LS Scienza dei Materiali - a.a. 2008/09

Fisica delle Nanotecnologie – part 1

Version 7, Oct 2008

Francesco Fuso, tel 0502214305, 0502214293 - fuso@df.unipi.it
http://www.df.unipi.it/~fuso/dida

Nanotecnologie: generalità, ambito, motivazioni; stato dell'arte (elettronica)

Outlook

- What is *nanotechnology?*
- What are the *components* of nanotechnology?
- What are the main *driving forces* for the development of nanotechnology?
- What is the *present status* of technology?
- Survey of *conventional* electronics:
  - Bipolar junctions (old-style technology);
  - Planar technology and MOS-FETs
- Some *limits and problems* in miniaturization and the need for new approaches

(Nano)technology

Technology: the ability to fabricate systems **useful** for some applications
Nano: fabricated systems are “small”-sized (*hard to figure out how small…)*

i.e., the ability to manipulate matter in order to fabricate systems (or structures, or devices) with a size in the **sub-micrometer** range

Technology uses techniques, but **it is not just a technical application**: basic science is involved as well in designing new techniques and new structures with improved functionalities

(Nano)technology is strictly connected with basic science, but **it is not just investigation/interpretation** of processes in the nano-world

[concepts from M.Wilson et al., Nanotechnology (Chapman&Hall, 2002)]

Components of nanotechnology

Nanotechnology shares topics with other disciplines,
but it **should not be confused** with:

-Chemistry, for the higher *control* of the involved processes;

-Materials science, for the specific interest in the “small world”;

-Physics, for the complexity of the systems under investigation;

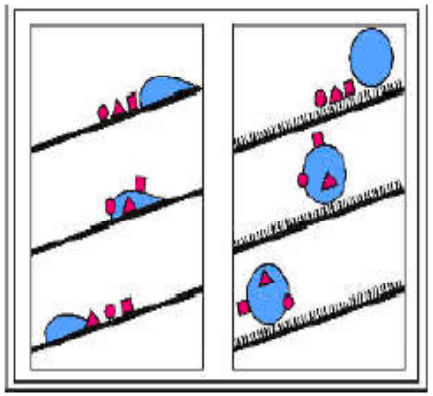
-Engineering, for the specific interest in new systems;

-Biophysics, (self assembly and replication) for the *artificial* systems

Nanotechnology is an “open” and strongly interdisciplinary field

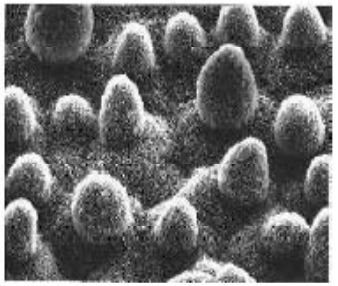
Nanotechnology in the natural world I

10 mm

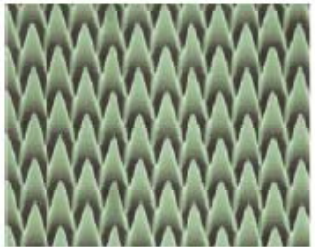


W. Barthlott, Univ. of Hamburg

On a smooth surface the contaminating particles are only moved by the water droplet (left). In contrast to that, on a rough surface they stick to the droplet rolling off the leaf thus being washed off (right).



Epicuticular wax



(Source: Metin Sitti, CMU)

A functional property (hydrofobicity) depends on the structural surface arrangement at the nanoscale

Nanotechnology in the natural world II

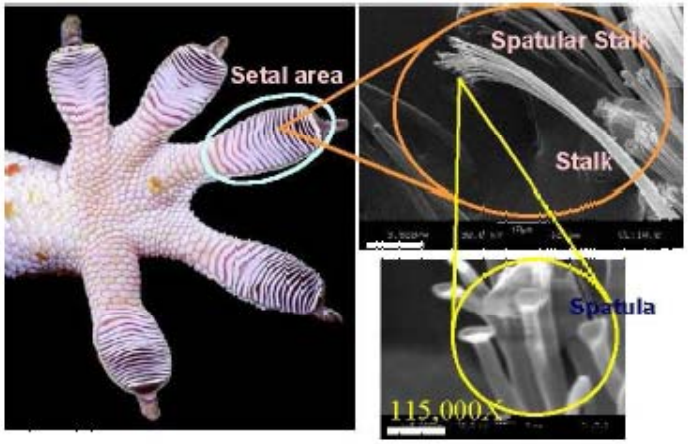
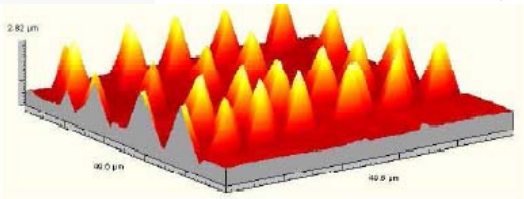


Figure 1: *Tokay* gecko foot-hair images: gecko foot bottom view (left image); zooming into one of the stalks (right upper image, bar indicates 10 μm), and zooming into spatulae and spatular stalks at the end of a stalk under SEM (right lower image, bar indicates 300 nm) (courtesy of Kellar Autumn).



(b)

Figure 3: 3-D AFM tapping mode image of (a) the AFM probe based indented flat wax surface, (b) molded and peeled off silicone rubber nano-hairs.

Synthetic Gecko Foot-Hair Micro/Nano-Structures for Future Wall-Climbing Robots

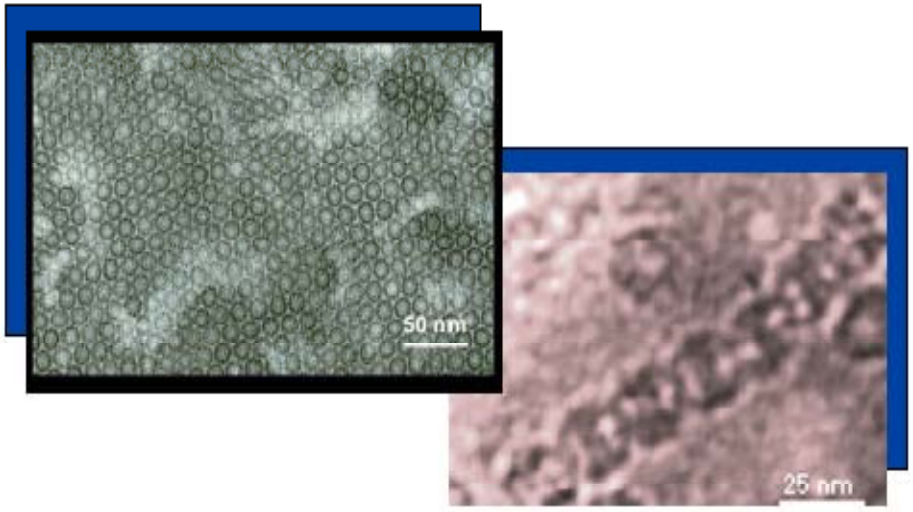
Metin Sitti[1] and Ronald S. Fearing[2]
[1] Dept. of Mechanical Engineering and Robotics Institute, Carnegie Mellon University, USA
[2] Dept. of EECS, University of California at Berkeley, USA

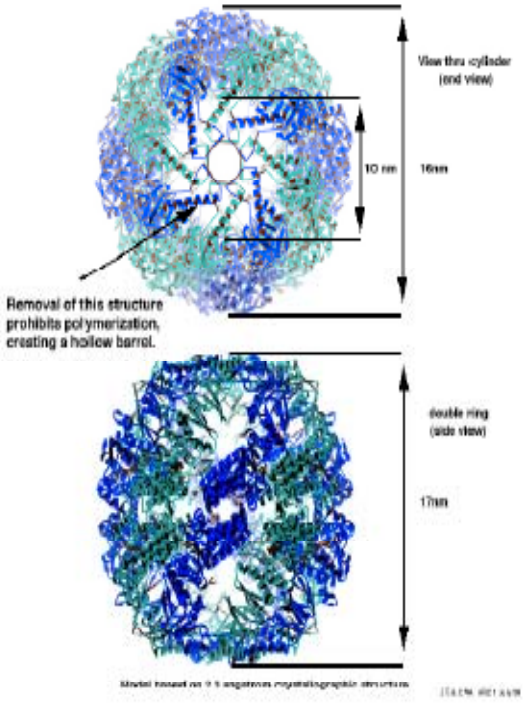
A functional/structural property (adhesion) depends on surface nanostructures (their artificial fabrication is under way)

Nanotechnology in the natural world III

- Heat shock protein (HSP 60) in organisms living at high temperatures ("extremophiles") is of interest in astrobiology

- HSP 60 can be purified from cells as a double-ring structure consisting of 16-18 subunits. The double rings can be induced to self-assemble into nanotubes.



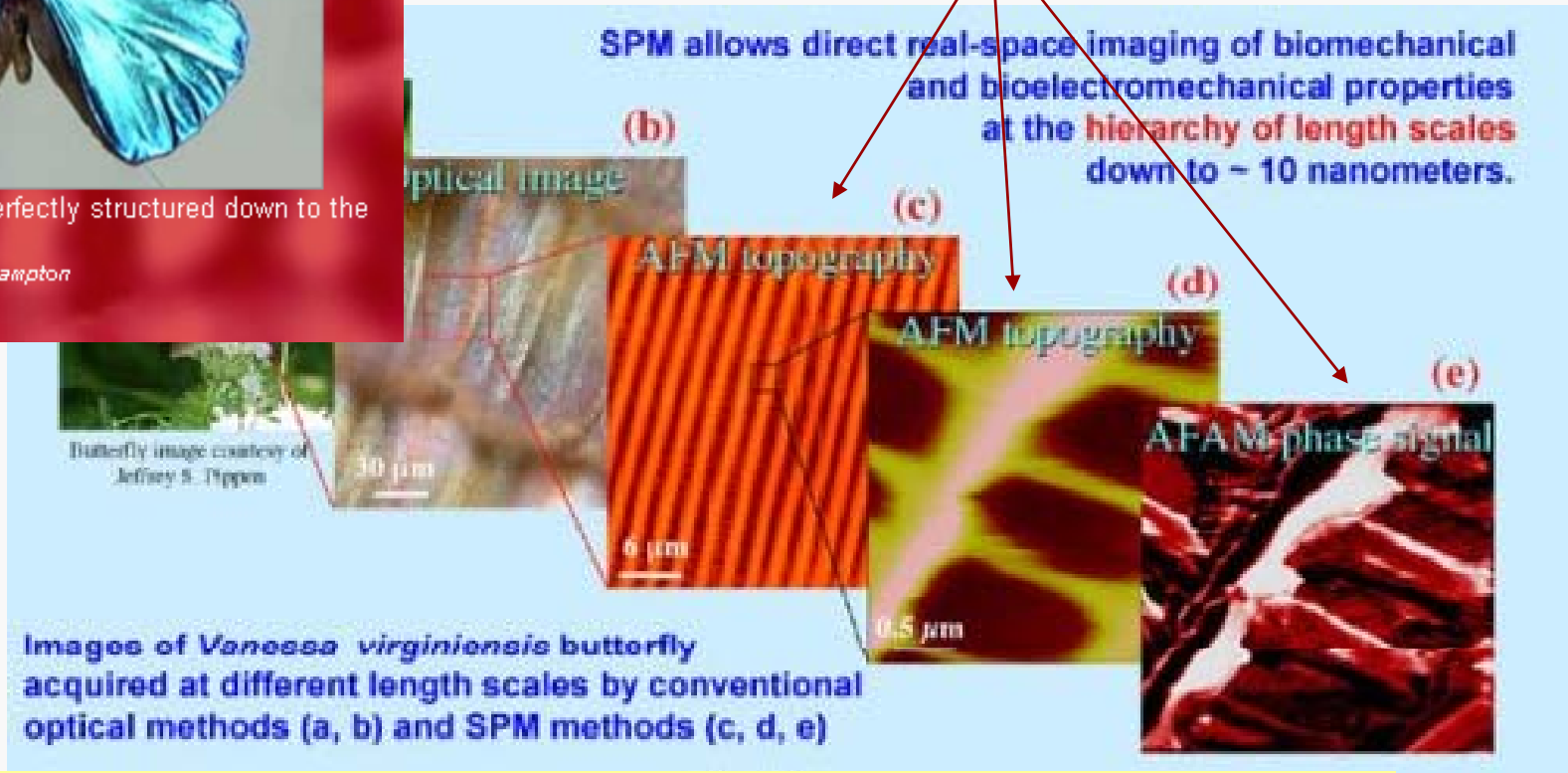


Termo-mechanical properties are enhanced when specific geometries are attained (at the nanoscale)

Nanotechnology in the natural world IV

Butterfly wings are perfectly structured down to the nano-scale.
Image: University of Southampton

Observations made possible thanks to the availability of new techniques



Images of *Vanessa virginiensis* butterfly acquired at different length scales by conventional optical methods (a, b) and SPM methods (c, d, e)

Very diverse features can be explained when considering the structure/morphology at the nanoscale

An historical example of nanotechnology

Lycurgus Cup in Roman times

Dr. Juen-Kai Wang

The glass appears green in daylight (reflected light), but red when the light is transmitted from the inside of the vessel.

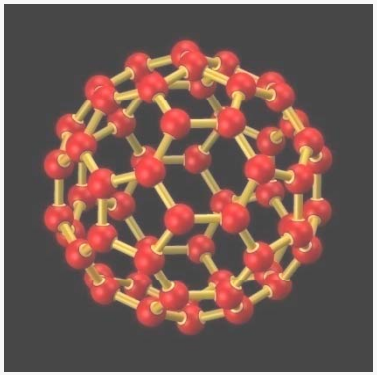
Interpretation:
"Nanostructured" glass (i.e., containing gold and silver nanoparticles)

The Lycurgus Cup, Roman (4th century AD), British Museum *(www.thebritishmuseum.ac.uk)*
*F. E. Wagner et al., Nature **407**, 691 (2000).*

A more recent example

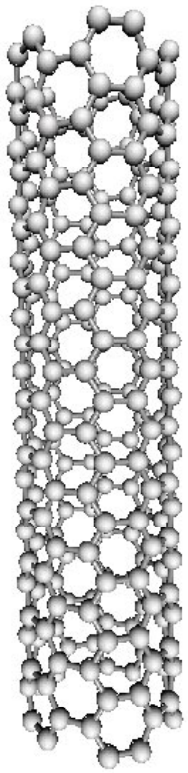
BASIC SCIENCE

NANOTECHNOLOGY



Fullerene (C_{60})

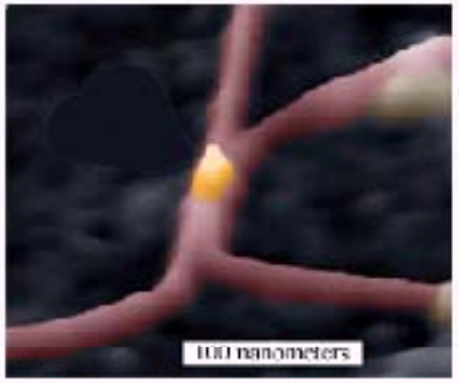
(Nobel Prize, mid 90's)



Single Wall
Carbon NanoTube
(90's)

Mesoscopic systems (interesting for their physico-chemical properties

An artificial system made of CNT and gold nanoparticle intended to be a prototypal single-electron device
(a couple of years ago)



There is plenty of room at the bottom…I

There's Plenty of Room at the Bottom

An Invitation to Enter a New Field of Physics

by Richard P. Feynman

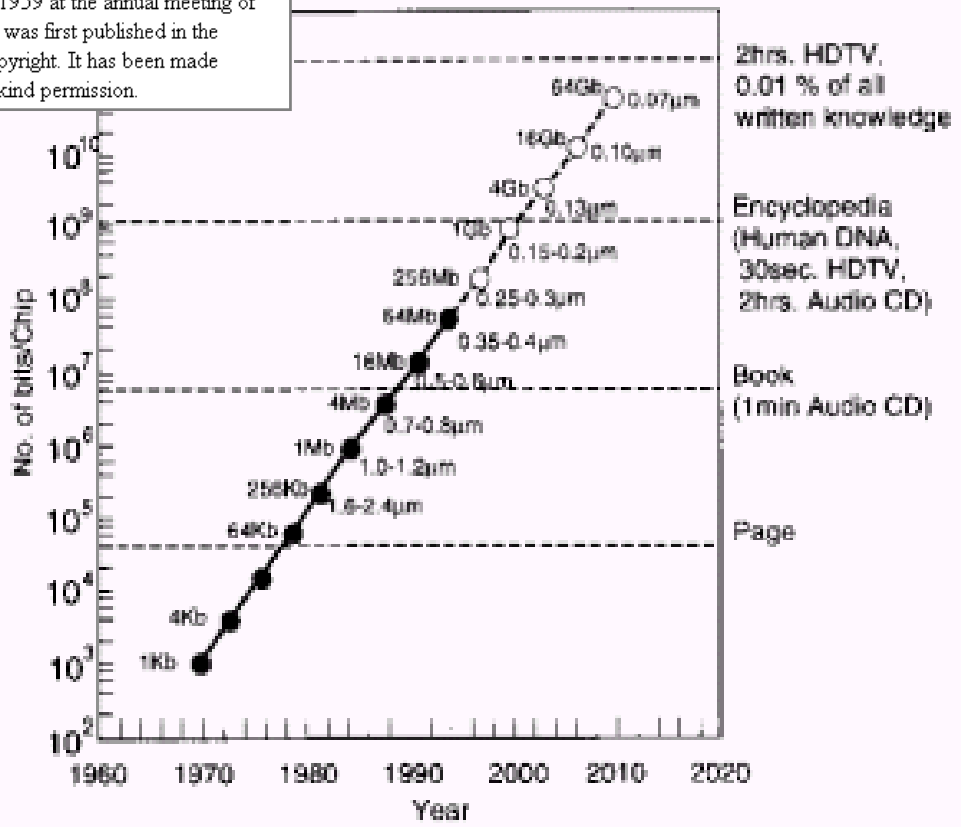
This transcript of the classic talk that Richard Feynman gave on December 29th 1959 at the annual meeting of the American Physical Society at the California Institute of Technology (Caltech) was first published in the February 1960 issue of Caltech's *Engineering and Science*, which owns the copyright. It has been made available on the web at http://www.zyvex.com/nanotech/feynman.html with their kind permission.

How do we write small?

Information on a small scale

Miniaturizing the computer

Miniaturization means increase of "power" in Information Technology

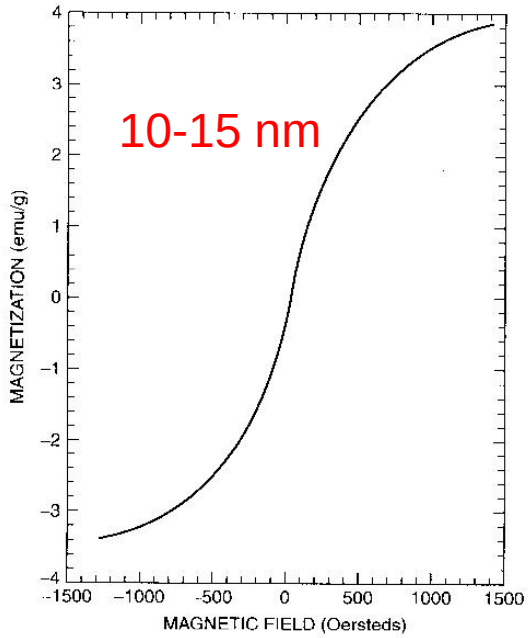


Increasing the information density: an example

Hard disk technology

In hard disks (magnetic), information is retained in the magnetization status of nanosized systems

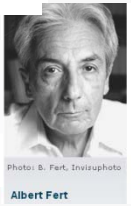
Reading small magnetic field variations requires sensitive systems based on nanostructured materials (GMR)

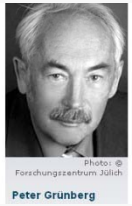


Magnetization vs magnetic field for a Co salt nanoparticle system

The Nobel Prize in Physics 2007

"for the discovery of Giant Magnetoresistance"


Albert Fert

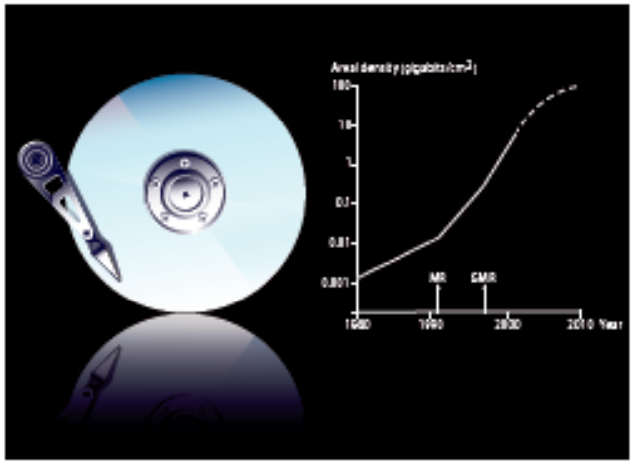

Peter Grünberg

The Nobel Prize in Physics 2007

This year's Nobel Prize in Physics is awarded to ALBERT FERT and PETER GRÜNBERG for their discovery of Giant Magnetoresistance. Applications of this phenomenon have revolutionized techniques for retrieving data from hard disks. The discovery also plays a major role in various magnetic sensors as well as for the development of a new generation of electronics. The use of Giant Magnetoresistance can be regarded as one of the first major applications of nanotechnology.

Better read-out heads for pocket-size devices

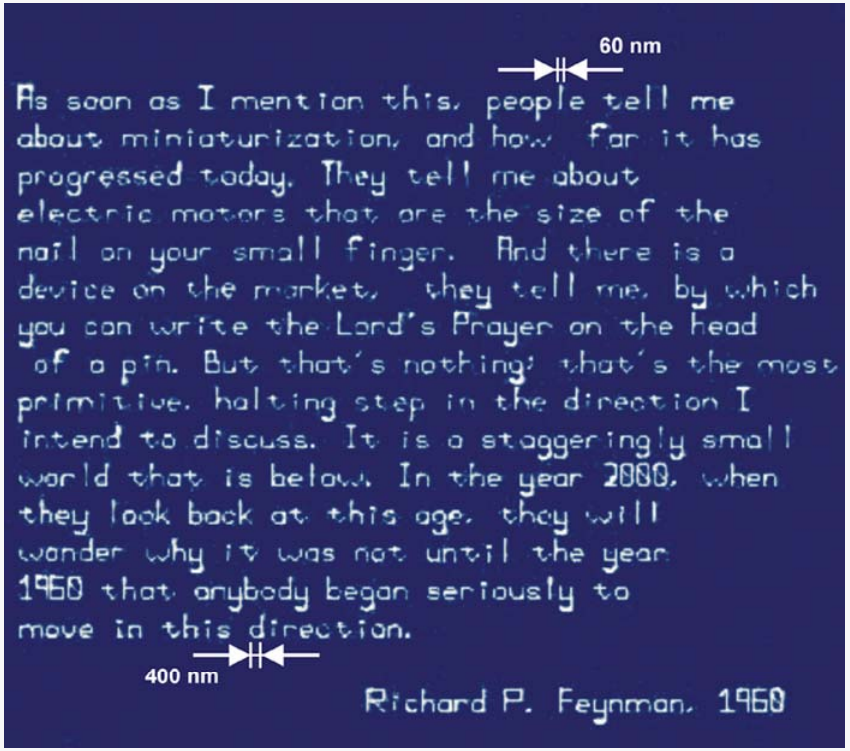
Constantly diminishing electronics have become a matter of course in today's IT-world. The yearly addition to the market of ever more powerful and lighter computers is something we have all started to take for granted. In particular, hard disks have shrunk – the bulky box under your desk will soon be history when the same amount of data can just as easily be stored in a slender laptop. And with a music player in the pocket of each and everyone, few still stop to think about how many cds' worth of music its tiny hard disk can actually hold. Recently, the maximum storage capacity of hard disks for home use has soared to a terabyte (a thousand billion bytes).



Diagrams showing the accelerating pace of miniaturization might give a false impression of simplicity – as if this development followed a law of nature. In actual fact, the ongoing IT-revolution depends on an intricate interplay between fundamental scientific progress and technical fine tuning. This is just what the Nobel Prize in Physics for the year 2007 is about.

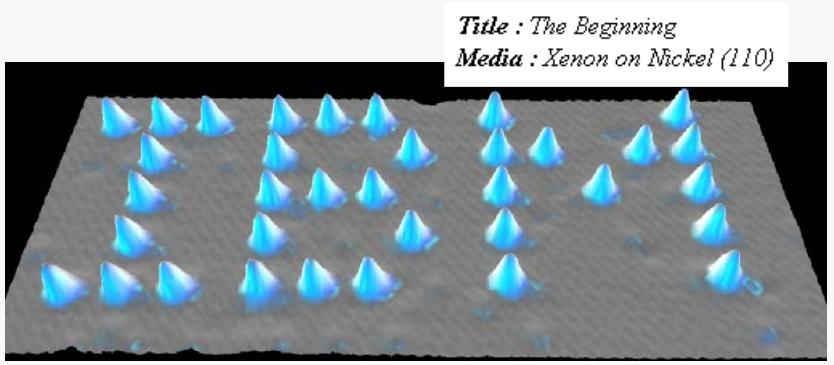
Towards less conventional implementations

Dip pen lithography



"Nanosized microfilm" displaying the initial part of the Feynman's speech

Single atom manipulation by STM



D.M. Eigler, E.K. Schweizer. Positioning single atoms with a scanning tunneling microscope. Nature 344, 524-526 (1990).

Strict interplay between basic science (e.g., fundamental phenomena occurring at the nanoscale) and applicative implementations

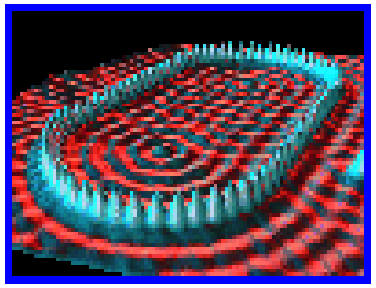
There is plenty of room at the bottom…II

Miniaturization by evaporation

Better electron microscopes

Atoms in a small world

Rearranging the atoms

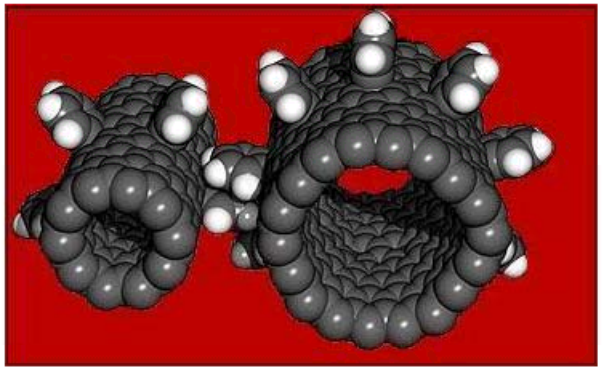


Title : *Stadium Corral*

Media : *Iron on Copper (111)*

IBM®

Miniaturization means new (*quantized*) functionalities exploitable in novel applications



A representation of nanogears made from graphitetubes billionths of a meter wide. (Picture from the NanoGallery, see references)

4 • NANOTECHNOLOGY

'nanotechnology is the principle of atom manipulation atom by atom, through control of the structure of matter at the molecular level. It entails the ability to build molecular systems with atom-by-atom precision, yielding a variety of nanomachines.'

Eric Drexler (1990)

Manipulation and control of the matter at the single atom level

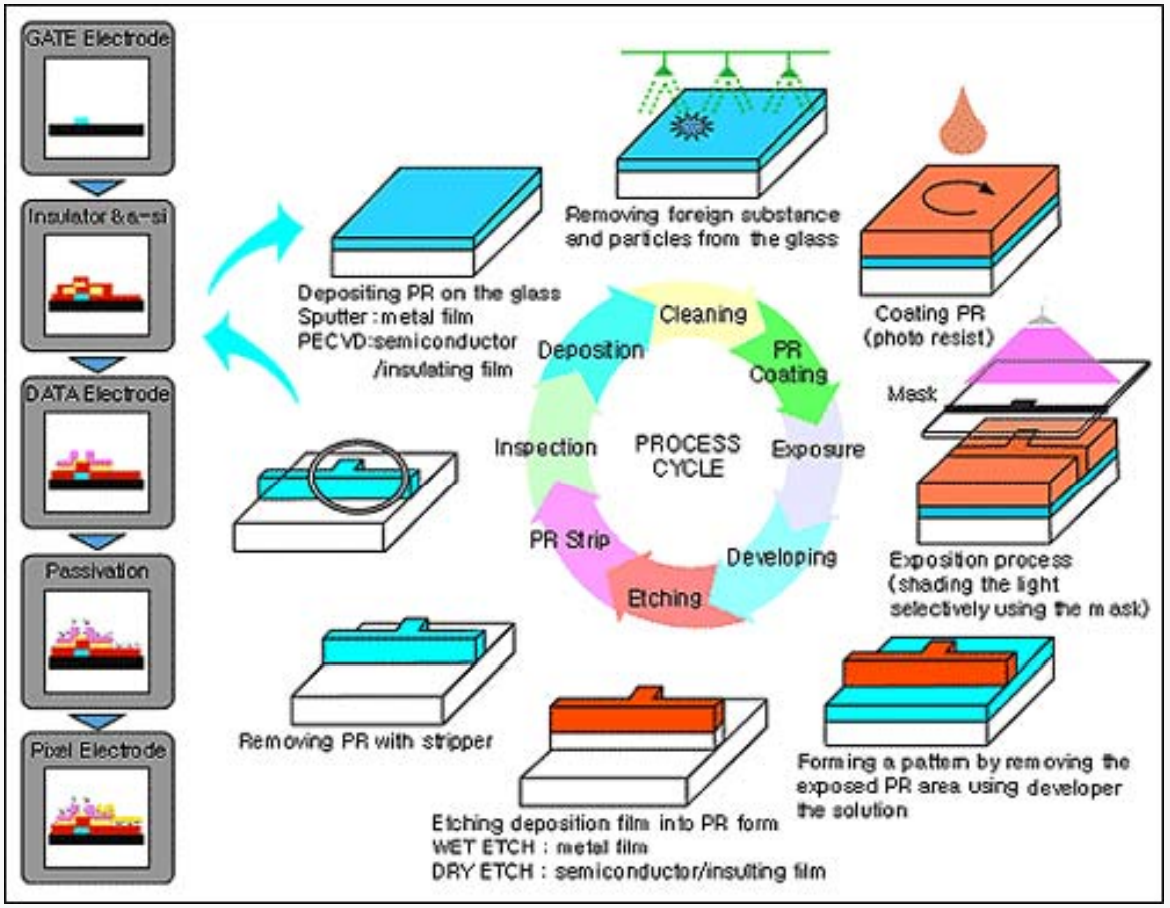
Nanomachines for, e.g., computation, drug dispensing, nanofluidics, …

Driving forces for nanotechnology I

Electronics devices:
they are typically (and *traditionally*) made of "small" structures

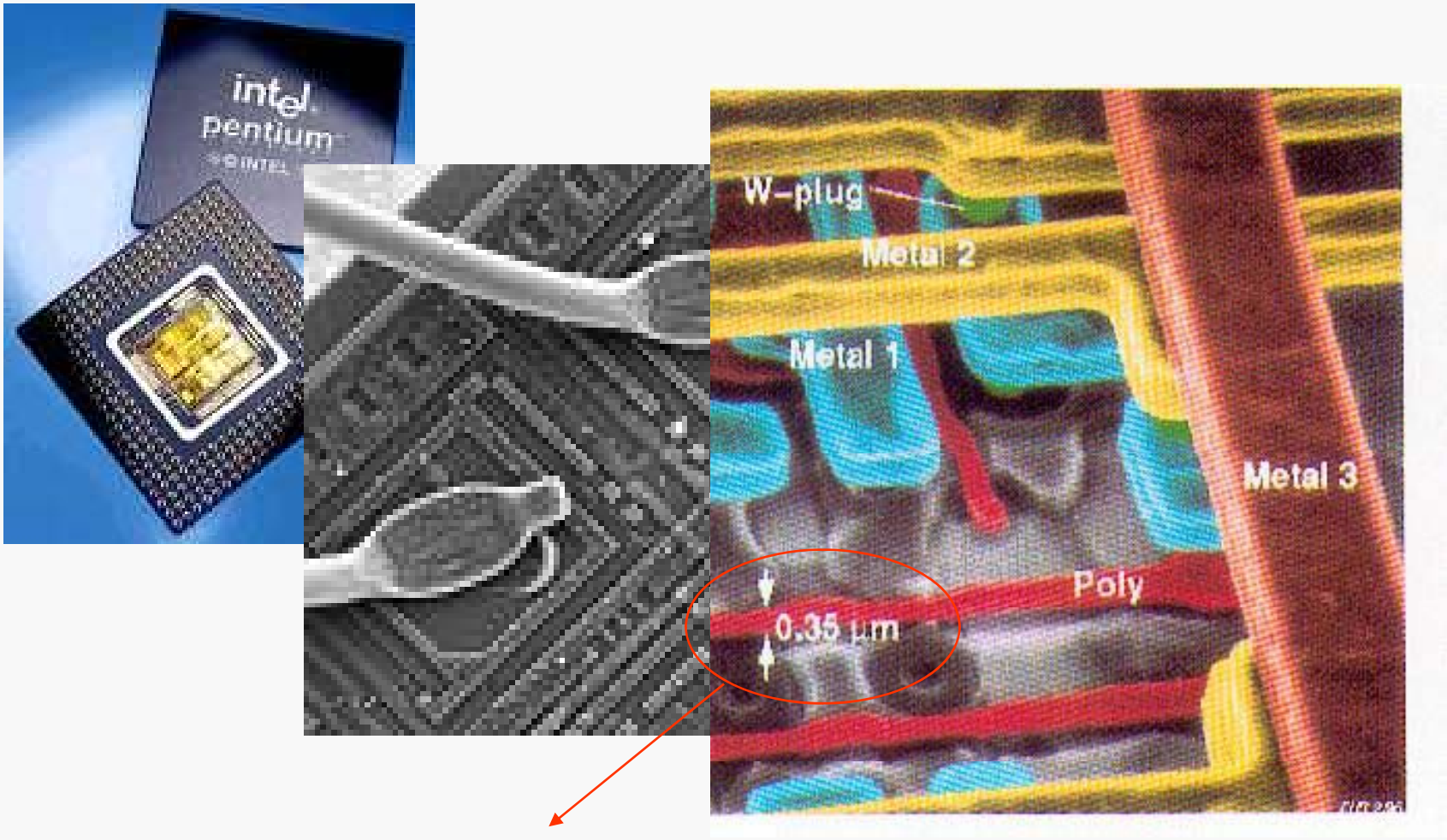
1. Thin films are deposited
2. A pattern is tranferred to the multilayered structure

Device components (resistors, capacitors, transistors, …) are so defined in an *integrated* structure



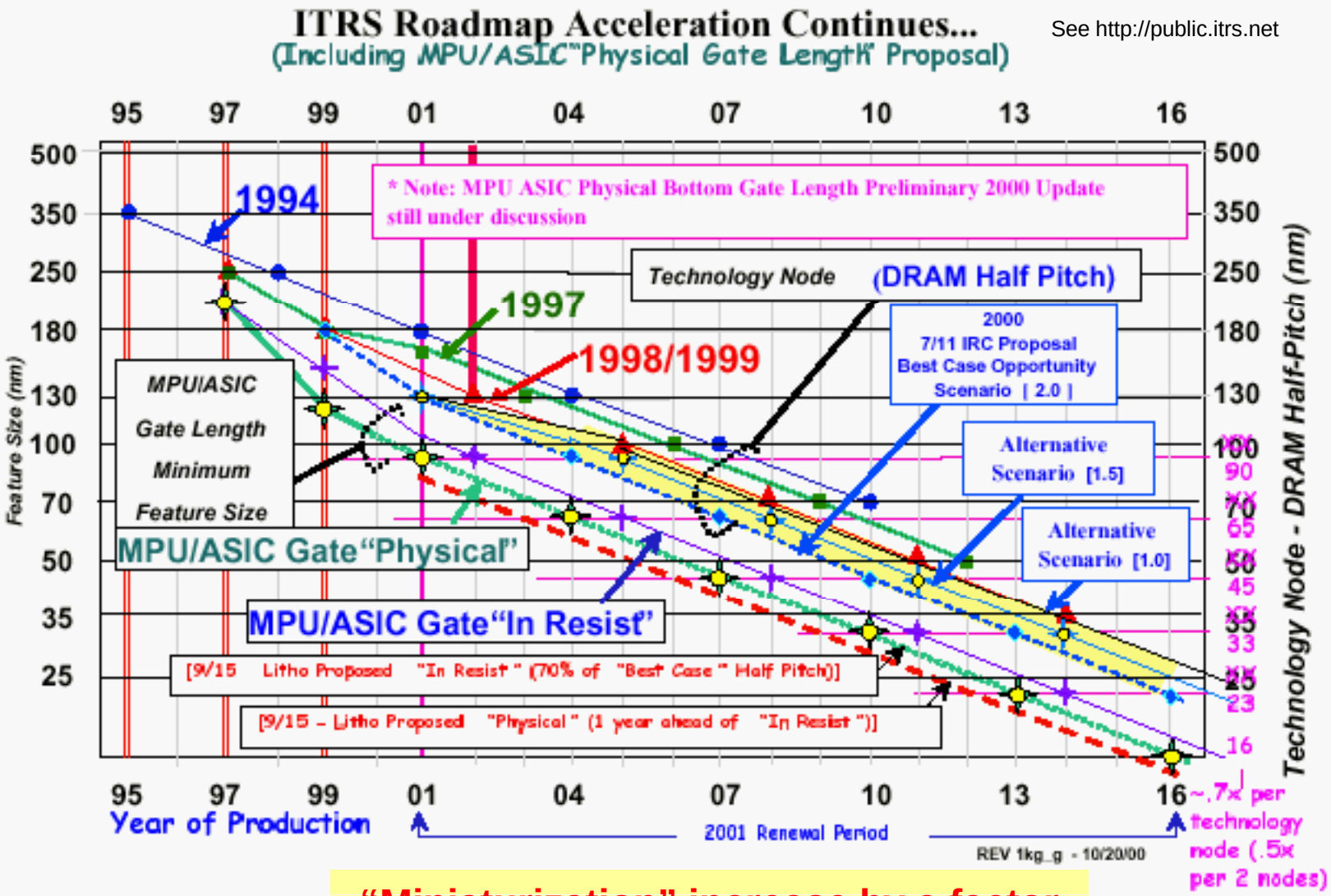
Traditionally, a *top-down* approach is adopted

Driving forces for nanotechnology II



Feature size
(typ., fwhm of the smaller device features)

The “Moore’s law”



“Miniaturization” increase by a factor 3-4 in a 3-4 year period (still true?)

The 2004 status of miniaturization (commercial)

How many transistors can dance on the head of a chip only 66 millimeters square? Over 58 million, thanks to IBM's sophisticated process technology that builds them just 90 nanometers wide. Such superior technology developments turbo-charge the G5 processor to speeds of up to 2.5GHz.

To get electronics so small requires miniaturization breakthroughs, and IBM's dedication to basic scientific research makes these advances possible. For instance, the company began researching copper as an interconnect method over 25 years ago, but the technique wasn't practical until just recently.

One in 58 Million. A transistor just 90nm wide (yellow) on substrate of SOI (blue) with copper interconnects (gray). Layers of nitride (brown) and oxide (green) insulate it from its brethren. Magnified 146,000 times.

So Small

Transistors on the PowerPC G5 hold a charge to let the system make logic decisions based on whether the transistor is on or off. Using a 90nm process for even greater performance, IBM builds these devices just .00000009 meters wide on a layer of silicon on insulator. The 58 million transistors themselves are connected by over 400 meters of copper wire that's less than 1/1000th the width of a strand of your hair. Tiny paths mean less time to complete a sequence, since the

http://www.apple.com

Feature size below 100 nm (nanoelectronics)

The very present (2008) status of miniaturization (commercial)

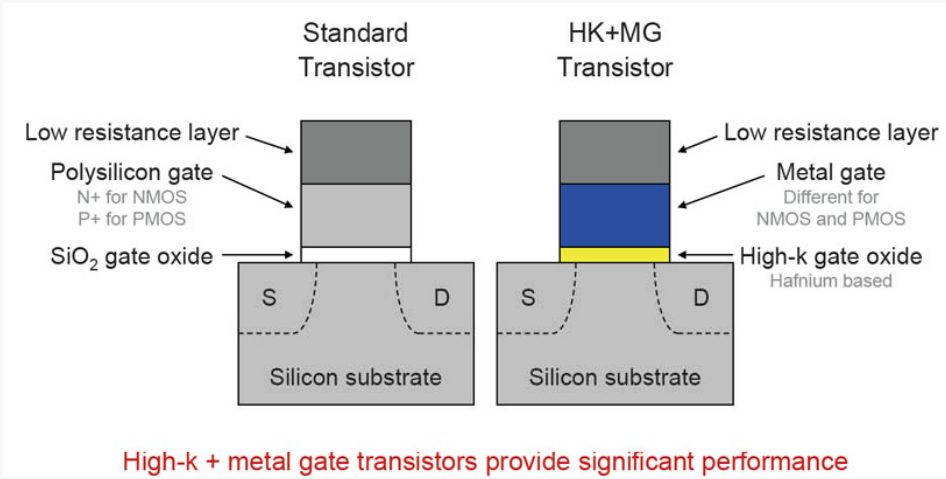
45 nanometer

From Wikipedia, the free encyclopedia

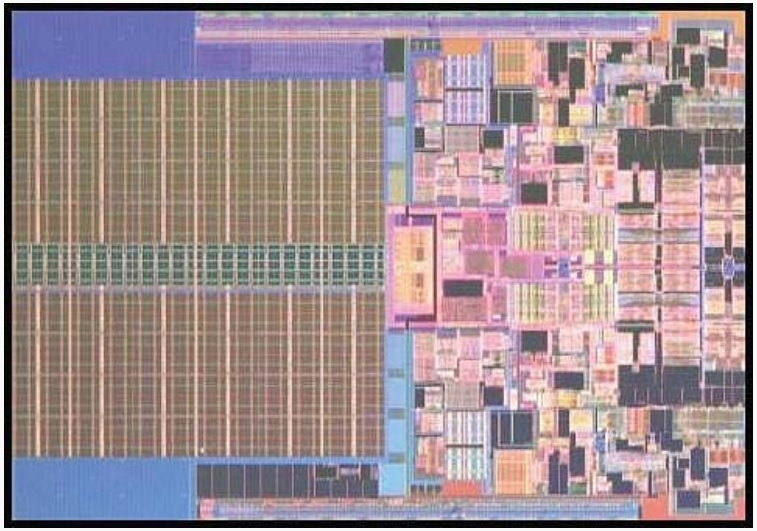
Per the International Technology Roadmap for Semiconductors, the **45 nm** technology node should refer to the average half-pitch of a memory cell manufactured at around the 2007-2008 time frame.

Matsushita and Intel started mass producing 45 nm chips in 2007, and AMD is targeting 45 nm production in 2008, while IBM, Infineon, Samsung, and Chartered Semiconductor have already completed a common 45 nm process platform. By the end of 2008, SMIC will be the first China-based semiconductor company to move to 45 nm, having licensed the bulk 45 nm process from IBM.

Many critical feature sizes are smaller than the wavelength of light used for lithography, i.e., 193 nm and/or 248 nm. A variety of techniques, such as larger lenses, are used to make sub-wavelength features. Double patterning has also been introduced to assist in shrinking distances between features, especially if dry lithography is used. It is expected that more layers will be patterned with 193 nm wavelength at the 45 nm node. Moving previously loose layers (such as Metal 4 and Metal 5) from 248 nm to 193 nm wavelength is expected to continue, which will likely further drive costs upward, due to difficulties with 193 nm photoresists.



High-k + metal gate transistors provide significant performance



Skilled and smart fabrication and material processing

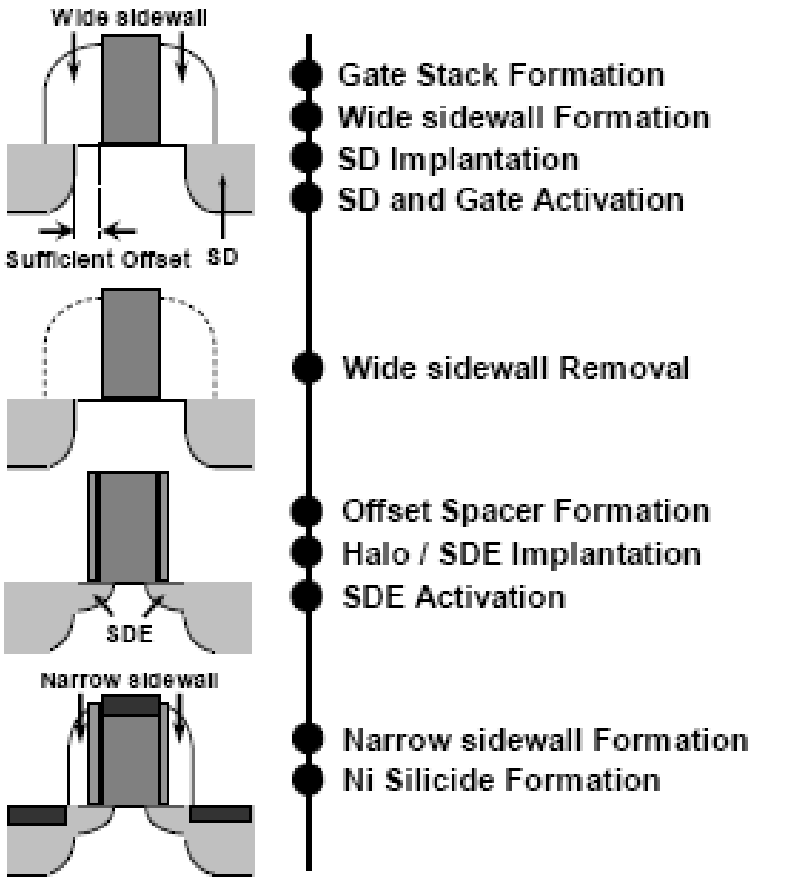


Fig.2 Process flow of reverse ordered SD/SDE formation. SD is formed by using wide sidewall which is disposed after SD formation. Narrow sidewall is formed for Silicide spacer after SDE formation.

High Performance CMOSFET Technology for 45nm Generation and Scalability of Stress-Induced Mobility Enhancement Technique

A.Oishi, O.Fujii, T.Yokoyama***, K.Ota***, T.Sanuki, H.Inokuma*, K.Eda*, T.Idaka*, H.Miyajima*, S.Iwasa*, H.Yamasaki*, K.Oouchi**, K.Matsuo*, H.Nagano*, T.Komoda, Y.Okayama, T.Matsumoto***, K.Fukasaku***, T.Shimizu*, K.Miyano*, T.Suzuki*, K.Yahashi*, A.Horiuchi***, Y.Takegawa, K.Saki*, S.Mori*, K.Ohno***, I.Mizushima*, M.Saito***, M.Iwai, S.Yamada, N.Nagashima*** and F.Matsuoka
System LSI Division, Semiconductor Company, Toshiba Corporation
*Process and Manufacturing Engineering Center, Semiconductor Company, Toshiba Corporation
**SoC R&D Center, Semiconductor Company, Toshiba Corporation
***Semiconductor Solutions Network Company, SONY Corporation
E-mail: ooishi@semicon.toshiba.co.jp
8 Shinsugita-cho, Isogo-ku, Yokohama 235-8522, Japan.
Phone: +81-45-770-3498 Fax: +81-45-770-3194

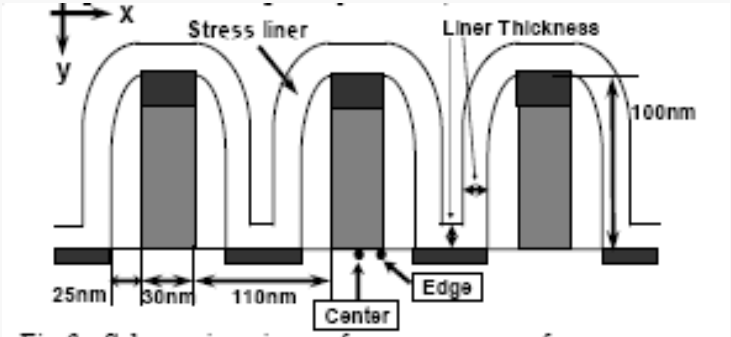


Fig.6 Schematic view of a structure for stress simulation. Minimum gate pitch which is often used for stacked gate circuit is assumed. It is assumed that stress liner deposition is conformal.

Striking results require combinations of different fabrication technologies

Technical and fundamental problems

- In order to maintain the progress of Moore's Law, the 2001 ITRS envisions **more aggressive scaling than projected in prior roadmaps**. For example, dynamic random access memory chips will feature critical dimensions of 90 nanometers in 2004, which is both smaller and sooner than the 100 nanometers projected for 2005 in the roadmap published just two years ago. Similarly, microprocessor transistor gate lengths – a critical dimension that affects the processor's speed – will be just 25 nanometers in 2007, six years sooner than expected in the 1999 version of the roadmap. (Note: a nanometer is one-billionth of a meter. A human hair is 100,000 nanometers in width, and a red blood cell is 5,000 nanometers in width.)

- **We are beginning to reach the fundamental limits of the materials used in the planar CMOS process**, the process that has been the basis for the semiconductor industry for the past 30 years. Further improvements in the planar CMOS process can continue for the next five to ten years by introducing new materials into the basic CMOS structure. However as the ITRS looks forward 10-15 years, it becomes evident that even with the introduction of new materials, most of the known technological capabilities of the CMOS device structure will approach or have reached their limits. In order to continue to drive information technology advances, it becomes necessary to investigate new devices that may provide a more cost-effective alternative to planar CMOS in this timeframe.

Da www.sia-online.org

The rate of increase in miniaturization has been growing fast in information technology

Main motivations:
- Increase of "power" (computing efficiency, information storage, time response, …)
- Decrease of power consumption, usually associated with miniaturization
- Commercial reasons (a huge market!)

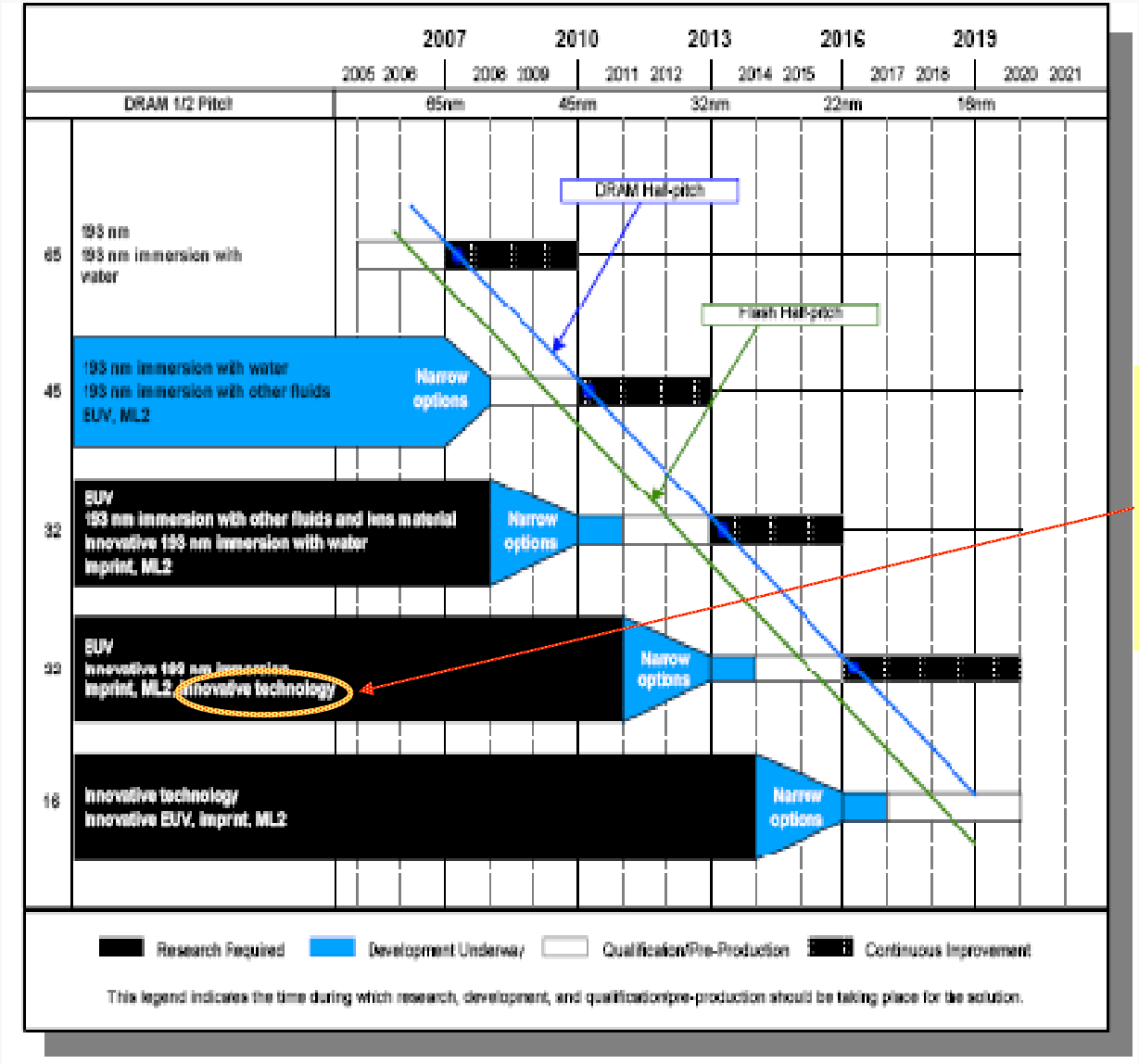
Technical limitations: lack of control in the manipulation, limitations inherent to the materials

Fundamental limitations: in the **techniques** (e.g., optical diffraction in lithography), in the **system operation** (e.g., *quantum* behavior)

Need for novel approaches

Need for new...

International Semicon Technol Roadmap



Alternative (new) technologies required simply to keep the pace of miniaturization

http://public.itrs.net

Our point of view I

Besides nanoelectronics, development of nanotechnologies is a crucial point in a huge variety of applicative areas, including, e.g.:
- improvement of mechanical/surface properties in coatings and structural materials (nanocomposites, …);
- realization of new and more efficient approaches for diagnostics (and even therapy) in biophysics, biomedicine (fluorescence markers, drug dispensers ,..);
- enhancement of data storage capabilities (DVD, hard-disks, …);
- enhancement of energy storage capabilities (fuel cells, …);
- realization of novel computation methods (quantum computers, …);
- design and fabrication of emitting devices (quantum-dot lasers, …);
- …

However, the “transition” from micro- to nanoelectronics appears as the most compelling and challenging area for:
- “historical” reasons;
- the inherent miniaturized nature of electronic devices;
- the ability to point out both limitations and new possibilities offered by extreme miniaturization

Our point of view II

Selected topics of interest

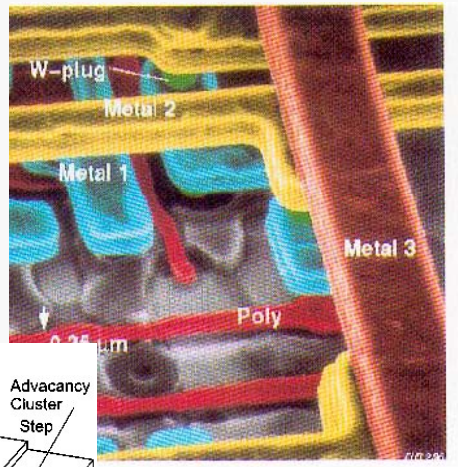
1. *A physical picture* of the **state-of-the-art** in (micro)technology for electronics

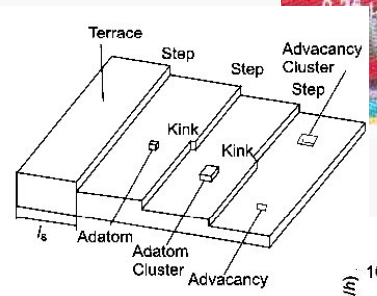
2. *Physical* methods for **fabrication** of nanostructures (i.e., evaporation, lithography, atom manipulation) and associated problems

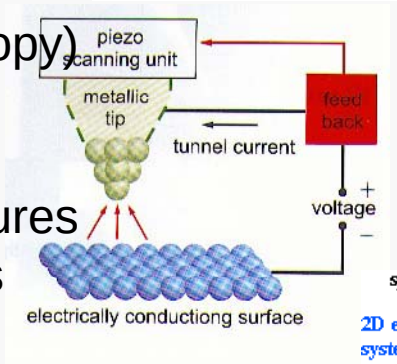
3. *Physical* tools for nanostructure **investigation** (i.e., probe microscopy) and new tools for fabrication

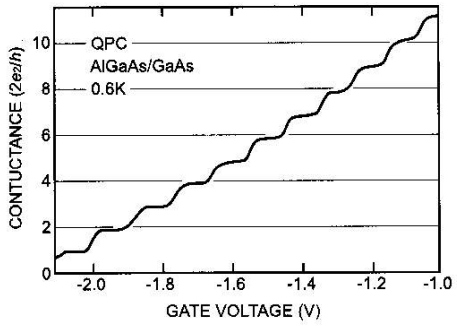
4. *Physical* **properties** of nanostructures (quantum confinement) and issues associated with miniaturization of electronic devices

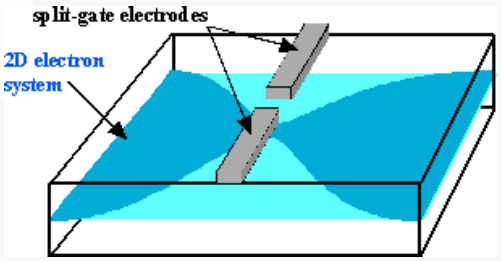
"Special" interest in *materials*











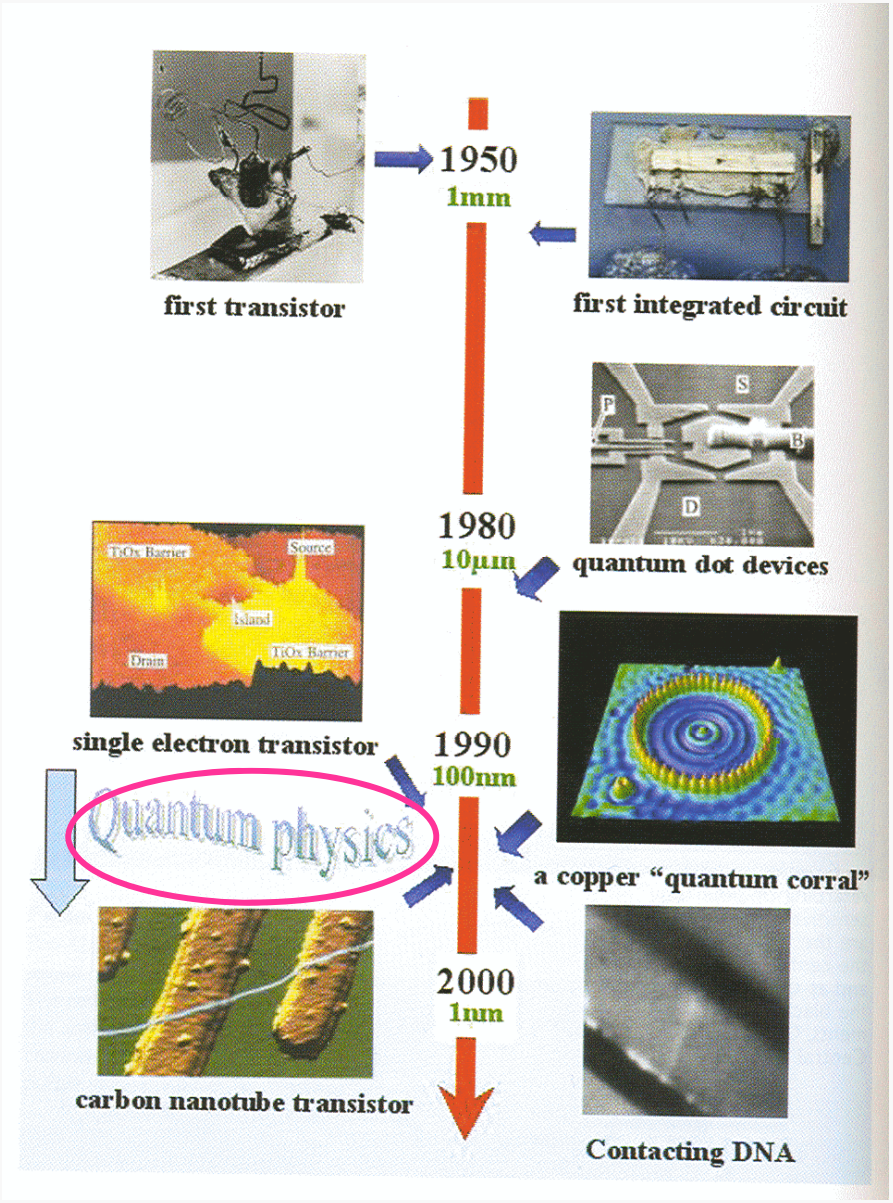
What is electronics made of

Electronics means the ability to control charge transport and requires:
- regions/materials where charge can move almost freely (e.g., interconnects, electrodes);
- regions/materials where charge motion is forbidden (e.g., dielectrics, capacitors);
- devices where charge motion is “controlled”

Conventional (micro)electronics achieves those tasks mostly by using inorganic materials (crystalline and/or amorphous conductors, semiconductors, dielectrics) and suitable architectures which, usually, can be (ideally) scaled down in size

Miniaturization may find osbtacles even when “conventional” (bulk-like) processes are concerned

Progress in “electronics”



Bipolar transistor

Planar (thin film) technology

Optical lithography

Very Large Scale Integration

Quantum confinement

Alternative materials
and technologies

???

The role of semiconductors in electronics

Most progresses experienced by electronics in the last decades rely on the availability of (solid-state) semiconductors

Semiconductor-based devices can be easily made to control charge flow, thus behaving as active devices

Silicon has been mostly used (along with binary/ternary solid-state mixtures, like GaAs) especially because of ease of fabrication (we will see some methods soon)

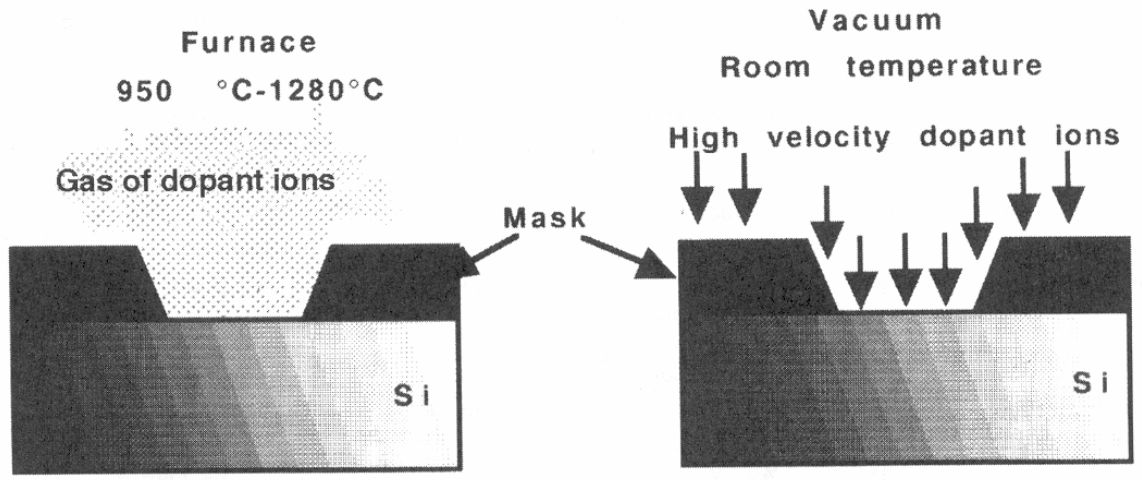
Presently, organics are investigated as “replacements” for silicon in many applications (we will see examples later on)

A fundamental feature of semiconductors used in electronics: possibility of (p- or n-) doping

Conventional doping techniques

Inter-crystal diffusion from vapor at high temperature

Ion bombardment at room temperature



Much more used, now!

Dopant uniformity and reproducibility	±5% on wafer, ±15% overall	±1% overall
Contamination danger	High	Low
Delineation	Refractory insulators and refractory metals, polysilicon	Refractory and nonrefractory materials, met
Environment	Furnace	Vacuum
Temperature	High	Low

Si-doping is a highly common technique which requires a mask for lateral definition
(a good example to start facing technological problems)

A few words on ion implantation

Ion Implantation

The principle method of doping today centers on high energy ion implantation. Implantation offers the advantage of being able to place any ion at any depth in the sample, independent of the thermodynamics of diffusion and problems with solid solubility and precipitation. Ion beams produce crystal damage that can reduce electrical conductivity, but most of this damage can be eliminated by annealing at 700 to 1000°C.

A beam of energetic ions 'implants' dopants into the substrate. Depth and dopant concentration are controlled by the acceleration energy and the beam current. The stopping mechanism of the ions involves nuclear collisions at low energy and electronic interactions at high energy.

The jargon associated with ion implantation includes:[25]

- Projected range R_p, i.e., average distance traveled by ions parallel to the beam
- Projected straggle, ΔR_p, i.e., fluctuation in the projected range
- Lateral straggle, $\Delta R//$, i.e., fluctuation in the final rest position, perpendicular to the beam
- Peak concentration, N_p, i.e., concentration of implanted ions at R_p

Dopant control at the atomic scale can be hardly achieved by conventional doping methods

The concentration profile, to a first order approximation, is Gaussian, i.e.,

$$N(x) = N_p \exp\left[-\frac{(x - R_p)^2}{2(\Delta R_p)^2}\right] \qquad 3.61$$

$$N_p = \frac{Q}{\sqrt{2\pi}\Delta R_p}$$

The range is determined by the acceleration energy, the ion mass, and the stopping power of the material. Orientation of the substrate surface away from perpendicular to the beam prevents channeling which can occur along crystal planes, ensuring reproducibility of R_p. Ion channeling leads to an exponential tail in the concentration vs. depth profile. This tail is due to the crystal lattice and is not observed in an amorphous material. The dose is determined by the charge per ion, zq; the implanted area, A; and the charge per unit time (current) arriving at the substrate. In other words:

$$\int i\,dt = Q \quad \text{and} \quad \frac{Q}{[zqA]} = \text{Dose (atoms/cm}^2\text{)} \qquad 3.62$$

The technique is now commonly used with penetration depths in silicon of As, P, and B typically being 0.5, 1, and 2 μm at 1000 kEV.

Thermal annealing at temperatures above 900°C is required to remove damage to the silicon lattice and to activate the implanted impurities. For deep diffusions (>1 μm), implantation is used to create a dose of dopants, and thermal diffusion (limited source) is used to drive in the dopant.

Ion beams can implant enough material to actually form new materials, e.g., oxides and nitrides, some of which show improved wear and strength characteristics.

M.Madou, Fundamentals of Microfabrication (CRC-Press, 1997)

Reminders on Fermi level in intrinsic semiconductors

Consider an intrinsic semiconductor, and let $n_c(E) = D_c(E)/V$ and $D_v(E)/V$ indicate the *density-of-states per unit volume in the conduction* and *valence bands*, respectively. At zero temperature, all the valence states are o and all the conduction states are empty, as schematically indicated in Fig. 2. A temperature T, a number of electrons from the valence bands are thermally exc the conduction bands; the occupancy probability of the allowed band structur of energy E is given by the Fermi–Dirac distribution function

$$f(E) = \frac{1}{e^{(E-\mu)/k_BT} + 1} ,$$

where μ is the chemical potential (the terms "chemical potential μ "and "Ferm E_F" are used by us as synonymous).

The density of electrons at temperature T in the conduction bands is given b expression

$$n_0(T) = \int_{E_c}^{\infty} n_c(E)\, f(E)\, dE = \int_{E_c}^{\infty} n_c(E)\, \frac{1}{e^{(E-\mu)/k_BT} + 1}\, dE ,$$

where the subscript 0 to the electron concentration is just to remind us tha quantity refers to thermal equilibrium. Similarly, the density of missing electro equivalently the density of holes) at temperature T in the valence bands is deter

ensity-of-states in the valence bands and the complementary to 1 of the Fermi- stribution function. We have the expression

$$(T) = \int_{-\infty}^{E_v} n_v(E)\,(1 - f(E))\, dE = \int_{-\infty}^{E_v} n_v(E)\, \frac{1}{e^{(\mu-E)/k_BT} + 1}\, dE . \qquad (2b)$$

grals in Eqs. (2) extend (in principle) to the whole energy regions where $n_c(E)$ E) are different from zero; in practice, as seen below, only the energy regions the band edges E_c and E_v are of relevance.

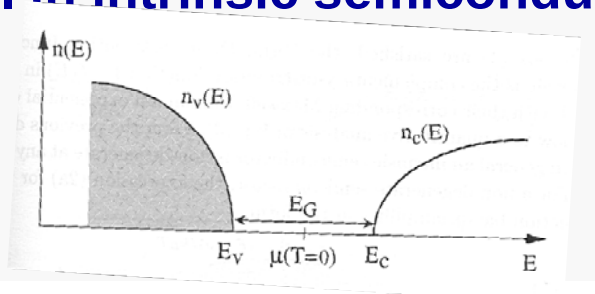
intrinsic semiconductor, the chemical potential $\mu(T)$ is determined by the ent that the number of electrons in the conduction bands equals the number left in the valence bands

$$n_0(T) = p_0(T) . \qquad (3)$$

olving explicitly the balance equation (3), we make the following considera- nce the distribution function $f(E)$ is a step function around $E = \mu$ at zero ture, and at the same time $n_0(T=0) \equiv p_0(T=0) \equiv 0$ in an intrinsic semicon- we have from Eqs. (2) that the chemical potential must *lie within the energy* ero temperature, i.e. $E_v < \mu\,(T=0) < E_c$. Furthermore, in the particular t the density-of-states of the semiconductor is symmetric with respect to the of the energy gap, the balance of electrons and holes obviously requires that ical potential, at any temperature, coincides with the middle of the band gap. he density-of-states in the valence and conduction bands (in the energy range k_BT around the band edges) are not specular, it is evident that a small shift hemical potential of order k_BT (towards the edge with lower density-of-states) ient to equalize the number of electrons and the number of holes.

miconductor (either intrinsic or extrinsic) is said to be *non-degenerate* if the l potential $\mu(T)$ lies within the energy gap and is separated from the band several k_BT (say $\approx 5\,k_BT$ or more); the *non-degeneracy conditions* are

$$< \mu(T) < E_c , \quad \text{with} \quad E_c - \mu(T) \gg k_BT \ \text{and} \ \mu(T) - E_v \gg k_BT . \qquad (4)$$



When conditions (4) are satisfied, the Fermi–Dirac distribution function $f(E$ Eq. (2a), as well as the complementary distribution function $1 - f(E)$ in Eq. (2b) be simplified with their corresponding Maxwell–Boltzmann exponential distribut

We pass now to a quantitative analysis of Eq. (3). From the previous discussio expect that in general an intrinsic semiconductor is non-degenerate at any temper of interest. For a non-degenerate semiconductor, the expression (2a) for the elec in the conduction bands simplifies in the form

$$n_0(T) = N_c(T)\, e^{-(E_c-\mu)/k_BT} ,$$

where

$$N_c(T) \equiv \int_{E_c}^{\infty} n_c(E)\, e^{-(E-E_c)/k_BT}\, dE .$$

Similarly, for a non-degenerate semiconductor, the expression (2b) for the holes le the valence bands takes the simplified form

$$p_0(T) = N_v(T)\, e^{-(\mu-E_v)/k_BT} ,$$

where

$$N_v(T) \equiv \int_{-\infty}^{E_v} n_v(E)\, e^{-(E_v-E)/k_BT}\, dE .$$

The quantities $N_c(T)$ and $N_v(T)$ are referred as the *effective conduction band* *valence band density-of-states*, respectively. Thus a non-degenerate semiconductor be schematized as a two-level system, where the whole conduction bands can be placed by a single level of energy E_c and degeneracy $N_c(T)$, and the whole val bands can be replaced by a single level of energy E_v and degeneracy $N_v(T)$.

The chemical potential in an intrinsic semiconductor is obtained by the requiren that expressions (5a) and (5c) coincide:

$$N_c(T)\, e^{-(E_c-\mu)/k_BT} = N_v(T)\, e^{-(\mu-E_v)/k_BT} ;$$

taking the logarithms of both members we have

$$\boxed{\mu(T) = \frac{1}{2}(E_v + E_c) + \frac{1}{2}\, k_BT\, \ln \frac{N_v(T)}{N_c(T)}} .$$

In an intrinsic (i.e., non-doped) non-degenerate (i.e., at moderate temp.) semiconductor, the Fermi level lies at mid gap energy

The effects of doping: reminders on Fermi level in doped semiconductors I

Fermi level and carrier density in doped semiconductors

Carrier concentration in n-type semiconductors

We consider now extrinsic semiconductors, containing donor impurities, or acceptor impurities, or both, and we wish to study their influence on the Fermi level and the free carrier concentrations. We consider first the case of semiconductors in which only donor impurities are present (n-*type semiconductors*). The density N_d of donor impurities is supposed to be uniform in the sample, and the binding energy of the donor levels is ε_d. The schematic representation of the energy levels and occupancy (at $T = 0$) is given in Fig. 7a.

In intrinsic semiconductors we have seen that the Fermi level lies (basically) at the middle of the energy gap (see Eq. 6). Doping with donors (or acceptor) levels is the most common method to change in a controlled way the position of the Fermi level within the energy gap. The presence of donor levels shifts the Fermi level from the middle of the energy gap toward the edge of the conduction band. Let us in fact define the temperature

$$k_B T_d \equiv \varepsilon_d \ ,$$

where T_d can be considered as the "ionization temperature" of the donor levels. If $T \ll T_d$ we expect that practically all donor levels are occupied and thus the chemical potential must be located in the energy range $E_d < \mu(T) < E_c$. If T is comparable with T_d we expect that most donor levels are ionized and $\mu(T)$ lies somewhat below the donor energy E_d, but still very near to the conduction band edge. At temperatures so high that the intrinsic carriers are much larger than the concentration of donor impurities, doping becomes uninfluential and we expect that the chemical potential approaches the middle of the bandgap. The chemical potential and the carrier concentration can be determined quantitatively from the knowledge of donor concentration,

G.Grosso and G.Pastori Parravicini, Solid State Physics (Academic, 2000)

density-of-states of the bulk crystal, and appropriate Fermi–Dirac statistics for band levels and donor levels.

The impurity states within the energy gap are described by localized wavefunctions; a donor level can this be empty, or occupied by one electron of either spin, but not by two electrons (of opposite spin) because of the penalty in the electrostatic repulsion energy. Due to this, the probability $P(E_d)$ that the level E_d is occupied by an electron of either spin is given by

$$P(E_d) = \frac{1}{(1/2)\, e^{(E_d-\mu)/k_B T} + 1} \ ; \qquad (19)$$

the above expression has been derived in Appendix III-C in the same way as the fundamental Fermi–Dirac statistics (1).

The chemical potential of the doped semiconductor is determined by enforcing the conservation of the total number of electrons as the temperature changes. In a semiconductor with N_d donor impurities per unit volume, the density $n_0(T)$ of electrons in the conduction band must satisfy the relation

$$n_0(T) = N_d\,[1 - P(E_d)] + p_0(T) \qquad (20)$$

where n_0 and p_0 are given by expressions (2). Eq. (20) is the straightforward generalization of Eq. (3); it states that the free electrons in the conduction bands are supplied by the thermal ionization of donor levels and by the thermal excitation of valence electrons. Eq. (20) can also be interpreted as an overall *charge neutrality condition* in the sample: the concentration n_0 of negative charges equals the concentration of ionized donor impurities plus the concentration of holes.

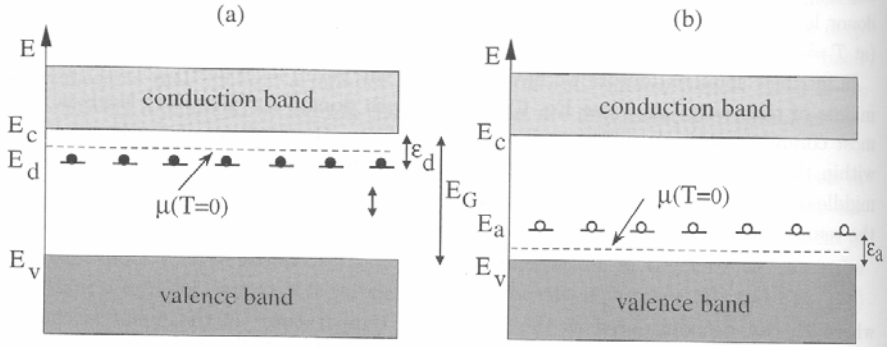


Fig. 7 (a) Schematic representation of the energy levels of a homogeneously doped n-type semiconductor at $T = 0$ (in abscissa any arbitrary direction in the homogeneous material can be considered). Typical energy values are $E_G = E_c - E_v \approx 1\,\text{eV}$ and $\varepsilon_d = E_c - E_d \approx 10\,\text{meV}$. The Fermi level at zero temperature lies at $(1/2)(E_d + E_c)$, which is the middle point between E_d and E_c. (b) Schematic representation of the energy levels of a homogeneously doped p-type semiconductor at $T = 0$; typical values of $\varepsilon_a = E_a - E_v$ are of the order of 10 meV. The Fermi level at zero temperature lies at $(1/2)(E_v + E_a)$, which is the middle point between E_v and E_a.

Reminders on Fermi level in doped semiconductors

Equation (20) can be solved (numerically) to obtain the Fermi level and hence the free carrier concentration. In the case the n-type semiconductor is non-degenerate (which is the ordinary situation, except for extremely high concentration of dopants), Eq. (20) can be simplified using Eqs. (5). We have:

$$N_c(T)\, e^{-(E_c-\mu)/k_BT} = N_d \frac{(1/2)\, e^{(E_d-\mu)/k_BT}}{(1/2)\, e^{(E_d-\mu)/k_BT} + 1} + N_v(T)\, e^{-(\mu-E_v)/k_BT} \,. \qquad (21)$$

This is a third order algebraic expression in $x = \exp(\mu/k_BT)$ that could be easily solved. We prefer to consider Eq. (21) in different regions of physical interest and handle it analytically.

(i) *Very low temperatures (or "freezing out region").* Consider the semiconductor at very low temperatures $T \ll T_d$. In this temperature region we certainly have

$$E_d < \mu(T) < E_c \,.$$

Thus the second term in the right hand side of Eq. (21) can safely be neglected; furthermore the denominator in the first term in the right-hand side of Eq. (21) can be taken as unity. We have thus

$$N_c(T)\, e^{-(E_c-\mu)/k_BT} = \frac{1}{2} N_d\, e^{(E_d-\mu)/k_BT} \,; \qquad (22a)$$

Low T →

taking the logarithm of both members we obtain for the Fermi level

$$\mu(T) = \frac{1}{2}(E_d + E_c) + \frac{1}{2} k_B T \ln \frac{N_d}{2\, N_c(T)} \,.$$

We can replace expression (22b) into equation (22a), and we obtain that the c… density in the conduction band is

$$n_0(T) = N_c(T)\, e^{-(E_c-\mu)/k_BT} = \sqrt{N_c(T)\frac{N_d}{2}}\, e^{-\varepsilon_d/2k_BT} \,.$$

Thus, the temperature dependence of the free electron carriers in n-type semicon… tors at temperatures $T \ll T_d$ has (approximately) the exponential form $\exp(-\Delta/k$… where Δ is *half the binding energy of the donor levels*. Notice that for high do… Eq. (22b) shows a tendency of $\mu(T)$ to increase and possibly to invade the conduc… band; in this situation we must consider directly the implicit equation (20) for … determination of the chemical potential.

(ii) *Saturation region.* Consider the semiconductor in the temperature region … $T \ll E_G/k_B$; we expect that (almost) all donor levels are ionized, while the ther… excitation of valence electrons is still negligible. We have

$$n_0(T) = N_c(T)\, e^{-(E_c-\mu)/k_BT} \cong N_d \,;$$

from the logarithm of both members, we have for the chemical potential

Intermediate T →

$$\mu(T) = E_c + k_B T \ln \frac{N_d}{N_c(T)} \,.$$

While the number $n_0(T)$ of majority carriers is essentially constant and equal N_d, … number of minority carriers is obtained by considering the mass-action law (7). In t… *saturation region*, characterized by all donor levels ionized, and at temperatures wh…

High T →

$n_i(T) \ll N_d$, we have

$$n_0(T) \cong N_d \quad \text{and} \quad p_0(T) \cong \frac{n_i^2(T)}{N_d} \,. \qquad (24c)$$

For instance, the intrinsic carrier concentration of silicon at room temperature is $n_i(T) \approx 10^{10}\,\text{cm}^{-3}$. In n-type silicon with donor concentration $N_d \approx 10^{14}\,\text{cm}^{-3}$, we have $n_0 \approx 10^{14}\,\text{cm}^{-3}$ and $p_0 \approx 10^{6}\,\text{cm}^{-3}$; in the above situation there are eight orders of magnitude in the difference between the concentration of majority carriers and of minority carriers. Notice also that in silicon $N_c(T) \approx 10^{19}\,\text{cm}^{-3}$, the chemical potential (24b) remains near the conduction band edge, but safely below it, so that the non-degeneracy conditions (4) are justified. As another example, consider an n-type GaAs crystal at room temperature with $n_i(T) \approx 10^{7}\,\text{cm}^{-3}$ and $n_0 \approx N_d \approx 10^{14}\,\text{cm}^{-3}$; in this case we have $p_0 \approx 1\,\text{cm}^{-3}$, a value fourteen orders of magnitude less than the majority carrier concentration.

(iii) *Intrinsic region.* If we increase further the temperature, the thermal excitation of valence electrons into the conduction band increases, and eventually the intrinsic situation is recovered. The temperature dependence of the density of free electron carriers in an n-type semiconductor is schematically summarized in Fig. 8.

Up to this point, impurities have been (tacitly) considered as isolated and independent; furthermore the doped semiconductor is assumed to remain non-degenerate, i.e. the Fermi level is several k_BT away from the band edges. As the concentration of dopants is increased new phenomena occur; for instance, the Fermi level may approach and invade the energy bands; the density-of-states of the semiconductor may be perturbed near the edges and a bandgap narrowing may result; the impurity levels may interact forming an impurity band, with effects on the conductivity of the sample; here, we do not enter in these and other interesting consequences of heavy doping in semiconductors.

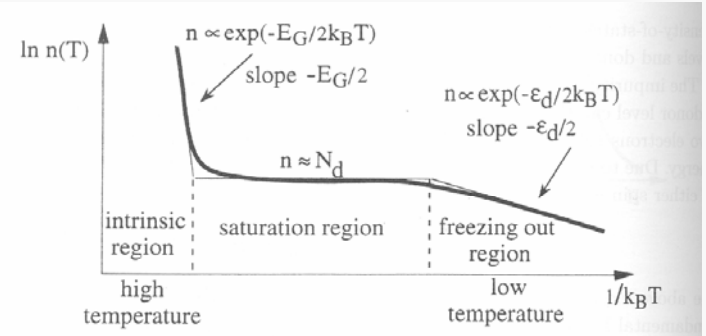
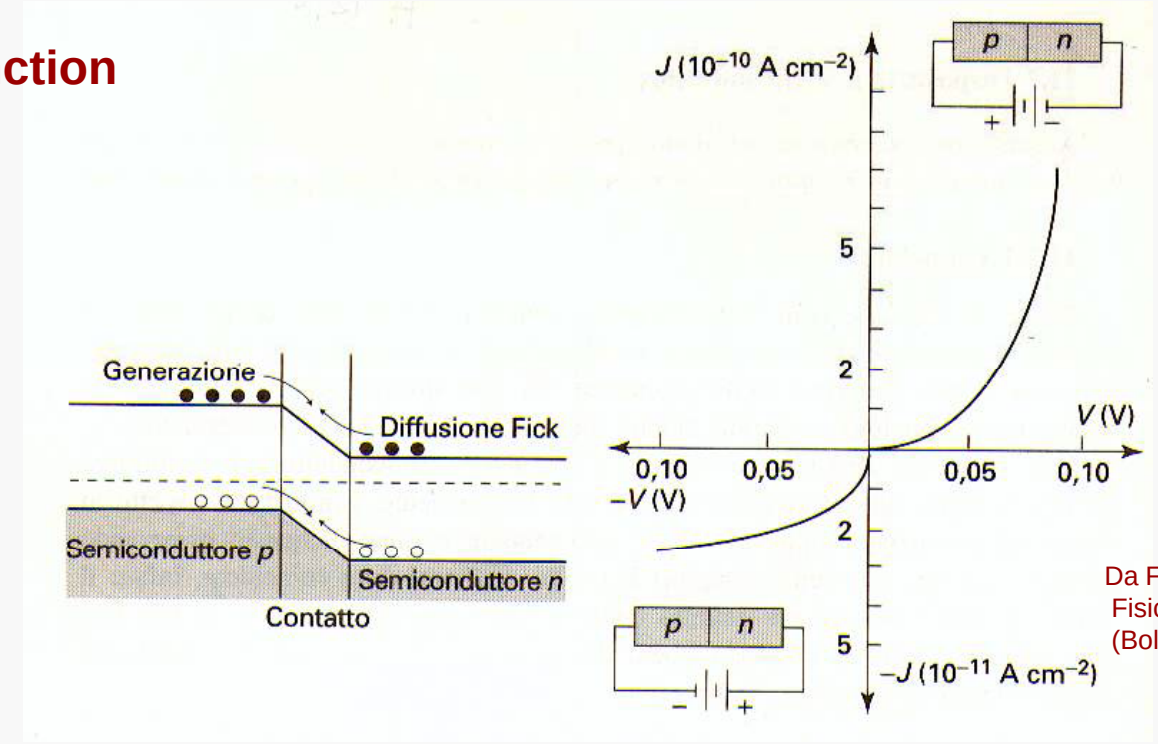


Fig. 8 Schematic variation of the electron concentration as a function of $1/k_BT$ in an n-ty… semiconductor with N_d donor impurities per unit volume.

Basics of *conventional* electronics I

Bipolar junction



Da F.Bassani, U.M. Grassano,
Fisica dello Stato Solido
(Bollati Boringhieri, 2000)

In the absence of an applied field, p and n charges are redistributed so to create a junction (a charge-free region similar to a capacitor)

$\Rightarrow$ the junction acts as a potential barrier for charges

$\Rightarrow$ transport is possible only when a direct polarization is applied

$\Rightarrow$ a rectifying behaviour is achieved

(note: a similar behaviour occurs also in metal/semiconductor – Schottky – junctions)

An historical look at the bipolar transistor

A current (e.g., BE) is used to control a current flow (e.g., CE)

Power consumption issues!

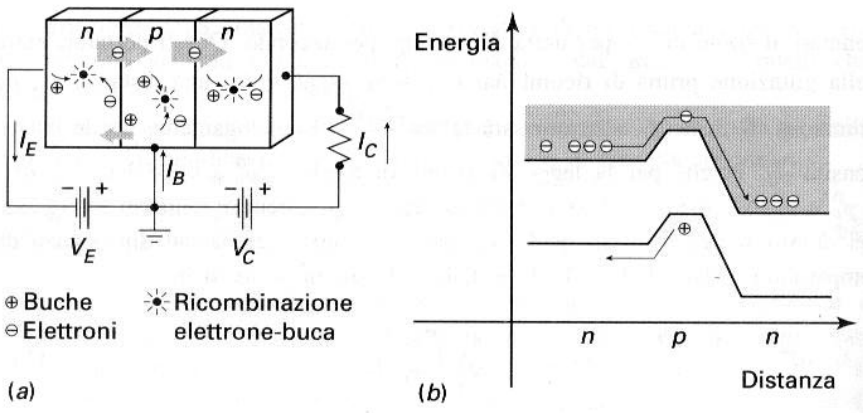


Figura 11.16
Transistor $n-p-n$, con i relativi simboli per indicare la corrente di emettitore (I_E), di collettore (I_C), e la corrente di base (I_B). (*a*) Indicazione degli stati di polarizzazione e dei flussi di corrente (buche ed elettroni). (*b*) Posizionamento delle bande in presenza di un campo (diretto per la giunzione $n-p$).

In un transistor $n-p-n$ si ha ad un estremo l'emettitore di elettroni, i quali entrano dal contatto nel semiconduttore n e all'altro estremo del secondo semiconduttore n vi è il collettore, mentre il semiconduttore p intermedio, molto più sottile degli altri, è chiamato base. All'equilibrio senza polarizzazione non si ha passaggio di corrente perché $I_g^0 = I_r^0$ a entrambe le giunzioni. Basta però applicare una differenza di potenziale tra il collettore e l'emettitore e controllare il potenziale della base per ottenere un'amplificazione di tensione. Illustriamo il funzionamento di tale transistor riferendoci alla fig. 11.16. In questo schema si hanno due circuiti. Uno è il circuito $e-b$ (emettitore-base) che è rettificante per le ragioni esposte precedentemente a proposito del diodo. L'altro è un circuito $b-c$ (base-collettore) che da solo lascerebbe passare poca corrente perché il potenziale è tale da aumentare la barriera di potenziale. In presenza del circuito precedente però molti più elettroni arrivano al semiconduttore p per l'effetto dell'abbassamento della barriera al confine $n-p$ e tali elettroni non trovano ostacoli a proseguire attraverso la zona n ed arrivare al collettore. Questo produce perciò amplificazione di potenza nel circuito $b-c$ rispetto al circuito $e-b$.

La corrente che passa per $n-p$ è I_e:

$$I_e = I_g^0\left(e^{\frac{eV_e}{kT}} - 1\right), \tag{11.64}$$

dove V_e è il potenziale dell'emettitore. La corrente che passa al collettore I_c sarà

$$I_c = I_e - I_b, \tag{11.65}$$

dove I_b, corrente di base, è piccola in ogni caso. Se la base è a terra si può ritenere $I_b \simeq 0$ e la corrente raccolta al collettore sarà data dalla (11.64). Non c'è in questo caso amplificazione di corrente tra emettitore e collettore, ma c'è grande amplificazione di tensione (o di potenza), perché la stessa corrente passa da un circuito d'ingresso a bassa impedenza (giunzione con polarizzazione diretta) ad un circuito d'uscita a grande impedenza (giunzione con polarizzazione inversa) e qui scorre attraverso una grande resistenza R_L. Dunque un transistor a base comune si comporta come un amplificatore di tensione (o di potenza).

Se il transistor è collegato con emettitore comune (a terra), si comporta come un amplificatore di corrente (vedi fig. 11.17). Come si è visto prima, quasi tutta la corrente I_e della giunzione emettitore-base (polarizzata direttamente) raggiunge il collettore, così si può scrivere

$$I_c = \alpha I_e \tag{11.66a}$$

Old-style technology

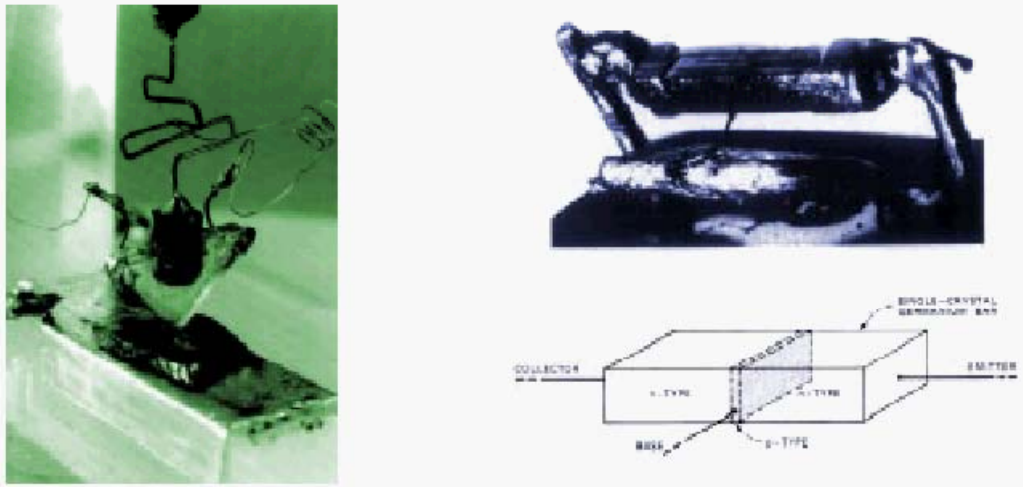


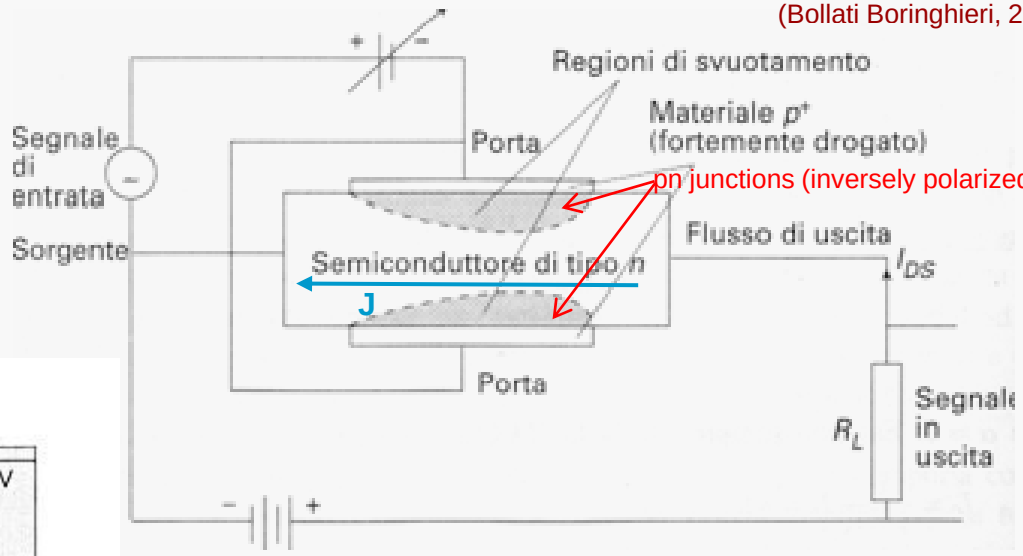
Figure 1 – The first transistors: (a) the point contact transistor of Brattain and Bardeen, 1947 (left); (b) the junction transistor of Shockley, Morgan, Sparks, and Teal, 1950 (right).

“Linearly shaped” technology did not allow for miniaturization

Basics of *conventional* electronics II

Da F.Bassani, U.M. Grassano,
Fisica dello Stato Solido
(Bollati Boringhieri, 2000)

Basic concept of a Field-Effect Transistor (FET)



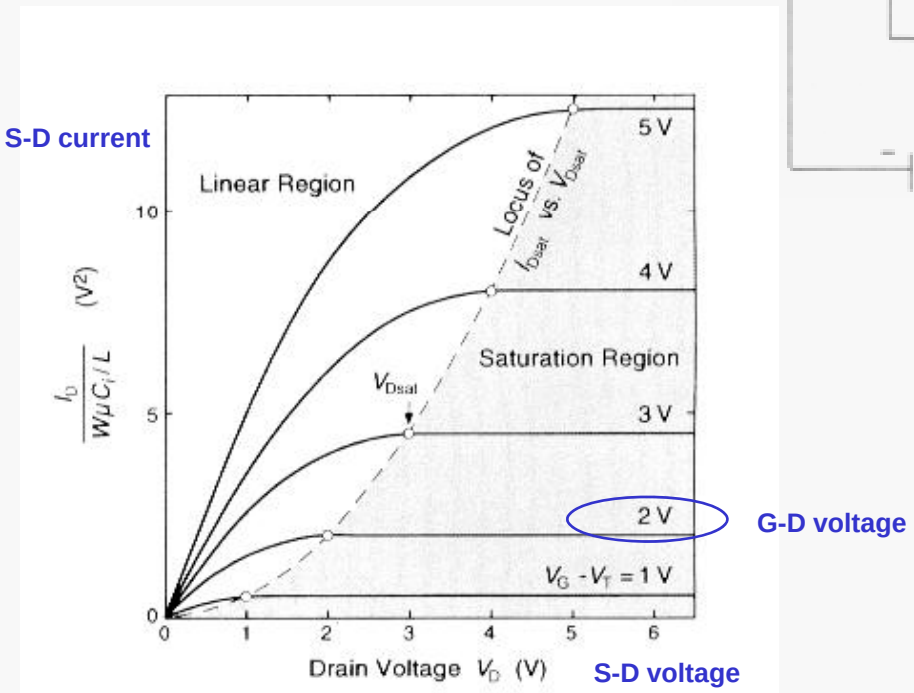


Figure 9: Idealized *I-V* curves of a MOSFET. The dashed line indicates the locus of I_{Dsat} vs. V_{dsat} [6].

The gate voltage is used to control the source-drain current (better behavior in terms of power consumption)

Towards planar technologies (’60s-‘70s)

Planar technology (thin film multilayers of different materials)

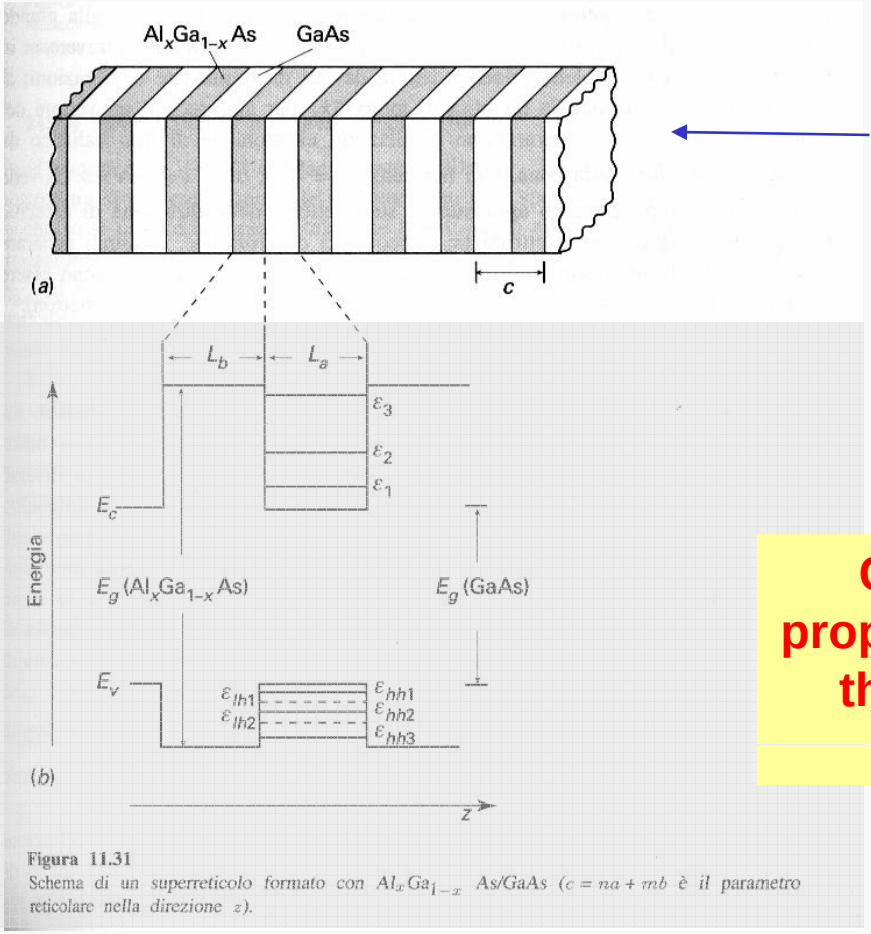


Figura 11.31
Schema di un superreticolo formato con $Al_xGa_{1-x}As/GaAs$ ($c = na + mb$ è il parametro reticolare nella direzione z).

Examples:

Semiconductor heterostructures

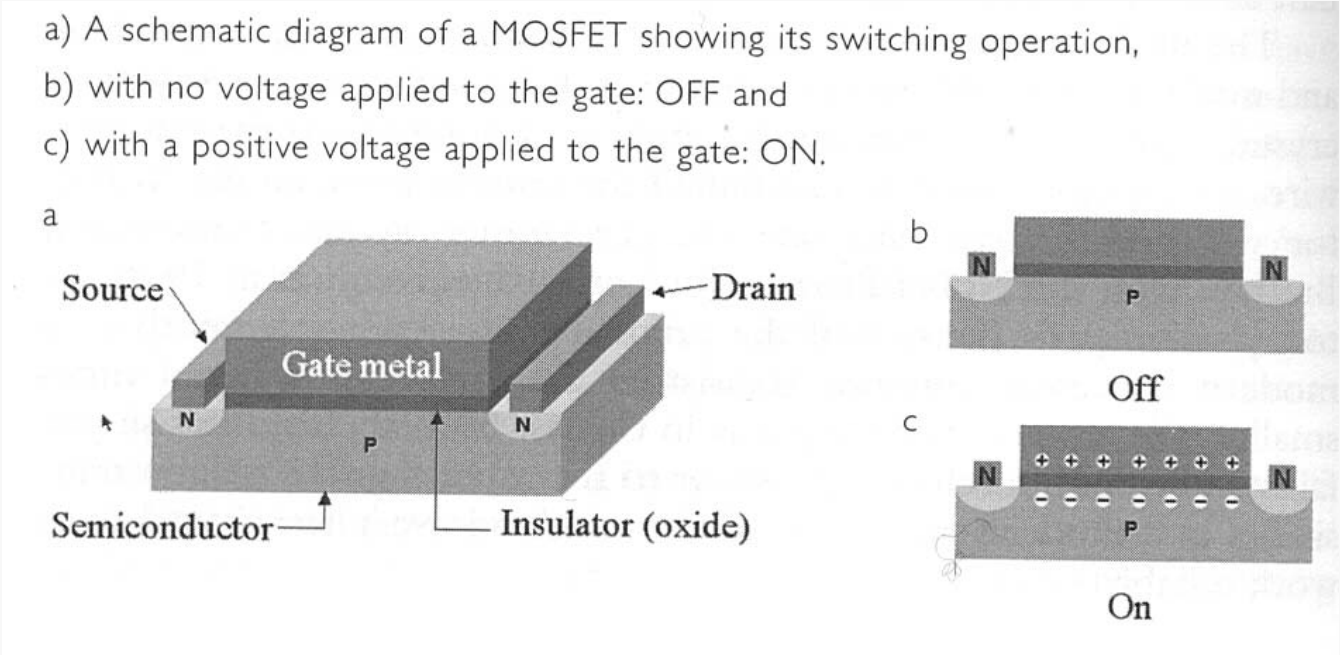
(Dielectric thin films

Thin multilayered structures)

Careful engineering of the material properties achieved by (one-dimensional) thickness control at the few (single!) atomic layer level

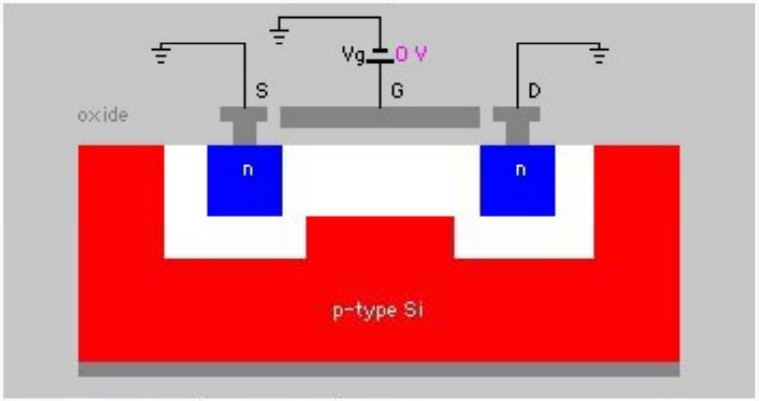
The MOS-FET I

MOS-FET architecture is compatible with planar technologies

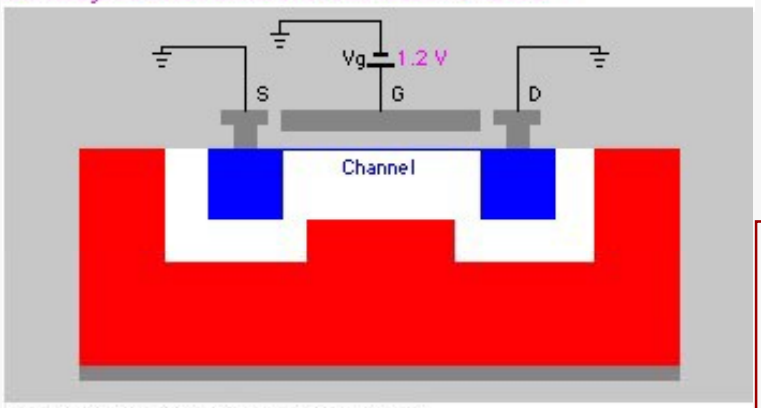


Key points:
- Metal-Oxide-Semiconductor (MOS) multilayer
- Thin dielectric (oxide) layer (E=V/d!!)
- Lateral definition of the structure
- Large scale integration possible

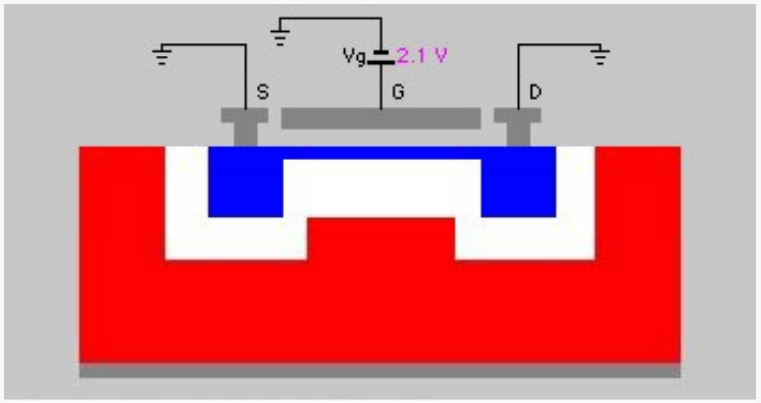
The MOS-FET II



Enhancement-mode (Normally-off) MOSFET
N-channel
Vg < Vt: gate bias is less positive than the threshold voltage.
Not enough electrons and no inversion channel is formed.



Enhancement-mode (Normally-off) MOSFET
N-channel
Vg > Vt: gate bias is more positive than the threshold voltage.
Sufficient electrons accumulate and forms the inversion channel.



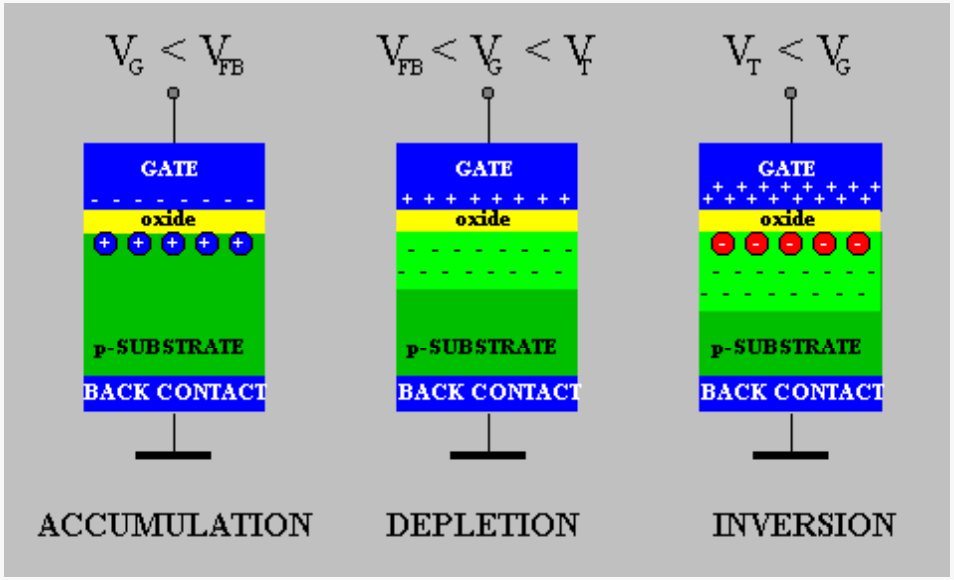
Enhancement-mode (Normally-off) MOSFET
N-channel
Vg > Vt: gate bias is more positive than the threshold voltage.
Sufficient electrons accumulate and forms the inversion channel.

In the inversion and depletion conditions the interface charge creates a channel for the transport from source to drain

MOS-FET requires additional electrodes (and a longitudinal field)

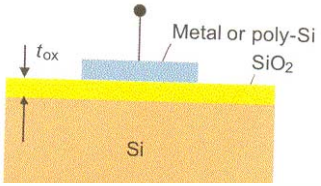
The MOS capacitor I

The electric field produced by applying a voltage to the metal (gate) rules the density of charge carrier at the semiconductor/oxide interface.
Example for p-doped semiconductor:
- accumulation: holes (positive carriers) are accumulated at the interface
- depletion: holes are depleted at the interface
- inversion: a thin layer of *almost free* electrons (negative carriers) is formed



http://people.deas.harvard.edu/~jones/es154/lectures/lecture_4/mosfet/mos_models/mos_cap/mos_cap.html

The MOS capacitor II: energy diagram



Applying an electric field:
- Fermi level is raised/depressed (depending on the sign)
- conductive/valence levels are “deformed” (band bending - depends on space)

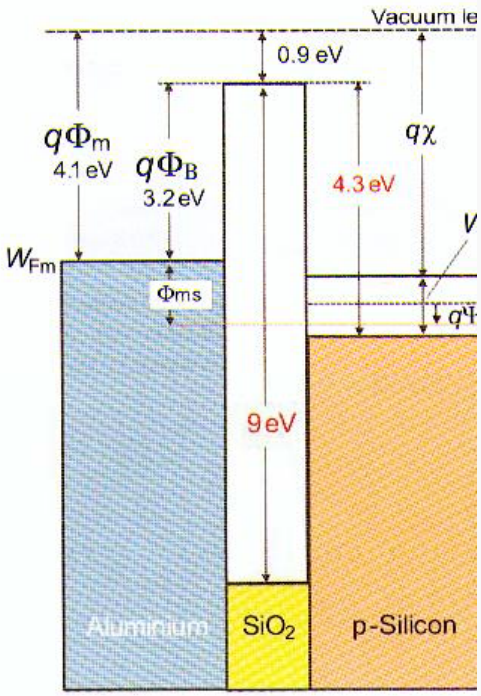
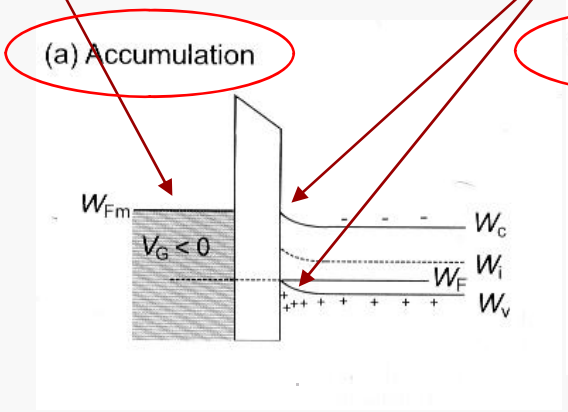
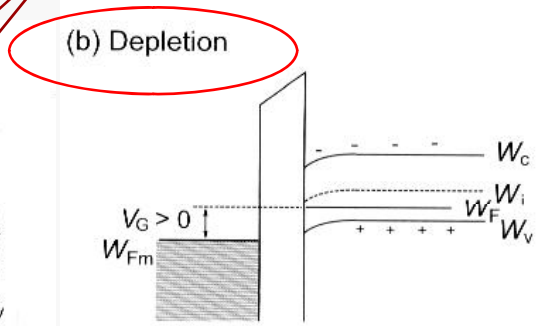
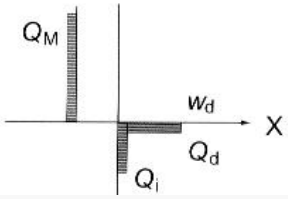
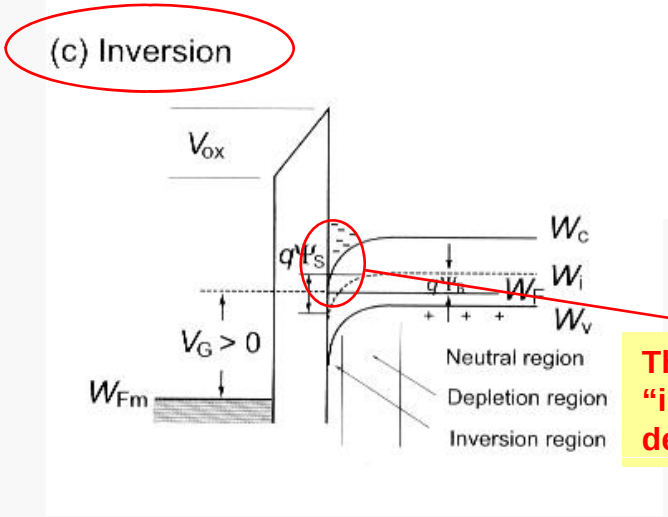


Figure 4: Energy-band diagram of the th components of a real MOS capacitor, con ing of an Al contact, silicon dioxide and p- silicon. $q\Phi_m$ denotes the work function of metal, $q\Phi_{ms}$ the workfunction difference versus p-Si, χ the electron affinity of the con, W_g the band energy, W_c the conducti band, W_v the valence band of silicon, $q\Psi_F$ difference between the intrinsic Fermi leve and the Fermi level W_F [5].









The cond. band crosses the “intrinsic” Fermi level ⇒ a thin, dense electron layer is formed

The MOS capacitor III

2.1 MOS Capacitor

Figure 3 shows the structure of a MOS capacitor with the three components, the metal or polysilicon contact, the silicon dioxide with a thickness t_{ox} and the silicon. The corresponding band diagram is shown in Figure 4. Due to the 9 eV bandgap of the silicon dioxide and the large band offsets relative to the silicon, the potential barrier between the conduction band of the silicon and the silicon dioxide is large ($\approx$ 3.1 eV). This barrier crucially controls possible charge transport through the dielectric layer in the presence of an applied voltage, and thus, determines the reliability of the dielectric-semiconductor interface. Frequently poly-Si is used as a contact material instead of a metal. For p-type poly-Si the work function is $\Phi_s = \chi + W_g/2q + \Psi_B \approx 4$ eV, where χ denotes the electron affinity, W_g the band gap energy, Ψ_B the difference between the Fermi potential W_F and the intrinsic potential W_i.

The energy band diagram of an ideal MOS capacitor with a p-type semiconductor is shown in Figure 5 ($q\Phi_{ms}$ is assumed to be zero, see Figure 4). When a negative gate potential $V_G < 0$ is applied the Fermi level of the metal increases and an electric field is created in the SiO_2, indicated by the slope of the conduction band of the SiO_2, and in the silicon. Because of the low carrier concentration the Si bands bend upwards at the SiO_2 interface, leading to an **accumulation** of excess holes. In order to conserve charge, an equivalent number of electrons is accumulated at the metal side of the MOS capacitor.

When a positive potential is applied at the gate contact, its Fermi level moves down leading to band bending in the silicon in the downward direction. As a consequence, the hole concentration near the interface decreases. This status is called the **depletion condition**. Charge neutrality requires the induction of an equivalent amount of positive charge at the metal-oxide interface Q_M as negative charge in the semiconductor Q_S, explicitly,

$$Q = -Q_M \text{ with } Q_S = Q_d \qquad (1$$

where Q_d originates from the ionized donor states. A further increase of the positive gate potential, enhances band bending such that at a certain gate potential the intrinsic Fermi level crosses the Fermi level as shown in Figure 5c. Energetically, it becomes now favourable for electrons to populate the newly created surface channel. The surface behaves like an n-type semiconductor where the doping was created by inverting the original p-type silicon with an applied field. This condition is called **weak inversion** and the corresponding onset gate voltage the threshold voltage V_T. The negative charge at the semiconductor interface Q_S consists of inversion charge Q_i (electrons) and ionized acceptors Q_d (Figure 5d)

$$Q = Q_i + Q_d \qquad (2$$

As indicated in Figure 5c, three regions develop within the semiconductor: a shallow inversion region, a depletion region with a maximum depth w_d and deeper in the substrate a neutral region. A further increase of the potential yields to **strong inversion** when the concentration of the electrons exceeds the hole concentration in the substrate ($Q_i > Q_d$). Then, the gate voltage V_G can be expressed by

$$V_G = V_{ox} + \psi_S = -\frac{Q_S}{C_{ox}} + \psi_S \qquad (3$$

where C_{ox} is the oxide capacitance per unit area and ψ_S is the surface potential, reflected by the band bending in Figure 5c. The surface potential and the total induced charge at the interface can be calculated by solving Poisson's equation with appropriate boundary conditions (see e.g. [6], [7]). Under extreme accumulation and inversion conditions, when V_G and V_{ox} are significantly larger than ψ_S, then Q_S can be approximated by (a)

$$Q = -C_{ox} V_G, \text{ with } C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}} \qquad (4)$$

since ψ_S is always less than W_g. Eq. (4) implies that the total induced charge at the interface increases with the gate capacitance (per unit area) C_{ox}. ε_{ox} denotes the permittivity and t_{ox} the thickness of the oxide layer.

The total capacitance of the MOS-capacitor C is a series combination of the oxide capacitance C_{ox} and the semiconductor capacitance C_S. Figure 6 shows a capacitance-voltage (C-V) curve for an ideal MOS capacitor at low and high frequencies, as well as under deep depletion conditions. Whereas C_{ox} is basically independent of the gate voltage, the semiconductor capacitance changes, due to the different charge states discussed above. At zero voltage the flat band capacitance C_{FB} is given by (b)

$$C_{FB} = \frac{1}{\frac{t_{ox}}{\varepsilon_{ox}} + \frac{L_D}{\varepsilon_S}} \qquad (5)$$

where L_D is the Debye length and ε_S the silicon permittivity. (For a real capacitor a voltage must be applied to flatten the bands, because $\Phi_{ms} \neq 0$ (see Figure 4)). At negative voltages an accumulation charge builds up with a capacitance $C_S = -dQ_S/d\psi_S$ (Figure 5c). Since ψ_S is limited to 0.1 to 0.3 V in accumulation the total capacitance rapidly reaches its saturation value C_{ox}. A small positive voltage produces a depletion layer which acts as a dielectric with a width w_d in series with the oxide. Thus, the total capacitance C is given by

$$C = \frac{1}{\frac{t_{ox}}{\varepsilon_{ox}} + \frac{w_D}{\varepsilon_S}} \qquad (6)$$

(c)

decreases rapidly to a minimum C_{min}. When the gate voltage reaches the threshold voltage $V_T = 2\psi_S$ an inversion layer starts to form and C increases again. Analogous to Eq. (2), the semiconductor capacitance C_S can be broken up into a depletion charge capaci-

R.Waser (Ed.), *Nanoelectronics and Information Technology* (Wiley-VCH, 2003)

The MOS capacitor IV: capacitance

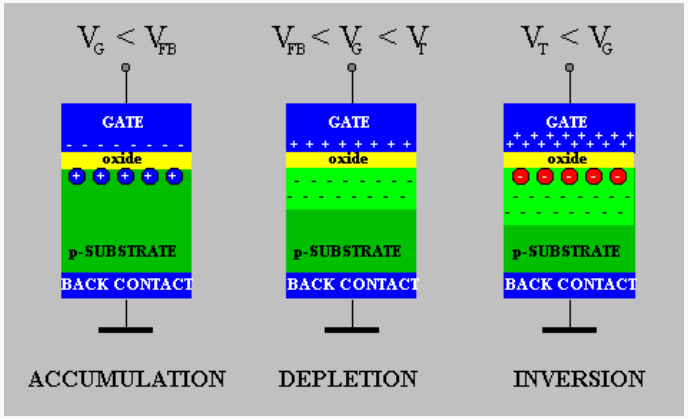
tance C_d and an inversion layer capacitance C_i. C_d and C_i are parallel capacitances in series with C_{ox}, and thus an increase of C_i increases the total capacitance as shown in Figure 6 (curve a). In contrast to the accumulation condition, under the inversion condition the surface potential ψ_S may increase to about 1.0 V. Consequently, the concomitant inversion capacitance C_i can become much larger than the depletion capacitance C_d. Under strong inversion w_d reaches its maximum when the semiconductor is effectively shielded from further penetration of the electric field by the inversion layer. C reaches its maximum value C_{ox}.

If the capacitance measurement is performed at higher frequencies (> 100 Hz), curve (b) in Figure 6 is obtained because the inversion charge arising from minority carriers cannot respond to high frequencies, unless the surface inversion channel is connected to a reservoir of minority carriers as in a MOSFET device. Thus, at high frequencies the inversion charge remains fixed at its dc value and the capacitance does not show an increase at larger V_G.

So far we have discussed only ideal MOS-structures. Real capacitors have undesirable charges within the oxide and at the dielectric/semiconductor interface. These may be mobile ionic charges, like K^+ or Na^+ ions, trapped charges in the SiO_2, fixed charges close to the interface and interface-state charges. Their densities have to be kept at a minimum. *C-V* measurements are sensitive to such defects, and thus are used to characterize the dielectric layers. Oxide charges will affect the threshold voltage and consequently the performance of the MOSFET. The Si/SiO_2 interface has excellent properties, making silicon the most important semiconductor material. The interface density of the state of the art thermally grown oxides is 2×10^{10} cm^{-2}/eV. However, fundamental limitations will arise when the thickness of the oxide layer becomes so thin that direct tunnelling through the ultrathin silicon oxide causes unacceptable leakage. Alternative gate dielectrics with higher permittivities solve this problem and will be discussed in this chapter.

Total capacitance is a series of the oxide and semiconductor capacitance

It depends on the gate voltage
(may be relevant for high-speed applications)



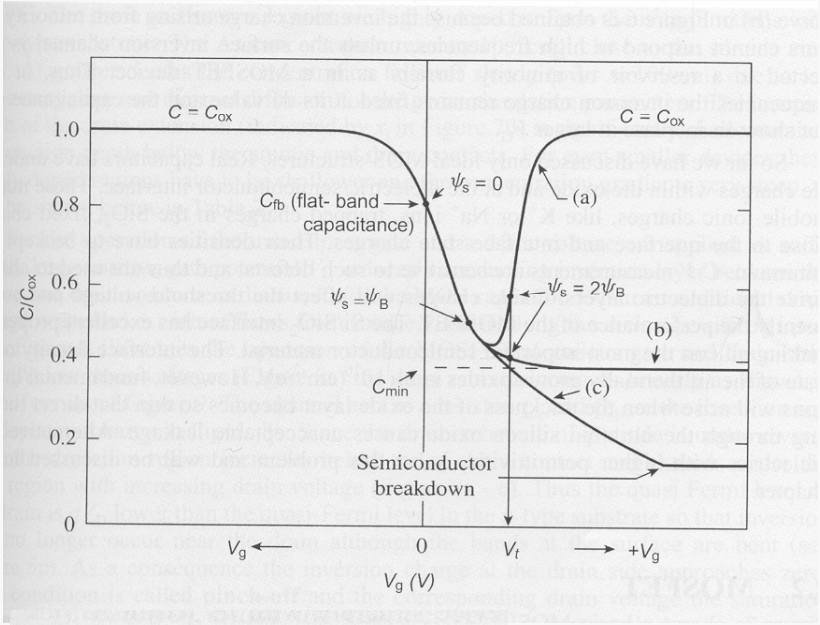


Figure 6: *C-V* curve of an ideal MOS capacitor under
(a) low frequency,
(b) high frequency and
(c) deep-depletion conditions [6].

The MOS-FET III

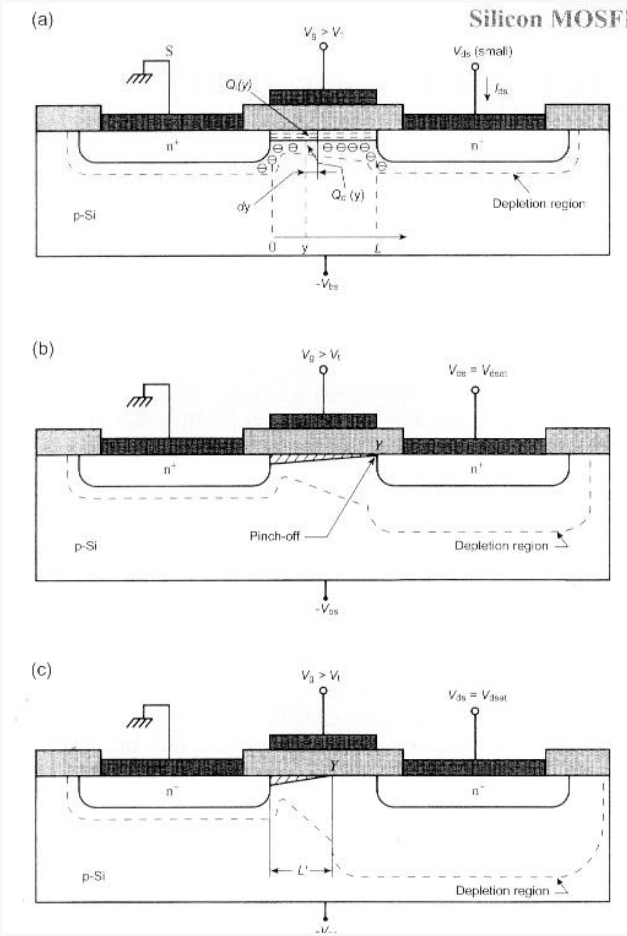
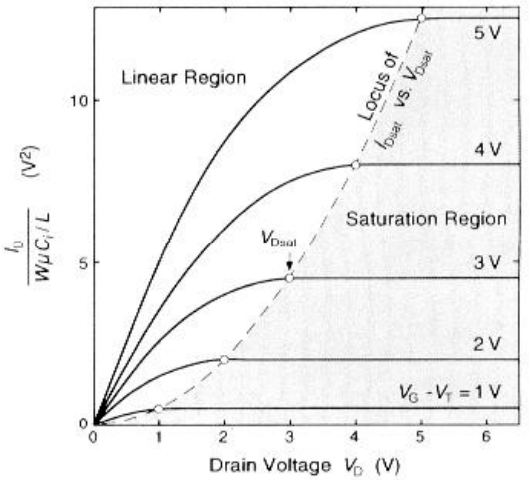


Figure 8: MOSFET operation at a gate voltage $V_g > V_t$ with increasing drain voltage V_{ds}.
(a) At low drain voltages the transistor is in the linear range. Q_i and Q_d are the inversion and the depletion charges, respectively.
(b) shows the pinch-off condition with the pinch-off point Y.
(c) saturation regime where the effective channel length is reduced to L' [6].



For the operation of the transistor a gate and a drain voltage are applied. A sufficiently large positive gate potential V_G induces a conducting inversion layer between the source and drain contacts as discussed for the MOS capacitor. When an additional drain voltage is applied, a current flows from source to drain along the dielectric/silicon interface. Figure 8 illustrates the operation of the MOSFET at various gate and drain voltages. At low drain voltages (Figure 8a) the drain current increases linearly as shown in the *I-V* curves of Figure 9. The channel acts as a resistor. Q_i and Q_d are the inversion and the depletion charge in Figure 8a. Since the drain-substrate n^+-p-diode is under reverse bias, the depletion region increases below the n^+-drain contact and extends under the gate region with increasing drain voltage (Figure 8a - c). Thus the quasi Fermi level of the drain is qV_D lower than the quasi-Fermi level in the p-type substrate so that inversion can no longer occur near the drain although the bands at the surface are bent (see Figure 5c). As a consequence the inversion charge at the drain side approaches zero. This condition is called **pinch-off** and the corresponding drain voltage the saturation drain voltage (Figure 8b). Since the channel resistance is increased, the drain current saturates (saturation region). The pinch-off point, determined by V_{Dsat}, moves towards the source contact with increasing drain voltage. The carriers now drift down the conducting channel and are injected into the surface depletion region at the pinch-off point near the drain (Figure 8c).

"Saturation" is achieved when no additional minority charges can be created

Material modulation doping

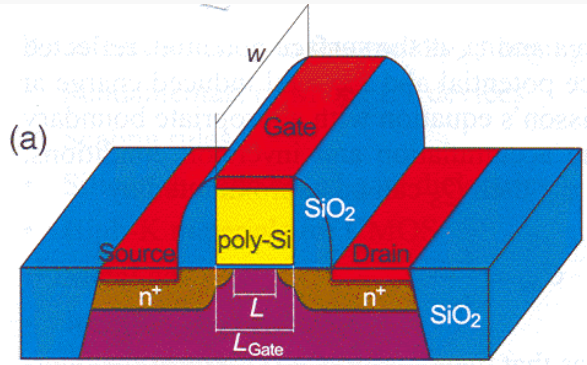
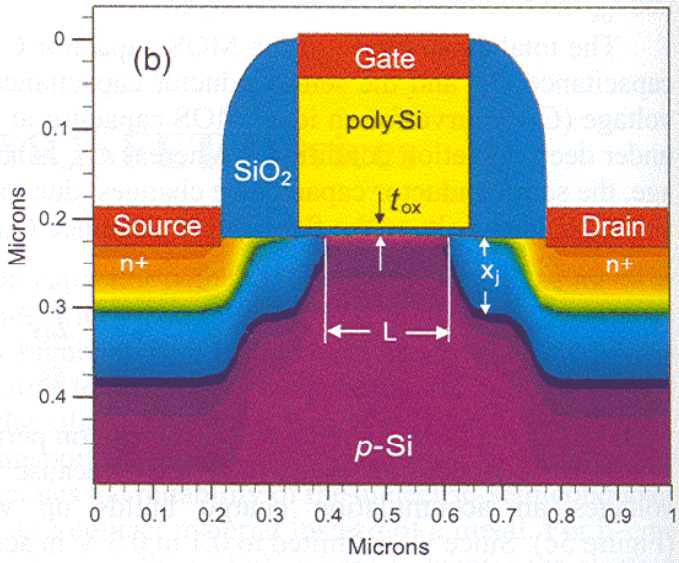


Figure 7:
(a) MOSFET device structure with the terminals, source, gate and drain. L_{Gate} denotes the (printed) gate length, L the channel length or physical gate length and w the gate width. The transistors are isolated with SiO_2 trenches on each side. The gate contact is isolated from source and drain with SiO_2 spacers on each side of the poly-silicon gate contact.
(b) Net doping profiles on a micrometer length and depth scale for a transistor with a gate length $L \cong 0.2$ μm and a gate oxide thickness t_{ox} as calculated with a device simulator (Silvaco). The colours reflect the net dopant concentrations in the Si, ranging from $\approx 10^{17}$ B cm^{-3} in the p-Si to $\approx 10^{20}$ As cm^{-3} near the source/drain silicide contacts. The depth of the n^+/p-junction at the extensions, indicated with x_j, is much shallower than the junction depth below the source and drain contacts (at the blue/dark red boundary).



Modulation doping for the semiconductor on a small scale

Field distribution and homogeneity

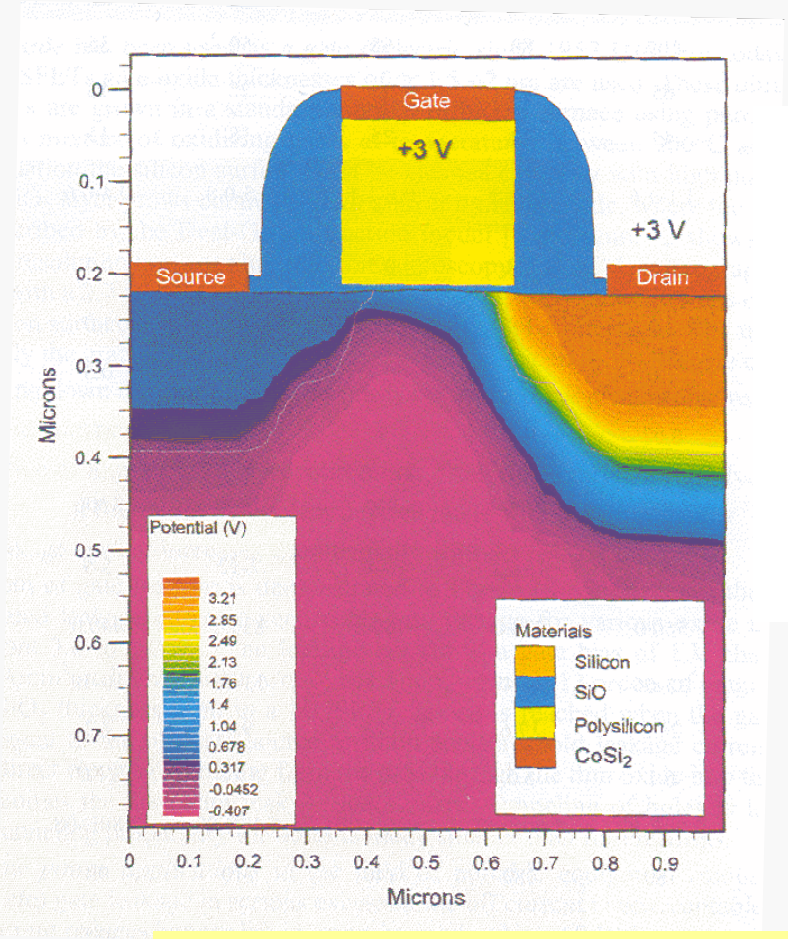


Figure 10: Simulated potential distribution of a 0.2 µm MOSFET with $V_G = 3$ V and $V_D = 3$ V. Near the drain region the potential lines are strongly affected by the drain voltage. The thin solid line indicates the n+/p-junctions.

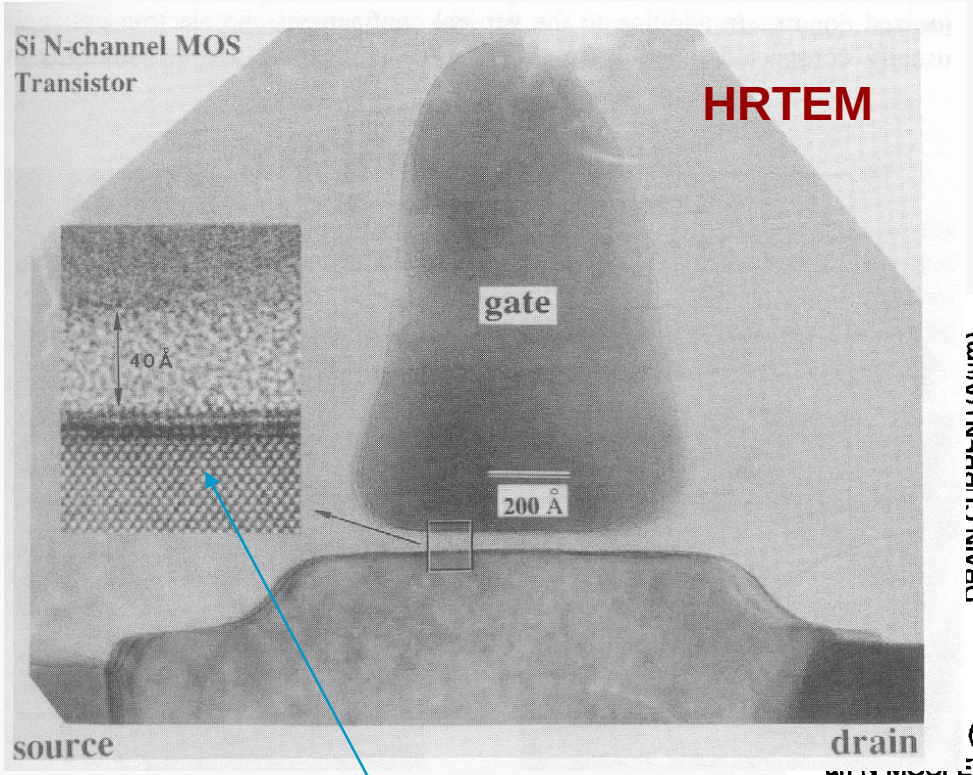
Relatively large field gradients in small regions

Need to accurately control the material at the nanometer scale

⇩

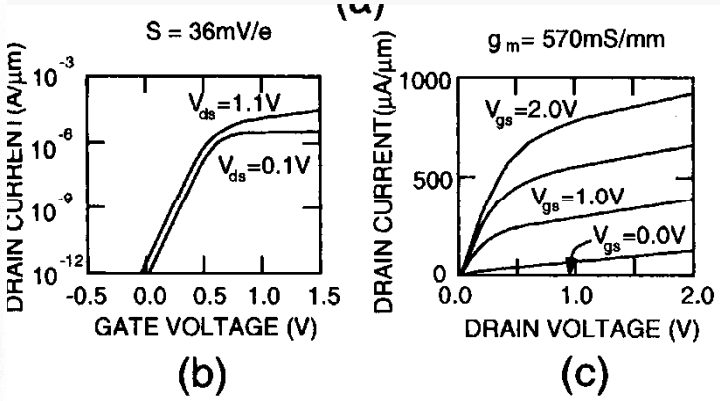
Limits in miniaturization

A “miniaturized” (sub-micron) MOS-FET



Crystalline nature of Si is detected by HRTEM

MOS-FET with nanosized features can be produced (we will see how)



(a) A high-resolution transmission electron micrograph of a cross-section of an N-MOSFET with a gate length of 0.13 μm adapted from[27] and through the courtesy of Y. Kim. The channel is less than 400 atoms long. The inset shows a lattice image of the channel region of this device. (b) and (c) represent the measured subthreshold and drain characteristics found at room temperature for an N-MOS transistor like that shown in (a). From these measurement it can be inferred that the transconductance is approximately 570 μS/μm, and the subthreshold slope, S = 36 mV per e-fold change in I_D or 84mV/decade, and the threshold voltage is V_t =0.45 V.

presently, down to 70-80 nm → ~ 5/0 nm, or below

G.Timp (Ed.), Nanotechnology (Springer, 1999)

Miniaturization issues

What happens if we start decreasing dimensions?

1. Reducing the gate (channel) length ⇒ **quantum conf.** (we'll mention it)
2. Reducing the involved amount of charge ⇒ **single electron** (we'll mention it)
3. Reducing thickness of oxide layer ⇒ **materials problems** (we'll mention it)
4. Reducing the overall size ⇒ **nanofabrication problems** (we'll mention it)

To be kept in mind:

whole set of issues must be addressed when trying to realize nanosized MOS-FETs

Conclusions

✓Nanotechnology is a wide area cross-related with many scientific and technical fields

✓Electronics is an important driving force

✓Filling the gap between micro- and nanotechnology has to face:
- Inherent limitations in scaling down the feature size;
- Material limitations due to the small size (ultra thin films);
- (Fundamental issues associated with quantum confinement)
- (Fundamental problems in the fabrication process)

✓New approaches are required for fabrication

✓New architectures are required for the device operation

✓*New functions can be achieved*