10^{-6} m | 10^{-5} m | 10^{-4} m | 10^{-3} m | 10^{-2} m | 10^{-1} m |

| atome, 0.1-1 nm | quantum dot, 5 nm | nanoparticle, 10 nm | nanotransistor, 20 nm | "Cooper" transistor, 1 µm | integrated circuit interconnects, 1 to 10 µm | microsystem, 10 to 100 µm | card chip, 1 cm | mobile phone, 10 cm | private vehicle, 2 m |

top-down

artificial world

B A guide to quantum physics

Quantum physics (historically known as quantum mechanics) covers a set of physical laws that apply at microscopic scale. While fundamentally different from the majority of laws that appear to apply at our own scale, the laws of quantum physics nevertheless underpin the general basis of physics at all scales. That said, on the macroscopic scale, quantum physics in action appears to behave particularly strangely, except for a certain number of phenomena that were already curious, such as **superconductivity** or superfluidity, which in fact can only explained by the laws of quantum physics. Furthermore, the transition from the validating the paradoxes of quantum physics to the laws of classical physics, which we find easier to comprehend, can be explained in a very general way, as will be mentioned later.

Quantum physics gets its name from the fundamental characteristics of quantum objects: characteristics such as the angular momentum (**spin**) of **discrete** or discontinuous particles called **quanta**, which can only take values multiplied by an elementary *quantum*. There is also a **quantum of action** (product of a unit of energy multiplied by time) called **Planck's cons-**
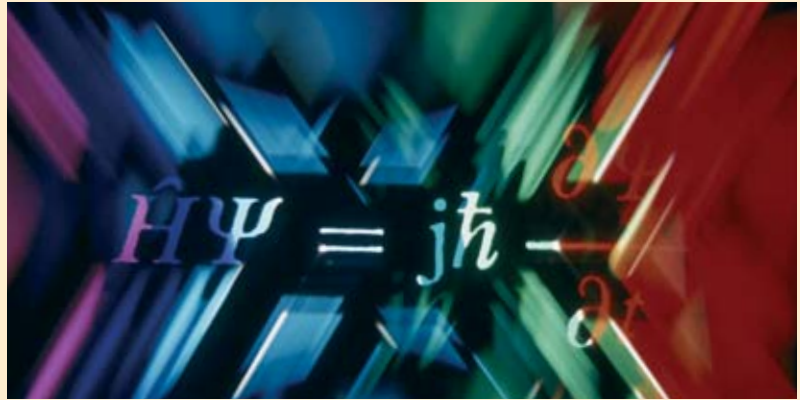
An "artist's impression" of the Schrödinger equation.

tant (symbolized as h) which has a value of 6.626×10^{-34} joule·second. While classical physics separates *waves* from *particles*, quantum physics somehow covers both these concepts in a third group, which goes beyond the simple wave-particle duality that Louis de Broglie imagined. When we attempt to comprehend it, it sometimes seems closer to waves, and sometimes to particles. A quantum object cannot be separated from how it is observed, and has no fixed attributes. This applies equally to a particle - which in no way can be likened to a tiny little bead following some kind of trajectory - of light (**photon**)

or matter (**electron**, **proton**, **neutron**, **atom**, etc.).

This is the underlying feature behind the **Heisenberg uncertainty principle**, which is another cornerstone of quantum physics. According to this principle (which is more *indeterminacy* than *uncertainty*), the position and the velocity of a particle cannot be measured *simultaneously* at a given point in time. Measurement remains possible, but can never be more accurate than *h*, Planck's constant. Given that these approximations have no intrinsically real value outside the observation process, this simultaneous determination of both position and velocity becomes simply impossible.

B (next)

At any moment in time, the quantum object presents the characteristic of *superposing* several states, in the same way that one wave can be the *sum* of several others. In quantum theory, the amplitude of a wave (like the peak, for example) is equal to a **probability amplitude** (or probability wave), a complex number-valued function associated with each of the possible sates of a system thus described as quantum. Mathematically speaking, a physical state in this kind of system is represented by a **state vector**, a function that can be added to others *via* superposition. In other words, the sum of two possible state vectors of a system is *also* a possible state vector of that system. Also, the product of two vector spaces is also the sum of the vector products, which indicates **entanglement**: as a state vector is generally spread through space, the notion of local objects no longer holds true. For a pair of entangled particles, *i.e.* particles created together or having already interacted, that is, described by the *product* and not the *sum* of the two individual state vectors, the fate of each particle is linked - entangled - with the other, regardless of the distance between the two. This characteristic, also called *quantum state entan-glement*, has staggering consequences, even before considering the potential applications, such as quantum cryptography or - why not? - teleportation. From this point on, the ability to predict the behaviour of a quantum system is reduced to probabilistic or statistical predictability. It is as if the quantum object is some kind of "juxtaposition of possibilities". Until it has been measured, the measurable size that supposedly quantifies the physical property under study is not strictly defined. Yet as soon as this measurement process is launched, it destroys the **quantum superposition** through the "collapse of the wave-packet" described by Werner Heisenberg in 1927. All the properties of a quantum system can be deduced from the equation that Erwin Schrödinger put forward the previous year. Solving the **Schrödinger equation** made it possible to determine the energy of a system as well as the **wave function**, a notion that tends to be replaced by the probability amplitude.

According to another cornerstone principle of quantum physics, the **Pauli exclusion principle**, two identical half-spin ions (**fermions**, particularly electrons) cannot simultaneously share the same position, spin and velocity (within the limits imposed by the uncertainty principle), *i.e.* share the same *quantum state*. **Bosons** (especially photons) do not follow this principle, and can exist in the same quantum state.

The coexistence of **superposition states** is what lends **coherence** to a quantum system. This means that the theory of **quantum decoherence** is able to explain why macroscopic objects, atoms and other particles, present "classical" behaviour whereas microscopic objects show quantum behaviour. Far more influence is exerted by the "environment" (air, background radiation, etc.) than an advanced measurement device, as the environment radically removes all *superposition of states* at this scale. The larger the system considered, the more it is coupled to a large number of degrees of freedom in the environment, which means the less "chance" (to stick with a probabilistic logic) it has of maintaining any degree of quantum coherence.

C Molecular beam epitaxy

Quantum wells are grown using Molecular Beam Epitaxy (from the Greek *taxi*, meaning order, and *epi*, meaning over), or MBE. The principle of this physical deposition technique, which was first developed for growing III-V semiconductor crystals, is based on the evaporation of ultra-pure elements of the component to be grown, in a furnace under ultra-high vacuum (where the pressure can be as low as $5 \cdot 10^{-11}$ mbar) in order to create a pure, pollution-free surface. One or more thermal beams of atoms or molecules react on the surface of a single-crystal wafer placed on a substrate kept at high temperature (several hundred °C), which serves as a lattice for the formation of a film called epitaxial film. It thus becomes possible to stack ultra-thin layers that measure a millionth of a millimetre each, *i.e.* composed of only a few atom planes.

The elements are evaporated or sublimated from an ultra-pure source placed in an effusion cell (or Knudsen cell; an enclosure where a molecular flux moves from a region with a given pressure to another region of lower pressure) heated by the Joule effect. A range of structural and analytical probes can monitor film growth *in situ* in real time, particularly using surface quality analysis and grazing angle phase transitions by LEED (*Low energy electron diffraction*) or RHEED (*Reflection high-energy electron diffraction*). Various spectroscopic methods are also used, including Auger electron spectroscopy, secondary ion mass spectrometry (SIMS), X-ray photoelectron spectrometry (XPS) or ultraviolet photoelectron spectrometry (UPS).

As *ultra-high-vacuum* technology has progressed, molecular beam epitaxy has branched out to be applied beyond III-V semiconductors to embrace metals and insulators. In fact, the vacuum in the growth chamber, whose design changes depending on the properties of the matter intended to be deposited, has to be better than 10^{-11} mbar in order to grow an ultra-pure film of exceptional crystal quality at relatively low substrate temperatures. This value corresponds to the vacuum quality when the growth chamber is at rest. Arsenides, for example, grow at a residual vacuum of around 10^{-8} mbar as soon as the arsenic cell has reached its set growth temperature.

The pumping necessary to achieve these performance levels draws on several techniques using ion pumps, cryopumping, titanium sublimation pumping, diffusion pumps or turbomolecular pumps. The main impurities (H_2, H_2O, CO and CO_2) can present partial pressures of lower than 10^{-13} mbar.

D The transistor, fundamental component of integrated circuits

The first transistor was made in germanium by John Bardeen and Walter H. Brattain, in December 1947. The year after, along with William B. Shockley at Bell Laboratories, they developed the bipolar transistor and the associated theory. During the 1950s, transistors were made with silicon (Si), which to this day remains the most widely-used semiconductor due to the exceptional quality of the interface created by silicon and silicon oxide (SiO_2), which serves as an insulator. In 1958, Jack Kilby invented the integrated circuit by manufacturing 5 components on the same substrate. The 1970s saw the advent of the first microprocessor, produced by Intel and incorporating 2,250 transistors, and the first memory. The complexity of integrated circuits has grown exponentially (doubling every 2 to 3 years according to "Moore's law") as transistors continue to become increasingly miniaturized.

The transistor, a name derived from transfer and resistor, is a fundamental component of microelectronic integrated circuits, and is set to remain so with the necessary changes at the nanoelectronics scale: also well-suited to amplification, among other functions, it performs one essential basic function which is to open or close a current as required, like a switching device (Figure). Its basic working principle therefore applies directly to processing binary code (0, the current is blocked, 1 it goes through) in logic circuits (inverters, gates, adders, and memory cells).

The transistor, which is based on the transport of electrons in a solid and not in a vacuum, as in the electron tubes of the old triodes, comprises three electrodes (anode, cathode and gate), two of which serve as an electron reservoir: the source, which acts as the emitter filament of an electron tube, the drain, which acts as the collector plate, with the gate as "controller". These elements work differently in the two main types of transistor used today: bipolar junction transistors, which came first, and field effect transistors (FET).

Bipolar transistors use two types of charge carriers, electrons (negative charge) and holes (positive charge), and are comprised of identically doped (p or n) semiconductor substrate parts
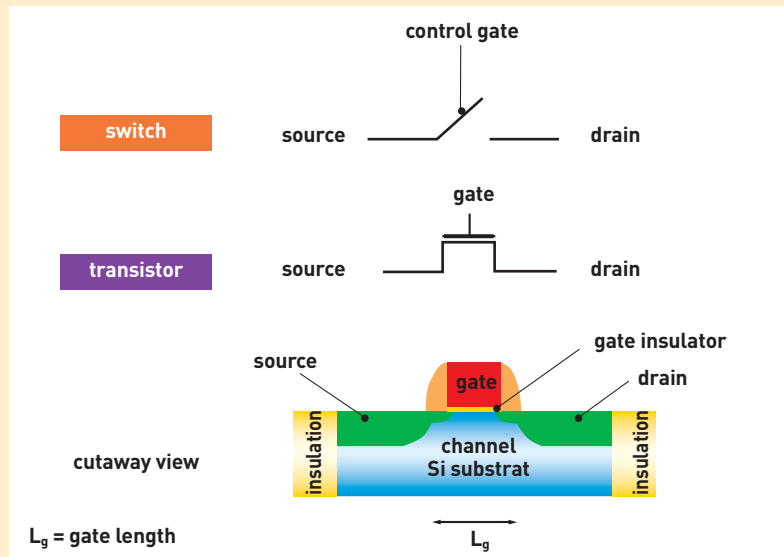

Figure.
A MOS transistor is a switching device for controlling the passage of an electric current from the source (S) to the drain (D) via a gate (G) that is electrically insulated from the conducting channel. The silicon substrate is marked B for Bulk.

separated by a thin layer of inversely-doped semiconductor. By assembling two semiconductors of opposite types (a p-n junction), the current can be made to pass through in only one direction. Bipolar transistors, whether n-p-n type or p-n-p type, are all basically current amplifier controlled by a gate current[1]: thus, in an n-p-n transistor, the voltage applied to the p part controls the flow of current between the two n regions. Logic circuits that use bipolar transistors, which are called TTL (for transistor-transistor logic), consume more energy than field effect transistors which present a zero gate current in off-state and are voltage-controlled.

Field effect transistors, most commonly of **MOS** (metal oxide semiconductor) type, are used in the majority of today's **CMOS** (C for complementary) logic circuits[2]. Two n-type regions are created on a p-type silicon crystal by doping the surface. These two regions, also called drain and source, are thus separated by a very narrow p-type space called the **channel**. The effect of a positive current on the control electrode, naturally called the **gate**, positioned over the semiconductor forces the holes to the surface, where they attract the few mobile electrons of the semiconductor. This forms a conducting channel between source and drain (Figure). When a negative voltage is applied to the gate, which is electrically insulated by an oxide layer, the electrons are forced out of the channel. As the positive voltage increases, the channel resistance decreases, letting progressively more current through.

In an integrated circuit, transistors together with the other components (**diodes**, condensers, resistances) are initially incorporated into a "chip" with more or less complex functions. The circuit is built by "sandwiching" layer upon layer of conducting materials and insulators formed by **lithography** (Box E, *Lithography, the key to miniaturization*). By far the most classic application of this is the microprocessor at the heart of our computers, which contains several hundred million transistors (whose size has been reduced 10,000-fold since the 1960s), soon a billion. This has led to industrial manufacturers splitting the core of the processors into several subunits working in parallel!

The very first transistor.
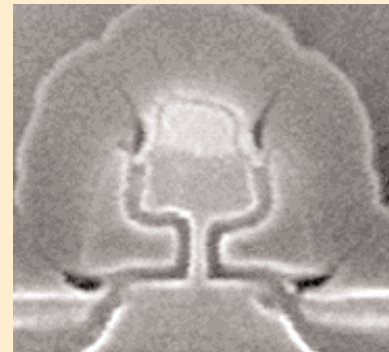
8 nanometre transistor developed by the Crolles2 Alliance bringing together STMicroelectronics, Philips and Freescale Semiconductor.

(1) This category includes **Schottky transistors** or **Schottky barrier transistors** which are field effect transistors with a metal/semiconductor control gate that, while more complex, gives improved charge-carrier mobility and response times.

(2) Giving **MOSFET** transistor (for Metal Oxide Semiconductor Field Effect Transistor).

E Lithography, the key to miniaturization

Optical lithography (photolithography) is a major application in the particle-matter interaction, and constitutes the classical process for fabricating integrated circuits. It is a key step in defining circuit patterns, and remains a barrier to any future development. Since resolution, at the outset, appears to be directly proportional to wavelength, feature-size first progressed by a step-wise shortening of the wavelength λ of the radiation used.

The operation works via a reduction lens system, by the *exposure* of a photoresist film to energy particles, from the ultraviolet (UV) photons currently used through to X photons, ions, and finally electrons, all through a mask template carrying a pattern of the desired circuit. The aim of all this is to transfer this pattern onto a stack of insulating or conducting layers that make up the mask. These layers will have been deposited previously (the *layering* stage) on a wafer of semiconductor material, generally silicon. After this process, the resin dissolves under exposure to the air (*development*). The exposed parts of the initial layer can then be etched selectively, then the resin is lifted away chemically before deposition of the following layer. This lithography step can take place over twenty times during the fabrication of an integrated circuit (Figure).

In the 1980s, the microelectronics industry used mercury lamps delivering near-UV (g, h and i lines) through quartz optics, with an emission line of 436 nanometres (nm). This system was able to etch structures to a feature-size of 3 microns (μm). This system was used through to the mid-90s, when it was replaced by excimer lasers emitting **far-UV** light (KrF, krypton fluoride at 248 nm, then ArF, argon fluoride at 193 nm, with the photons thus created generating several electronvolts) that were able to reach a resolution of 110 nm, pushed to under 90 nm with new processes.

In the 1980s, the CEA's Electronics and Information Technology Laboratory (Leti) pioneered the application of lasers in lithography and the fabrication of integrated circuits using excimer lasers, and even the most advanced integrated circuit production still uses these sources.

Photolithography section in ultra-clean facilities at the STMicroelectronics unit in Crolles (Isère).

Artechnique

The next step for high-volume production was expected to be the F_2 laser (λ = 157 nm), but this lithography technology has to all intents and purposes been abandoned due to complications involved in producing optics in CaF_2, which is transparent at this wavelength. While the shortening of wavelengths in exposure tools has been the driving factor behind the strong resolution gain already achieved, two other factors have nevertheless played key roles. The first was the development of polymer-lattice photoresists with low absorbance at the wavelengths used, implementing progressively more innovative input energy reflection/emission systems. The second was enhanced optics reducing diffraction interference (better surface quality, increase in numerical aperture).

Over the years, the increasing complexity of the optical systems has led to resolutions actually *below* the source wavelength. This development could not continue without a major technological breakthrough, a huge step forward in wavelength. For generations of integrated circuits with a lowest resolution of between 80 and 50 nm (the next "node" being at 65 nm), various different approaches are competing to offer particle projection at ever-shorter wavelengths. They use either "soft" X-rays at extreme ultraviolet wavelength (around 10 nm), "hard" X-rays at wavelengths below 1 nm, ions or electrons.

The step crossing below the 50 nm barrier will lead towards low-electron-energy (10 eV)-enabled nanolithography with technology solutions such as the scanning tunnelling microscope and molecular beam epitaxy (Box C) for producing "superlattices".
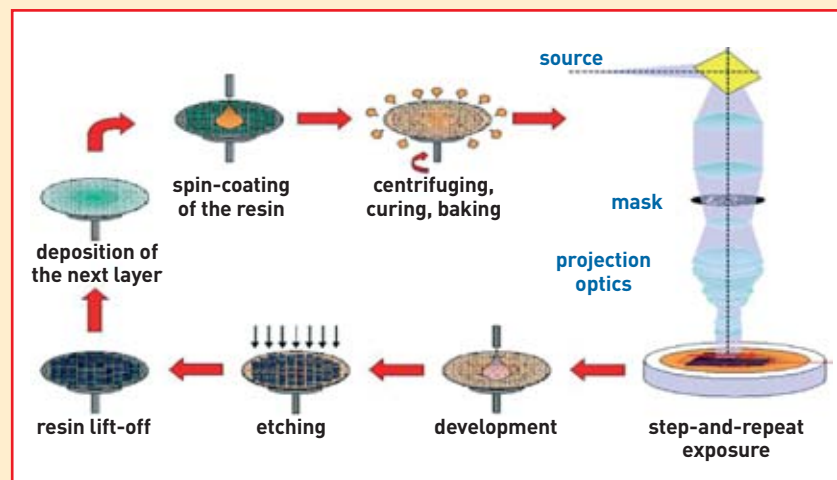

Figure. The various phases in the lithography process are designed to carve features out of the layers of conducting or insulating materials making up an integrated circuit. The sequences of the operation are laying of a photoresist, then projecting the pattern on a mask using a reduction optics system, which is followed by dissolution of the resin that is exposed to the light beam (development). The exposed parts of the initial layer can then be etched selectively, then the resin is lifted away before deposition of the following layer.

G The tunnel effect, a quantum phenomenon

Quantum physics predicts unexpected behaviour that defies ordinary intuition. The tunnel effect is an example. Take the case of a marble that rolls over a bump. Classical physics predicts that unless the marble has enough kinetic energy it will not reach the top of the bump, and will roll back towards its starting point. In quantum physics, a particle (proton, electron) can get past the bump even if its initial energy is insufficient, by "tunnelling" through. The tunnel effect makes it possible for two protons to overcome their mutual electrical repulsion at lower relative velocities than those predicted by classical calculations.

Tunnel effect microscopy is based on the fact that there is a finite probability that a particle with energy lower than the height of a potential barrier (the bump) can still jump over it. The particles are electrons travelling through the space between two electrodes. These electrodes are a fine metal tip terminating in a single atom, and the metal or semiconductor surface of the sample. In classical physics a solid surface is considered as a well-defined boundary with electrons confined inside the solid. By contrast, in quantum physics each electron has wave properties that make its location uncertain. It can be visualized as an electron cloud located close to the surface. The density of this cloud falls off exponentially with increasing distance from the solid surface. There is thus a certain probability that an electron will be located "outside" the solid at a given time. When the fine metal tip is brought near the surface at a distance of less than a nanometre, the wave function associated with the electron is non-null on the other side of the potential barrier and so electrons can travel from the surface to the tip, and vice versa, by the tunnel effect. The potential barrier crossed by the electron is called the tunnel barrier. When a low potential is applied between the tip and the surface, a tunnel current can be detected. The tip and the surface being studied together form a local tunnel junction. The tunnel effect is also at work in Josephson junctions where a direct current can flow through a narrow discontinuity between two superconductors.

In a transistor, an unwanted tunnel effect can appear when the insulator or grid is very thin (nanometre scale). Conversely, the effect is put to use in novel devices such as Schottky barrier tunnel transistors and carbon nanotube assemblies.