Ultimate MOSFET as a nanoscience object

The "ultimate" field-effect transistor, a transistor in which electrons can be controlled one-by-one, is a first-choice subject of study in mesoscopic physics, at the fringe of both quantum physics and the physics of macroscopic objects.

Electrons in **field-effect transistors** smaller than 50 nanometres (nm) produced from **silicon** technology can be controlled one by one. The mechanisms of electron transport, which have appeared only recently in these FET devices, are in fact well known to **mesoscopic** physics, where researchers working with low temperatures have no particular difficulties in understanding certain physical limits of how nanotransistors operate. However, for physicists, these well-controlled, very-small-scale samples are also very stable **quantum** dots that are relatively original compared to more extensively-studied assemblies.

Three reasons for miniaturization

The basic component of processors is the silicon field-effect transistor, which acts as an ultra-small-scale switch (Box D). In this three-point device, the current between the first two points, **drain** and **source**, is controlled by a voltage charge applied to the third, the **gate**. At any given drain voltage, the drain-source current has to be highest when the gate is **polarized** and lowest when the gate voltage is zero. The miniaturization of the field-effect transistor is driven by cost imperatives, the need to integrate a great many functions on a reduced

D The transistor, fundamental component of

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.

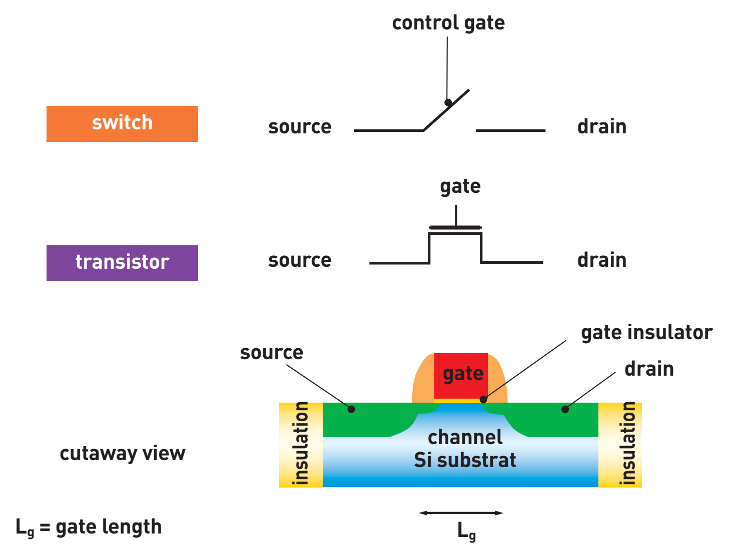


Figure.
A MOS transistor is a switching device for controlling the passage of an electric current from the source (S) to the drain (D) via a gate (G) that is electrically insulated from the conducting channel. The silicon substrate is marked B for Bulk.

The transistor, a name derived from *trans*fer and res*istor*, 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 surface, and the increase in circuit speed. It should be borne in mind that a shorter physical source-drain length gives a shorter switching time between the two logical states (on and off). Transistors with a drain-source distance typically lower than 10 nm are currently in the prototype stage. These components count among the smallest individually-connected objects that are stable, reproducible and controllable. While they are still huge compared to atoms (around 0.1 nm) and molecules (around 1 nm), the active region of these electrical devices can still only accommodate a few electrons. This borderline region raises several questions, particularly the possibility of controlling individual electrons.

The ultimately-miniaturized field-effect transistor is in fact an object in which there is no electron available for transport between drain and source at zero gate voltage, and an electronic state available with a polarized gate. Studying ultra-small field-effect transistors at low temperatures reveals the conditions under which the object can be engineered and what performance could be expected.

The Coulomb blockade for counting electrons

Matter contains many electrons, but only a very small number of them that are delocalized conduct electrical current. The underlying principle behind the field-effect transistor is to use the **capacitance** controlled by the gate voltage to increase the number of **conduction electrons** between source and drain (this region is called the **channel**). As electron density increases, the current, which is proportional to the electron flow through the channel, gets stronger. In an ultra-small field-effect transistor, the number of electrons can be counted thanks to a phenomenon called **Coulomb blockade**, which competes with the field-effect.

The Coulomb blockade stems from **Coulomb repulsion** which "feels" an electron as it approaches a charged object. Generally, the channel as a whole is charged, as there is an integer of electrons, whereas the positive charge varies constantly with gate voltage. There is, therefore, a small charge imbalance - less than the elementary electric charge e - except for the **discrete** values of the gate voltage. The only condition is

integrated circuits

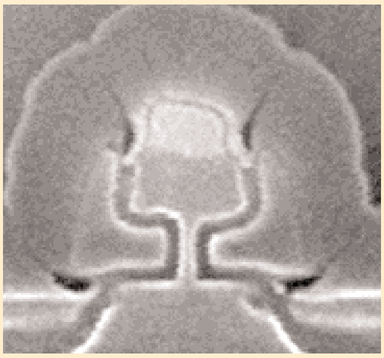
separated by a thin layer of inversely-doped semiconductor. By assembling two semiconductors of opposite types (a p-n junction), the current can be made to pass through in only one direction. Bipolar transistors, whether n-p-n type or p-n-p type, are all basically current amplifier controlled by a gate 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!


Lucent Technologies Inc./Bell Labs

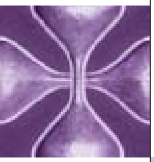
The very first transistor.

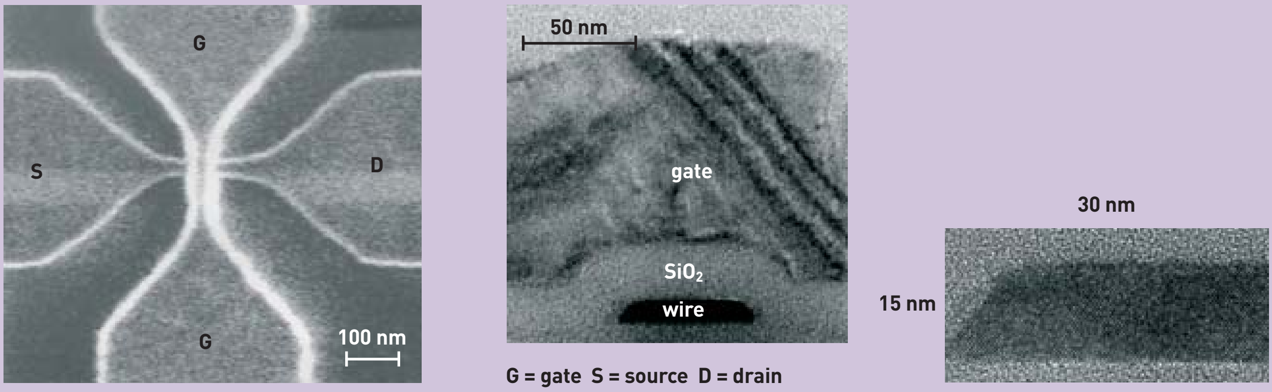

STMicroelectronics

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).





CEA/DSM/DRFMC-DRT/Leti

The simplest single-electron transistor is a silicon field-effect component with input resistances of around 100 kilohms.

that there is a clear integer of electrons in the channel, which is tested when resistances at the inputs between channel and drain-source, which in a way “insulate” the channel, are sufficiently strong, and the temperature is lower than the charging energy $e^2/2C$, where C is total channel capacitance. Ultra-narrow devices etched into ultrathin films naturally display high input resistances, whereas capacity, which decreases with channel size, can reach a few meV in the samples developed by a team of researchers from the CEA Electronics and Information Technology Laboratory (Leti: Laboratoire d'électronique et de technologie de l'information) and Department of Fundamental Research on Condensed Matter (DRFMC: Département de recherche fondamentale sur la matière condensée). The Coulomb blockade thus prevents the electrons from crossing the channel, except for discrete gate voltage values which, modelled simply, are regularly spaced at an interval of ΔVg= e/Cg, where Cg is capacitance between gate and channel.

At low temperatures though, it has been observed that the current does not increase monotonically with gate voltage, thus with electron density, but displays oscillations (Figure). The number of oscillations gives the number of electrons in the channel.

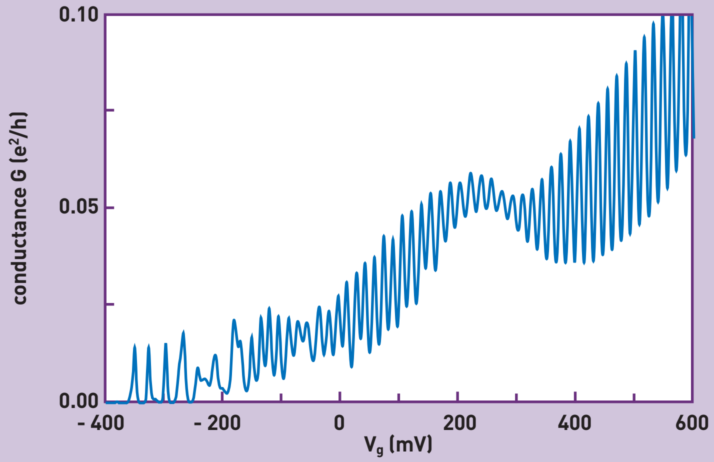


Figure.
Drain-source conductance in quantum units (proportional to the current between drain (D) and source (S) in relation to gate (G) voltage V_g in a field-effect transistor (Illustration 1a). The measurement temperature is 4.2 K. The measured current oscillates regularly at high gate voltage (at lower temperatures, the oscillations become regularly-spaced resonances). Each oscillation corresponds to the addition of an electron into the region under the gate in the etched-out silicon-on-insulator wire (Illustrations 1b and 1c). The electrons are confined by the resistance of the wire not covered by the gate. At negative gate voltage, the first oscillations are irregularly spaced, this characterizing quantum effects on the small numbers of electrons under the gate (joint Leti-DRFMC project).

Every extra electron carries its cost

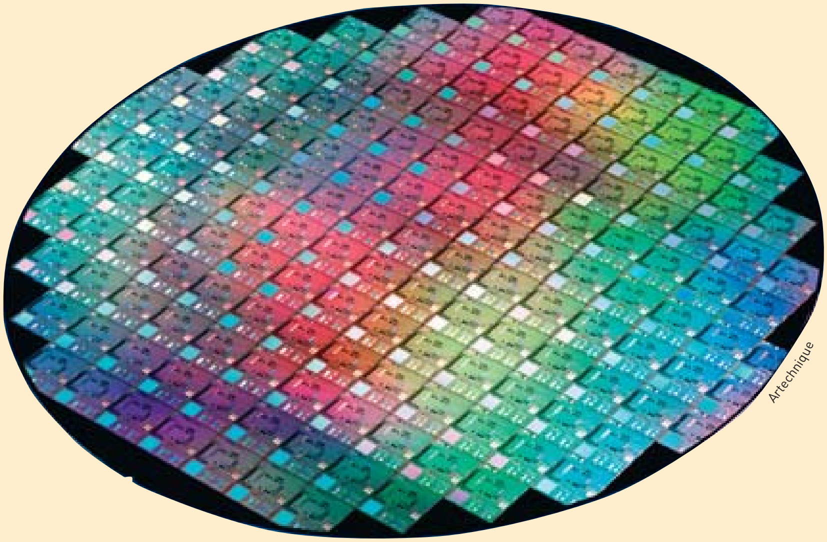
When this figure climbs into the tens, the oscillation period is particularly regular, and can be given by e/Cg. However, for the first electrons, this is not the case. For example, adding a third or a fourth electron does not cost the same energy, as the spatial organization of the electron wave functions is highly dependent on electron number. Their Coulomb interaction, magnetization, and kinetic energy fluctuate according to their number, in the same way as for atomic electrons and their layer structure (although their confinement potentials are very different). This is a characteristic pattern of electron structures similar in size to the electron wavelength. A further hurdle, in relation to atoms, is that artificial devices that appear to be identical in fact have imperceptible differences, and the gate voltage for the first electron and those that follow actually varies between devices. Similarly, the current carried by these first electrons fluctuates according to microscopic features in the contacts.

Irregularities in Coulomb oscillations for the first electrons therefore have to be separated between common characteristics and characteristics that change between samples. Silicon technology can be used to compare samples according to well-controlled variables - some particularly innovative - and obtain series of stable, reproducible, ultra-small samples.

This is a major challenge, since at around 10 nanometres, the physics of the devices becomes particularly complicated in comparison with both atoms and macroscopic objects. Experiments led in this fringe zone should help to expand our understanding of this mesoscopic regime.

> Marc Sanquer
Materials Science Division
CEA Grenoble Centre

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


Artechnique

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

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 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.

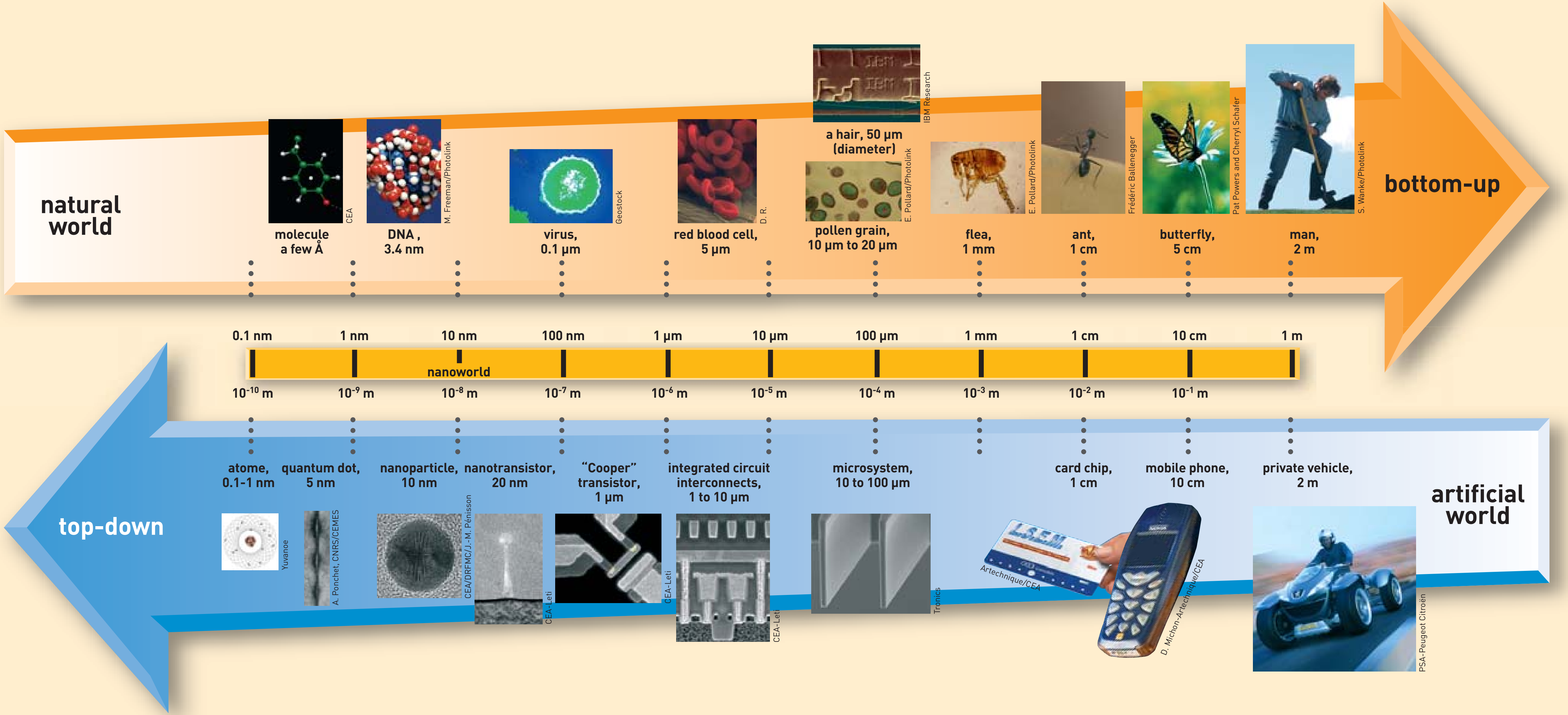
The nanosciences can be defined as the body of research into the physical, chemical or biological properties of nano-objects, how to manufacture them, and how they self-assemble by auto-organisazation.

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.

---

* 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.

A (next)



natural world

bottom-up

molecule
a few Å
CEA

DNA ,
3.4 nm
M. Freeman/Photolink

virus,
0.1 µm
Geostock

red blood cell,
5 µm
D. R.

a hair, 50 µm
(diameter)
IBM Research

pollen grain,
10 µm to 20 µm
E. Pollard/Photolink

flea,
1 mm
E. Pollard/Photolink

ant,
1 cm
Frédéric Ballenegger

butterfly,
5 cm
Pat Powers and Cherryl Schafer

man,
2 m
S. Wanke/Photolink

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

nanoworld

top-down

atome,
0.1-1 nm
Yuvanoe

quantum dot,
5 nm
A. Ponchet, CNRS/CEMES

nanoparticle,
10 nm
CEA/DRFMC/J.-M. Pénisson

nanotransistor,
20 nm
CEA-Leti

"Cooper"
transistor,
1 µm
CEA-Leti

integrated circuit
interconnects,
1 to 10 µm
CEA-Leti

microsystem,
10 to 100 µm
Tronics

card chip,
1 cm
Artechnique/CEA

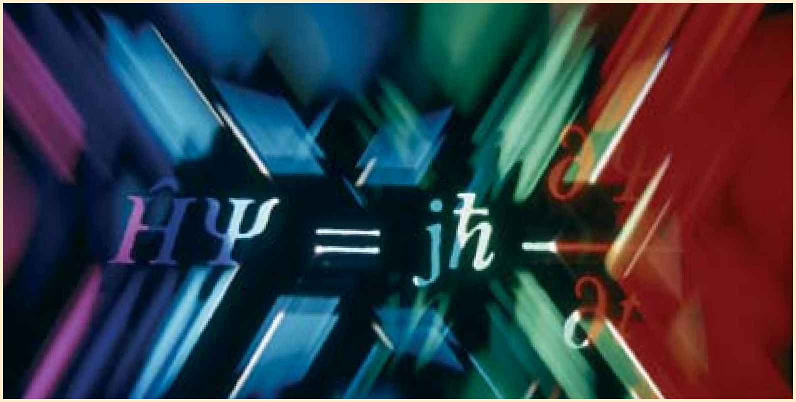
mobile phone,
10 cm
D. Michon-Artechnique/CEA

private vehicle,
2 m
PSA-Peugeot Citroën

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.



D. Sarraute/CEA

An "artist's impression" of the Schrödinger equation.

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 constant** (symbolized as h) which has a value of 6.626 x 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 entanglement*, 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.

TO FIND OUT MORE:

Étienne Klein, *Petit voyage dans le monde des quanta*, Champs, Flammarion, 2004.

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 turbo-molecular 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.

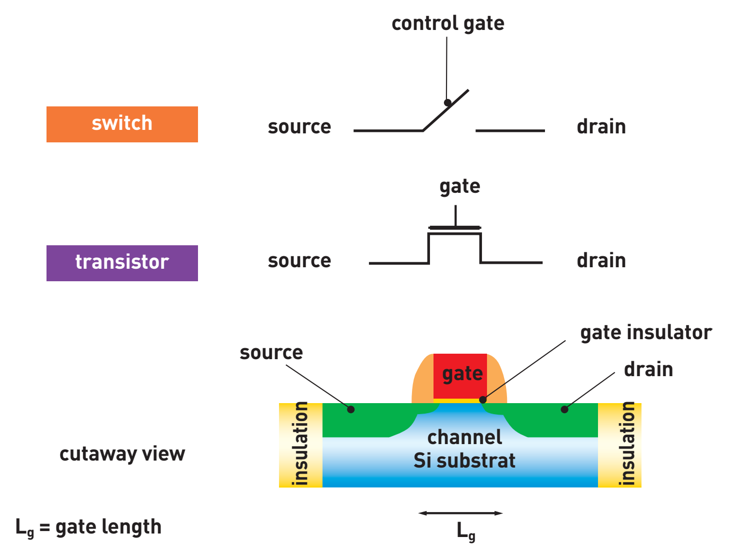


Figure.
A MOS transistor is a switching device for controlling the passage of an electric current from the source (S) to the drain (D) via a gate (G) that is electrically insulated from the conducting channel. The silicon substrate is marked B for Bulk.

The transistor, a name derived from *trans*fer and res*istor*, 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

D (next)

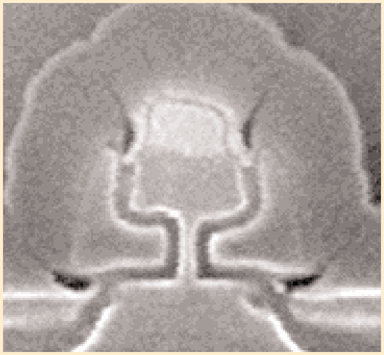
separated by a thin layer of inversely-doped semiconductor. By assembling two semiconductors of opposite types (a p-n junction), the current can be made to pass through in only one direction. Bipolar transistors, whether n-p-n type or p-n-p type, are all basically current amplifier controlled by a gate 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!


Lucent Technologies Inc./Bell Labs

The very first transistor.


STMicroelectronics

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

Artechnique

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

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.

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".

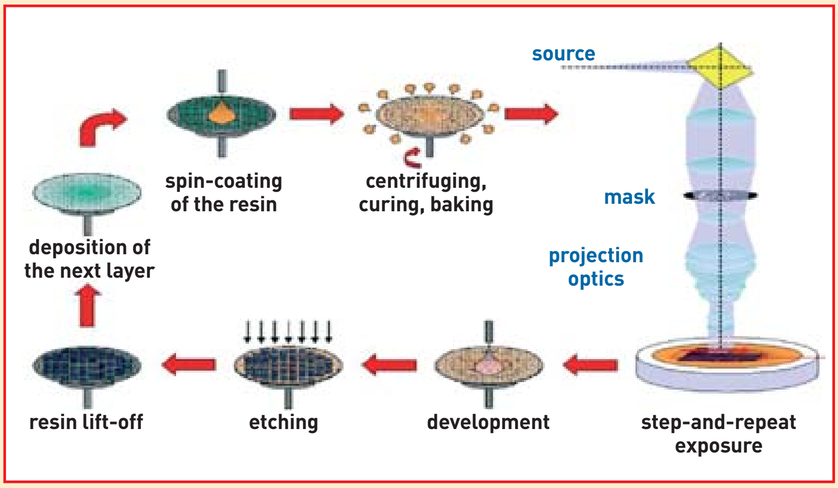


Figure. The various phases in the lithography process are designed to carve features out of the layers of conducting or insulating materials making up an integrated circuit. The sequences of the operation are laying of a photoresist, then projecting the pattern on a mask using a reduction optics system, 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

Q**uantum 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.