Some Semiconductor Products and Components

Here are tables listing products, components and services that need some amount of semiconductor technology (***ST***) in this module

First let's look at obvious and not so obvious products and their components:

Illustration

Product	Abbreviation	(Major) Components	Remarks
Personal computer	***PC***	Integrated circuits (***IC***) or "Chips"	There is much more **ST** inside a **PC** then just chips!
Almost anything with a battery inside or a power connection. (Cars, communication entertainment, satellites, **MRT** machines, ...)	World Economy	Integrated circuits (***IC***) or "Chips", optoelectronics	The major enabling business for nearly everything
Solar cells	Save the world	Cheap large-area semiconductors	See table below
Microsystems	**MEMS**	Si-based special chips	Axiomatic product: Air bag sensor, "**DLP**" in beamers
Diode Lasers		**III-V** compound Heterojunction	There is one in every **DVD** player.
Lights	Save the world, part 2	**III-V LED**'s	Will soon replace the old-fashioned light bulb.
Radio frequency identification device	**RFID**	Very cheap chip (organic semiconductor?), printed antenna	You may have one implanted soon.
LCD Flat panel displays	**LCD** **OLED**	transistor matrix plus "liquid crystals" or organic light emitting *diodes*)	First **OLED**'s are on teh market
...	....	....	....

Here is a table detailing the product "Solar Cell" some more:

Product=Solar Cell	Remarks
"Bulk" Solar Cells	
General definition	The light absorbing part is relatively thick (**> 100 µm**); no substrate for mechanical stability is needed. Nearly all solar cells on roof tops now (=**2007**) fall into this category; and they are practically always made from **Si**.
Single crystalline **Si**	Best Mass product in terms of solar cell efficiency (about **21 %** by now)
"Multi"crystalline **Si**	Slightly worse and somewhat cheaper than single crystalline **Si**, but market leader
GaAs solar cell	Very good . And very expensive. Some space applications
Thin film Solar Cells	
General definition	Typically **(0.5 - 30) µm** film on substrate.
Amorphous Si	The classic thin film cell; found in most small-scale applications (watches, pocket calculators, ...). Major performance problems even after > 20 a ***R&D***
Nanocrsytalline Si	The solar cell of the future? Not yet in production
$CuIn_xGa_{1-x}Se_yS_{1-y}$	So-called "***CIGS***" solar cells. Mass production has started.
CdTe	Some production; future unclear
Organic semiconductors	Max. efficiency at present (**2007**) around **5 %**. We shall see.
Specialities	
TiO_2 plus	The so-called "Grätzel Cell". Sintered **TiO_2** nanoparticles coated with organic molecules absorb the light; carrier transport occurs via the **TiO_2** and a Redox electrolyte
Tandem cells	Several thin layers of semiconductor with different band gaps on top of each other. Very high efficiency and price.
Concentrator cells	"Small" high-efficiency cell mounted in the focus of a parabolic mirror. Saves on (expensive) semiconductor on the expense of mirror and positioning device, Only usable in countries with lots of direct sun light (not in S.-H.). Gets hot!

Germanium

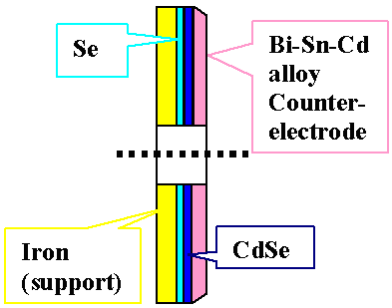
- What is Ge good for? Now and in the past? An interesting story could be told here.
  - Why don't you try to do that?

Advanced

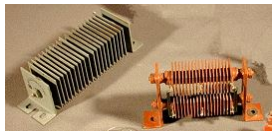
Selenium

Advanced

- What is (and was) **Se** good for? An interesting story can be told here.
  - **Se** was good for making rectifiers for **250 V DC**, always needed for running the vacuum tubes in radios and early **TVs**. The rectifying junction consisted of **Se - CdSe**, and it worked without anybody knowing why.
  - Many tablets as shown below were switched in series (with a central screw) as shown in the top picture or, in somewhat later developments, contained in a flat housing



Se - CdSe rectifier tablet



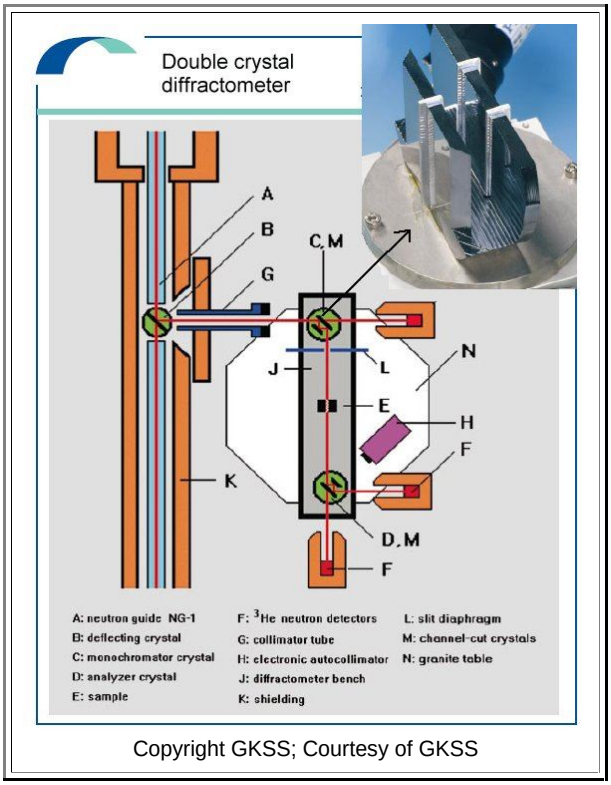
Selenium based rectifiers for **250 V**.
Bottom: Famous Siemens&Halske bridge rectifier
© *Jan Wüsten, used with permission.*

- This is just one example of semiconductor technology that existed before semiconductor theory was "invented"
  - Other examples are the **Cu**-Oxide rectifier ("Kupferoxydul") and, most important, the early **"crystal detector" radio**.
- Working with **Se** rectifiers had one side effect that is practically unknown with today's semiconductors. Whenever the contraption blew up (which happened every now and then) it emitted the most unpleasant smell imaginable because some **Se** compounds where produced (**HSe** is the stinkiest gas known to humankind, and all other **Se** gases are similar in that respect).
  - That helped to diagnose the cause of malfunctioning radios and gave rise to a word play only possible in German: Selengleichrichter = Selen-gleich-riecht-er.

Exotic Uses of Silicon

Advanced

- Here are a few of the more *exotic* uses of **Si**. I keep this brief; details might be filled in later.
  - The **kilogram standard in Paris** is getting lighter - nobody knows why. Efforts are under way to use a perfect **Si** crystal as the new standard. In addition, this would also serve to determine Avogadro's constant to a better accuracy.
  - Very precise mirrors, possibly coated with a sequence of layers to act as **X-ray mirrors**, need practically atomically flat substrates of considerable lateral extension. Your run-of-the-mill **300 mm Si** wafer is already much closer on this target than anything you can make with all possible tricks from a piece of glass or metal, for example.
  - The picture below shows a **neutron diffractometer** using **Si** crystals in three places. The inset shows on of the "slotted" rather large **Si** crystals. The principle is that only neutrons with a wavelengths meeting the Bragg condition will be reflected.



Copyright GKSS; Courtesy of GKSS

  - There are things like "**Photonic crystals**" and other novel optical components made from **porous Si**.
  - There are even **high explosives** made from porous **Si**; also all kinds of sensors, electrodes for micro fuel cells, and much more. For some details activate the link
  - Finally and ironically: **Mechanical watches** of the expensive to extremely expensive variety (Jaeger-LeCoultre,...) have started to use **Si** parts inside. Of course, using **Si** for making a simple "watch" chip and a **SiO_2** crystal for the frequency normal allows to make a much better watch at a tiny fraction of the costs of a mechanical one.

III-V Highlights

Here are a few Highlights concerning emerging uses of III-V semiconductors

- Northrop Grumman Corp. (Redondo Beach, Calif.) is claiming a new world record for transistor speed with an *indium phosphide*-based high-electron-mobility transistor (**InP HEMT**). The device has a maximum frequency of operation of >**1 THz** (**1000 GHz**). Researchers at Northrop Grumman's Space Technology sector, led by Richard Lai, detailed how they developed the terahertz-speed transistor in a technical paper delivered at the 2007 International Electron Devices Meeting (IEDM), held recently in Washington, D.C

  Semiconductor International, 12/21/2007

- Fujitsu announced a **GaN** based ***HF*** power chip (around **100 GHz**, **1.3 W**) in **2010**; here is a link to the article.
  Semiconductor International, Oct./Nov. 2010

- In "Solid State Technology", Jan. 2012, some interesting information concerning LED production is given. Highlights are:

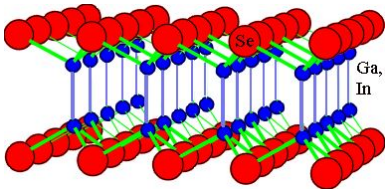
  - Yield for **1 mm^2** (**GaN** based) devices is around **25 %**.
  - Largest cost factor is he package and the Gold!
  - Average size moved form about **0.5 mm^2** to **1 mm^2** because more than doubling of light output foe same packaging costs. However, yield problem..
  - Fab cost around 100 Mil. Euro; half of that for ***MOCVD***.

Advanced

Layered Semiconductors

Advanced

- Compounds like **InSe**, **GaSe** or **Bi_2Se_3** are called "**layered semiconductors**" .
  - The reason is shown in the picture below. Considering the bonding situation, the two elements will form two-dimensional layers if the average number of bonds is three (= (**4** (**Ga**, **In**) + **2** (**Se**)**/2**). A **GaSe** or **InSe** layer as shown below thus has no "**dangling bonds**" and no urgent reason to form a three-dimensional crystal.
  - Note that there is a certain degree of polytypie with regard to the arrangement of the layers, but we won't go this deeply into the matter here.



Structure of **III-Se** (**III = Ga**, **In**, ...)

- A crystal is formed by stacking the layers; the bonding between the layers can only be weak and of the van-der-Waals (**vdW**) type
  - Accordingly, mechanical properties will be extremely non-uniform. It is easy to shear the crystal in the **vdW** plane, but not in strongly bonded planes and so on.
- Since the **vdW** bonded plane has a low density of dangling bonds, but still a perfectly ordered arrangement of atoms, it is ideal for growing epitaxial films of other materials.
  - The atoms of the film to be grown, upon striking the surface of the layered semiconductor substrate, are not strongly bounded and can move around, but their arrangement will still be influenced by the ordered array of atoms "below" - and they might follow this order, forming a single crystal.
  - A single-crystal layer growing on a substrate formed by some other crystal usually experiences increasing strain with increasing thickness because the lattice constants never match perfectly. A relaxation of this misfit stress is necessary at some critical thickness and misfit dislocations are commonly introduced. With **vdW** bonded planes this may not happen - the growing layer "simply" shears of the substrate, expanding or contracting; whatever is as required to release the stress, because it is only loosely bound to the substrate. Coherency to the substrate is lost in this case, but that may not be important since the layer can keep growing to considerable thickness as a single crystal
- As far as semiconducting properties go, here are a few numbers:

Type	Lattice	Band Gap	Remarks
GaSe	hex.	2-2.1 eV, direct	
InSe	hex.	1.2-1.3 eV, direct	

Exercise 2.1-1

Quick Questions / Class Exercises to

2.1 General Chemistry and Structure

We had the following class exercises:

- Ponder the history of "***LCD***" flat panel displays.
- Why does the world need saving? How shall it be done?
- Why do we use crystals, preferably single crystals, preferably "perfect" single crystals in case of doubt?
- Supply examples for critical parameter - component couplings.
- What makes a semiconductor interesting for technology?
- How would you like your bandgap, Sir?
- What does it mean to dope a semiconductor in reality?
- Come up with **2 - 3** examples where product requirements transfer to shape / structure requirements.
- Can you still afford it if your present product is hugely successful? - What could that mean?
- Provide examples (and criteria) for the products listed

Solution

Exercise 2.2-1

All Quick Questions / Class Exercises to

2.2 Silicon

We had the following class exercises:

- What is the approximate lateral size of *one* transistor in an **IC**?
- Why are there no **16 GB** memory chips *now*.?
- What properties should a semiconductor have for making **IC**'s?
- What exactly produced complexity and market growth rates of **30 %** for more than **30** years (for **Si IC**'s)?
- Where will it end?
- What do you know about **MEMS**?
- Are there any other uses of **Si** you know off (or can find quickly)?

Solution

Exercise 2.3-1

All Quick Questions / Class Exercises to

2.3 III-V Semiconductors

We had the following class exercises:

- What band gaps of which type would we like to have for **III-V** properties?
- What are the requirements for substituting light bulbs by **LED's**? Compare your answers to the list in the script.
- What are the requirements for substituting light bulbs by **LED's**? What is the state of the art?
- Amend and discuss the list of optoelectronic products given so far.

Solution

Exercise 2.4-1

Quick Questions to

2.4 Other Semiconductors and Related Products

- Here are some quick questions to check if you got the basic facts:
  - What are the strengths an weaknesses of **Ge**?
  - Why is **SiC** both very desirable and very difficult as a semiconductor material?
  - What are **II-VI** semiconductors and *chalcogenides*= Why are these materials of some interest?
  - Name at least two important semiconductors of the above groups and what they are used for.
  - What are *organic semiconductors* and their important uses?

- There is no "solutions" - you just need to think about it or look it up!

Exercise 2.6-1

All Quick Questions / Class Exercises to

2. Semiconductor Materials and Products

Subchapter 2.1: General Chemistry and Structure

- Ponder the history of "***LCD***" flat panel displays.
- Why does the world need saving? How shall it be done?
- Why do we use crystals, preferably single crystals, preferably "perfect" single crystals in case of doubt?
- Supply examples for critical parameter - component couplings.
- What makes a semiconductor interesting for technology?
- How would you like your bandgap, Sir?
- What does it mean to dope a semiconductor in reality?
- Come up with **2 - 3** examples where product requirements transfer to shape / structure requirements.
- Can you still afford it if your present product is hugely successful? - What could that mean?
- Provide examples (and criteria) for the products listed

Subchapter 2.2: Silicon

- What is the approximate lateral size of *one* transistor in an **IC**?
- Why are there no **16 GB** memory chips *now*.?
- What properties should a semiconductor have for making **IC**'s?
- What exactly produced complexity and market growth rates of **30 %** for more than **30** years (for **Si IC**'s)?
- Where will it end?
- What do you know about **MEMS**?
- Are there any other uses of **Si** you know off (or can find quickly)?

Subchapter 2.3: III-V Semiconductors

- What band gaps of which type would we like to have for **III-V** properties?
- What are the requirements for substituting light bulbs by **LED's**? Compare your answers to the list in the script.
- What are the requirements for substituting light bulbs by **LED's**? What is the state of the art?
- Amend and discuss the list of optoelectronic products given so far.

Subchapter 2.4: Other Semiconductors and Products

- What are the strengths an weaknesses of **Ge**?
- Why is **SiC** both very desirable and very difficult as a semiconductor material?
- What are **II-VI** semiconductors and *chalcogenides*= Why are these materials of some interest?
- Name at least two important semiconductors of the above groups and what they are used for.
- What are *organic semiconductors* and their important uses?

Solution to

All Quick Questions / Class Exercises to

2.1 General Chemistry and Structure

Ponder the history of "***LCD***" flat panel displays.

- Well - no. You do that yourself. Here just a few hints:
  - The first commercial color flat panel displays larger than the black-and-white stuff in cheap watches and based on **LCD** technology hit the market around **1990** - and they were small and incredibly expensive.
  - You need two things for a large **LCD**: liquid crystal technology and a lot of transistors in a matrix to control every pixel of the display.
  - Flat panel displays are much larger than the biggest wafers available. So how do you make a transistor matrix?

Why does the world need saving? How shall it be done?

- Two words are sufficient: **Climate catastrophe** and dwindling **resources of oil**. However, if nothing is done about that rather soon, it's not the world that will need saving, only the humans presently over populating in ever increasing numbers the surface of the planet.
- If the humans are to be saved, the magic word is: *solar energy*. In all its forms - direct heat, water power, wind and solar cells.
- Semiconductor technology is instrumental. Not only for making plenty, efficient and cheap solar cells, but also for controlling the other solar energy sources and for distributing power generated by untold millions of small power plants over large distances.

Why do we use crystals, preferably single crystals; preferably "perfect" single crystals in case of doubt?

- Because then we know that the electronic properties are the best we can get. If we need to adjust them to values other than those of the perfect crystal, we can try to do so. Dirty words like "Fermi level pinning" will not make our efforts useless.

Supply examples for critical parameter - component couplings.

- Fast transistors need high *mobilities* = small defect densities.
  Solar cells need large *life times* or diffusion lengths = small defect densities.
  Power devices need to sustain *high field strengths* and need very uniform conductivity and thus uniform doping.
  .....

What makes a semiconductor interesting for technology?

- Bandgap size and type.
  Properties of defect states.
  Can it be easily p- and n-doped?
  Production potential. Perfect or at least decent single crystals available for littel money? Thin films on suitable substrates achievable?
  Are materials needed for technology plentiful and cheap? Extremely dangerous or simple to handle and and to dispose off?

How would you like your Bandgap, Sir?

- Well, thanks for asking. But that depends on my present taste:
  - Rather small for **IR** detectors or thermoelectric devices like Peltier elements for active cooling.
  - Around or just above **1 eV** for common and cheap electronics.
  - Around **1.5 eV** if I want to make high-efficiency solar cells.
  - Direct and exactly **1,325 eV** if I want to make light with that energy.
  - Rather large if I have high temperature applications in mind.

What does it mean to dope a semiconductor in reality?

- To shift the Fermi energy close to the band edges. This can be done by introducing defect levels close to the band edge, but this will only work if you do not have a lot of other defect states in the band gap already.

Come up with **2 - 3** examples where product requirements transfer to shape / structure requirements.

- The overwhelming product requirement for a solar power station is to have a huge area (and decent efficiency) of the solar cells. This transfers to making the solar cells thin to save material and not to make them single crystals to save costs.
- Flat panel displays need a matrix of transitors to address single pixels. Sizes are kow in the $\mathbf{m^2}$ region

Can you still afford it if your present product is hugely successful? - What could that mean?

- You are presently producing **1 GB** memories - **DRAM's**, **SRAM's**. Flash, whatever, and making a lot of money. Your engineers have been working on the next generation, the **4 GB** memory, and have made a few functioning prototypes. Should you start to dig a hole, throw some concrete in it and built a **4 GB** factory to the tune of roughly $\mathbf{5 \cdot 10^9}$ **€**? Or wait a bit longer, enjoying the money coming in instead of spending it on a new factory that may not be needed so soon?
  That is what that question could mean for just one product example.
- You should be as sure as you can be that two conditions are met before you built the new factory:

  1. Your **4 GB** memory chip sells for less than **4** times the price of your **1 GB** memory. Otherwise you customers buy four **1 GB** chips instead of your new chip.
  2. The market for memory *bits* per year should grow at least **4**-fold during your **4 GB** production time. If it would be always constant, for example, the number of **4 GB** chips you could sell would only be ¼ of the number of **1 GB** chips. That means your income is far too small to recover the cost of your **4 GB** factory (that was far more expensive than your **1 GB** factory!)...
- In total: Being able to *make* the new product does not mean that your are going to be able to *afford* making the new product. Semiconductor technology always has a money component, too!

Provide examples (and criteria) for the products listed

- You do that!

Solution to

All Quick Questions / Class Exercises to

2.2 Silicon

Illustration

- What is the approximate lateral size of *one* transistor in an **IC**?
  - Given that state-of-the-art transistors have smallest dimensions of **32 nm**, one could guess that that the total lateral area would be in the order of **(100 - 200) nm^2**. Be that as it may, an answer of "below **1 µm^2**" is acceptable
- Why are there no **16 GB** memory chips *now*?
  - Because that would be a **16 ×8 = 128 Gbit** chip. With a chip area of **1.5 cm^2** with **1.28 cm^2** for the memory cells (the rest is needed for the periphery), one memory cell would be
    1.28 cm^2/1.28 × 10^{11} = 1.28 × 10^{14}nm^2/1.28 × 10^{11} = 10^3 nm^2.
    In other words, one memory cell, needing at least one transistor, would be about **30 × 30 nm^2** and that is not yet (**2008**) feasible.
- What properties should a semiconductor have for making **IC**'s?
  - Well. there is no end to the list:
    - Bandgap not too small to make it too temperature sensitive around ***RT***, or too large to allow easy doping and contacts plus some conductivity. So **Ge** is out now but **GaAs** etc, are still in. **Si** is OK, but a somewhat higher bandgap would be preferable.
    - Large defect free and "cheap" single crystals must be possible. Pretty much all semiconductors but **Si** are out.
    - A process compatible dielectric must exist with very high breakdown field strength and suitable dielectric constant. It's very difficult to beat **SiO_2** here. Big advantage for **Si**.
    - Precise doping establishing a precise conductivity must be possible. Many semiconductors meet this requirement, but **Si** is one of the easiest to dope.
    - The mobility of the carriers should be large for high-frequency applications. **Si** is mediocre in this respect, but still good enough for most applications.
  - ***Conclusion:*** There is no other semiconductor out there that comes even remotely close to **Si** as the best compromise.

- What exactly produced complexity and market growth rates of **30 %** for more than **30** years (for **Si IC**'s)?
  - There is only *one* reason: The **price per function** (e.g. **€** per bit of memory space, per logical operation, per numbers of operations per second, per kWh, ...) comes down substantially.
- Where will it end?
  - Nobody knows. The absolute limit is the size of an atom ≈ **0.3 nm**. How many atoms does it take to make a transistor = switch? A few hundred, occupying a volume of **< 10 nm^3** are enough. So we are a far cry from the limit with our present **2-**dim. integration of logical switches.
  - But maybe the cost limit (see question above) is reached before the technical limits are even close? Check the link for details
  - Wherever and whenever the limits will be felt, one thing is quite likely: Moore's law, maybe with a somewhat slower rate of growth, will go on for at least until **2015** because all essential ingredients are already in the laboratory stage.

- What do you know about **MEMS**?
  - Probably nothing yet. That will change. But you have definitely seen **MEMS** in action and very likely own **MEMS** products:
  - Many **beamers** contain a MEMS chip with roughly **1 Million** little mirrors that can be individually addressed to process one pixel of the light you see on the projection screen.
  - The *acceleration sensors* in your car that tell the microprocessor that it is time to inflate the air bags, as well as the *gyros* (sensors for rotational speed) that allow the microprocessor to calculate how to individually brake the wheels in order to keep you in line when you start to slip, are typical **MEMS** products.
- Are there any other uses of **Si** you know off (or can find quickly)?
  - Let's see: We had solar cells and MEMS. What else?

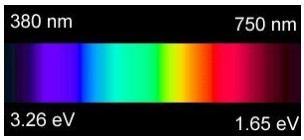
- **Si** wafers are the cheapest *perfect substrates* for all kinds of applications (in particular optics, including X-ray "optics").
- **Si** single crystals are used as perfect Bragg diffractors to produce, e.g. monochromatic Neutron beams.
- If you make nanoporous **Si** and fill the pores with something containing a lot of oxygen (e.g. **$KMnO_4$**), you have produced a high explosive with up to **3** times more bang per **kg** than TNT. A big project was launched to explore that as an integrated fuse for blowing up air bags.
- Si "nanwires" might be the ideal anode for better Li ion batteries.
- And so on - activate this link for some details.

Solution to

All Quick Questions / Class Exercises to

2.3 III-V Semiconductors

- What band gaps of which type would we like to have for **III-V** properties?
  - We want to cover regular optics, meaning whatever one can do with **visible light**. The human eye can perceive wavelengths from **750 nm** or **1.65 eV** (= red) to **380 nm** or **3.26 eV** (= violet)



Visible spectrum as far as it can be reproduced on a screen.

  - So we like to have semiconductors with a *direct* band gap and with every conceivable band gap energy between **1.65 eV** and **3.26 eV**
  - We also *know* that **infra red** (**IR**) optics is important - at the very least for communication via **optical fibres**. This happens at a wavelength of about **1 500 nm** or an energy of about **0.83 eV**. This means that we would also like to commandeer *direct* semiconductors with band gaps ranging from **1.65 eV** down to at least **0.8 eV**.
  - Moreover, we know that there is not only a lot of demand for **white light** but that there are already "white" **LED's** in, e.g., flashlights and car headlights. The easiest way to make white light out of monochromatic light (that we will always get from a **LED**) is to produce **ultraviolet light** (**UV**) that is converted to white light via some intermediate fluorescent material like in any "fluorescent" lighting fixture. This means that we also want to have *direct* semiconductors with band gaps ranging from **3.26 eV** up to **???** - well. at least **4 eV**.
  - *Summing up*: We need to have *direct* semiconductors with band gaps ranging from **4eV** - **0.8 eV**
- What are the requirements for substituting light bulbs by **LED's**? What is the state of the art?
  - First we remember: "Light "bulbs", i.e. lighting fixtures already do exist in many variants. If we want to replace what has been around for more than a century, we must be *cheaper* or *better*. We can be if we look at the weak points of existing "standard" lighting with some kind of "bulbs" where light is produced because something is very hot.

    - **LED's** can have much better efficiencies, i.e. you get more light per Watt.
    - **LED's** can have a much larger lifetime, i.e. you change them after **10** years instead of after just a (felt) few months.
  - The problem is to have high power (not the same as high efficiency) and all colors.
  - State of the art: Look in the Internet and you see: **LED** lighting is just now starting to replace conventional lighting in special applications. It is expected to go far in replacing general lighting in the next **10** years or so.
- Amend and discuss - the list of optoelectronic products
  - The most important *optoelectronic* products besides **LED's** are:

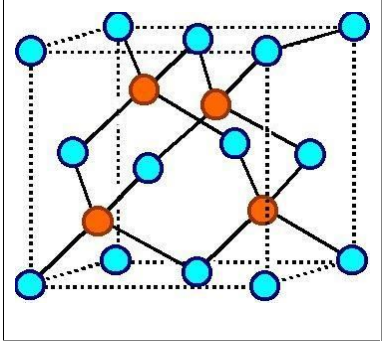
    - **Semiconductor Lasers**. No bar-code reader, ***CD*** or ***DVD*** or whatever comes next without them.
    - **Optical sensors**. The output of the fibre optic cable spanning the Pacific or Atlantic is modulated light. Something must convert it back into modulated voltage.
    - **Displays**. Large display boards, e.g. on Times Square in Mamhatan, consistof hundred thousands of individual **LED's**. Small display in your cell phone or (soon) **TV** screens consist of millions of *integrated* **organic LED's** in an ***OLED*** display.

Zinc Blende and Wurtzite

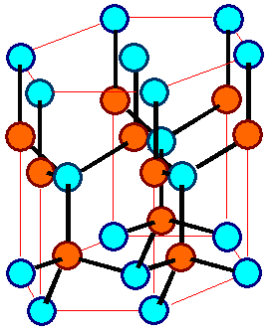
Illustration

Here are the two most important **crystal structures** for semiconductors.

- They are often referred to by the historical names "**Zinc blende**" from the German "Zinkblende" = **α-ZnS**, a rather ubiquitous mineral. The name "**Sphalerite** " also comes form the German: "Sphalerit", which, as was the custom of the time, stems from the Greek "sphaleros" meaning ***treacherous*** or malicious because it is easy to confuse it with other minerals.
- **Wurtzite** was and is the name of the **β-ZnS** modification - the hexagonal high-temperature variant. The named after the French chemist C. A. **Wurtz** (*** 1817,** † **1884**), which gives us an idea of how old those names are.



Zinc blende or sphalerite or Diamond structure



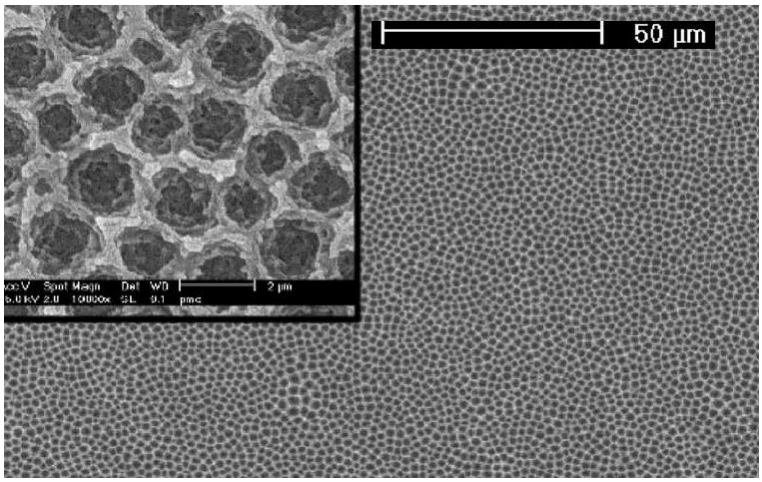
Wurtzite or hexagonal structure
Red lines are not showing bonds.

- Of course we see immediately that what many call *Zinc blende* or *sphalerite* is simply an **fcc** lattice with two atoms in the base: atom **A** at **(0,0,0,)** and atom **B** at **(½, ½, ½)**. *Wurtzite*, of course has an hexagonal lattice and ….*find it out yourselve!!!* ... atoms in the base. Check the links about lattice and base in general and hexagonal lattices in particular.
- Equally of course we notice that we have close packing, i.e. we simply have a case of **ABCABC...** stacking of the base, or a **ABAB...** stacking.
- From the viewpoint of tetrahedrally coordinated bond angles, it is simply a matter of going "cis" (⌐¬) or "trans" (⌐⌙).

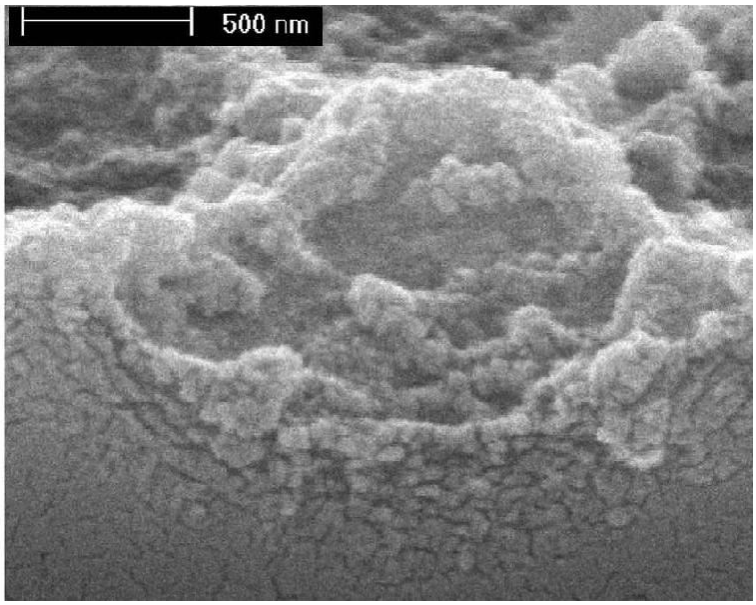
Microcracking in Anodic Oxides on Si

Advanced

- Let's do an extremely simple experiment that produces a "thin film" with quite mysterious properties:
  - We take a piece of **Si** and immerse it into some electrolyte that can only react with the **Si** by forming **SiO_2**; acetic acid or (diluted) **HNO_3** or whatever, will be fine in principle.
  - Now we take a big power supply (a "galvanostat") and run a constant (anodic) current through the **Si** - electrolyte interface. The necessary chemical reaction occurring is the formation of **SiO_2** - nothing else is possible.
  - So we start to build up a thin layer of **SiO_2** on top of the **Si** substrate. But **SiO_2** is an insulator, so it will impede current flow. Our power supply must keep the current constant, so it will crank up the voltage as the **SiO_2** layer gets thicker. Eventually, the voltage will go to infinity.... ?
- Interestingly, this is *not* what happens. The voltage will go up in the first few minutes or so, but then it will stay more or less constant. The thickness of the **SiO_2** layer formed, will also stay more or less constant. Moreover, it will, on occasion, form interesting self-organized structures as shown below. Now we have a puzzle!
- First we have to realize that the **SiO_2** layer will be under tremendous compressive stress. Why, you might ask. The temperature, after all, is room temperature and constant throughout the experiment.
  - Well, when a certain volume of **Si** is oxidized, you essentially put an oxygen atom in between two **Si** atoms. In consequence, if you oxidize **1 cm^3** of **Si**, you get almost **2 cm^3** of oxide. Since the oxide film being formed cannot expand in lateral directions, and is perfectly brittle at room temperature, a very large compressive stress builds up rather quickly. At a thickness of about **12 nm** it is so large that something must happen; the thickness just can't go up very much anymore.
  - On the other hand, as long as a constant current flows through the interface, we are producing oxide at a constant rate. The thickness thus should increase linearly with time.
- "Obviously", we must have a mechanism that allows to get rid of oxide at the same rate it is being formed so the thickness can stay halfway constant.
  - Any suggestions? The best, researchers could come up with so far is that the oxide "cracks off" somehow. But nobody really knows how that is supposed to happen.



"Thin" layer of anodically formed oxide on **Si**. Not the self-organized pattern and the "grainy" structure of the oxide



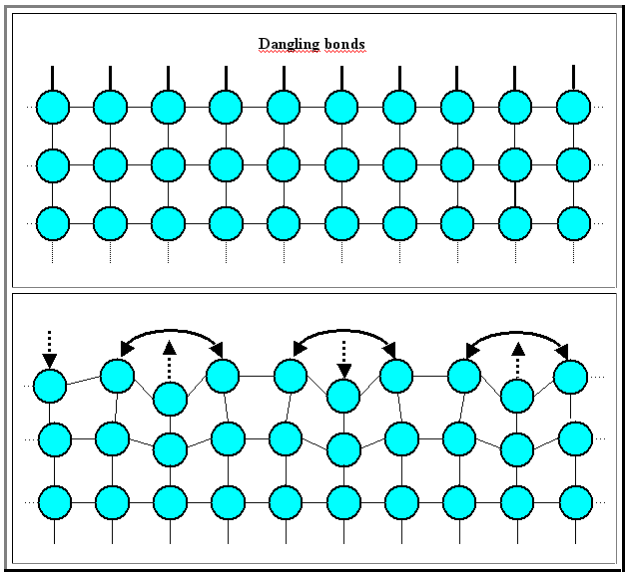
Cross-section through on "dimple. The interface between the **Si** and the **SiO_2** is not visible;
it is somewhere at the bottom of the picture. Note the "microcracks"

S. Frey, B. Grésillion, F. Ozanam, J.-N. Chazalviel, J. Carstensen, H. Föll, and R.B. Wehrspohn,
"Self-Organized Macrostructures in anodically formed mesoporous silica",
Electrochem. Sol. State Lett. 8(9), B25 (2005).

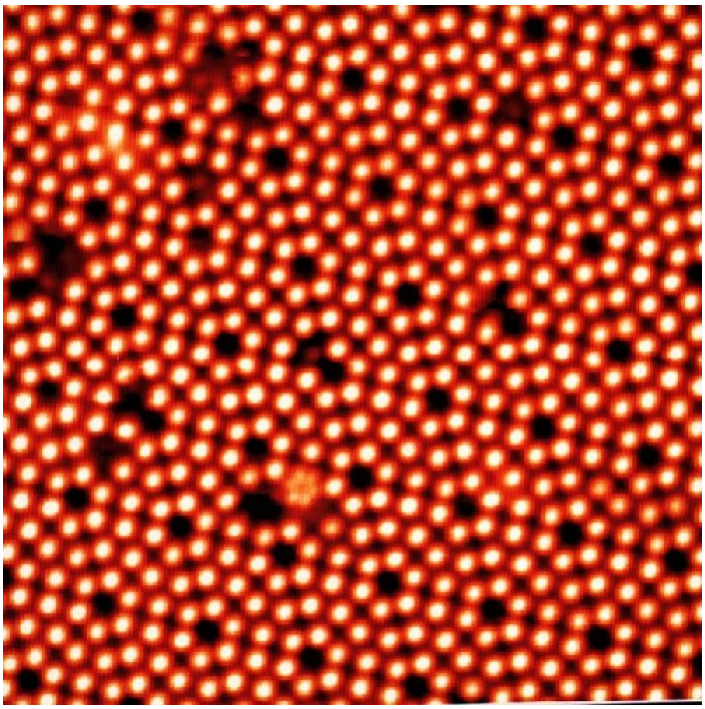
Surface Reconstruction

Advanced

- So you are an *almost* perfect **Si** crystal, or any other crystal, for that matter. The "almost" refers to the fact that you have a surface somewhere, which qualifies as a defect among purists.
  - This is painful because your surface atoms cannot bond to other atoms as they would like to do; in stark contrast to your atoms in the bulk. That's why that surface has a certain amount of surface energy, coming straight from the unsatisfied bonds of the atoms at the surface, which we will now call "**dangling bonds**". This situation is shown in the first picture below.
  - Can you do something about about your surface that lowers the energy (actually the free enthalpy, but entropy is not all that important here)? Of course you can - just get dirty. Oxidize, add a layer of water molecules, take whatever you find in the atmosphere and that forms a bond with your atoms, gaining some energy in the process. That you and pretty much any other crystal will do this is proven by the fact that most materials in air actually are covered with a thin layer of something.
  - If you have been out in single confinement in a vacuum chamber, grab any molecule coming by on occasion. You then reduce the number of molecules in the vacuum and lower the pressure, and people may abuse you by actually taking you as a **gettering layer** in order to improve their vacuum (that's how, for example, "**ion getter pumps**" work!).



- However, people being only human after all, on occasion like to see things stark naked and thus might strip you of all your nice surface layers and then put you into ultra high vacuum, where nothing fashionable comes by for a long time. What are you going to do *now*?
  - You use an old trick of thermodynamics: You spend some (elastic) energy first that allows you to rearrange the outermost atom layers a bit in such a way that they can now form some additional bonds with their neighbors. Maybe not very good bonds, but one has to take what one can get.
  - One possible way of doing that is shown in the second picture. Note that you can't do that in one dimension only. In the picture the surface bonds with just one arrowhead must be perceived as coming out of the screen.
- Is the way this has been drawn the way it will actually happen? Most certainly not! There are, if you think about it, innumerable ways of achieving such a **surface reconstruction**, and only *one* variant can be the best in terms of energy gain.
  - It is quite unlikely that a Professor, being only human after all too, will hit on the best way by just fooling around with balls and sticks.
  - You, however, being a smart crystal, will automatically do the right thing and reconstruct your surface in such a ways that the surface energy is minimized. And that can become quite complicated as an old friend, the reconstructed **{111} Si** surface, proves:



Courtesy of Omicron

Reconstructed **Si {111}** surface as seen with the **scanning tunneling microscope** (***STM***)

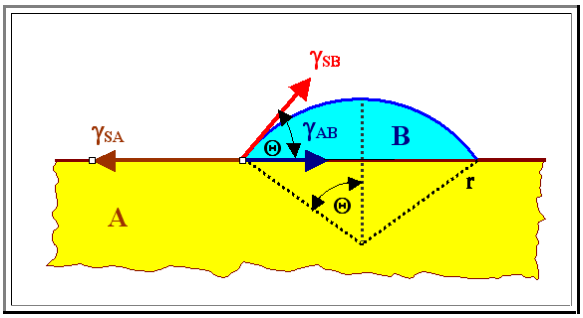
What we learn from this are two points:

- **1.** If you really have a clean surface of a crystal as a substrate for your thin film deposition, chances are that it is reconstructed. This will, of course, influence to some degree what will happen in the beginning.
- **2.** It is pretty hard to know (from theory) if a given surface reconstructs, and if it does, what it will look like.

Nucleation Theory

Advanced

- We want to minimize the free enthalpy G_{nuc} of some nuclei - a cluster of atoms (or molecules) of kind **B** on a substrate **A**.
  - One thing to do, just to make live easier, is to minimize the surface area of the cluster of **B** atoms. We can do this by assuming that the cluster is a little sphere since the surface of a sphere has the smallest surface per volume ratio.
  - But that would defy the purpose so we just assume that the cluster can take any shape, and if it can do that it would behave like a liquid. What we would get then is the shape of some liquid more or less "**wetting** " some substrate - water on glass or something like that.
  - So we consider a situation like the one shown in the picture below:



- Next we have to discuss a bit the meaning of the surface or interface energies γ. The basic definition is that it is simply the (differential) work **d*W*** needed to increase the surface or interface area by **d*A***. So we have

$$\gamma = \frac{dW}{dA} = \frac{dW}{dx \cdot dy}$$

- In other words, the surface energy has the dimension of a stress or tension, and we may simply interpret it as the *force* pulling back when you try to increase the surface in one dimension.
  - This is symbolized by the arrows in the picture above.
- Obviously, in equilibrium, we need the forces to cancel and this gives us immediately a rather important equation:

$$\gamma_{SA} = \gamma_{AB} + \gamma_{SB} \cdot \cos\Theta$$

- Obviously, if we know the three surface or interface energies involved, we can easily calculate the contact angle or **wetting angle** Θ of the cluster. However, before we discuss this relation in more detail, let's first write down the free enthalpy G_{nuc} of the situation. We have

$$G_{nuc} = a_3 \cdot r^3 \cdot G_V + a_1 \cdot r^2 \cdot \gamma_{SB} + a_2 \cdot r^2 \cdot \gamma_{AB} - a_2 \cdot r^2 \cdot \gamma_{SA}$$

  - The meaning is clear. G_{nuc} is the energy (OK, free enthalpy change) of the system containing a **B**-cluster with a size given by the radius ***r*** (without the cluster we put it at zero), G_V is the free enthalpy of the cluster, and the a_i are the coefficients that relate the volume of the cluster (the "cap"), the surface area of the cluster and the interface area to the radius ***r*** (that describes completely the geometry of the arrangement). From (boring, but necessary) spherical geometry we find:

$$a_1 = 2\pi \cdot (1 - \cos\Theta)$$

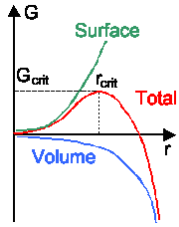
$$a_2 = \pi \cdot \sin^2\Theta$$

$$a_3 = \pi/3 \cdot (2 - 3\cos\Theta + \cos^3\Theta)$$

- The first thing to note is that the equation for ΔG_{nuc} has cubic and quadratic terms in r and thus will give the general behavior shown in the backbone and in the Introduction to Materials Science I.
  - Thermal equilibrium is achieved if $dG_{nuc}/dr=0$; and this will give us the critical radius r_{crit} because it is just the radius at the minimum of $G_{nuc}(r)$ and the critical free enthalpy $G_{crit}(r_{crit})$. Going through the differentiation gives

$$r_{crit} = -2 \cdot \frac{(a_1 \cdot \gamma_{SB} + a_2 \cdot \gamma_{AB} - a_2 \cdot \gamma_{SA})}{3a_3 \cdot G_V}$$

$$G_{crit} = 4 \cdot \frac{16\pi \cdot (\gamma_{SB})^3)}{3 \cdot (G_V)^2} \cdot \frac{2 - 3\cos\Theta + \cos^3\Theta}{4}$$



  - Looking at the denominators of the equation above we see that we have a small critical radius r_V or a small G_{crit} if a lot of volume energy is gained, i.e. if G_V is large, as it should be.
  - Looking at the nominator, we see for r_{crit} the same kind of reckoning up surface and interface energies as we had for the forces in the equation on top. Just multiply with the respective areas expressed in the a_i's, and you get energies. We have thus a direct relation between the contact angle Θ and the critical radius r_{crit} or the critical free enthalpy G_{crit}; the second equation for G_{crit} already expresses this fact.
- The contact angle Θ is obviously a quantity of major concern. It is a kind of direct account balance for the three competing interface energies, giving the energetically best mix of interface areas for a given cluster size.
  - Θ can vary between $\Theta=0^o$, meaning that the cluster would spread into a thin layer, and $\Theta=180^o$, meaning that the cluster would form a perfect little ball touching the substrate just at one point.
  - The second term in the equation for G_{crit} is only a function of the wetting angle and looks like this (You can generate curves like that yourself in this module).
  - Note that for $\Theta=0^o$ (or 0π) the second fraction in the equation for G_{crit} is zero, giving $G_{crit}=0$. For $\Theta=180^o$, (or π) the second fraction is $=1$ and G_{crit} then increases with the third power of the surface energy of the material to be deposited.

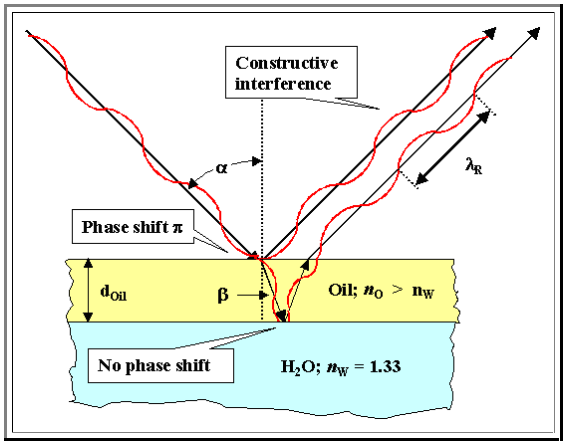
- Enough of basic nucleation theory a this point. We simply note two major points:

  - The formalism shown her is easily expanded to more complex situations (e.g. where stress and strain comes in) and, if carried out, allows to obtain valuable information about what you must expect to happen when you try to deposit **B** on **A**.
  - A lot of relevant information can already be obtained by just considering the wetting angle Θ

Interference Colors

Here is a quicky about interference colors from oil films (or any other transparent thin film). You may also want to look at

- "Interferenz und Beugung"
- "Wellen und Phasen"
- "Eigenschaften von Wellen und Teilchen"
- "Basic Optics"



- We have an oil film with thickness d_{Oil} and an index of refraction n_{Oil} that is larger than that of water (n_W **= 1.33**). An incoming wave with wavelength λ_R hits the liquids at an angle α and is reflected at the same angle and diffracted with an angle β.
- The phase of the wave reflected at the surface of the oil film "jumps" **180°** or by π - just believe it, it you don't know the reason for that.
- The diffracted wave propagates with a wavelength λ_{oil} given by $\lambda_{oil} = \lambda_R / n_{Oil}$ and is reflected without a phase jump at the oil - water boundary. At the oil - air boundary it is refracted and runs "parallel" to the wave that was directly reflected at the oil surface.

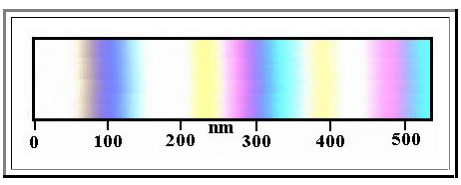
The total light reflected consists of whatever is left over after the waves reflected at both surfaces / interfaces interfered with each other.

- Only for constructive interference as drawn above, a sizeable reflection will be noticed. The equation for constructive interference of the **m**-th order is

$$2 n_{Oil} \cdot d_{Oil} \cdot \cos\beta = (m - ½) \cdot \lambda_R$$

- If you think about that for a bit, you will realize that you can measure the thickness of arbitrarily thin oil films by making β large, i.e. for "glancing" incidence. However, if you think a tiny bit harder, you will realize that there are all kinds of problems coming up in real life if you make β too large.
- Nevertheless, by doing experiments along the line drawn above, we should be able to measure the thickness of thin (transparent) films down to a fraction - **1/10th...1/20th** - of the wavelength used in the experiment, i.e. down to **10 nm - 30 nm** without much problems. Accepting problems (meaning paying much money for more sophisticated equipment) we can measure thicknesses in the **1 nm** region with optical means.

If we look at a **Si** wafer with a thin layer of SiO_2 on it, the color we will see correlates to the thickness roughly like that:



- With a little bit of experience you can make an educated guess at the thickness; but you must be sure you get the order right.
- More important than guessing a the thickness (a simple optical instrument can do that much better than you) is that you see if the thickness of the layer is uniform, because small changes of the thickness changes the color, and the eye is quite sensitive to that.

Exercise 3.1-1

Class Exercises and Quick Questions to

3.1 Thin Films - General

Here are some class exercises and quick questions:

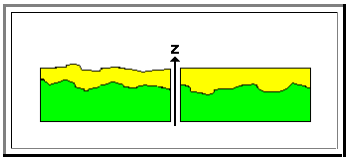
- Interference causes the color of a thin film and betrays its thickness? *Explain!*
- Give examples of what "thin" could mean in relation to *intrinsic* length scales. Provide (and discuss briefly) some intrinsic lengths, in particular with respect to semiconductors
- Give a few numbers for the meaning of "*thin*":
  - Thickness of a human hair ≈≈ **????**
  - Thickness of a gate oxide in an integrated transistor ≈≈ **????**
  - Thickness of antireflection layers of optical lenses ≈≈ **????**
  - Thickness of a thin film solar cell ≈≈ **????**
  - Other examples you can come up with **???**
- Give some examples of thin film applications *outside* of semiconductor technology.
- Give the equation for the capacity ***C*** of a parallel plate capacitor with plate area ***A*** for a maximum voltage of **10 V**. How can you achieve maximum capacity and what are the limits? Hint: Consider field strength and relvant intrinsic length scales.

Exercise 3.2-1

Quick Questions to

3.2 Thin Films - Mechanical Properties

Here are some quick questions:



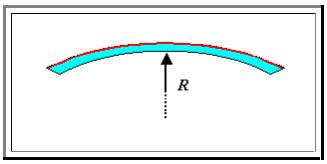
- How would you define the roughness of the two thin films shown? Give an equation if possible and differentiate between the two cases.
- Give examples for a thin layer of material **B** on substrate **A** for which you would expect good or bad adhesion, respectively: Give reasons for you expectation.
- The "surface" energy of glass is around γ**(Glas) 300 mJ/m^2**, for a metal we might have γ**(Metal)** $\approx$ **2100 mJ/m^2**. You deposit a noble metal. On which substrate would you expect better adhesion?
- Give an example of how one could measure the adhesion strength of a thin film.
- A thin layer on some substrate is either under strong tensile or strong compressive stress. Discuss and scetch what might happen if

  1. The interface energy is small, i.e. the adherence is weak.
  2. The interface energy is large.

  In the second case something must "give" Discuss possible mechanisms of strain relaxation
- Consider a thin **SiO_2** film on a thick **Si (10 × 10) mm^2** substrate that was made at **1.100 ^{0}K**. The thermal expansion coefficients are

  - $\alpha_{Si} = 3 \cdot 10^{-6}\ K^{-1}$
  - $\alpha_{quartz} = 0.6 \cdot 10^{-6}\ K^{-1}$

  What will be the stress and strain in the film at **300 ^{0}K**? How does the elastic energy stored in the film scale with its thickness d_z?
- Besides a mismatch in the coefficient of thermal expansion, other mechansim can produce stress in thin films, too. Do you know any or can you think of any?



- The red thin layer (thickness d_B) on the blue circular **Si** wafer substrate (thickness $d_A >> d_B$) is under compressive stress σ; the wafer thus is warped with a radius of curvature = ***R***.
  What would ***R*** be proportional to?
  Hint: It is a two-dimensional problem.
- Imagine, that in the picture on the right you deposited the red film also on the backside. What would the radius of curvature be now? The stress in the two layers? Now imagine that yopu keep processing this system until, for exampe, you get a transitor. Thjis involves structuring the top layer, i.e. etching part of the layer off. When would you take the layer on one side off, if you don't need it for whatever you try to produce?

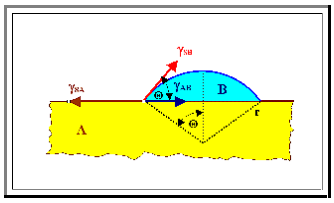
Exercise 3.3-1

Quick Questions to

3.3 Nucleation and Growth

Here are some class exercises and quick questions:

- What happens when first incoming atoms hits the surface of the substrate? Give at least 4 different possibilities.
- Where would you expect the first imcoming atoms to be solidly bound? Use the proper terminology.
- Define "sticking coefficient". Discuss the dependenc of the sticking coefficient for a given system on the precise substrate condition for a given substrate.
- Explain briefly the major methods for investigations of the nucleation of thin films on substrates.
- Explain how you get from interface energies to forces, and from forces to the wetting angle Θ
- Discuss and name the two major growth mode following from extreme values of Θ
- Discuss and name a third major growth mode



Exercise 3.4-1

Quick Questions to

3.4 Structure, Interface and Some Properties

Here are some quick questions:

- What is epitaxial growth? Consider the possibility of epitaxial growth; giving possible conditions (e.g. with respect to structures, lattice constants, ...) and use simple pictures:
  - **A** on **A**.
  - **A** (**fcc**) on **B** (**fcc**).
  - **A** (**fcc**) on **C** (**hex**).
  - **A** (**fcc**) on **B** (**fcc**) with intermediate layer.
  - ….
- **B (fcc; (100))** with lattice constant a_A is deposited on **A (fcc; (100)** with $a_B = 0.95\ a_A$. Sketch the structure for
  - Thickness of **B** only a few atomic layers.
  - Thickness of **B > 50 nm**
- *Difficult!* Sketch a pure edge misfit dislocation network on a **{100}** interface plane for a misfit of **10 %** for the case of
  - Burgers vector of the dislocations is $\underline{b} = \underline{a}$**<100>**.
  - Burgers vector of the dislocations is $\underline{b} = \underline{a/2}$**<110>**.
- *Difficult!* What would happen if the (square) network of misfit dislocations on a **{100}** type interface would be changed from edge dislocations to screw dislocations?
- What are the energetic reasons for introducing misfit dislocations into epitaxial interface it the layer thickness is larger than a critical thickness? What determines the critical thickness?
- Sketch the curve for the critical thickness d_{crit} in a d_{crit} - misfit diagram, Try to give approximate numbers.
- Enumerate and discuss *structures* obtainable with thin films but not (easily) with bulk materials. Give examples for applications.
- Give reasons why thin film properties can be quite different from bulk properties; give examples.
- Name some technologically extremely important special thin film properties; discuss with actual numbers.

Exercise 3.6-1

All Class Exercises and Quick Questions to

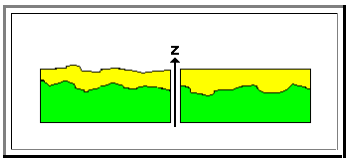
3. Thin Films

Subchapter 3.1: Thin Films - General

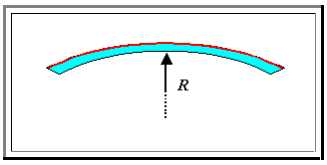
- Interference causes the color of a thin film and betrays its thickness? *Explain!*
- Give examples of what "thin" could mean in relation to *intrinsic* length scales. Provide (and discuss briefly) some intrinsic lengths, in particular with respect to semiconductors
- Give a few number for the meaning of "thin":
  - Thickness of a human hair ≈≈ **????**
  - Thickness of a gate oxide in an integrated transistor ≈≈ **????**
  - Thickness of antireflection layers of optical lenses ≈≈ **????**
  - Thickness of a thin film solar cell ≈≈ **????**
  - Other examples you can come up with **???**
- Give somer examples of thin film applications outside of semiconductor technology.
- Give the equation for the capacity ***C*** of a parallel plate capacitor with plate area ***A*** for a maximum voltage of **10 V**. How can you achieve maximum capacity and what are the limits? Hint: Consider field strength and relvant intrinsic length scales.

Subchapter 3.2: Mechanical Properties

- How would you define the roughness of the two thin films shown? Give an equation if possible and differentiate between the two cases.

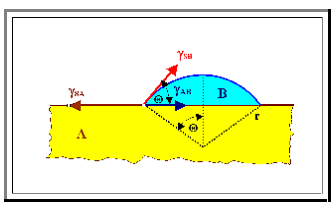


- Give examples for a thin layer of material **B** on substrate **A** for which you would expect good or bad adhesion, respectively: Give reasons for you expectation.
- The "surface" energy of glass is around γ**(Glas) 300 mJ/m^2**, for a metal we might have γ**(Metal)** ≈ **2100 mJ/m^2**. You deposit a noble metal. On which substrate would you expect better adhesion?
- Give an example of how one could measure the adhesion strength of a thin film.
- The red thin layer (thickness d_B) on the blue circular **Si** wafer substrate (thickness $d_A >> d_B$) is under compressive stress σ; the wafer thus is warped with a radius of curvature = ***R***.
  What would ***R*** be proportional to?
  Hint: It is a two-dimensional problem.



Subchapter 3.3: Nucleation and Growth

- What happens when first incoming atoms hits the surface of the substrate? Give at least 4 different possibilities.
- Where would you expect the first imcoming atoms to be solidly bound? Use the proper terminology.
- Define "sticking coefficient". Discuss the dependenc of the sticking coefficient for a given system on the precise substrate condition for a given substrate.
- Explain briefly the major methods for investigations of the nucleation of thin films on substrates.
- Explain how you get from interface energies to forces, and from forces to the wetting angle Θ



- Discuss and name the two major growth mode following from extreme values of Θ
- Discuss and name a third major growth mode

Subchapter 3.4: Structure, Interface and Some Properties

- What is epitaxial growth? Consider the possibility of epitaxial growth; giving possible conditions (e.g. with respect to structures, lattice constants, ...) and use simple pictures:
  - **A** on **A**.
  - **A** (**fcc**) on **B** (**fcc**).
  - **A** (**fcc**) on **C** (**hex**).
  - **A** (**fcc**) on **B** (**fcc**) with intermediate layer.
  - ….
- **B (fcc; (100))** with lattice constant a_A is deposited on **A (fcc; (100)** with $a_B = 0.95\ a_A$. Sketch the structure for
  - Thickness of **B** only a few atomic layers.
  - Thickness of **B > 50 nm**
- *Difficult!* Sketch a pure edge misfit dislocation network on a **{100}** interface plane for a misfit of **10 %** for the case of
  - Burgers vector of the dislocations is $\underline{b} = \underline{a}$**<100>**.
  - Burgers vector of the dislocations is $\underline{b} = \underline{a/2}$**<110>**.
- *Difficult!* What would happen if the (square) network of misfit dislocations on a **{100}** type interface would be changed from edge dislocations to screw dislocations?
- What are the energetic reasons for introducing misfit dislocations into epitaxial interface it the layer thickness is larger than a critical thickness? What determines the critical thickness?
- Sketch the curve for the critical thickness d_{crit} in a d_{crit} - misfit diagram, Try to give approximate numbers.
- Enumerate and discuss *structures* obtainable with thin films but not (easily) with bulk materials. Give examples for applications.
- Give reasons why thin film properties can be quite different from bulk properties; give examples.
- Name some technologically extremely important special thin film properties; discuss with actual numbers.

Some Thermal Expansion Coefficients

Illustration

The following table lists some thermal expansion coefficients:

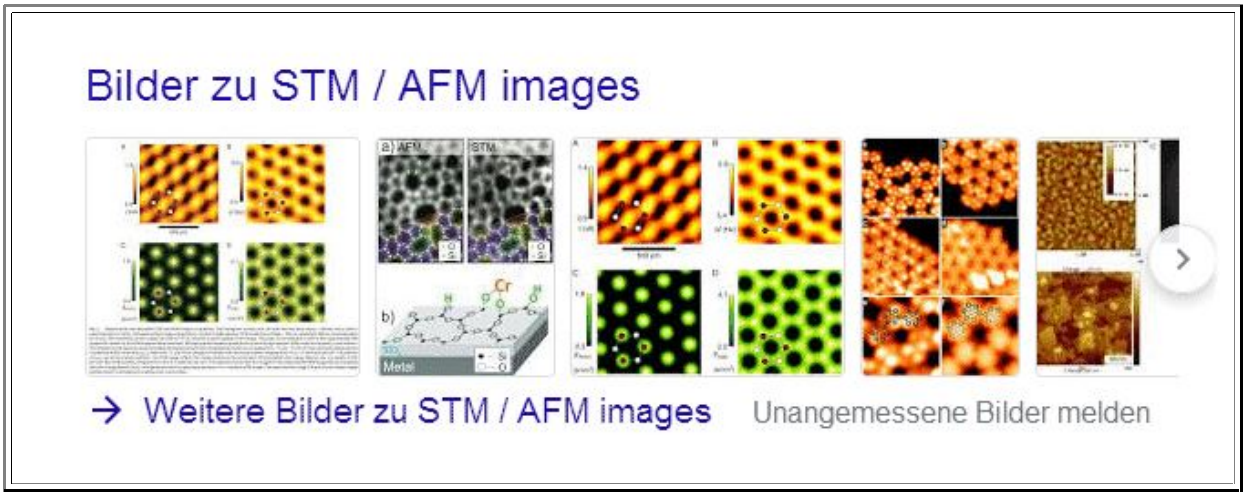
- Numbers are mostly from Wikipedia. Don't just believe it! Always counter check numbers found somewhere by at least going to one more source, e.g. from this link

Group	Metal	$\alpha \cdot 10^{-6}$/ at 20°C		Element	$\alpha \cdot 10^{-6}$/ at 20°C		Various Materials	$\alpha \cdot 10^{-6}$/ at 20°C
Ib	Cu	16.5		Antinomy	10.5		Aramid	-4.1
	Ag	19.5		Beryllium	12.3		Concrete	6 - 14
	Au	14.2		C (Diamond)	1.3		Bronze	17.5
IIa	Mg	26.0		Germanium	6.0		Ice	51
	Ca	22		Silicon	2.0		Rubber	160 - 220
IIb	Zn	36.0					Glas	4.7 - 7.6
	Cd	41					Quartz (amorphous)	0.5
	Hg	0.182					Ceramic ("Zerodur")	0.1
IIIa	Al	23					Granite	3.0
IVa	Sn	26.7					Graphite	2.0
	Pb	29.3					NaCl	40
VIIIb	Fe	12.2					Carbon fibre	-0.5
	Ti	10.8					Brass	18.4
	Ni	13.0					Nylon	120
Vb/VIb	Ir	6.5					PVC	50 - 240
	Cr	6.2					Steel (stainless)	14.4 - 16
	Mo	5.2					**Liquids**	
	W	4.5					Alcohol	1.1
VIIIb	Rh	9.8					Glycerin	0,49
	Pd	11.2					Water	0,21
	Pt	9.0					Acetic Acid	1.07
You find the values for all elements in the links in this periodic table								

Some STM / AFM Pictures

Illustration

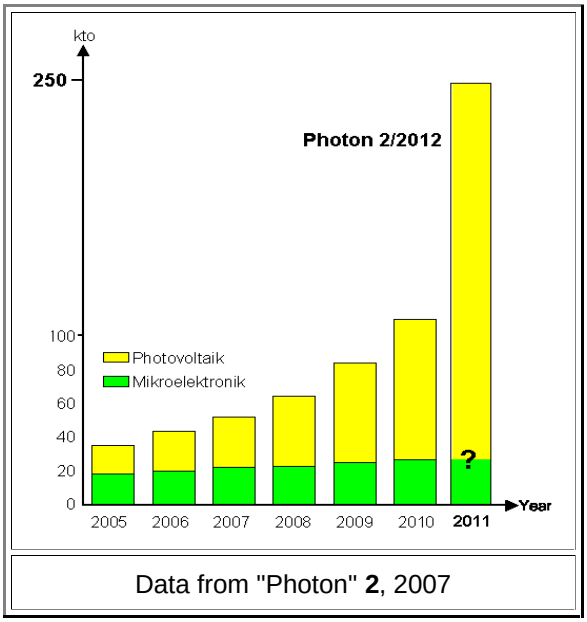
I've given up competing with the internet. Search there; you get something like this:



The Great Si Crisis in 2007

Advanced

- While today's buzzwords like "**energy crisis**", "**climate change**" and "greenhouse effect" have finally penetrated even the most simple minds on these planets (even George W. **Bush** knows the meaning of this words by now), more sophisticated minds knew about this already in the **70**ties of the last century.
  - That "the future" would only have a future if we would manage in time to switch from burning coal and carbohydrates to energy supplies relying on "solar" - wind, water, biofuel, solar heat and direct solar electricity via solar cells - was clear to everybody who knew a minimum about energy conservation laws and the effect of greenhouse gases. One might consider nuclear power in this context, but we won't.
  - What happened, or better what did not happen after the first and second "oil crisis" in the **70ties** (look it up in the Net), is history now. It can be safely predicted that in years to come, historians looking back on this chapter of global history will take a dim view of some present day heroes.
- Anyway, by some cunning politics and good old dumb luck, Germany started a thriving solar industry a few years after the millennium; a few other countries (foremost Japan), did the same.
  - *Predictably*, real money was made with the most old-fashioned technology: more or less "classical" solar cells made from bulk **Si** - either from (round) single crystal wafers or from so-called (square-shaped) multi-crystalline slices. Starting around **2003**, factories were built like crazy, and new companies like Q-Cells or Solarwold turned almost over night into stock market stars worth hundreds of Million **€**.
  - *Predictably*, electronic grade **Si** (even somewhat less "good" **Si** from minor suppliers) turned scarce and prices shot up. Solar cell companies without reliable sources of **Si** (i.e. without long-term contracts with major suppliers) had to give up; while others grew with growth rates only limited by how fast you could hire people and built factories (about **40 %** per year).
  - Given the growth rates of the solar industry *and* of the microelectronic industry and projecting that on the **Si** production capacities of the existing companies (just about a handful), it was pretty clear that electronic grade **Si** would become scarce around **2005**. However, given the general experience in big business, it was also clear that not much would happen *before* this predicted scarcity was really felt.
- Now, in Feb. **2007**, big articles about the **Si** crisis appear rather regularly in the relevant magazines and lots of money is diverted to electronic grade **Si** production.
  - Here are a few facts; mostly taken from the Feb. **2007** "Photon".
- Major **poly-Si** (always electronic grade now) producers like Wacker (Germany) MEMC (USA) and others are building new plants for **poly-Si** production.
  - Besides just cranking up the established production with the **"Siemens" process"** as described in the backbone; new variants of the basic **CVD** process are being tried and moved to production. Extremely simplified, instead of having a thin **Si** rod grow by depositing **Si** in a relatively small reactor, "**Si** dust" is fed into the top of a huge heated tower, where it slowly sinks down in a strong upward draft provided by silane and hydrogen gas (plus the doping gases) fed into the tower at the bottom.
  - The dust particles grow while sinking; if all goes well, you can eventually shovel out some **Si *gravel*** at the bottom.
  - Sound simple, is not. Just the **PH_3** or **B_3H_6** inside a decent sized reaction tower would be sufficient to wipe out a big city if it ever gets out.....
  - As a very raw but interesting number in this context we note that a production capacity of **10.000 to/a** will cost you **$(0.5 - 1) \cdot 10^9$ €** initial investment; you also need the (very secret and closely guarded) know-how.
- Here is a prediction of how much **Si** we will need in years to come. The average growth rate is about **25 % /a**; accelerating as time goes by.



Data from "Photon" **2**, 2007

Here is some of the present (Nov. **2007**) gossip around the **Si** crisis:

- "Orkla ASA" (Norwegian company) builds a **323 Mio €** plant for producing solar **Si** by direct cleaning of metallurgical **Si**. Capacity is **5.000 to/a** in **2009**. "ELKEM", a Orkla daughter company, is the world's largest producer of metallurgical **Si**.
- "Dow Corning" (USA company; not much previous **Si** experience) announces end of **2006** (as a surprise) that it will produce **1.000 to/a** of solar-grad **Si** by some process for cleaning metallurgical **Si** (in Brazil).
- **4** other companies are making similar, but less specific announcements.
- About **20** more companies (including Chinese) are suspected to secretly work on some **Si** production process based on cleaning metallurgical grade **Si**.
- **1 kg** of solar **Si** presently costs about **50 €** (if you have a good contract with a supplier), on the "spot market" you may have to shell out **150 €/kg**. Prices have gone up **30 %** and more during the last **3** years.
- Some numbers from the **May 2008** "Photon": Average price now about **70 $ /kg**; Spot market price up to **515 $/kg**; production costs about **36 $/kg**. Looks like the crisis is still with us. About **70** companies now make solar **Si**, about **100** more have announced to get into this market.
- The projected **110 kto/a** in **2010** is still not enough for the expected market for solar cells. This may mean, that the thin film solar cells, which so far are just a niche production, may come into there own.

Obviously, the new game in town is to take cheap metallurgical grade **Si** (at around **1 €/kg**) and to clean it "directly", avoiding the costly (**CVD**) Siemens process. It's actually not all that new; I have worked along those general lines around **1991** in the Siemens labs in Munich myself.

So how is it done? Few people know - whatever is going on is ultra secret, and all the players in this field have their own little "dirty" or better "cleaning" tricks. Before we look at some of those tricks, we will look at the money side of things a bit more.

The only definition of what defines "**solar grade" Si** as a subgroup of "electronic grade **Si**" is how good the solar cell will be that is manufactured with the stuff; in other words what kind of cell efficiency η (electric power out/ Light power in (in %)) you can get with your process.

A few rough numbers to that are: If the best efficiency you can get is below **η = 13%** - forget it. Even if your **Si** would be for free, it's too expensive. On the other hand, if you can get an **η = 15%**, you can ask for about **40 €/ kg**. This just shows how dramatically important the cell efficiency is (and how difficult to increase it relative to the best you can do).

In other words: If you make or loose huge amounts of money by going into the **Si** business depends on many variables most of which you don't know all that well.

Last, let's look at some ideas of how to clean metallurgical grade **Si** cheaply and efficiently. First, we have to realize that what works for some impurity many not work for some others, so we have to group impurities in e.g. metals, doping elements, **O** and **C**, and so on. A few typical processes might be

- Melt the stuff and throw some slag-formers in the melt (e.g. **$CaSiO_2$**) Doping elements like **B** and **P** might prefer to be in the slag, which you then spoon off with a ladle (haha).
- Melt the stuff and blow gases through the melt. With some luck (or special knowledge) - see above).
- Solidify your melt in a suitable way ("directional solidification") and use "segregation" to move a lot of the (metallic) impurities into that part of the melt that solidifies last (which you throw out then).
- Bath your (finely crushed) **Si** in suitable acids and lyes and dissolve the dirt sitting on the surface. If you crush material full of defects (particularly grain boundaries) after some treatment that drove your impurities into the defect sites, and if you crush it in such a way that fracture occurs at grain boundaries, you may now chemically remove a lot of the dirt.
- Keep our energy needs down. Melting a **kg** of **Si** takes a certain amount of **kWhr**; so does heating your equipment. The Siemens process runs up a bill of about **100 kWhr / kg**; cleaning metallurgical grade **Si**

may keep it a **15 kWhr/kg**.

- Don't just take any metallurgical grade **Si** - make your own with "optimized" dirt in there. Find a source of relatively clean coal and quartz - at least with respect to the more trickier impurities.

Let's stop here. If you get my drift, you will now understand that there are many interesting and demanding jobs out there for engineers who know their electronic materials, and that the **Si** crisis can and will be solved very quickly. Exactly how remains to be seen.

As an afterthought: The situation has many parallels to the introduction of steel mass production of the second half of the **19**th century. People were fighting impurities, had no clear idea of exactly what was needed, and found a lot of working solutions that were only understood later.

Jan Czochralski

Advanced

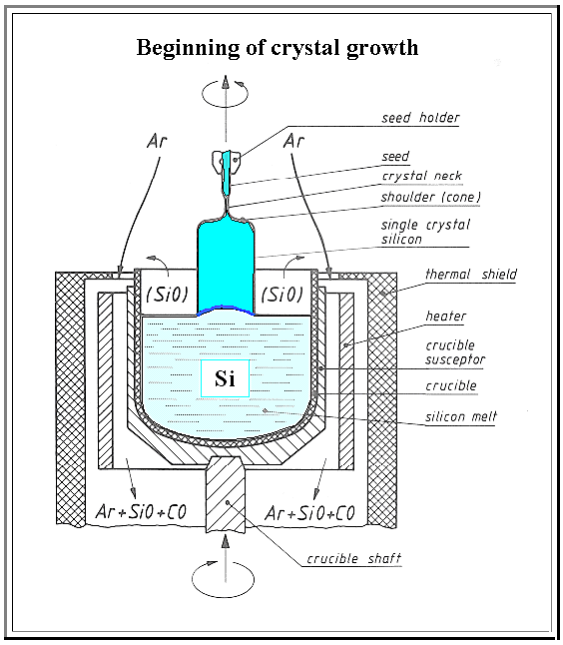
- While everybody in the semiconductor world knows the "*Czochralski grown crystals*", (almost) nobody knows Jan **Czochralski**.
  - He was born **1885** in Kcynia, a small town in Western Poland, which then was part of Prussia; i.e. the German empire. Around **1900** he moved to Berlin, **1907** he worked as an engineer for the "Allgemeine Elektrizitäts Gesellschaft (AEG), a formerly large and famous company, defunct since about **1980** (the last remnants were swallowed by Daimler-Benz). During this time he also studied Chemistry and Metallurgy at the "Charlottenburg Technical University" and Fine arts at the Berlin university.
  - **1917** he organized the well-equipped research laboratory of "Metallbank und Metallurgische Gesellschaft"; and he was the director of that institution until 1928.
  - **1929** he moved back to Poland to become a Professor of Metallurgy and Metal Research in in the Chemistry Department of the Technical University in Warsaw.
  - All the time he was involved in metallurgical research, active in many scientific organizations and advising companies and other professional entities. The second world war put an end to all this - J. Czochralski was active in supporting his co-workers and the polish underground army.
  - After the war he returned to his native town of Kcynia were he run a small drug firm. He died in **1953** and is buried in the family tomb in Kcynia.
- He was quite well-known for many contributions in metallurgy, especially for this studies of the velocity of metal crystallization, and, of course, for his method of monocrystalline growth, published in **1916**.
  - "**Czochralski grown crystals**" are a cornerstone of modern material science, and J. Czochralski will be remembered for his insights leading to this method.
  - But like many scientists, he was also interested in literature, music and painting and not afraid of branching out, as witnessed by his activities after the war.
  - He is one of the most famous Polish scientists and the Foundation for Materials Research Development in Poland established the Prof. Jan Czochralski Gold Medal for achievements in materials science.
- *Note added* after consulting an Internet article from *Dr Pawel Tomaszewski* that is no longer in the Net.
  - What I did not know was that the discovery of the Czochralski crystal growth methods was one of those **accidental discoveries**; here a quote from the article mentioned above
  - "DISCOVERY OF THE CZOCHRALSKI METHOD
    The Czochralski method of growing single crystals brought Jan Czochralski his greatest publicity. The method was developed in **1916** and was initially used to measure of crystallization rate of metals. The method was developed as a *result of an accident* and through Czochralski's careful observation.
    One evening he left aside a crucible with molten tin and returned to writing notes on the study carried out on a crystallization study. At some moment, lost in thoughts, instead of dipping his pen in the inkpot, he *dipped it in the crucible* and withdrew it quickly. He observed then a thin thread of solidified metal hanging at the tip of the nib. The discovery was made!
    The nib slot, in which crystallization was initiated, was replaced by a special narrow capillary and in some cases by a seed of the growing crystal. Czochralski checked later that the crystallized wire was a single crystal. The crystals obtained in that way had diameters of about a millimeter and lengths up to **150 cm**. Czochralski published a paper on the study of the rate of crystallization of tin, zinc and lead, and the maximum rate of pulling of a crystal was recognized as the characteristics of the crystallizing material (Ein neues Verfahren zur Messung des Kristallisationsgeschwindigkeit der Metalle [A new method for the measurement of crystallization rate of metals], Z. phys. Chem. 92, 219-221 (1918); the paper was received in the editorial office on 19 August 1916)."

The Science and Art of Si Crystal Growth

Advanced

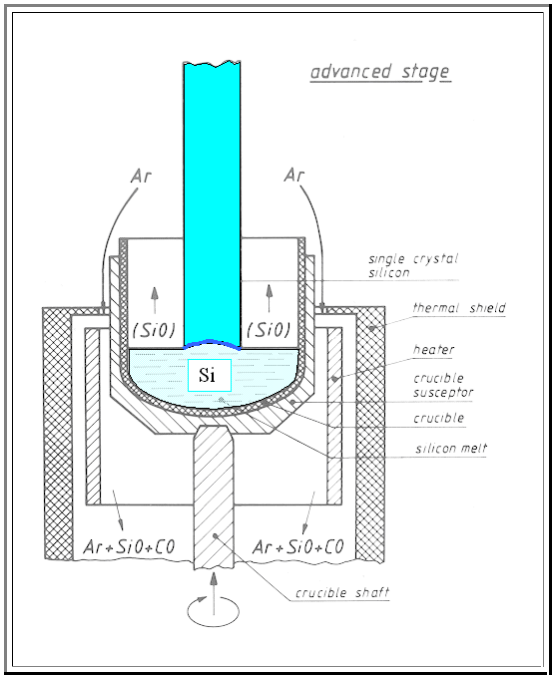
There are a few major points in Crochalski (or **CZ**) crystal growth that shall be outlined briefly. In particular, we will discuss

- The "**Dash**" **process** (or "necking") for the production of dislocation free crystals.
- **Bulk microdefects**, or the question of what happens to the point defects present in thermal equilibrium upon cooling down?
- **Convection** in the melt and the incorporation of oxygen.



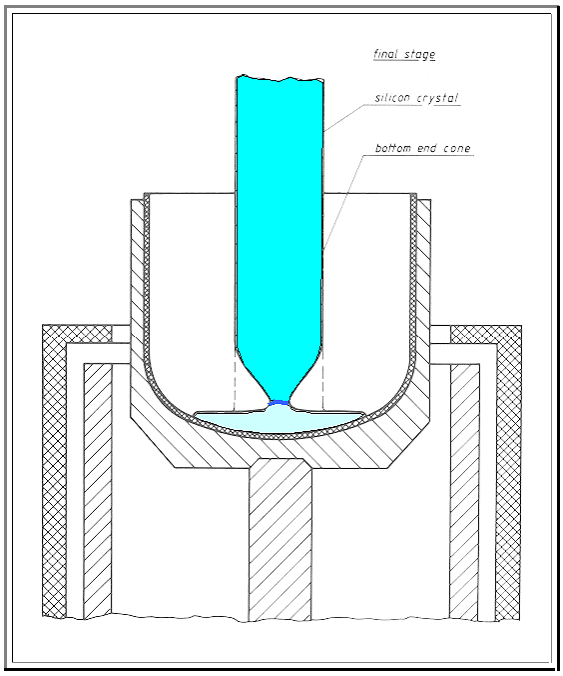
Dash Technique

- How do you obtain perfectly dislocation free crystals? Always considering that your seed crystal may not be dislocation free, and even if it is, the unavoidable thermal shock upon dipping the seed crystal in the melt will almost certainly produce some dislocations (this is real easy at temperatures close to the melting point).
- You just do two things:
  1. Reduce the strain as driving force for dislocation movement and multiplication, and
  2. Make sure that existing dislocations are never lined up in growth direction, so that they will sooner or later terminate at the surface.
- Both points are satisfied if you just make the first part of the growing crystal very thin, i.e. you form a "neck" as shown on the left.
- This technique was pioneered by W.C. **Dash** in **1959**. It was one of the many "little" inventions necessary to allow Si technology.



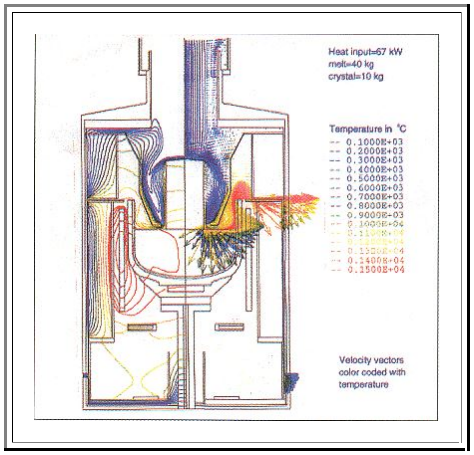
Bulk Micro Defects

- Just after crystallization, the crystal is at a temperature close to the melting point, and thus contains the maximum equilibrium concentration of point defects (vacancies and interstitials). As it cools down, the equilibrium concentration goes down to practically zero at room temperature.
- But how? In normal crystals the point defects disappear at grain boundaries or dislocations - whatever internal sinks there are. In dislocation free single crystalline **Si** there is only the surface available as sink - and for most point defects, the surface is far away and will never be reached by diffusion.
- What will happen is that some, if not most, of the point defects form small clusters or agglomerates, the so-called bulk micro defects, (***BMD***), ***COPs*** (crystal originated particles or pits), or, as they were called in the eighties, swirl defects.
- They are unavoidable, and they are bad for **ICs**. All you can do is to try and make them very small. Manipulating these defects is one oft the recurrent themes in **Si** crystal growth, made difficult by the fact that - quite ironically - we know far less about point defects in **Si** than in metals.



Convection Effects

- Even if crystal and crucible would not be rotated, there would be convection in the melt, driven by density gradients coupled to temperature gradients. The only way to totally avoid this would by to switch off gravity. This is the reasoning behind all those "*Very Important*" experiments with crystal growth in space.
- Don't fall for it. First, you also can suppress convection with a strong magnetic field (its actually done for large diameter crystals) at an incredibly tiny fraction of the prize for space-grown crystals, and secondly, why should you suppress convection?
- Well, uncontrolled convection is bad for many reasons. Lets just look at one: The melt streaming by the crucible walls dissolves some **SiO_2** and later deposits it into the crystal. The **O** - concentration then is not only higher compared to convection-free melts, but its radial distribution mirrors the flow pattern - and nothing is worse than inhomogeneities in **Si** crystals.
- Does this mean that you do not want convection? No - all it means that want to have convection that is precisely right for your crystal at every stage of its growth. If you have that, you can control **O** - concentration and keep the doping concentration about constant through the length of the crystal despite the effects of segregation.



Science and Art

- So, in growing a large **Si** crystal, you have a number of buttons to fiddle with. which you must turn in the right directions as your crystal grows. Consider the major parameters:
  1. Growth speed,
  2. Rotation of the crystal,
  3. Rotation of the crucible - same or opposite direction,
  4. Magnetic field (strength and direction)
  5. Power fed to crucible.
- Changing anyone of this parameters a little bit will influence everything in a way not easily cast into formulas - diameter of the crystal, **BMD** size and distribution, **O**-concentration and distribution, dopant concentration and distribution, and much more.
- Growing good crystals therefore certainly was (and to some extent still is) an **art**.
- Only in the nineties of the **20.** century, theory and simulation progressed enough to be of help. Now it is indispensable.
- The picture on the left shows one result: the complete temperature distribution in the whole machine (including, of course, melt and crystal) and the flow field in the melt (little arrows). Now you can change some parameters and see what will happen without actually growing (and later destroying for analysis) an expensive crystal.

Float Zone Crystal Growth

Advanced

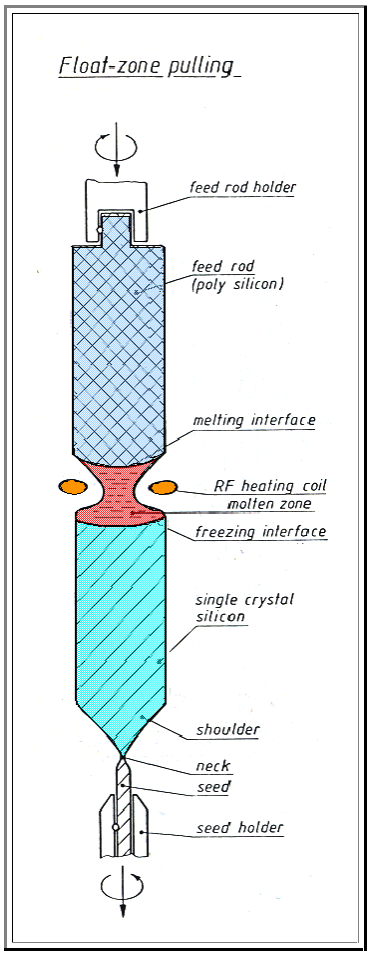
The basic idea in float zone (**FZ**) crystal growth is to move a liquid zone through the material. If properly seeded, a single crystal may result.

- The method was first used for purification (**zone melting**), taking advantage of the small segregation coefficients of many impurities. The impurities contained in the feed material would then prefer to remain in the melt and thus could be swept to the end of the feed stock.

- If properly done, the newly crystallizing material could be obtained as a single crystal. Again, it was the **Siemens AG** that pioneered the use of **FZ** crystal growth for the production of high-quality **Si** single crystals.

- Since the melt never comes into contact with anything but vacuum (or inert gases), there is no incorporation of impurities that the melt picks up by dissolving the crucible material as in the **CZ** crystal growth method. This is especially true for oxygen, which can not be avoided in **CZ** crystal growth. **FZ** crystals therefore are always used when very low oxygen concentrations are important.

- The problem of **FZ** crystal growth is clear, however, if one looks at the drawing: How do we keep the liquid **Si** from just collapsing? If it would only be held in place by *surface tension*, the maximum diameter of crystals possible in this way would be about **20 mm** - not very useful. There are, fortunately, other stabilizing mechanisms, and drawing the liquid zone through a "hole" - as indicated - also helps. Still, for large diameter crystals the difficulties grow rapidly and **FZ** crystal growth is rarely (if at all) used for diameters larger than **150 mm**.

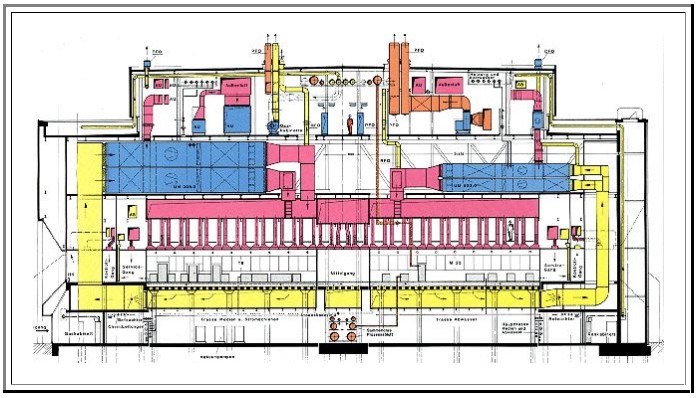


Cleanrooms

Advanced

- First. let's look at the cross-section of a typical "class **1**" cleanroom. Class **1** means roughly that there will be at most **1** particle per **foot^3** (about **30** liters) larger than **0,2** µm or so in the air.

.



  - Even in the small illustration you can see that the "actual" cleanroom where people make chips, is a small part of the building (the whitish portion just above the lower yellow part).
  - Everything colored is just for moving air around, keeping its temperature and humidity constant, add some fresh air from the outside and to get rid of "spent" air.
- A particular interesting place in a cleanroom building is the "basement" right under the actual cleanroom. It houses a large part of the "equipment", e.g. pumps, liquid and gas inlets, outlets, and cleaning parts, transformers, power equipment, heaters etc. It also houses miles of tubing for delivering away and taking gases and liquids. Some pictures:

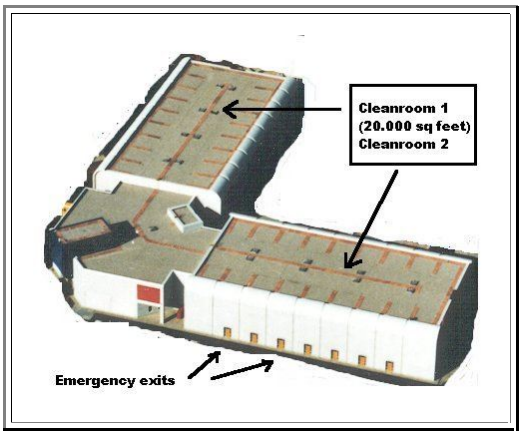
Water treatment

Pumps and piping

- Real cleanrooms are shown in the next two pictures (from the Siemens compound in München-Perlach)

- The building on the lower left are cleanrooms; the "little" one (**1000 m^2**) to the left (with the blue topping) was the **1 µm** research line, the two bigger ones (at right angles; **2000 m^2** each) were used for the development of the **4Mbit** and **16 Mbit DRAM** and for the pilot production.

Below an enlargement



- The yellow emergency exits indicate the actual cleanroom. There are several stories above, and two stories - not visible of course - below.
- The connecting building houses parts of the common infrastructure:
  - Air intake and initial processing
  - Water plant
  - Recycling and cleaning of liquids
  - Cleanroom control
  - Main entrance for heavy equipment
  - Shipping and receiving

Multiple Choice Test zu

4. Getting Started

Start Multiple Choice

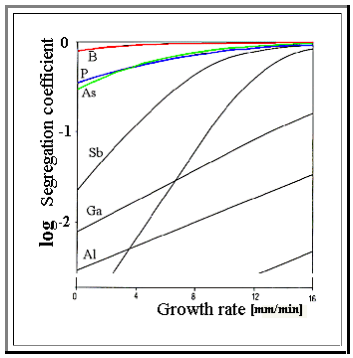
Exercise 4.1-1

Quick Questions to

4.1 Input to Si Processing in an Industrial Environment

Here are some quick questions:

- List (and discuss briefly) some essential inputs to a chip factory.
- What is the essential process for producing raw (= metallurgical) **Si** and what is the major use for this **Si**?
- Go through the essential of **Si** single crystal growth by the **CZ** technique. Give numbers and discuss in-situ doping, keeping the crystal dislocations-free, and any remaining problems.
- Describe shortly the essentials of how to obtain clean, doped poly-**Si** as needed for single crystal growth
- Where and why is a **CVD** process involved in making electronic grade **Si**?
- Describe the phenomenon of segregation. How does it impact **Si** crystal growth?
- Given the diagram on the right, discuss:
  - What a segregation coefficient of , e.g., $\mathbf{10^{-2}}$ means in terms of the concentration in the crystal in the beginning and the end of the crystal growth process if the initial concentration in the melt is $\mathbf{10^{-6}}$
  - Why you prefer **As** to **Sb** as a dopant during crystal growth .



- Why is extreme flatness an essential condition for standard **Si** wafers?
- Why is it possible to keep wafers completely free of dislocations, but not of "microdefects" = agglomerates of point defects?

Exercise 4.2-1

Quick Questions to

4.2 Other Semiconductor Crystal Growth Technologies

Here are some quick questions:

- Describe some problems encountered (and the solutions) when growing **III-V** single crystals.
- What are the incentives for trying to get **SiC** "to work"? Describe the specific problems encountered when growing **SiC** single crystals.
- Provide and describe major products not based on single crystal semiconductors.
- Explain the follwong abbreviations and give possible uses: **a-Si:H**, **µc-Si:H**, **CIGS**.

Exercise 4.3-1

Quick Questions to

4.3 Infrastructure

Here are some quick questions:

- "Cleanroom class **100**" means roughly....?
- Describe the effects of *particles on*, and *contamination in* a chip.
- The electrically crucial area of an integrated transisitor is **(50 × 300 ×300) nm^3**. The lattice constant of **Si** is roughly **0.5 nm**; there are **....** atoms per elementary cell. An **Au** concetration of **1 ppb** "kills" the transistor.
  - How many **Au** atoms can you tolerate in your transistor?
  - If you touch a gold ring and **1** out of **1** billion surface atoms gets stuck on your finger - how many, roughly, are now on your skin?

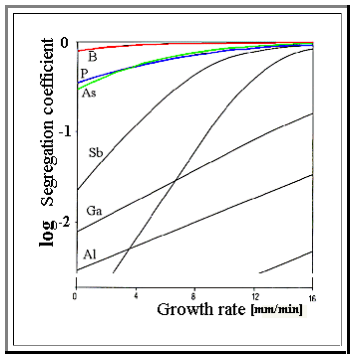
Exercise 4.4-1

All Quick Questions to

4. Getting Started

Subchapter 4.1: Input to Si Processing in an Industrial Environment

- List (and discuss briefly) some essential inputs to a chip factory.
- What is the essential process for producing raw (= metallurgical) **Si** and what is the major use for this **Si**?
- Go through the essential of **Si** single crystal growth by the **CZ** technique. Give numbers and discuss in-situ doping, keeping the crystal dislocations-free, and any remaining problems.
- Describe shortly the essentials of how to obtain clean, doped poly-**Si** as needed for single crystal growth
- Where and why is a **CVD** process involved in making electronic grade **Si**?
- Describe the phenomenon of segregation. How does it impact **Si** crystal growth?
- Given the diagram on the right, discuss:
  - What a segregation coefficient of , e.g., $\mathbf{10^{-2}}$ means in terms of the concentration in the crystal in the beginning and the end of the crystal growth process if the initial concentration in the melt is $\mathbf{10^{-6}}$
  - Why you prefer **As** to **Sb** as a dopant during crystal growth .



- Why is extreme flatness an essential condition for standard **Si** wafers?
- Why is it possible to keep wafers completely free of dislocations, but not of "microdefects" = agglomerates of point defects?

Subchapter 4.2: Other Semiconductor Growth Technologies

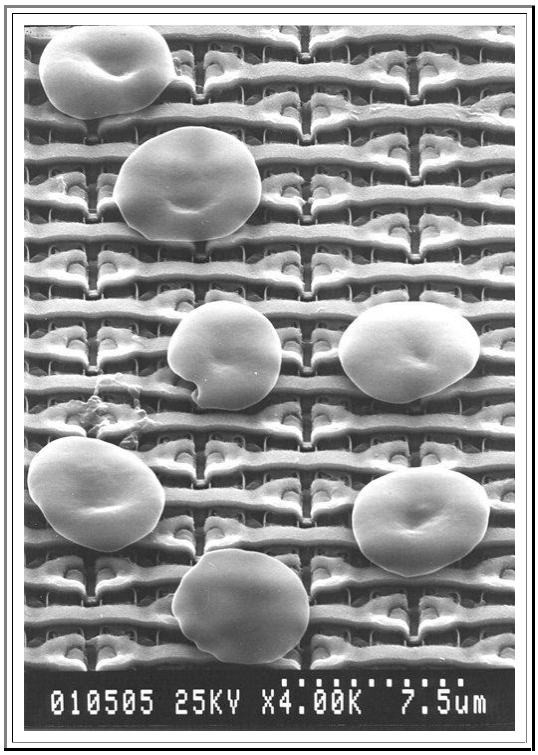
- Describe some problems encountered (and the solutions) when growing **III-V** single crystals.
- What are the incentives for trying to get **SiC** "to work"? Describe the specific problems encountered when growing **SiC** single crystals.
- Provide and describe major products not based on single crystal semiconductors.
- Explain the follwong abbreviations and give possible uses: **a-Si:H**, **µc-Si:H**, **CIGS**.

Subchapter 4.3: Infrastructure

- "Cleanroom class **100**" means roughly....?
- Describe the effects of *particles* on, and *contamination* in the chip.
- The electrically crucial area of an integrated transisitor is **(50 × 300 ×300) nm^3**. The lattice constant of **Si** is roughly **0.5 nm**; there are .... atoms per elementry cell. An **Au** concetration of **1 ppb** "kills" the transistor. How many Au atoms can you tolerate in your transistor? If you touch a gold ring and **1** out of **1** billion surface atoms gets stuck on your finger - how many roughly are now on your skin?

Particles on Chips: Blood Cells

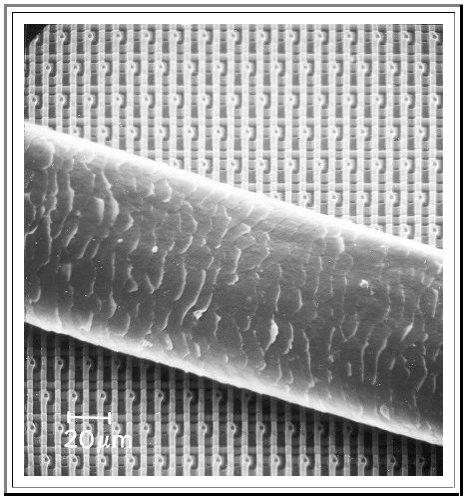
Here is the large size picture of red blood cells on a **1 Mbit** memory



Illustration

Particles on Chips: Hair

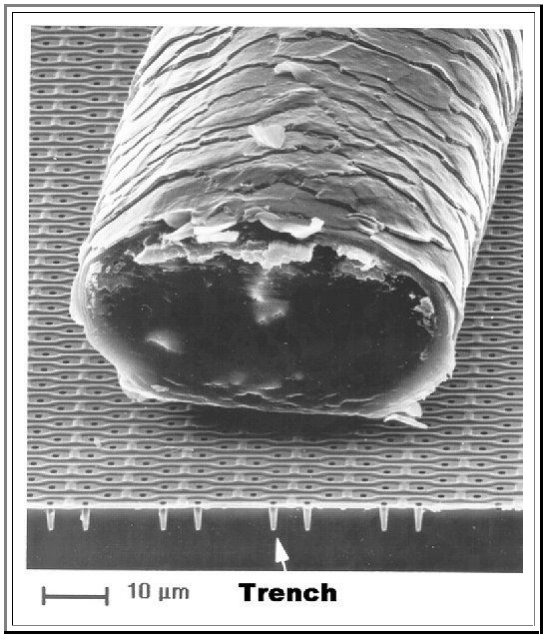
Here is the large size picture of a hair (from a female) on a **256 kbit** memory chip. Just one of the little flakes of the hair would be enough to cover one memory cell.



Illustration

Particles on Chips: More Hair

Here is another the large size picture of a hair on a **4Mbit** memory chip. It also show hows structure sizes decreased - compare with the hair on a **256 kbit DRAM**.

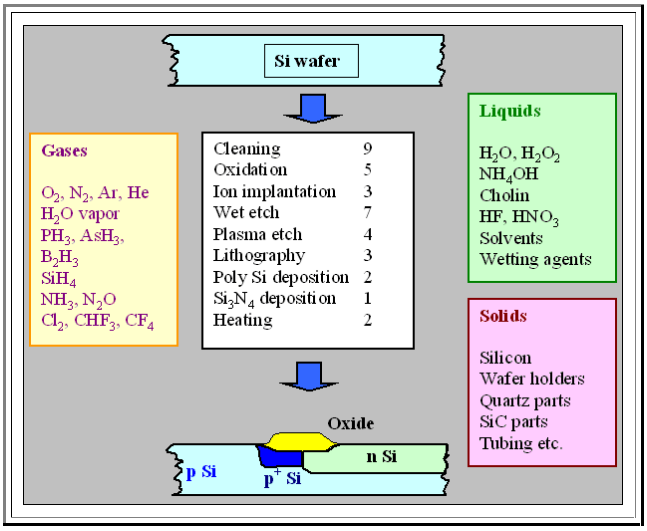


Illustration

Processes and Materials I

Illustration

Here is the list of processes and materials needed for the **16 Mbit DRAM** (an about **1999**) in a graphic way.



Note that any material that comes in contact with the wafer or with materials that will come into contact with the wafer, is an **electronic material** - utmost care has to be taken in selecting the right stuff!

- Wafer holders or tweezer, e.g., can totally ruin a wafer by leaving minute amounts (far below the detection limit) of heavy metals (most notably **Fe**) on the wafer if they are unsuitable (Never, really never, touch the wafer with a metal tweezer!).
- Gas pipes may corrode internally if made from the wrong metal and thus contaminate the gas flowing through it with traces of impurities - your factory then will only produce garbage.

Note also that some of the most dangerous inorganic chemical are used!

- ***HF*** (hydrofluoric acid) will cause heavy tissue and especially bone damage already by its vapors - you do not even have to touch it to get severly damaged.
- ***PH_3*** (**phosphine**) and ***AsH_3*** (**arsine**) are among the most toxic gases known to mankind; minute amounts are deadly (**PH_3**, in fact, was used as a poison gas in world war I).

To continue, use the link

Commercial Wafer Specifications

Illustration

Here are the specification for Si wafers from one of the worlds top companies, **Wacker Siltronic**, as they appear in the Internet in Nov. **2000**.

- **Notice**: Concentrations here are in **cm^3**. The conversion to parts per .. is simple:
  The atomic density of Si is **4.96 · 10^{22} cm^{-3}** or about **5 · 10^{22} cm^{-3}**.
  1 ppm** thus corresponds to **5 · 10^{16} cm^{-3}.
- The lowest concentration given in the table (look for it) is **5 · 10^{10} cm^{-3}**; it corresponds to **1 ppt** or **10^{-12}**.
- Surface concentrations **[S]** (given in **cm^{-2}**) are converted to volume concentrations **[V]** by
  [S] = [V] /a with **a =** lattice constant (**= 0,5431 nm**) or, more precise for single crystals, distance between the crystallographic planes. With **a** ≈ **0,5 nm = 5· 10^{-8}** cm, we have
  [V] = 5 · 10^{16} cm^{-3} = 1 ppm corresponds to **S = 10^8 cm^{-2}**.
- Many specifications relate to the "flatness" of the wafers and the perfection of the surface; the abbreviations used are
  LLS (sometimes also abbreviated **LPDs**): *Localized Light Scattering Defect*; this relates to a detection method of **sub-µm** size surface imperfections (resulting from bulk microdefects)
  SFQR: *Site flatness quality requirements* (??): Definitely a measure of flatness in a region comparable to the size of a single chip
  (The rest: Who knows?)
- Here is a link with precise defininitioons of geometrical parameters:
  http://www.freiberger.com/english/products/geom_parameters.php

Polished & Epitaxial Wafers for IC Applications

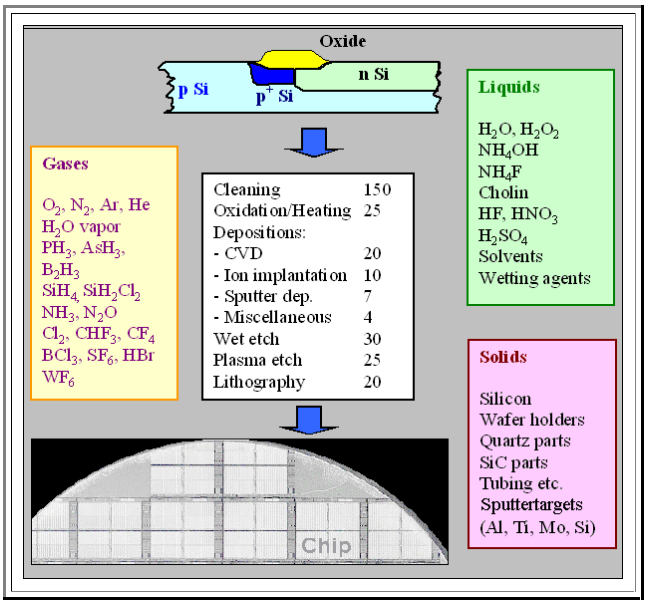
Crystal / Bulk			300mm	200mm	150mm	125mm	100mm
Growth Technique *)			CZ	CZ	CZ	CZ	CZ
Orientation			1-0-0	1-0-0	1-0-0 / 1-1-1	1-0-0 / 1-1-1	1-0-0 / 1-1-1
Orientation Tolerance		degree	± 0.2	± 0.2	± 0.5	± 0.5	± 0.5
Off Orientation		degree	0	0 - 4	0 - 4	0 - 4	0 - 4
Dopant			Boron / Phosphorus	Boron / Phosphorus	Boron / Phosphorus	Boron / Phosphorus	Boron / Phosphorus
Resistivity Target Range	pol prime - Boron	Ohmcm	0.5 - 50	0.5 - 50	0.5 - 50	0.5 - 50	0.5 - 50
	pol prime - Phosphorous	Ohmcm	1.0 - 50	1.0 - 50	1.0 - 50	1.0 - 50	1.0 - 50
	epi substrate - Boron	Ohmcm	0.006 - 50	0.006 - 50	0.006 - 50	0.006 - 50	0.006 - 50
Radial Resistivity Variation	Boron typical 1-0-0 / 1-1-1	%	< 10	< 5	< 5 / < 6	< 6 / < 10	< 8 / < 9
	Phosph. typical 1-0-0 / 1-1-1	%	< 15	< 15	< 12 / < 25	< 12 / < 25	< 15 / < 25
Oxygen Target Range ± Tol.	pol prime - Boron 1-0-0		4.8 - 7.8 x 10^{17} ± 0.5	5 - 7.8 x 10^{17} ± 0.5	5.8 - 8.9 x 10^{17} (± 0.6 - 0.8)	5.8 - 8.9 x 10^{17} (± 0.5 - 1.0)	5.8 - 8.9 x 10^{17} (± 0.8 - 1.2)
	1-1-1	at cm^{-3}	NA	NA	5.8 - 8.9 x 10^{17} (± 0.7 - 1.0)	6.2 - 8.9 x 10^{17} (± 0.5 - 1.0)	5.9 - 8.9 x 10^{17} (± 0.8 - 1.5)
	pol prime - Phosph. 1-0-0	ASTM F121-83	4.8 - 7.8 x 10^{17} ± 0.5	6 - 7.5 x 10^{17} ± 0.5	5.8 - 8.9 x 10^{17} (± 0.6 - 0.8)	5.8 - 8.9 x 10^{17} (± 0.5 - 1.0)	5.8 - 8.9 x 10^{17} (± 0.8 - 1.2)
	1-1-1		NA	NA	5.8 - 8.9 x 10^{17} (± 0.7 - 1.0)	6.2 - 8.9 x 10^{17} (± 0.5 - 1.0)	5.9 - 8.9 x 10^{17} (± 0.8 - 1.5)
Radial Oxygen Variation	typical	%	< 10	< 5	< 6	< 6	< 5 - 10
Bulk Metal Concentration	Fe	at cm^{-3}	≤ 5.0 x 10^{10}	≤ 5.0 x 10^{10}	≤ 1.0 x 10^{11}	≤ 1.0 x 10^{11}	≤ 1.0 x 10^{11}
Bulk Carbon Concentration	measured on wafer	at cm^{-3}	≤ 2.0 x 10^{16}	≤ 2.0 x 10^{16}	≤ 2.0 x 10^{16}	≤ 2.5 x 10^{16}	≤ 2.5 x 10^{16}

Polished Wafers / Substrates			300mm			200mm			150mm			125mm	100mm
Surface Metals	Cu / Cr / Fe / Ni	at cm^{-2}	≤ 1.0 x 10^{10}			≤ 2.5 x 10^{10}			≤ 5.0 x 10^{10}			≤ 5.0 x 10^{10}	≤ 5.0 x 10^{10}
	Al / Zn / K / Na / Ca	at cm^{-2}	≤ 5.0 x 10^{10}			≤ 1.0 x 10^{11}			≤ 2.0 x 10^{11}			≤ 2.0 x 10^{11}	≤ 2.0 x 10^{11}
LLSs (Frontside) *)	size	µm	> 0.2	> 0.16	> 0.12	> 0.2	> 0.16	> 0.12	> 0.3	> 0.2	> 0.12	> 0.3	> 0.3
	pol prime	# per wafer	< 30	< 40-300	< 200-10^3	< 15-35	< 20-120	< 70-600	< 15	< 30	NA	< 15	< 15
	UltraFlat (150 mm)	# per wafer	NA	NA	NA	NA	NA	NA	< 5	< 10	< 150	NA	NA
	monitor	# per wafer	< 30	< 60	< 100	< 15	< 20-65	< 130-700	NA	NA	NA	NA	NA
Diameter Tolerance		mm	± 0.2			± 0.2			± 0.2			± 0.2	± 0.2
Warp	polished - without layer	µm	< 50			< 20			< 30			< 30	< 30
Wafer / Substrate Thickness	Standards	µm	775			725			375 / 525 / 625 / 675			375 / 525 / 625	300 / 375 / 525
Thickness Tolerance		µm	± 25			± 15			± 15			± 15	± 15
GBIR = TTV (Std | *UltraFlat*) *)		µm	< 4			< 3.5			< 5.0		*< 2.5*	< 5.0	< 5.0
GFLR = TIR (Std | *UltraFlat*) *)		µm	NA			< 2.0			< 2.0		*< 1.2*	< 2.0	< 2.0
Local Flatness *) (Std | *UltraFlat*)	SFQR / STIRmax, s.b.f.	µm	< 0.25			< 0.25			< 0.5		*< 0.3*	NA	NA
	SFQD / SFPD, s.b.f.	µm	< 0.18			< 0.18			< 0.3		*< 0.2*	NA	NA
	SBIR / STIRmax, b.r.	µm	NA			< 0.7			< 1.0		*< 0.6*	< 1.0	< 1.0
Standard Site Size		mm²	25 x 25			25 x 25			15 x 15			15 x 15	15 x 15

Processes and Materials II

Illustration

- Here is the rest of the processes
  - A few more materials are needed, especially solids in the form of "sputter targets".
  - A great total of about **450** process- and control steps are needed.

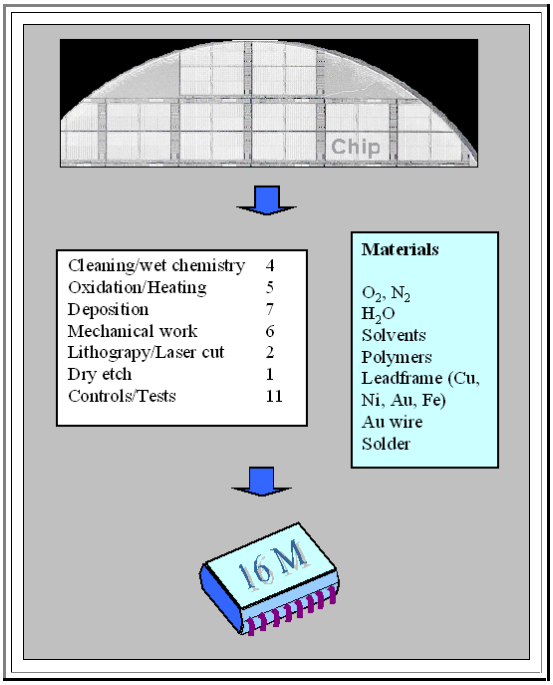


- We don't have a chip yet - we only have unpackaged chips on a wafer.
  - Next, the wafer has to be cut and the chips that work (this needs a measurement) are packaged.
  - Packaging, although not needing processes at very small dimensions, is not simple either.
  - **To continue, use the link**

Processes and Materials III

- Packaging is a completed process in its own right
  - It needs very special materials - even the lowly black plastic that dominates the appearance of chips is a sophisticated material!
  - It also includes extensive testing of the chip.

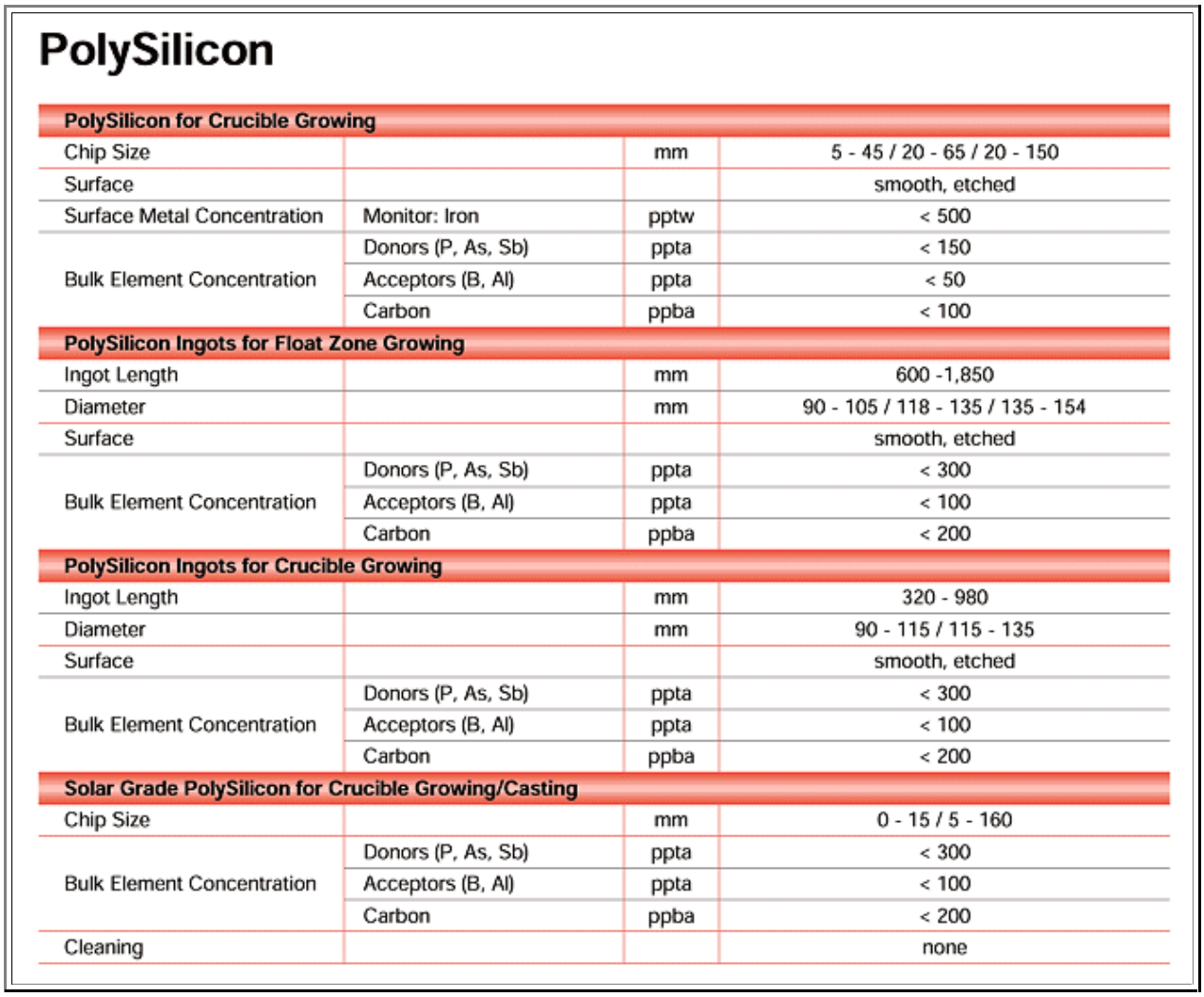
Illustration



Commercial Poly-Silicon Specifications

Illustration

- Here are the specification for poly-silicon from one of the worlds largest suppliers, **Wacker Siltronic** as they appear in the Internet in Nov. **2000**.
  - *Notice*: The "**w**" or "**a**" behind the concentration denotes **w**eigth or **a**tomic parts per **m = million**, **b = billion**, **t = trillion**.



PolySilicon

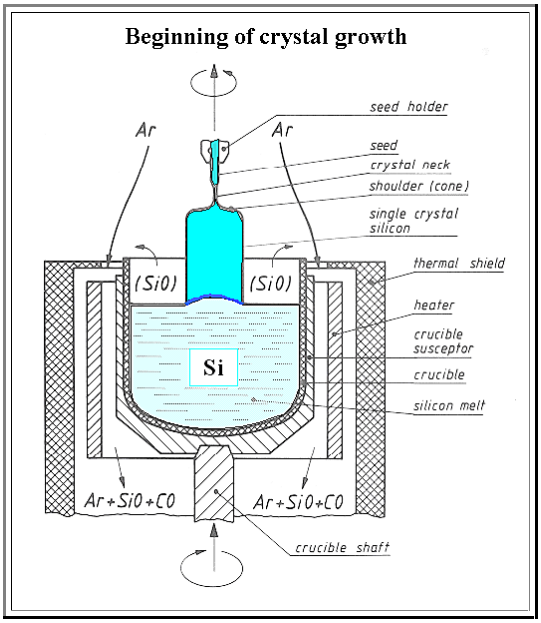
PolySilicon for Crucible Growing			
Chip Size		mm	5 - 45 / 20 - 65 / 20 - 150
Surface			smooth, etched
Surface Metal Concentration	Monitor: Iron	pptw	< 500
Bulk Element Concentration	Donors (P, As, Sb)	ppta	< 150
	Acceptors (B, Al)	ppta	< 50
	Carbon	ppba	< 100
PolySilicon Ingots for Float Zone Growing			
Ingot Length		mm	600 -1,850
Diameter		mm	90 - 105 / 118 - 135 / 135 - 154
Surface			smooth, etched
Bulk Element Concentration	Donors (P, As, Sb)	ppta	< 300
	Acceptors (B, Al)	ppta	< 100
	Carbon	ppba	< 200
PolySilicon Ingots for Crucible Growing			
Ingot Length		mm	320 - 980
Diameter		mm	90 - 115 / 115 - 135
Surface			smooth, etched
Bulk Element Concentration	Donors (P, As, Sb)	ppta	< 300
	Acceptors (B, Al)	ppta	< 100
	Carbon	ppba	< 200
Solar Grade PolySilicon for Crucible Growing/Casting			
Chip Size		mm	0 - 15 / 5 - 160
Bulk Element Concentration	Donors (P, As, Sb)	ppta	< 300
	Acceptors (B, Al)	ppta	< 100
	Carbon	ppba	< 200
Cleaning			none

- Here some production information:
  - According to "Solid State Technology" July **2005**, the production numbers are as follows:
    - Total production **2005**: **26.000.000 kg**; about **2/3** for microelectronics, **1/3** for photovoltaics.
    - Expected production **2006**: **29.000.000 kg**.
  - Right now (end of **2007**) there is a tremendous shortage of poly **Si** because the solar cell industry grows so fast, that poly **Si** production cannot keep pace, see also the link.
  - Expected shortfalls:
    - **2005**: **4.000.000 kg**
    - **2006**: **6.000.000 kg**.
    - **2007**: **12.000.000 kg**.
    - **2008**: **20.000.000 kg**.
  - The expected shortfalls result to a large extent from a growth growthraterate of **40 %** for photovoltaics and from technical and financial difficulties to crank up production at a high rate. However, alternative processes for solar **Si** production are expected to come on-line in **2006**..

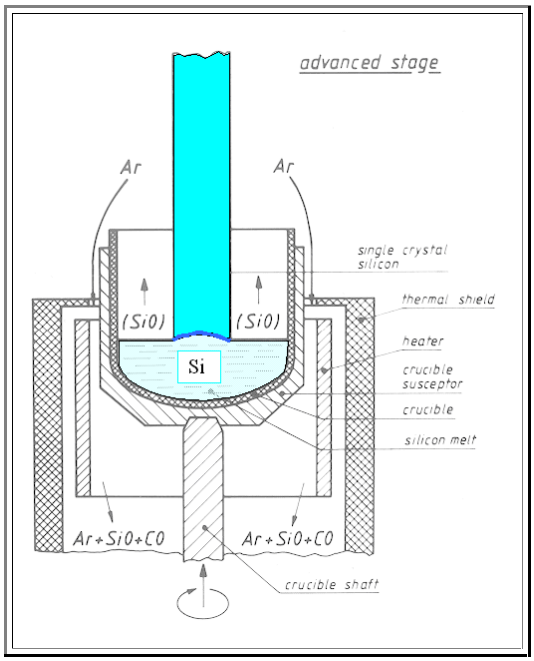
Czochralski Crystal Growth Process

Illustration

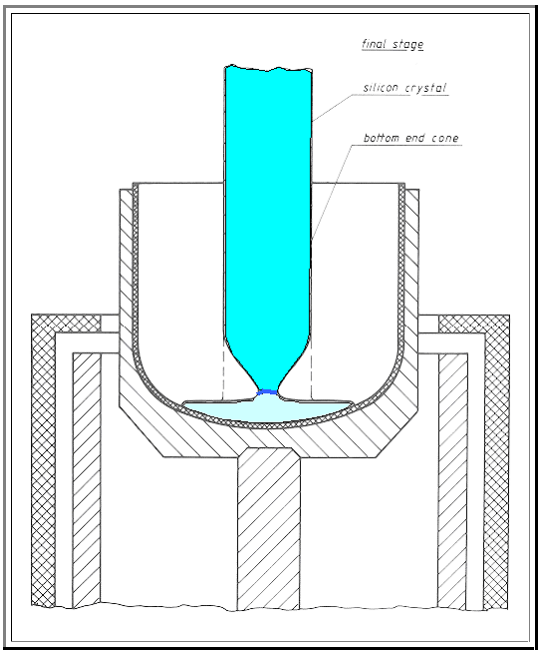
- You start growing a "*Czochralski crystal*" by filling a suitable crucible with the material - here hyperpure correctly doped Si pieces obtained by crushing the poly-**Si** from the Siemens process. Take care to keep impurities out - do it in a clean room - and use hyperpure silica for your crucible.
  - Make sure that the inside of the machine is very clean too and that the gas flow - the gas you introduce but also the SiO coming from the molten **Si** because parts of the crucible dissolve - does not interfere with the growing crystal.
  - Dissolve the Si in the crucible and keep its temperature close to the melting point. Since you cannot avoid temperature gradients in the crucible, there will be some *convection* in the liquid **Si**. You may want to suppress this by big magnetic fields.
  - Insert your *seed crystal*, adjust the temperature to "just right", and start withdrawing the seed crystal. For homogeneity, *rotate* the seed crystal and the crucible. Rotation directions and speeds and their development during growth, are closely guarded secrets!
  - First pull *rather fast* - the diameter of the growing crystal will decrease to a few mm. This is the "**Dash process**" ensuring that the crystal will be dislocation free even though the seed crystal may contain dislocations.
  - Now decrease the growth rate - the crystal diameter will increase - until you have the desired diameter and commence to grow the commercial part of your crystal at a few mm/second.



- As your crystal grows, the impurity concentration (including the dopants if you do not watch out) will increase in the melt (due to segregation) and therefore also the percentage incorporated into the crystal. The temperature profile of the whole system will also change - you are now deeper down in the crucible and the crystal cools off a little more slowly. All these factor influence the homogeneity of the crystal.
  - The radial and lateral doping level is influenced - it will not stay constant without some special measures
  - The concentration of impurities, especially interstitial oxygen, may change. In general, the concentration increases from "head" to "tail".
  - Crystal lattice defects still present (essentially agglomerates of the point defects present in thermal equilibrium at high temperatures) may change in size and distribution.
- You do not want this - you want a crystal where all this factors are constant - everywhere!
  - So you must do something - change the rotation speeds, the temperature, the growth speed - whatever.
  - This is where crystal growing becomes an art - and you will not find much literature about this. This is the tricky and secret part: Changing all important parameters continuously so that the crystal is homogeneous!



- Now the crystal is nearly finished. You do not want to use up all the **Si**, because the "last drop" contains all the impurities not yet incorporated because of their small segregation coefficients.
  - But you cannot simply pull out the crystal after the desired length has been reached. The thermal shock of the rapidly cooling end would introduce large temperature gradients in the crystal which in turn produce stress gradient - plastic deformation (easy in **Si** at high temperatures) will take place and this means dislocation are nucleated and driven into the crystal.
  - The dislocation will even run up into the formerly dislocation free part of the crystal, destroying your precious Silicon.
  - So you withdraw gradually by just increasing the pulling rate a little bit which will lead to a reduced diameter. The crystal then ends in an "end cone" similar to the "seed cone".

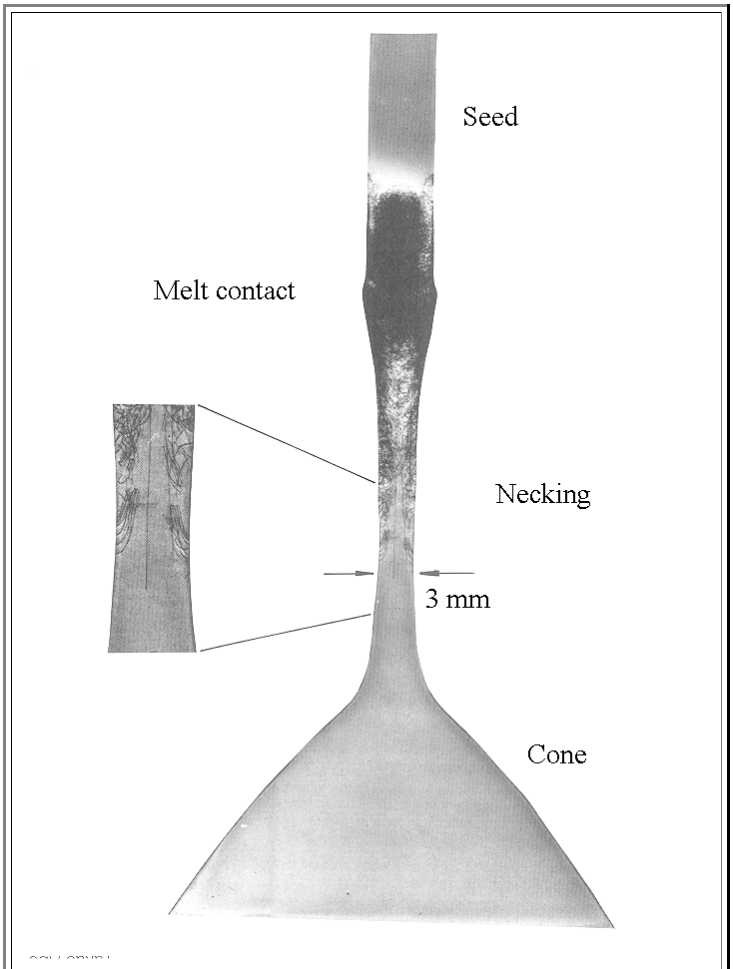


  - The finished product can be seen in a different link.

Necking

Illustration

- Here is an X-ray topograph of the first part of crystal growth
  - An **X**-ray topograph is similar to a transmission electron microscope image - it shows the interior of the sample and dislocations are visible as dark lines.
- You see a dislocation-free seed crystal, followed by region full of dislocations. This is unavoidable because dipping a solid seed in a melt that has by definition a higher temperature, always causes a "thermal shock" with stress and strain and therefore plastic deformation.
  - The the diameter of the now growing crystal is made as small as possible (it still must be able to carry the weight of the finished crystal - up to **250 kg** or so). This is the "**necking**" or Dash process.
  - The dislocations disappear after a few **cm**, the question is why? The picture almost shows it. For the usual **<100>** oriented crystal, the glide planes of the dislocations (the **{111}** planes) are all inclined to the growth direction, and the dislocations, still feeling some stress, will simply move out of the crystal.
  - This is where the art part comes in - or better came in. Keep enough stress to move the dislocations, but not that much that new ones will be generated.



Courtesy of Wacker / Siltronic; Burghausen, Germany

Silicon Crystal

Illustration

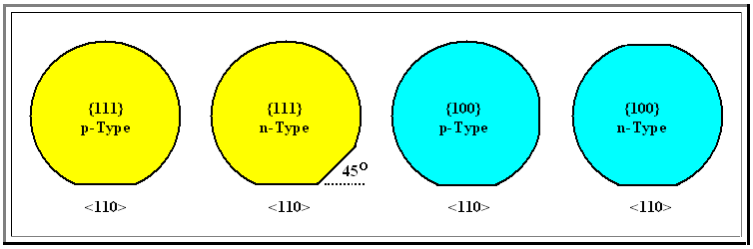
- Here is a picture of a state-of-the-art **200 mm Si** crystal as they are grown by the thousands for present day (**2000**) chip manufacture.
  - While it does look like an oversized chromium-plated salami, it is a much more sophisticated product (and much more expensive).

  - Note that this huge crystal is hanging on a rather thin **Si** seed crystal (see inset). This seed crystal does not only have to support the weight of the crystal, but also the torque needed to rotate the crystal during its growth.

Wafer Flats

Illustration

- In the old times (up to the **150 mm** wafer diameter era), wafers had **flats**, and the flats told you two things:
  1. The doping type of the wafer (**n**- or **p**-type
  2. The orientation of the wafer: **{100}** or **{111}**
  - While this is trivial information, consider: All wafers, whatever doping type or crystal orientation, look exactly the same! As soon as a wafer has been removed from its box that carried this and other information, you can't see anymore what you got. You also cannot measure it easily (and without destroying the wafer).
  - And if something goes wrong here (and things that can go wrong will go wrong some day), it may be a horribly expensive mistake! If you feed wafers of the wrong doping kind into the line, it will really, really cost you - probably your job.
- So here is the convention

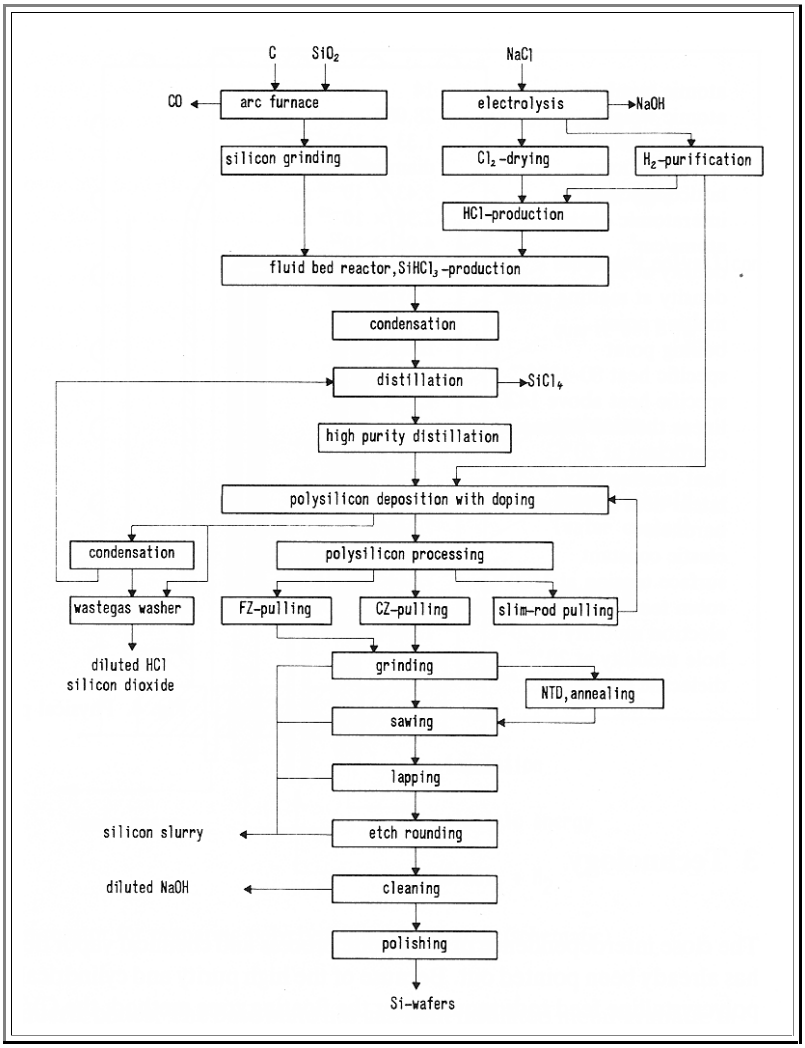


- But beware! Wafer manufacturers will produce whatever the customer wanted, and after **n**-type **Si** went out of style for most mass-produced chips, the only reason for a flat was to allow the patterns to be made to be aligned with a crystallographic direction.
  - You then ordered your **p**-type **{100}** wafers with only one flat in the **<110>** direction. One reason was that the wafer would easily cleave along this and the respective perpendicular direction.
  - So wafers with diameters larger or equal to, say, **100 mm** and just *one* flat are more likely **{100}** **p**-type than the "proper" **{111}** **p**-type. The picture in the backbone, e.g., shows **p** -type **{100}** **150 mm** wafer!
  - And wafers with diameters larger or equal to, say, **200 mm**, probably will have no flat at all, but just a small "**notch**" - simply because you loose too much expensive area by cutting of a flat.
- So, how can you tell what you have - if you don't trust the one flat there is, or if there is none! There are extremely simple ways of checking:
  - *Checking doping type*: Take an Voltmeter and measure the *thermovoltage* between a hot tip (tale a soldering iron) and a room-temperature tip pressed on the wafer somewhere. Its *sign* will tell you if the wafer is **n**- or **p**-type. Which is which results from hard thinking or from checking a known piece of **Si**.
  - *Checking orientation*: Break your wafer. If the pieces tend to be rectangular, it was **{100}**.
- In either case your wafer is now "dead", i.e. no longer usable for making **IC's.**

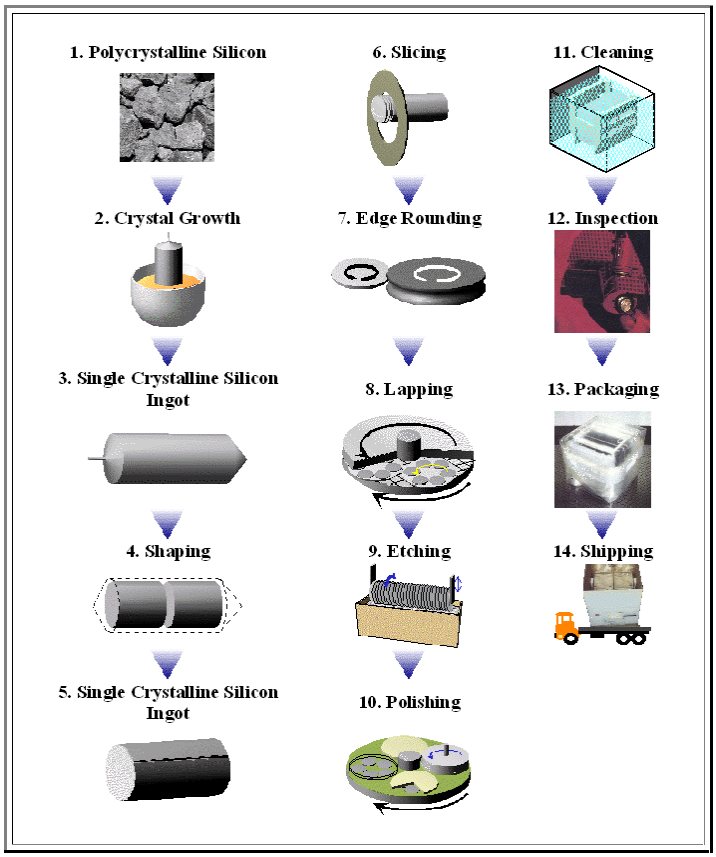
Complete Wafer Process

Illustration

- First a schematic process flow graph is given for everything - from sand to wafer.
  - It includes side processes not covered in the backbone modules which, however, are self explanatory.

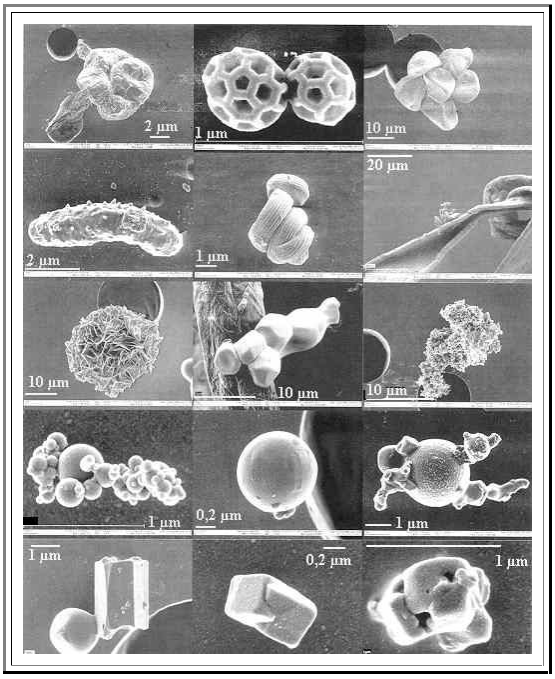


- Next, we have a pictorial representation of just crystal growth and wafer production. Again, it contains some self-explanatory processes not introduced before.


1. Polycrystalline Silicon
2. Crystal Growth
3. Single Crystalline Silicon Ingot
4. Shaping
5. Single Crystalline Silicon Ingot
6. Slicing
7. Edge Rounding
8. Lapping
9. Etching
10. Polishing
11. Cleaning
12. Inspection
13. Packaging
14. Shipping

The Air We Breath

Here is a picture of the actual air we breath; it was assembled by the Max-Planck-Institut für Chemie in Mainz/ Germany



What we have are

- Biological objects (pollen and such) in the first two rows and middle of the third row.
- A silica particle (from the desert?); left third row; and an unknown particle right, third row.
- Ash particles in the fourth row.
- anorganic salt crystals (we think) in the fifth row.

.

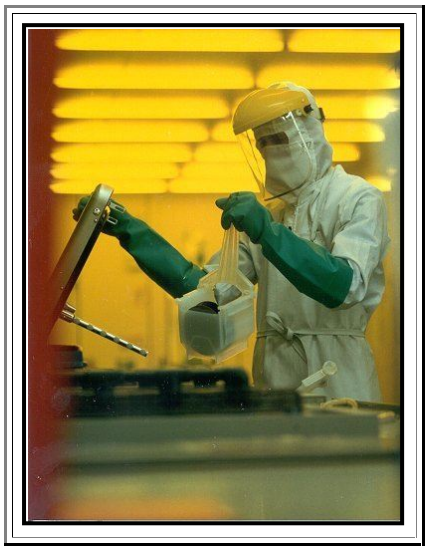
Illustration

Cleanroom Garments

Illustration

Below some pictures from inside a cleanroom. We also see some pieces of equipment.

General view of work in a clean room



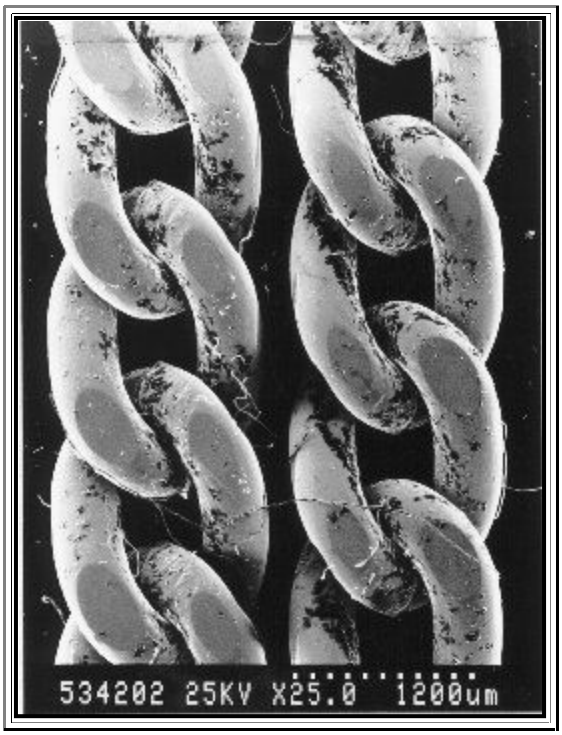
Inserting wafers into a "cleaner"; a kind of washing machine
for wafers employing extremely aggressive chemicals.
Special protection (heavy gloves and a face shield) are worn.

Looking along the mechanism for moving wafers into a furnace.
The orange glow from the furnace tubes is visible.

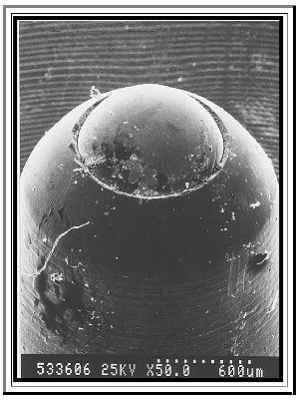
Humans as Source of Particles

Here a few examples of humans as source of particles. The magnification is typically low; even particles not visible under these conditions would be deadly

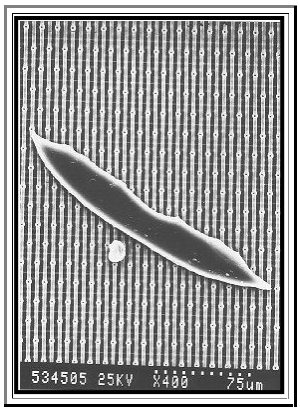
Illustration



Dirt on a necklace

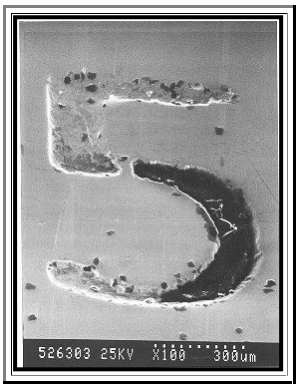


Tip of a ball point writer

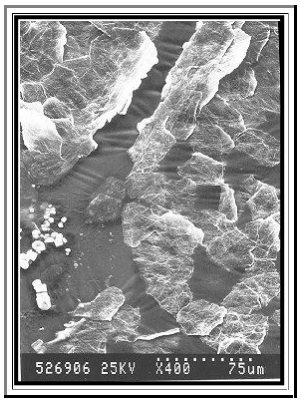


A piece of nail polish on a **256 k DRAM** chip

Hair with hair spray



Inside of finger ring



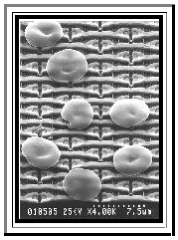
Flakes of skin

Particles on Chips

Illustration

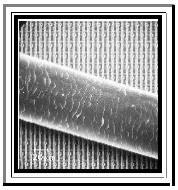
Here you can see a collection of ***SEM*** pictures which not only illustrate graphically the "particle" problem in making chips, but also have a certain esthetical appeal

Click on the link for an enlarged view.



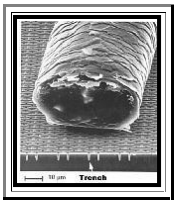
Red blood cells on a **1 Mbit** memory

(large size)



Hair (from a female) on a **256 kbit** memory chip. Just one of the little flakes of the hair would be enough to cover one memory cell.

(large size)

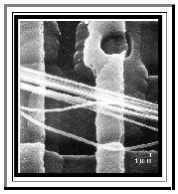


Hair on a **4Mbit** memory chip. - just to show how structure sizes decrease.

(large size)

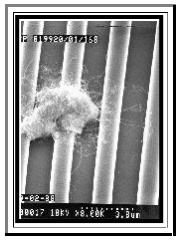
Here a metal particle, probably a tiny drop of **Al** that was burnt of by a an electrical discharge in a sputtering machine and hit the **Si** as a solidified droplet . It was coated with **Al** which was subsequently structured by etching. Four conducting lines are now short circuited.

(large size)



Mother nature is still ahead when it comes to small structures. Here we see spiderwebs on a **256 k DRAM**. A typical strand of spider silk consists of several individual strings with diameters around **0.2 µm**.

(large size)



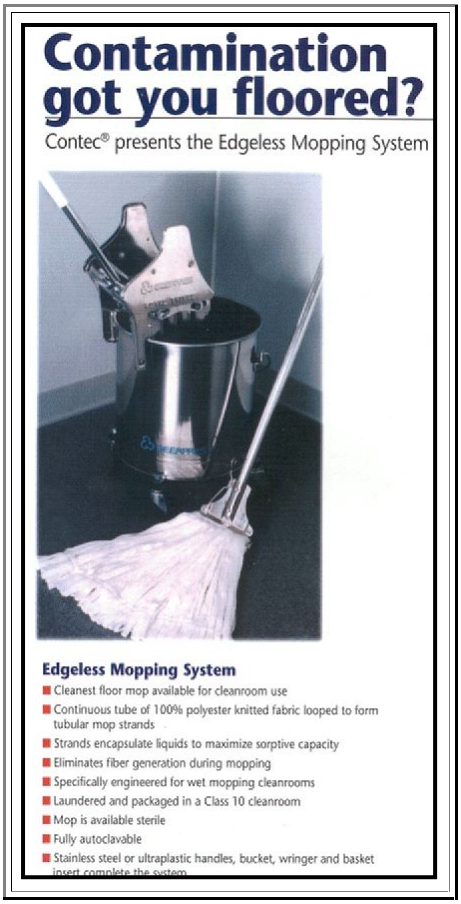
Here is a particle of unknown nature. Whatever it is, it will kill a chip.

(large size)

Electronic Materials in Unexpected Places

Illustration

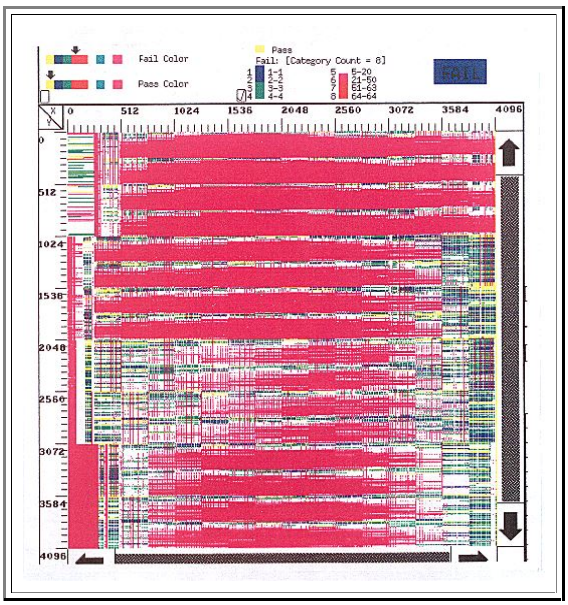
Below an add (from "Semiconductor International"; June **2000**) that shows that nothing is too unimportant to qualify as electronic materials,

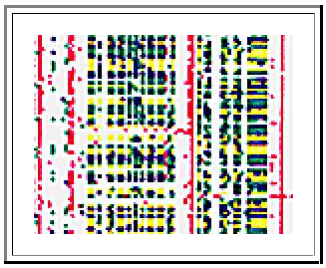


Emergence of a New Chip

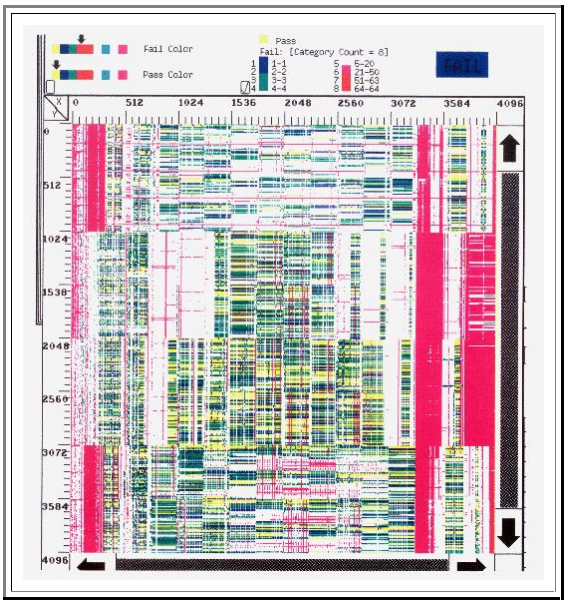
Illustration

- After much designing, simulation, process trials and test chips comes the moment when the first product chip containing the new generation will enter the line and becomes processed. Processing may take several month because many processes do not work too well or are done in prototype machinery. Here is a real example from the development of a **16 Mbit DRAM** at Siemens.
  - If the batch of wafers does not die a sudden death at some shaky process (or because somebody simply drops them), "**First Silicon**" will eventually emerge. This is a big moment in the life of the team - now we will see if several **100** man years and many million **$$** were well spent.
  - The test crew starts to measure the function of the memory with the chips still on the wafer. The results are plotted in a matrix mimicking the matrix of memory cells on the chip. The color given to a pixel encodes the quality of the individual memory cell. Yellow means fully functional, all other colors, especially red, mean trouble.
  - Below is a plot for the best chip from the "first Silicon". The insert on the right shows that single pixels are resolved in principle. How was the feeling with a result like that?

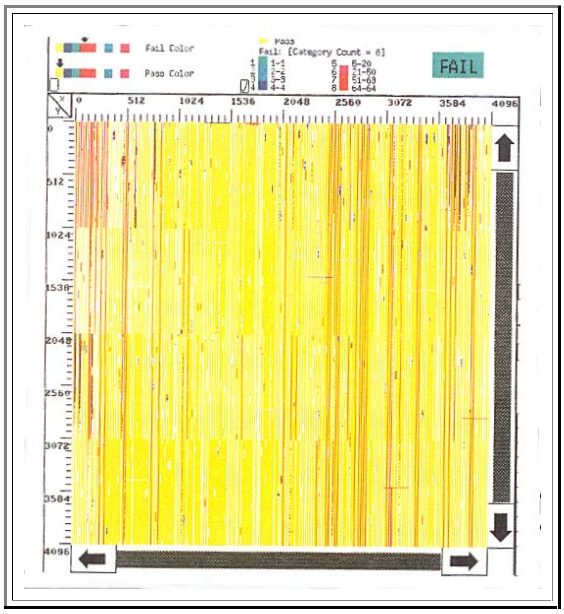




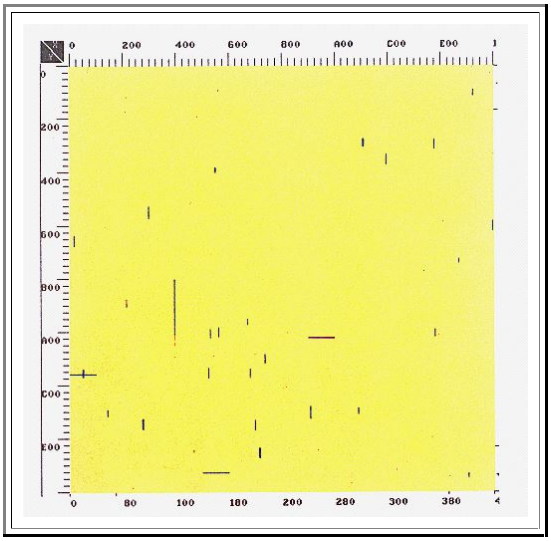
  - The feeling was : *Get out the champagne*! There are fully functional memory cells on this chip! Not too many - but who cares. This was good! More often than not, nothing whatsoever works on first Silicon. A few hundred cells that work are enough to measure all critical parameters of the circuits and to get the desperately needed data for optimizing design and processes.
- Next comes *second Silicon*. Those wafers probably were processed to some degree when first Silicon came out, so only back end processes can now be optimized. But there is improvement: A little bit more yellow, but far less deep red!
  - The structure in these plots - pixels of the same color arranged in lines - simply reflects the fact that a number of memory cells share the same bit- or word line. If this line is interrupted, all its memory cells fail.



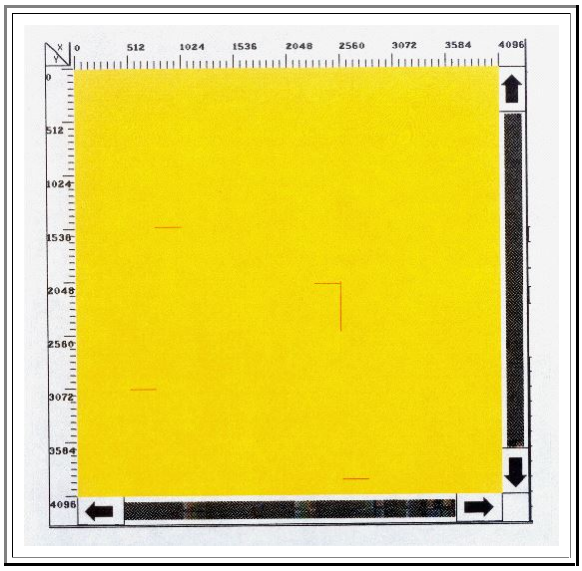
- Here we have the first redesign - all the things learned from first Silicon are now incorporated into the chip.
  - And now the yellow color dominates. While this looks pretty good, we are still a million or so cells short of total success.



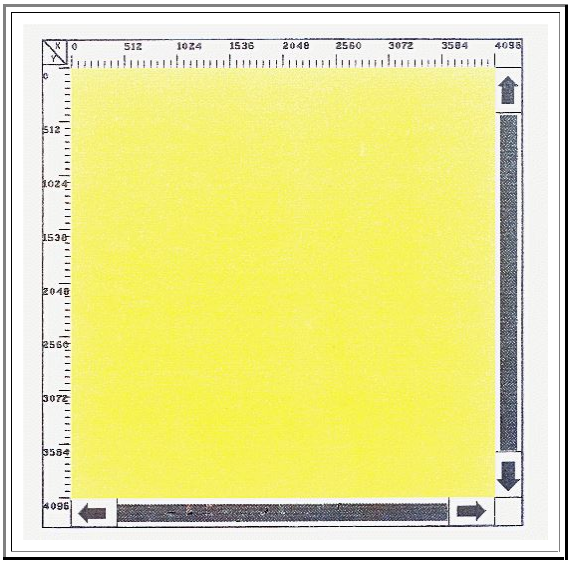
- Now the process stabilizes and a new game starts: Get rid of the few but noticeable problems in the memory matrix. This is tough!



- Now we won: Only a few strongly correlated failures - no single bit problems. This can be solved - with luck - by activating some redundant memory cells that will take over the function of the failed cells



- Triumph!! Using the built in redundancy, a fully functional memory matrix was obtained for the first time - in the third full process cycle, a very good score.

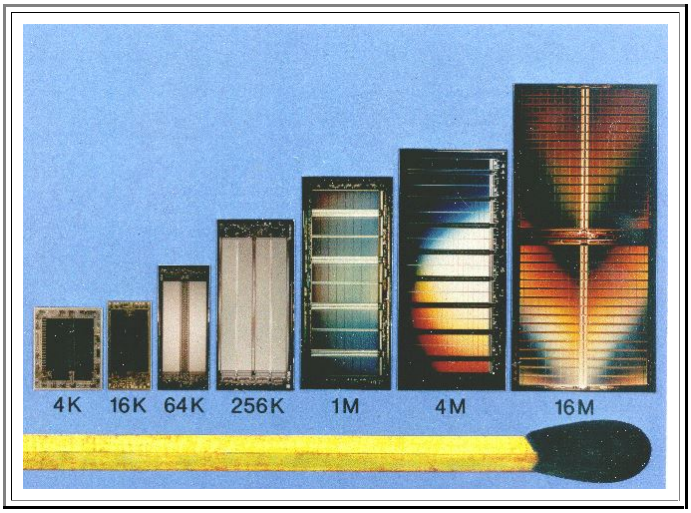


- Exactly two chips out of some thousand candidates are fully functional. The project goals have been met.
- All that remains to be done now is to crank up the yield to, say, **15 %** functioning chips per wafer so that production can start.
  - *That was a joke*: Now the really *hard* work starts. It will take considerably more time and money to "just improve the yield" than it took to get to this point in the chip development.

Development of the Chip Size

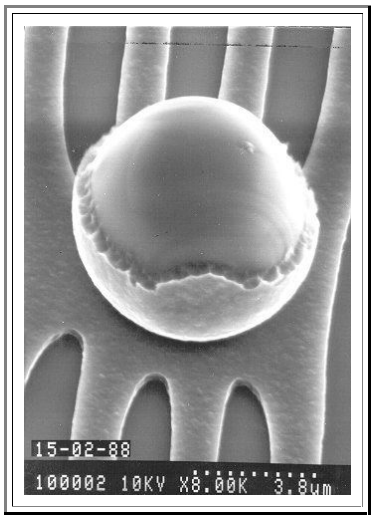
- Here is a very graphical picture of the development of the chip size for **DRAMs** up the beginning of the nineties.
  - You must be aware, however, that chip sizes are not static. As soon as your first product is finished, you try like crazy to reduce the chip size.
  - That is almost the only way to make money: Make *more* chips with the established (already payed for) technology.

Illustration



Particles on Chips: A Metal Ball

Here is the large size picture of a metal particle, probably a tiny drop of **Al** that was burnt of by a an electrical discharge in a sputtering machine and hit the **Si** as a solidified droplet . It was coated with **Al** which was subsequently structured by etching. Four conducting lines are now short circuited.

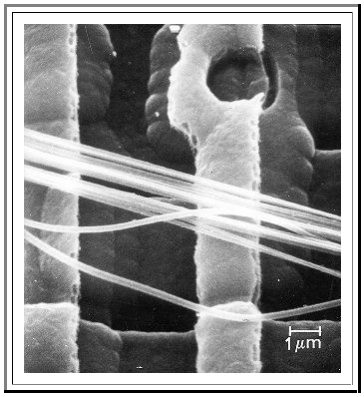


Illustration

Particles on Chips: Hair

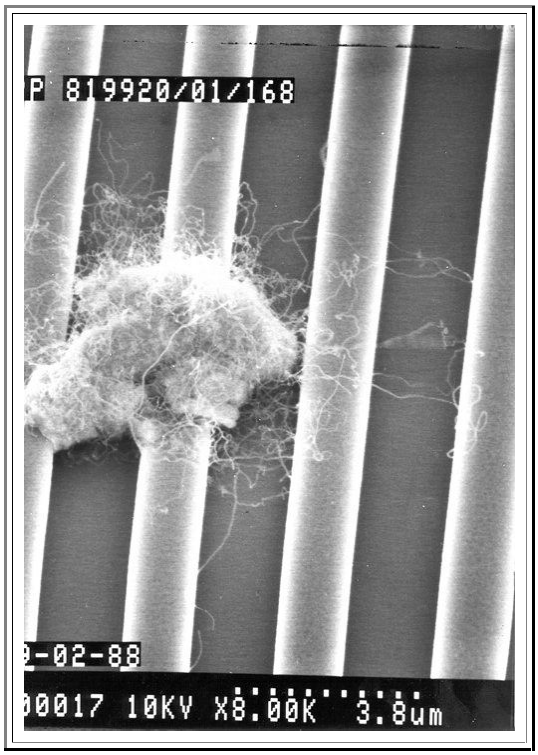
Illustration

- Here is the large size picture of some strands spider silk on a **256 k DRAM**. It shows that Mother Nature was still ahead then (**1988**) with regard to small structures. A typical strand of spider silk consists of several individual strings with diameters around **0.2 µm**
  - However: Nowadays (**2002**) we are doing better than that. Minimal dimensions on state-of-the-art chips are **0.13 µm**.



Particles on Chips: Unknown Object

Here is the large size picture of something unknown (probably biological). Whatever it is, it will kill a chip.



Illustration

The Polyimide Story

- Here is a (slightly edited) article about the development of polyimides, which appeared in 2005 in the MATERIALS SCIENCE AND TECHNOLOGY NEWSLETTER.
  - It is one of the rare articles that also mentions what went wrong. It gives a good idea how difficult it is in the real world out there to introduce a new matetial in a complex product.
  - Here is the link to the Journal;
- Copyright courtesy of **Dr. Lacombe**, the author of this story.

Advanced

MATERIALS SCIENCE AND TECHNOLOGY NEWSLETTER

Vol. 2, No. 2

Dr. Robert H. Lacombe
Chairman
Materials Science and Technology
CONFERENCES, LLC
3 Hammer Drive
Hopewell Junction, NY 12533-6124
Tel. 845-897-1654
FAX 212-656-1016
E-mail: rhlacombe@compuserve.com

EDITORIAL NOTES

During the week of November 7, 2005 MST CONFERENCES will be holding the 4th in its series of symposia on Polyimides and High Temperature Polymers along with the 2nd International Symposium on Adhesion Aspects of Thin Films in Savannah Georgia. Apropos of this event, this issue of the Newsletter will be focusing on these two topics and their interrelation. The following narration gives some reminiscences of our experiences with these topics as they played out in the mainframe computer business during the late 1970' s and 1980's.

Polyimides, Stress and Adhesion: or There are no Bad Materials, Just Bad Uses of Good Ones:

Prelude to polyimides as electrical insulators

Subsequent to the DuPont company's patenting of the polyimide materials, one of the first major applications was for electrical insulators capable of sustaining high temperature conditions. To those of us who were first exposed to the properties of engineering thermoplastics in the 1950's, the polyimide materials were a marvelous revelation. Those who recall the early days of manufactured goods made of plastic materials will remember that if subjected to high temperatures, say the flame from a match, these items would either melt into a viscous goo or possibly also ignite and burn. In addition, articles made from the early resins were highly prone to fracture as in the case of children's toys made from polystyrene. On the positive side, however, the early plastics were very easy to process into a large variety of useful shapes by either molding or casting methods. Thus one could easily and very cheaply manufacture useful household items such as knives, forks, cups and spoons. So cheaply in fact that these became throwaway items. One could even get around the brittleness problem by using polycrystalline materials such as polyethylene or polypropylene. These resins could be drawn or extruded into flexible sheets or tubing for use in packaging or plumbing applications. However, these materials also had the problem of temperature stability and would either distort or melt at elevated temperatures at or above 100C.

Thus it seemed that the price one paid for low cost and ease of manufacture was limited thermal-mechanical performance. Now, however, along came the polyimide materials which were clearly polymeric in nature but which also possessed extraordinary thermal stability previously unheard of in a polymer. The earliest polyimide to be produced in commercial quantities was the KAPTON® (The basic resin goes under the name pyromellitic dianhydride oxydianaline or PMDA-ODA for short) film material produced by the DuPont company. Here was a flexible plastic film that you could take a match to and it would not burn or melt. Under a nitrogen environment this material could be shown to be stable up to 400C in a Thermo-Gravimetric (TGA) experiment. This was truly astounding behavior for a polymer and it was felt that here was a class of materials that could surmount the thermal-mechanical limits of the ordinary thermoplastics. However, high performance came at a price. It turned out that in order to process the KAPTON material one had to use concentrated sulphuric acid or some other equally potent solvent. It

seemed there would be no free lunch. If you wanted the high performance thermal-mechanical properties you had to sacrifice ease of processing.

All was not lost though since there was another way to make polyimide films without resorting to extreme processing techniques. It turns out that nearly every polyimide has a twin sister known as a polyamic acid. The polyamic acid has a chain architecture very similar to its polyimide sibling except that instead of the refractory imide group one has an acid moiety which gives the chain much more flexibility and also makes it much more soluble in relatively benign solvents such as N-methy-pyrrolidone (NMP). Thus one first casts a film of the amic acid material out of NMP and then after drying the amic acid can be converted to the polyimide by curing at an elevated temperature. With this processing strategy it seemed as though the ultimate miracle had finally been achieved. Now one had a relative simple and benign process for achieving high performance films and coatings.

It was at this stage in the late 1960's and early 1970's that the microelectronics industry started to take a serious interest in using the polyimide materials as high performance insulators for semiconductor devices and advanced multichip modules. Up to that time the insulating material of choice was sputtered glass. This material is essentially amorphous SiO_2 which is the purest form of glass known and forms the basis of nearly all glass artifacts. The glassware in your cupboard is basically SiO_2 doped with boron and other materials which are used to give the final artifact a number of desired properties such as color, density, ... etc. SiO_2 was also a natural for the microelectronics industry since nearly every silicon wafer comes with a thin layer of SiO_2 due to the natural oxidation behavior of the pure silicon. This thin oxide layer is the primary reason that silicon is the primary semiconductor material used for microchips in the world today even though one can in principle use materials such as gallium arsenide to make significantly faster devices. gallium arsenide does not come with a built in insulating layer whereas the oxide layer on the silicon serves as a convenient and highly efficient built in insulator which is crucial to the manufacture of integrated circuits. All circuits require three separate materials: a semiconductor, a conductor and an insulator each of which is critical in fabricating the final device. By chance SiO_2 came as a made to order insulator on top of every silicon wafer which could be used as is or enhanced in thickness for wiring up transistors and other devices in integrated circuits.

Amorphous SiO_2 was indeed a wonderful insulator material but it came with its own set of drawbacks which became painfully apparent as soon as one wanted to make multilevel structures. Among the list of problems with using SiO_2 were the problems that such coatings were subject to pin hole defects if they were too thin and if too thick they were susceptible to cracking also. However, one of the biggest problems was the fact that SiO_2 coatings were conformal, i.e. they closely replicated the underlying surface. This is a killer issue when making multilevel circuits. The first level is no problem since one is building on top of a flat silicon wafer. After the first layer is built it has a complex topology of metal lines and other structures which need to be covered over before the next layer can be built. This is where the problem with SiO_2 starts, since after coating, all of the hills and valleys of the underlying circuitry will be replicated leaving a lumpy surface. This surface must somehow be flattened before the next layer can be built. One would much rather have a coating that filled in all of the hills and valleys leaving a completely flat surface the same way that water filling a rough basin leaves a perfectly flat surface regardless of the topology of the underlying lake bed. In the jargon of the industry what one wants is a planarizing coating material that will leave behind a flat smooth surface after coating.

Enter the polyimide materials. Since these materials go down as a viscous liquid, they tend to planarize the underlying topology much better than SiO_2. The degree of planarization is not perfect but is nonetheless much much better than what one gets with SiO_2. Thus it looked as though the polyimides were a natural to replace SiO_2 as the insulator of choice for multilevel circuits, since not only did they planarize much better they also had a significantly lower dielectric constant which for the circuit designers meant that one could go to higher wiring densities before the problem of crosstalk between metal lines became a problem. Thus it was that in the early to mid 1970's very serious development efforts got underway within the microelectronics industry to use the polyimides as replacement materials for SiO_2 in multilevel wiring structures. It was also at this time that the staff of MST got their first exposure to the polyimides and the host of problems that came along with them.

Professor J. E. Gordon has neatly summed up the nature of dealing with new materials with this insightful comment: "A deep, intuitive appreciation of the inherent cussedness of materials and structures is one of the most valuable accomplishments an engineer can have. No purely intellectual quality is really a substitute for this" ("Structures or why Things Don't Fall Down", J. E. Gordon (Da Capo, Press, 1978) p. 63). We were about to learn the truth of this of this statement firsthand as we started upon a development program to use the polyimide materials as insulator layers in multilevel wiring structures. But first a little introduction to the essential architecture of mainframe computers.

Wiring needs of a mainframe computer

All digital computer devices are what might BE called "vertically integrated". That is they form a layered

structure. At the lowest level, the basic calculations and data manipulations are carried out by transistors, diodes and resistors on silicon chips at the scale of roughly 1 micrometer and for the most advanced devices less than half that dimension. This is referred to as the chip level. A problem immediately arises due to the fact that humans cannot directly read data stored at the 1 micrometer scale. The digital signals must be scaled up to a coarser size or in technical jargon fanned out in order to be interpreted. The way this is done is to package the silicon chip on a larger substrate, typically but not necessarily a ceramic block, where the wiring structure is fanned out to roughly 100 times the dimension that exists on the chip. This is the second layer comonly referred to as the second level of wiring or the first level of packaging. If this is still not enough, yet a third level of wiring is employed in a third layer whereby the ceramic block is plugged into an epoxy based board which fans out the wiring density another 10 to 100 times. By this stage one usually achieves a wiring density that can be dealt with by human fingers which can then plug the wires into useful devices such as disk drives, display screens and keyboards. The polyimide materials were being considered for application at both the first and second level of wiring and this is where the story starts to get interesting. First, however, we need to digress briefly on the overall architecture of mainframe computers.

In the late 1960' and early 1970's the IBM company was having great success packaging single chips on ceramic substrates roughly 3/4" square. These ceramic modules would then be plugged into epoxy boards which could further be mounted into racks which would then constitute a mainframe computer. This was all well and good but there was still much room for improvement. In the first place all those cards and boards mounted in racks were large messy affairs taking up a lot of room and requiring extensive support structures such as cooling fans and power supplies. Another even more serious problem was the fact that by mounting each chip on a single ceramic module one was setting up a communications barrier of sorts between the chips. For example, say that a central processing circuit on one chip needs a piece of data on a remote memory chip. The electrical signal which transfers the data has to go from the chip to the ceramic module, thence to the supporting card, from there to the ceramic module supporting the memory chip and then to actual chip itself which sends the desired data back by the same route. It quickly became apparent to the electrical designers that the main bottleneck standing in the way of improving machine speed was the highly spread out packaging structure which single chip modules required. One idea that came up was to make the machine as one huge integrated circuit on a 3 inch wafer. This never flew for a number of reasons not the least of which was the problem of doing 1 micrometer lithography accurately over a span of 3 inches. The next level of thinking said that if we cannot cram the whole machine onto a single wafer at least we can mount all the chips on a single substrate and thereby eliminate one full level of packaging. At this stage the concept of a multichip module was born. Not only that, this concept was vigorously implemented resulting in multichip ceramic modules with approximately 40 levels of wiring buried in the ceramic and supporting up to 100 chips. It is hard to convey just how successful this multichip module concept was. By eliminating one full level of packaging the IBM company was able to manufacture machines in the early 1980's using chip technology that had been fully developed and amortized by the mid 1970's. These machines were cheaper than and out performed those of competitors which were using the latest expensive chip technology in order to achieve similar performance. I think one can safely say that this technology made a major contribution to the roughly $4billion/year profit that the company was earning in those days.

Implementing polyimides: or development programs always work out better in slide presentations than in the development lab

Part one: prelude to selecting a material

If it is true that in the fashion industry one can never be thin enough and in business never rich enough then in the computer industry ones machines can never run fast enough. Already in the late 1970's we were looking for ways to improve the performance of the large multichip substrates. The material being used at that time was essentially alumina (Al_2O_3) ceramic. This was the same material used for the single chip modules and was an industry standard insulator for all applications requiring stability at high temperatures. Its main strengths are its excellent insulating properties and outstanding thermal-mechanical durability. Its main weakness is its relatively high dielectric constant and its highly refractory nature which forces the use of equally refractory metals such as molybdenum as the embedded electrical conductor. This arises from the fact that the alumina and the chosen metal must be co-fired together in a furnace in order to achieve the final multilayered structure. Since firing alumina requires temperatures on the order of 2000 C only the refractory metals such as molybdenum can be used. Other metals such as copper and aluminum simply cannot take the heat!

From the electrical design point of view this was a double whammy both in terms of achievable signal speed and wiring density. As a rule of thumb the fastest and densest circuits are achieved by using a low dielectric constant insulator combined with a high conductivity metal. The alumina/molybdenum pair were quite mediocre performers on both counts. However, a solution to this dilemma quickly presented itself due to the dual electrical functions performed by the packaging substrate. At the most basic level the substrate has to provide power for all of the chips. The power signal does not have to be exceptionally

fast and the requirement is readily met by the alumina/molybdenum combination. At the next level comes the signal wiring which must transfer data among all of the chips. The machine performance is critically dependent on the speed of this wiring and this is where alumina/molybdenum fails to make the grade. The resolution to the problem now pops out. Make the basic substrate out of alumina/molybdenum to handle the power requirements and then build thin film polyimide/metal wiring on top to handle the signal wiring where the metal can now be a high conductivity material such as copper or aluminum.

The stage was now set for the polyimide materials to make a major contribution to high level packaging technology. All that had to be done was to select the appropriate material and make it work. In all of the group meetings and slide presentations it looked like a done deal. However, the cussedness of the real material world was about to enter the stage in a most decisive way.

The first step of course was to choose an appropriate polyimide material. This is a far more critical step than might be appreciated at first. There are essentially a near infinite number of polyimides to choose from depending on the details of the backbone architecture and any side groups that might be added. Nonetheless, a choice had to be made and made quickly since in the microelectronics industry any development program has to "hit the ground running" as the paramilitary people would put it. Requirement number one was that whatever material was going to be used had to be commercially available, which greatly pared down the field of choices. The final choice was made even easier since there was one polyimide that had already been experimented with at the chip level in making 1 micrometer thick insulator layers. This was a commercial product that went under the name "Skybond 703"®. The original application was apparently as a high temperature glue for high performance fighter aircraft. This material had all the desirable properties required of an insulator material including a low dielectric constant but most importantly it planarized very well. What cinched the case for the packaging laboratory, however, was the fact that this material already had a track record of sorts. Thus the program went forward to fabricate thin film wiring onto of ceramic substrates using the Skybond material. Everything looked fine since the only change being made was that 10 to 15 micrometer coatings would have to be fabricated instead of 1 micrometer layers which the chip people were using. Now how could simply making thicker coatings cause any problems? (Answer: The tendency of coatings yo crack and delaminate scales linearly with the coating thickness. This comes directly out of elementary fracture mechanics)

Part two: revenge of the material gods

In retrospect it is hard to believe that one could have found a worse material for the application at hand. In essence had we systematically screened all of the polyimide materials and then on the basis of a rigorous analysis tried to select the worst possible material I doubt that we could have beaten the Skybond material. As far as I can remember not a single viable substrate was made. The coatings all had massive cracking problems due to a number of thermal-mechanical problems that went entirely overlooked. The first problem was that Skybond had nearly all the mechanical properties of window glass. The strain at break was less than 2% and if you looked at a typical stress-strain curve and suppressed the scales on the axes the diagram was indistinguishable from that which a glass sample would give. This by itself does not necessarily imply disaster since many glue type materials are mechanically brittle but still perform adequately. However, the second problem that descended was the fact that the thermal expansion of the Skybond material was some 10 times larger than that of the ceramic onto which it was coated. This should have set off alarm bells since, as mentioned above, the polyimide materials have to be cured at an elevated temperature which in the case of Skybond was roughly 400 C. Oh Oh, could be trouble ahead. A little arithmetic immediately reveals that on cooling from the curing temperature at 400 C to room temperature at 20 C the part in question will have undergone a thermal excursion of roughly 380 C which, when multiplied by the thermal expansion difference between the polyimide and the ceramic, gives rise to a rather substantial thermal strain. There is a handy little equation called the membrane formula which allows one to quickly estimate the expected stress level in a coating which is subjected to these process conditions. In a nutshell it goes like this:

$$s = E \times a \times DT/(1-v)$$

In this equation E is the modulus of the polyimide, thermal expansion difference between the polyimide and the ceramic, T the temperature excursion and Poisson's ratio of the polyimide. In this example E = 3GPa, = 1/3, a = 30×10^{-6}/C and DT = 380 C. Inserting these values into our little formula gives an expected biaxial stress of = 0.051 GPa (GigaPascal) or in more convenient notation 51 MPa (Megapascal). Now what does this value 51 MPa mean? Looking at a stress strain diagram for Skybond one immediately notices that the ultimate tensile strength of the material is close to 70 MPa. Thus upon coating Skybond onto a ceramic substrate and curing at 400 C one has a uniform coating that is already stressed to about 3/4 of its ultimate breaking strength. Not enough to cause a problem you say. This is true. The properly fabricated blanket coatings almost never gave a problem. Unfortunately one cannot stop at a blanket coating. One also has to drill via holes in the film and fill them with metal in order to electrically connect to the wiring below. However, by doing so one creates what is know as a stress riser or point of stress concentration. A fairly elementary analysis demonstrates that close to such a hole the local stress field

increases by a factor of 2 to 3 which automatically puts the stress level over the top for Skybond. (A detailed discussion of the stress state near a via hole can be found in the following: "Stresses in thin Polymeric Films: Relevance to Adhesion and Fracture", Robert H. Lacombe in Surface and Colloid Science in Computer Technology, Ed. K. L. Mittal (Plenum Press, New York, 1987) p.179) This was in fact what was observed in all cases. As soon as the parts emerged from the plasma etcher where the via holes were bored, ubiquitous radial cracks were observed emanating from nearly every via structure. In fact nearly any kind of imperfection in the coating such as particle contamination or a defect in the substrate would also act as a stress riser and thereby give rise to cracking of the polyimide.

Part 3: revenge of the adhesion gods

OK, so we learned that it was not a good idea to go with a brittle material and maybe also we would have to pay more attention to the thermal mechanical properties of the polyimide as well as its planarizing properties and whether or not it is commercially available. Quite after the fact I looked into the history of the development process and thereby came across a memo describing the material property selection criteria. In this document was a table listing about a dozen different polyimides along with columns specifying their electrical properties, how well they planarized, whether they were commercially available ... and so on. There was only one column listing a thermal-mechanical property and that was the thermal expansion coefficient. However, the memo went on to point out that, since all of the polyimides listed had roughly similar thermal expansion coefficients, this property would thereby not be considered. Oh well, at least someone was starting to worry about the thermal-mechanical properties of the polyimide but somehow could not quite make a case for paying close attention to this detail. This was hardly surprising since in the early days of the program the entire team consisted of either chemists or electrical engineers who were really not in a position to appreciate the subtleties of stresses in thin films.

It quickly became clear that using a brittle material was not going to work and as soon as someone familiar with the mechanical properties of polyimides was brought on board it was pointed out that there are mechanically tough materials available. In fact the PMDA-ODA material which forms the basis of the KAPTON films mentioned above was one such material. Unlike the other polyimides which were essentially amorphous materials the PMDA-ODA material was quasi crystalline in that it had an ordered liquid crystalline type of morphology due to the length and rigidity of the backbone repeat unit. The stress strain diagram of this material looks almost like a rubber with a strain at break approaching 100% or better. It is the semi-ordered morphology that explains the toughness of the material. If you have an amorphous glassy material like the standard polyimides, and you apply a tensile load to it, there is not much that the material can do to accommodate the applied load. The induced stress in the material must be supported by the inter-chain van der Waals attractive forces and these are not very strong and are very short range to boot. Thus one gets a very low strain at break. With the PMDA-ODA material, however, the initial effect of the applied load is to start pulling apart and stretching out the semi-crystalline phase and only after this is accomplished does the load shift entirely to the inter-chain bonding forces. This effect is most dramatic in polycrystalline polymers such as polyethylene which can stretch out to 200% or better.

Thus, by going over to the PMDA-ODA material the problem of mechanical cracking was eliminated but unfortunately the thin film stresses remained since PMDA-ODA has roughly the same thermal expansion behavior as the other polyimides. The immediate consequence of this was that instead of cracking the material delaminated. The gods of adhesion now extracted their revenge. Coating PMDA-ODA on nearly anything such as silicon or SiO_2 tends to give poor adhesion due mainly to the chain rigidity, limited scope for chemical interaction and the pervasive presence of high residual stress levels. Thankfully, the chemists were now able to step in and alleviate this problem with the use of coupling agents such as the silane materials. The silanes improved the adhesion of PMDA-ODA tremendously and allowed the program to progress to the point where thin films with full via structures could be fabricated.

Part 4: Twilight of the polyimide program

I would like to report that at this stage the program went forward ending in a triumphal success. However, this was not to be. Development programs involving advanced materials technologies are complex and chaotic processes and the final result tends to be determined as much by luck and circumstances as by engineering skill and design expertise. Too much time and resources had been dissipated with too little in the way of results coming out of it. In addition, an even more ominous threat came onto the scene. The ceramics group which had developed and implemented the alumina based technology was not standing idly by. They essentially came upon a ceramic material which could be sintered at under 1000 C and thus could be co-fired with copper. Not only that, this material also had a low dielectric constant. Thus coupling a low dielectric constant with a high conductivity metal the electrical designers could easily show that an all ceramic/metal module was possible with greatly improved signal speed and wiring density. Two other circumstances, however, were what really sealed the fate of the polyimide program. The major circumstance was essentially political. The ceramics group

was essentially the party in power having already developed and implemented the alumina/molybdenum technology and as a consequence held most of the positions of authority and clout within the organization. The second circumstance which provide the "coup de grâce" was the fact that the limited resources available allowed only one program to go forward. Thus ended the polyimide program.

Epilog

As fate would have it the material and adhesion gods were just beginning to show their wrath and were now about to truly demonstrate the cussedness of real materials. Whereas it was true that the new ceramic material had much better electrical properties than alumina, its thermal-mechanical properties were a disaster, even worse than Skybond. In order to achieve a low sintering temperature the new material had to have a high glass content and this essentially gave it all the properties of window glass. Worse yet, as is well known to all those who work with glass, it had very poor adhesion to copper. As a consequence cracking and delamination problems descended with a vengeance. In retrospect all of this could have been anticipated in advance but one always has to remember the chaotic nature of real world development programs. Time, resources and expertise were in very short supply and the need to press forward unrelenting. Leaving out the gruesome details, the ultimate consequence of these technology failures were truly tragic. For the multichip module program the critical upshot was that a technology which should have been delivered in 1985 did not materialize until approximately 1990. This is a 5 year delay which in the microelectronics industry amounts to an eternity. One can never afford that kind of delay. While the development lab was wrestling with the daemons unleashed by the material and adhesion gods, microprocessors and personal computers based on them were steadily advancing. These devices were selling for a few thousand dollars whereas mainframe computers were going for closer to a million. In order to be competitive the mainframe had to offer much greater computing power. A factor of 2 or 3 was not good enough since PCs were so cheap one could just let a few of them run overnight. No, the mainframe had to offer a bare minimum of at least a factor of 10 to 100 greater compute power. In essence the only way a customer is going to shell out mainframe level dollars is if the mainframe would handle work loads the PC's could not. One had to deliver what is called an "enabling" technology. One that could deliver what the competition could not. If we could have delivered the advanced multichip module technology in 1985 as planned then mainframes 100 times faster than PC's would have emerged. Delaying the technology for 5 years made that dream impossible.

Ending on a more upbeat note, the polyimide materials have nonetheless done rather well in the electronics industry as a whole. In particular they have found widespread use in flex circuitry whereby integrated circuit modules are mounted directly onto a flexible polyimide film similar to the KAPTON material mentioned previously. All of the wiring is electodeposited copper which can be made to adhere very well to polyimide and, since everything is flexible, problems with cracking are avoided as well. This technology has found widespread use in everything from CD Players to ink jet print heads.

In closing it is appropriate to reiterate what is I believe the universal theme of this essay "there are no bad materials only bad uses of good ones". All materials are essentially indifferent to our needs and aspirations, they obey the laws of physics and chemistry very rigidly and that is that. Any material whatever can find successful application so long as we understand its inherent nature and physical limitations. The Skybond material works quite well as a high temperature glue so long as you understand its limits. Developing such an understanding for polyimides and other high temperature polymers is one of the major goals of the upcoming MST symposia on HIGH TEMPERATURE POLYMERS and ADHESION ASPECTS OF THIN FILMS.

Basic Bipolar Transistor

- For the purpose of this basic module, we simply take the contents of the "Bipolar Transistor" module from the *Semiconductor Hyperscript*.
  - There you will always find the newest version; the module is reproduced below.
  - It is about as basic as it can be - just assumming that you know the *basics about **pn**-junctions*.
  - If you remember **pn**-junctions diodes only vaguely (or not at all), turn to the diode parts of the Semiconductor Hyperscripts and check the links from there.
- If you understand German; this link will bring you to the relevant parts of the Hyperscript "Einführung in die Materialwissenschaft **II**"

Bipolar Transistors: Basic Concept and Operation

- We are not very particularly interested in **bipolar transistors** and therefore will treat them only cursory.
  - Essentially, we have two junctions diodes switched in series (sharing one doped piece of **Si**), i.e. a **npn** or a **pnp** configuration, with the *added condition* that the middle piece (the **base**) is *very thin*. "Very thin" means that the base width d_{base} is much smaller than the diffusion length L.
- The other two doped regions are called the **emitter** and the **collector**.
  - For transistor operation, we switch the emitter - base (**EB)** diode in forward direction, and the base - collector (**BC**) diode in reverse direction as shown below.
  - This will give us a large forward current and a small reverse current - which we will simply neglect at present - in the **EB** diode, exactly as described for diodes. What happens in the **BC** diode is more complicated and constitutes the principle of the transistor.
  - In other words, in a **pnp** transistor, we are injecting a lot of holes into the base from the emitter side, and a lot of electrons into the emitter from the base side; and vice versa in a **npn**- transistor. Lets look at the two **EB** current components more closely:
- For the *hole* forward current, we have in the simplest approximation (ideal diode, no reverse current; no **SCR** contribution):

$$j_{hole}(U) = \frac{e \cdot L \cdot n_i^2}{\tau \cdot N_{Acc}} \cdot \exp -\frac{e \cdot U}{kT}$$

  - and the relevant quantities refer to the *hole* properties in the ***n** - doped base* and the doping level N_{Acc} in the ***p** - doped emitter*. For the electron forward current we have accordingly:

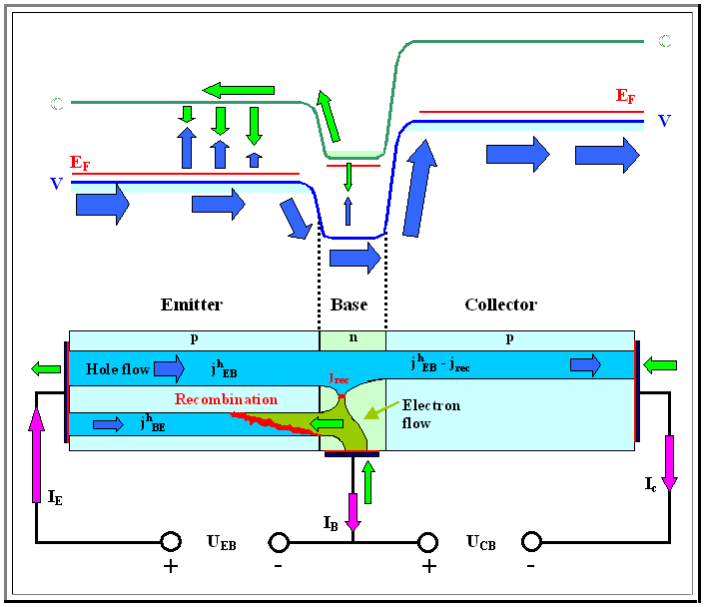
$$j_{electron}(U) = \frac{e \cdot L \cdot n_i^2}{\tau \cdot N_{Don}} \cdot \exp -\frac{e \cdot U}{kT}$$

  - and the relevant quantities refer to the *electron* properties in the ***p** - doped emitter* and the doping level N_{Don} in the ***n** - doped base*.
  - The relation between these currents, i.e. $j_{hole}/j_{electron}$, which we call the **injection ratio** κ, then is given by

$$\kappa = \frac{\frac{L_h}{\tau_h \cdot N_{Ac}}}{\frac{L_e}{\tau_e \cdot N_{Don}}} = \frac{N_{Ac}}{N_{Don}}$$

  - Always assuming that electrons and holes have identical lifetimes and diffusion lengths.

- The *injection ratio* κ is a prime quantity. We will encounter it again when we discuss optoelectronic devices! (in a separate lecture course).
- For only one diode, that would be all. But we have a second diode right after the first one. The holes injected into the base from the emitter, will diffuse around in the base and long before they die a natural death by recombination, they will have reached the other side of the base
  - There they encounter the electrical field of the base-collector **SCR** which will sweep them rapidly towards the collector region where they become majority carriers. In other words, we have a large hole component in the reverse current of the **BC** diode (and the normal small electron component which we neglect).
  - A band diagram and the flow of carriers is shown schematically below in a band diagram and a current and carrier flow diagram.



- Let's discuss the various currents going from left to right.
  - At the *emitter contact*, we have two hole currents, j_{EB}^{h} and j_{BE}^{h} that are converted to electron currents that carry a negative charge away from the emitter. The technical current (mauve arrows) flows in the opposite direction by convention.
- For the *base current* two major components are important:
  - **1.** An electron current j_B^e, directly taken from the *base contact*, most of which is injected into the emitter. The electrons are minority carriers there and recombine within a distance ***L*** with holes, causing the small hole current component shown at the emitter contact.
  - **2.** An internal recombination current j_{rec} caused by the few holes injected into the base from the emitter that recombine in the base region with electrons, and which reduces j_B^e somewhat. This gives us

$$j_{BE}^{h} = j_B^e - j_{rec}$$

  - Since all holes would recombine within ***L***, we may approximate the fraction recombining in the base by

$$j_{rec} = j_{EB}^{h} \cdot \frac{d_{base}}{L}$$

- Last, the current at the *collector contact* is the *hole* current $j_{EB}^{h} - j_{rec}$ which will be converted into an electron current at the contact.
- The external terminal *currents* I_E,I_B, and I_C thus are related by the simple equation

$$I_E = I_B + I_C$$

A bipolar transistor, as we know, is a *current amplifier*. In black box terms this means that a small current at the the *input* causes a large current at the *output*.

- The input current is I_B, the output current I_C. This gives us a current amplification factor γ of

$$\gamma = \frac{I_C}{I_B} = \frac{I_E}{I_B} - 1$$

- Lets neglect the small recombination current in the base for a minute. The emitter current (density) then is simply the total current through a **pn**-junction, i.e. in the terminology from the picture $j_E = j_{BE}^h + j_B^e$, while the base current is just the electron component j_B^e.
- This gives us for I_E/I_B and finally for γ:

$$\frac{I_E}{I_B} = \frac{j_{BE}^h + j_B^e}{j_B^e} = \kappa + 1$$

$$\gamma = \frac{I_E}{I_B} - 1 = \kappa + 1 - 1 = \kappa = \frac{N_{Ac}}{N_{Don}}$$

Now this is really easy! We will obtain a large current amplification (easily **100** or more), if we use a lightly doped base and a heavily doped emitter. And since we can use large base - collector voltages, we can get heavy power amplification, too.

- Making better approximations is not difficult either. Allowing somewhat different properties of electrons and holes and a finite recombination current in the base, we get

$$\gamma = \frac{\frac{L_h}{\tau_h \cdot N_{Ac}}}{\frac{L_e}{\tau_e \cdot N_{Don}}} \cdot \left(1 - \frac{d_{base}}{L}\right) \approx \frac{N_{Don}}{N_{Ac}} \cdot \left(1 - \frac{d_{base}}{L}\right)$$

- The approximation again is for identical life times and diffusion lengths.

Obviously, you want to make the base width d_{base} small, *and* keep L large.

Real Bipolar Transistors

Real bipolar transistors, especially the very small ones in integrated circuits, are complicated affairs; for a quick glance on how they are made and what the **pnp** or **npn** part looks like, use the link.

Otherwise, everything mentioned in the context of real diodes applies to bipolar transistors just as well. And there are, of course, some special topics, too.

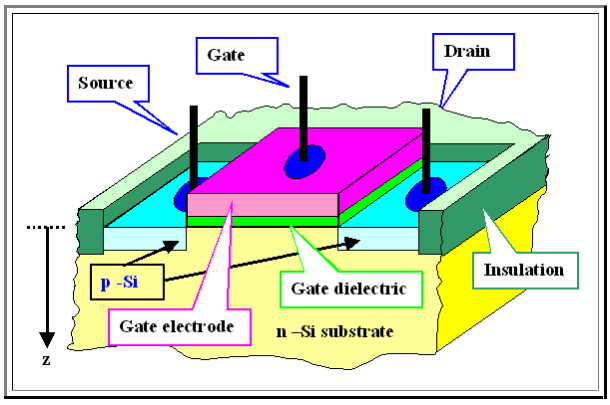
- But we will *not* discuss this any further, except to point out that the "small device" topic introduced for a simple p-n-junction now becomes a new quality:
- Besides the length of the emitter and collector part which are influencing currents in the way discussed, we now have the **width of the base region** d_{base} which introduces a new quality with respect to device dimensions and device performance.
- The numerical value of d_{base} (or better, the relation d_{base}/L), does not just change the device properties somewhat, but is the *crucial* parameter that brings the device into existence. A transistor with a base width of several **100 µm** simply is not a transistor, neither are two individual diodes soldered together.

- The immediate and unavoidable consequence is that at this point of making semiconductor devices, *we have to make things real small*.
  - Microtechnology - typical lengths around or below **1 µm** (at least in one dimension) - is mandatory. There are no big transistors in more than two dimensions.
  - Understanding *microscopic* properties of materials (demanding quantum theory, statistical thermodynamics, and so on) becomes mandatory. *Materials Science and Engineering was born*.

Basic MOS Transistor

Qualitative Description

- The basic concept of a ***MOS* Transistor** transistor is simple and best understood by looking at its structure:



  - It is always an *integrated* structure, there are practically no single individual **MOS** transistors.
- A **MOS** transistor is primarily a switch for digital devices. Ideally, it works as follows:
  - If the voltage at the **gate electrode** is "*on*" , the transistor is "*on*", too, and current flow between the **source** and **drain** electrodes is possible (almost) without losses.
  - If the voltage at the gate electrode is "*off*", the transistor is "*off*", too, and no current flows between the source and drain electrode.
- In reality, this only works for a given *polarity* of the gate voltage (in the picture above, e.g., only for negative gate voltages), and if the supply voltage (always called U_{DD}) is not *too small* (it used to be **5 V** in ancient times around **1985**; since then it is declining and will soon hit an ultimate limit around **1 V**).
  - Moreover, a **MOS** transistor needs *very thin* **gate dielectrics** (around, or better below **10 nm**), and *extreme control* of materials and technologies if real **MOS** transistors are to behave as they are expected to in "ideal" theory.
- What is the working principle of an "ideal" **MOS** transistor?
  - In order to understand it, we look at the behavior of carriers in the **Si** under the influence of an external electrical field under the gate region.
- Understanding **MOS** transistor *qualitatively* is easy. We look at the example from above and apply some source-drain voltage U_{SD} in either polarity, but *no gate voltage yet*. What we have under these conditions is
  - A **n**-type **Si** substrate with a certain equilibrium density of electrons $n^e(U_G = 0)$, or $n^e(0)$ for short. Its value is entirely determined by doping (and the temperature, which we will neglect at the present, however) and is the same everywhere. We also have a much smaller concentration $n^h(0)$ of holes.
  - Some **p**-doped regions with an equilibrium concentration of holes. The value of the hole concentration in the source and drain defined in this way also is determined by the doping, but the value is of no particular importance in this simple consideration.
  - Two **pn**-junctions, one of which is polarized in forward direction (the one with the positive voltage pole), and the other one in reverse. This is true for any polarity; in particular one junction will *always* be ***biased*** in reverse. Therefore *no source-drain current I_{SD} will flow* (or only some small reverse current which we will neglect at present).
  - There will also be no current in the forwardly biased diode, because the **n-Si** of the substrate in the figure is not electrically connected to anything (in reality, we might simply ground the positive U_{SD} pole and the substrate).
- In summary, for a gate voltage $U_G = 0$ **V**, there are no currents and everything is in equilibrium. But now apply a *negative voltage* at the gate and see what happens.
  - The electrons in the substrate below the gate will be electrostatically repelled and driven into the substrate. Their concentration directly below the gate will go down, $n^e(U)$ will be a function of the depth coordinate z .

$$n^e = n^e(z) = f(n^e(0), U)$$

  - Since we still have equilibrium, the mass action law for carriers holds anywhere in the Si, i.e. .

$$n^e(z) \cdot n^h(z) = n_i^2$$

- With n_i = **intrinsic carrier density in Si = const.(U,z)**
- This gives us

$$n^h(z) = \frac{n_i^2}{n^e(z)}$$

- In other words: If the electron concentration below the gate goes down, the hole concentration goes up.

If we sufficiently decrease the electron concentration under the gate by cranking up the gate voltage, we will eventually achieve the condition $n^h(z=0) = n^e(z=0)$ right under the gate, i.e. at $z = 0$

- If we increase the gate voltage even more, we will encounter the condition $n^h(z) > n^e(z)$ for small values of z, ie. for $z_c > z > 0$.

In other words: Right under the gate we now have *more holes than electrons*; this is called a state of **inversion** for obvious reasons. **Si** having more holes than electrons is also called *p-type Si*. What we have now is a **p**-conducting **channel** (with width z_c) connecting the **p**-conducting source and drain.

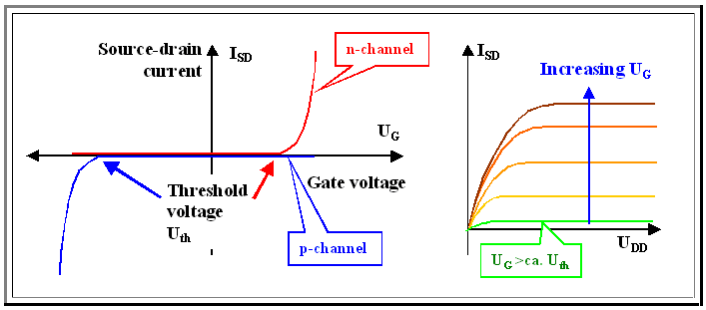
- There are no more **pn**-junctions preventing current flow under the gate - current can flow freely; only limited by the ohmic resistance of contacts, source/drain and channel.

Obviously, while cranking up the gate voltage *with the right polarity*, sooner or later we will encounter inversion and form a conducting channel between our terminals which becomes more prominent and thus better conducting with increasing gate voltage

- The resistivity of this channel will be determined by the amount of **Si** we have inverted; it will rapidly come down with the voltage as soon as the **threshold voltage** necessary for inversion is reached.

If we reverse the voltage at the gate, we attract electrons and their concentration under the gate increases. This is called a state of **accumulation**. The **pn** junctions at source and drain stay intact, and no source - drain current will flow.

- Obviously, if we want to switch a **MOS** transistor "on" with a *positive* gate voltage, we must now reverse the doping and use a **p**-doped substrates and **n**-doped source/drain regions.
- The two basic types we call "**n-channel MOS**" and "**p-channel MOS**" according to the kind of doping in the channel upon inversion (or the source/drain contacts).
- Looking at the electrical characteristics, we expect curves like this:



- The dependence of the source-drain current I_{SD} on the gate voltage U_G is clear from what was described above, the dependence of I_{SD} on the source-drain voltage U_{SD} with U_G as parameter is maybe not obvious immediately, but if you think about it a minute. you just can't draw currents without some U_{SD} and curves as shown must be expected qualitatively.

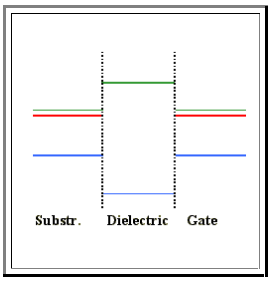
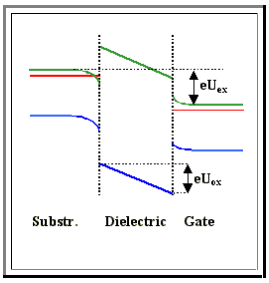
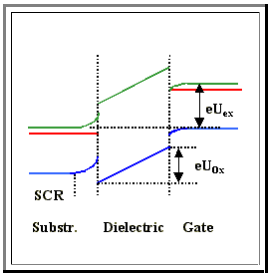
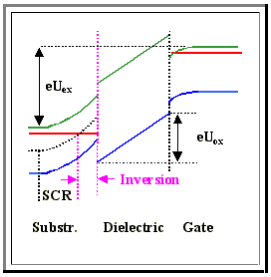
What can we say *quantitatively* about the working of an **MOS** transistor?

- What determines the threshold voltage U_{th}, or the precise shape of the $I_{SD}(U_{th})$ curves? Exactly how does the source - drain voltage U_{SD} influence the characteristics? How do the prime quantities depend on material and technology parameters, e.g. the thickness of the gate dielectric and its dielectric constant ϵ_r or the doping level of substrate and source/drain?
- Plenty of questions that are, as a rule, not easily answered. We may, however, go a few steps beyond the qualitative picture given above.

Some Quantitative Considerations

The decisive part is achieving *inversion*. Lets see how that looks like in a band diagram. To make life easier, we make the gate electrode from the same kind of **n-Si** as the substrate, just highly doped so it is as metallic as possible - we have the same kind of band diagram then to the left and right of the gate dielectric

- Lets look schematically what that will give us for some basic cases:

Voltage at the gate	Conditions in the Si	Voltage drop	Charge distribution
Zero gate voltage. "**Flat band**" condition			
	Nothing happens. The band in the substrate is perfectly flat (and so is the band in the contact electrode, but that is of no interest).	We only would have a voltage (or better potential) drop, if the Fermi energies of substrate and gate electrode were different	There are no *net* charges
Positive gate voltage. **Accumulation**			
	With a positive voltage at the gate we attract the electrons in the substrate. The bands must bend down somewhat, and we increase the number of electrons in the conduction band accordingly. (There is a bit of a space charge region (**SCR**) in the contact, but that is of no interest).	The voltage drops mostly in the oxide	There is some *positive charge* at the gate electrode interface (with our **Si** electrode from the **SCR**), and *negative charge* from the many electrons in the (thin) accumulation layer on the other side of the gate dielectric.
Small negative gate voltage. **Depletion**			
	With a (small) negative voltage at the gate, we repel the electrons in the substrate. Their concentration decreases, the hole concentration is still low - we have a layer depleted of mobile carriers and therefore a **SCR**.	The voltage drops mostly in the oxide, but also to some extent in the **SCR**.	There is some *negative charge* at the gate electrode interface (accumulated electrons with our **Si** electrode), and *positive charge* smeared out in the the (extended) **SCR** layer on the other side of the gate dielectric.
Large negagive gate voltage. **Inversion**			
	With a (large) negative voltage at the gate, we repel the electrons in the substrate very much. The bands bend so much, that the Fermi energy (red line) is in the lower half of the band close to the interface. In this region holes are the majority carriers, we gave *inversion*. We still have a **SCR**, too.	The voltage drops mostly in the oxide, but also to some extent in the **SCR** and the inversion layer.	There is more *negative charge* at the gate electrode interface (accumulated electrons with our **Si** electrode), some *positive charge* smeared out in the the (extended) **SCR** layer on the other side of the gate dielectric, and a lot of *positive charge* from the holes in thin inversion layer.

- Qualitatively, this is clear. What happens if we replace the (highly **n**-doped) **Si** of the gate electrode with some metal (or **p**-doped **Si**)?
  - Then we have *different Fermi energies* to the left and right of the contact, leading to a *built-in potential* as in a **pn**-junction. We will than have some band bending at zero external voltage, flat band conditions for a non-zero external voltage, and concomitant adjustments in the charges on both sides.
  - But while this complicates the situation, as do unavoidable fixed immobile charges in the dielectric or in the **Si**-dielectric interface, nothing new is added.
- Now, the decisive part is achieving inversion. It is clear that this needs some minimum threshold voltage U_{th}, and from the pictures above, it is also clear that this request translates into a request for some *minimum charge* on the capacitor formed by the gate electrode, the dielectric and the **Si** substrate.
  - What determines the amount of charge we have in this system? Well, since the whole assembly for any distribution of the charge can always be treated as a simple capacitor C_G, we have for the charge of this capacitor.

$$Q_G = C_G \cdot U_G$$

  - Since we want U_{th} to be small, we want a *large gate capacitance* for a large charge Q_G, and now we must ask: What determines C_G?
- If all charges would be concentrated right at the interfaces, the capacitance *per area unit* would be given simply by the geometry of the resultant plate capacitor to

$$C_G = \frac{\epsilon\epsilon_0}{d_{Ox}}$$

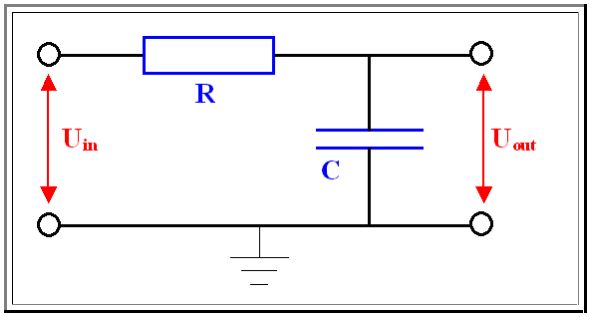
  - With d_{Ox} = thickness of the gate dielectric, (so far) always silicon dioxide SiO_2.
- Since our charges are somewhat spread out in the substrate (we may neglect this in the gate electrode if we use metals or very highly doped **Si**), we must take this into account.
  - In electrical terms, we simply have a second capacitor C_{Si} describing the effects of spread charges in the **Si**, switched in series to the geometric capacitor which we now call **oxide capacitance** C_{Ox}. It will be rather large for concentrated charges, i.e. for accumulation and inversion and small for depletion.
  - The total capacitance C_G then is given by

$$\frac{1}{C_G} = \frac{1}{C_{Ox}} + \frac{1}{C_{Si}}$$

- For inversion and accumulation, when the most of the charge is close to the interface, the total capacitance will be dominated by C_{Ox}. It is relatively large, because the thickness of the capacitor is small.
  - In the depletion range, C_{Si} will be largest and the total capacitance reaches a minimum.
  - In total, C_G as a function of the voltage, i.e. $C_G(U)$ runs from a constant value at large positive voltages through a minimum back to about the same constant value at large positive voltages. The resulting curve contains all relevant information about the system. Measuring $C_G(U)$ is thus the first thing you do when working with **MOS** contacts.
  - While it is not extremely easy to calculate the capacitance values and everything else that goes with it, it can be done - just solve the Poisson equation for the problem.
- All things considered, we want C_{Ox} to be *large*, and that means we want the dielectric to be *thin* and to have a *large* dielectric constant - as stated above without justification.
  - We also want the dielectric to have a large breakdown field strength, no fixed charges in the volume, no interface charges, a very small **tg δ**; it also should be very stable, compatible with **Si** technology, and cheap.
  - In other words, we wanted SiO_2 - even so its dielectric constant is just a mediocre **3.9** - for all those years of microelectronic wonders. But now (**2001**), we want something better with respect to dielectric constants. Much work is done, investigating, e.g., CeO_2, Gd_2O_3, ZrO_2, Y_2O_3, $BaTiO_3$, **BaO/SrO**, and so on. And nobody knows today (**2002**) which material will make the race!

RC Time Constant

- This will be a short statement of very elementary facts from basic physics or electrical engineering.
  - *Any* "wire" used to carry electricity has a resistance ***R*** (we will not consider superconductors here) - just measure it via ***R = Ul/l***. Its resistance measured in Ohms (O) might be small, but it is never zero.
  - *Any* "wire" that serves its purpose is isolated from the environment by some dielectric (air or some insulation material) having some dielectric constant ϵ_r
  - At some distance from *any* wire is always another conductor at some arbitrary potential. Even if our hypothetical device consist of a wire only, there is always the "earth" somewhere at (by definition) zero potential. It follows: *Any* wire has some capacitance ***C*** to something else. Again, this unintentional **parasitic capacity** might be small, but it is never zero. Moreover, whatever value it has, it is proportional to the dielectric constant ϵ_r of the dielectric in question.
- *Any* wire (at not too high frequencies) thus can be described by an equivalent circuit diagram that looks like this.



- What happens if we put a digital signal on one end of the wire, which we call the "input" end? Ideally, that means that the voltage or better potential on the wire (the not grounded upper "wire" in the picture above) jumps from **0V** to, say, **5V** instantaneously. What happens at the output?
  - Ideally, the potential would also go up suddenly after a certain time t_0 which is dictated by the speed of light ***c***, because nothing can move faster than that. we thus have $t_0 = l/c$ with ***l*** = length of the wire, and ***c*** = speed of light "in" the wire, whatever that means.
  - However, as soon as we raise the potential at the input, we have to put charge in the parasitic capacitor ***C***. The voltage at this capacitor is the output voltage, and for a voltage ***U*** we need to have the charge $Q = C \cdot U$ stored in the capacitor.
  - For that a current has to flow into the capacitor, and that current will be restricted by the resistance ***R***. Note that for just transmitting information as jumps in the potential, current flow would not really be necessary, but our parasitic capacitor needs charges flowing in (and later out) of it, if its potential is to change.
- It is easy to describe quantitatively what happens. For the current ***I*** flowing during the charging of the capacitor we have two equations.

$$I = C \cdot \frac{dU_{out}}{dt} = \frac{U_{in} - U_{out}}{R}$$

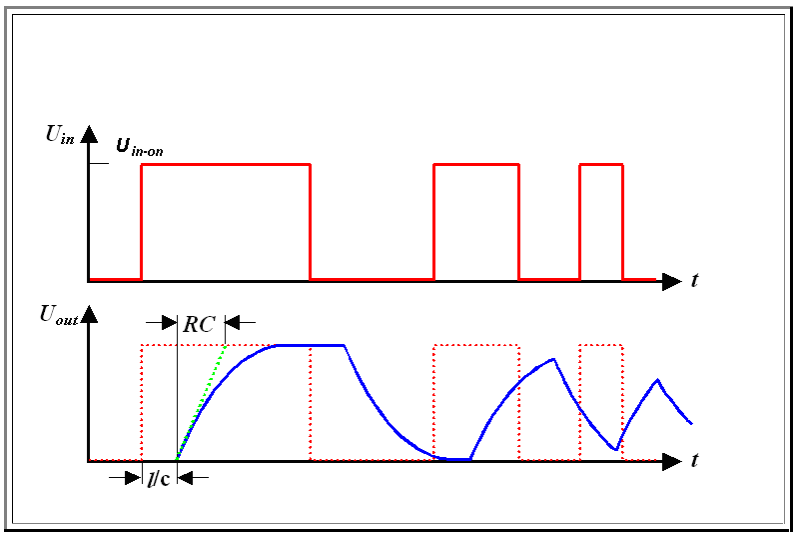
  - This is an easy differential equation. We will just look at the solutions for the two cases of U_{in} going "up" to $U_{in\text{-}on}$ at $t = 0$ and at U_{in} going "down from " to $U_{in\text{-}on}$ to **0**; again at $t = 0$.
  - For the "going up" case we obtain

$$U_{out} = U_{in\text{-}on} \cdot \{1 - \exp{-(t/RC)}\}$$

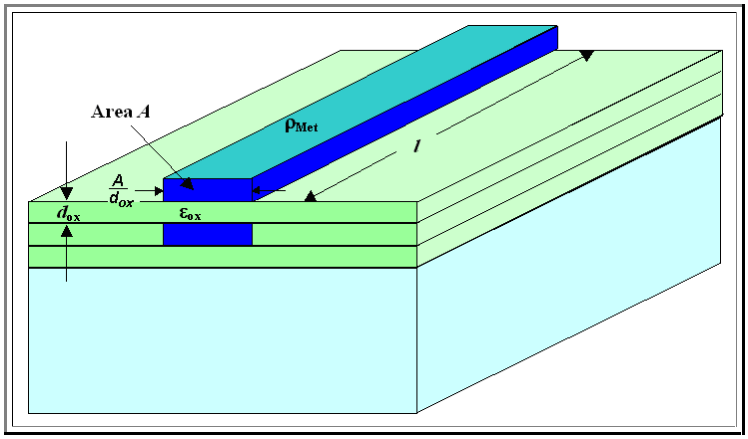
  - This describes an output voltage that increases as a function of time with the **time constant** ***RC***; i.e. after the time ***RC*** the output voltage is $U_{out} = U_{in\text{-}on} \cdot (1 - 1/e) = 0.63 \cdot U_{in}$.
- If we look at what happens if the input voltage is "going down", i.e.switched to zero (after it has been on long enough to make sure the output voltage is equal to the input voltage), we obtain in the same way:

$$U_{out} = U_{in\text{-}on} \cdot \exp{-(t/RC)}\}$$

  - The output voltage thus decays with the time constant ***RC*** from the "on" value to zero.
- Putting all of this together for some input signals gives the following picture:



- What we have now is that at the output end of a "wire" with length ***l***, the voltage will start to go up after a time $t_0 = l/c$ with **c** = speed of light, i.e. the speed of electromagnetic wave propagation in the wire (**c** might not be exactly the speed of light in vacuum, but that is not the important part here).
  - The voltage then will start to increase, after the **time constant *RC*** it will be at about **2/3** of the input voltage. After a few time constant it is identical to the input voltage.
  - If the input signal goes "down" again, the output signal will follow, again with a delay of t_0 and then with a decay time again given by the time constant ***RC***.
- Now consider that we are discussing digital applications. The output voltage will be interpreted as either "**0**" or "**1**"; let's say the low potential = **0**, the high potential = **1**.
  - During the transients of the voltage, we are neither here nor there, but our electronic circuitry typically will interpret a potential **< *U*/2** as **0**, **> *U*/2** as **1**. That simply means that the time it takes for a signal to travel through a wire with the length ***l*** and then to be recognized by what ever follows as a "**1**" is t_0 plus about one time constant ***RC*** of the wire. That means we have an additional delay time given by the "undesirable" properties of the wire.
  - Even worth! The picture makes clear, that as soon as you start to transmit signals shorter than a few time constant ***RC***, your output signal can't even follow the input any more. In other words, at input frequencies much larger than about **1/*RC***, you run into big problems.
- In plain words: You cannot run your circuitry at frequencies **> 1/*RC*** - give or take a factor of **2 - 3**.!
- Where does that leave us? How large is ***RC*** typically for the "wires" in an integrated circuit. Let's look at a simplified situation:



- Let's look at a typical **Al** or **Cu** conductor on a chip with a length of **1 cm**, a cross-sectional area of **0.5 µm²** and the specific resistivity of a typical metal of about **2 µΩcm**. The total resistance of our "wire" then is

$$R = \frac{\rho \cdot l}{A} = \frac{2 \cdot 10^{-6}\ \Omega\text{cm}^2}{0.5 \cdot 10^{-8}\text{cm}^2} = 400\ \Omega$$

  - The capacitor in the picture is formed by the two longish "plates", each with an area $l \cdot w = l \cdot A/d_{Ox}$ (***w*** is the lateral extension of the conductor strip; i.e $w = A/d_{Ox}$). With typical numbers like $d_{Ox} = 300$ nm, dielectric constant $\epsilon = 3.7$ (the value for SiO_2), the capacity of the wires then is

$$C = \epsilon\epsilon_0 \frac{A \cdot l}{d^2_{Ox}} = \frac{3.7 \cdot 8.85\,10^{-12}\ A \cdot s \cdot 0.5 \cdot 10^{-12}\ m^2 \cdot 10^{-2}\ m}{9 \cdot 10^4 \cdot 10^{-18}\ V \cdot m \cdot m^2} = 1.82 \cdot 10^{-12} \frac{A \cdot s}{V} = 1.82\ pF$$

- For our time constant ***RC*** we obtain

$$R \cdot C = 400 \cdot 1.82 \cdot 10^{-12} \frac{A \cdot s \cdot V}{V \cdot A} = 7.28 \cdot 10^{-10}\ s$$

- The maximal frequency we can transmit through this "wire " then would be $\nu_{max} = 1/RC = 1.36$ **GHz**. This may appear a bit "handwaving" or just a rough estimation; nevertheless, the problem should be clear.
  - On a real chip, signals travel through many wires; possibly far shorter then **1 cm**, and possibly not all the way atop another wire, so the capacity could be somewhat smaller. On the other hand, on a real chip, theer are also neighboring wires to the left and the rigth, which increases the capacity.
  - Nowadys, we always use a multilevel metalization scheme and being smart, we will keep the small wires on the lower levels quite short, and give the long wires (i.e. for power supply) on the upper levels a large cross-section - keeping ***R*** down in both cases. Nevertheless, a clock frequency of **4 GHz**, standard nowadays on many mass-produced microprocessors, is an amazing feat considering the number from above.
  - Essentially, the specific resistivity ρ of the conductor and the dielectric constant ϵ of the "**intermetal dielectric**" limit the maximal frequency of the chip after all tricks of an optimized geometry have been exhausted.
- Decreasing ρ by a factor of about **1.6** by switching from **Al** to **Cu** started to make a lot of sense around the year **2000**.
  - Replacing the ubiquitous SiO_2 with an $\epsilon \approx 3.7$ by a so-called "**low k**" dielectric with an ϵ of **1.5** or so would make a lot of sense right now (**2005**), unfortunately, nobody knows exactly how to do it (despite a Billion **$** or so, that have been spent on the search for a low-k material up to now).

Multiple Choice Test zu

5.1. Basic Considerations for Process Integration - Summary

Start Multiple Choice

Multiple Choice Test zu

5.2. Chips on Wafers

Start Multiple Choice

Multiple Choice Test zu

5. Process Integration - Summary

Start Multiple Choice

Standard Exercise 5.1-1

Integrated Transistors

Answer the following questions, make simple drawings and explain key features. Give approximate number whereever possible

- **a)** Draw the cross section of an *integrated bipolar transistor* (please denote, which type you have drawn) and denote the function and doping of all layers.
- **b)** Shortly describe the individual steps during the *fabrication* of a bipolar transistor.
- **c)** Draw a cross section of *two* adjacent *integrated MOS transistors* (please denote, which type you have drawn) including the metallization layer. Denote in the drawing the materials and the purpose of the decisive layers! Include typical lateral and vertical dimensions for the gate dielectric!
- **d)** Give key requirements for the *dielectrics* in a **MOSFET**!
- **e)** Draw qualitatively the source-drain-current I_{SD} vs. the gate voltage U_G(for both polarities at the gate) for the one of the transistors which you have drawn in **c)**
  If you have drawn two different transistors in **c)** then please indicate for which transistor you are drawing the characteristic!
- **f)** Suggest two measures to lower the *threshold voltage* of the **MOS** transistors (with explanation)!

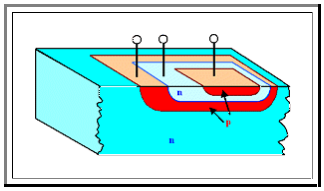
Exercise 5.1-6

Quick Questions to

5.1 Basic Considerations for Process Integration

Basic Considerations for Process Integration

- Draw a cross section of an integrated **pnp**-bipolar transistor! Denote in the drawing all doping types and the purpose of the layers!
- Compare an integrated bipolar transistor made with or without an epitaxial layer. Describe advantages and problems of either approach.
- Give a plot of the doping atom concentration as a function of depth for the kind of transistor shown. Describe qualitatively but with rough numbers as far as possible



  - Indicate the **pn** junctions in your drawing. Discuss their depth and distance in terms of process stability.
  - Describe how the doping could be administered.
  - Discuss possible problems encountered, in particular if the substrate doping is rather high.
- What function has a "buried layer" in bipolar technology?
- Draw a cross section of two integrated **p-MOSFET** transistor! Denote in the drawing the materials and the purpose of the decisive layers! Include typical lateral and vertical dimensions! Give key requirements for the dielectrics!
- Give a schematic drawing of a two-level metallization; make a list of the essential process steps and enumerate the materials used in each step.
- What exactly is a field oxide needed for?
- What is the difference between **MOS** and **CMOS**? Compare a **MOS** and **CMOS** inverter for this.
- Draw a schematic cross-section of two complementary **CMOS** transistors next to each other. Indicate the major difference to **MOS**.

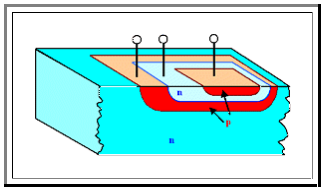
Exercise 5.4-1

All Quick Questions to

5. Integrated Circuits - Process Integration

Basic Considerations for Process Integration

- Draw a cross section of an integrated **pnp**-bipolar transistor! Denote in the drawing all doping types and the purpose of the layers!
- Compare an integrated bipolar transistor made with or without an epitaxial layer. Describe advantages and problems of either approach.
- Give a plot of the doping atom concentration as a function of depth for the kind of transistor shown. Describe qualitatively but with rough numbers as far as possible

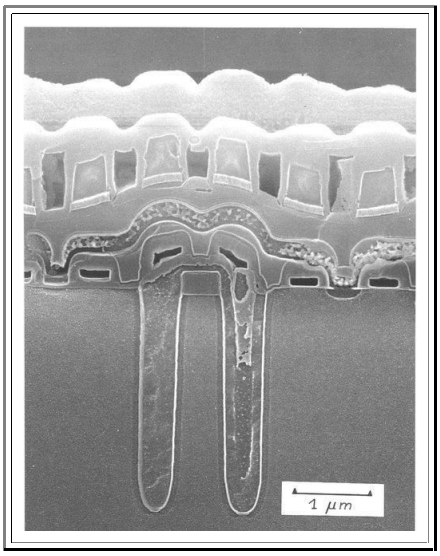


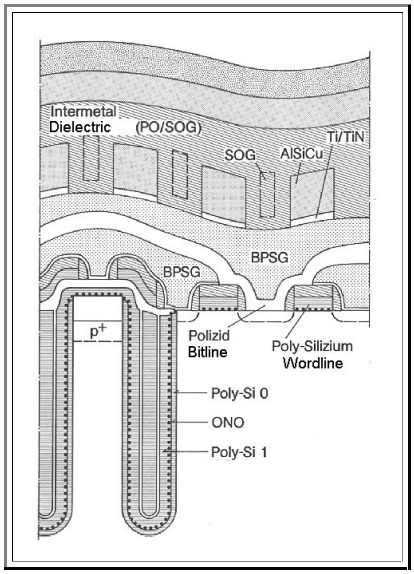
  - Indicate the **pn** junctions in your drawing. Discuss their depth and distance in terms of process stability.
  - Describe how the doping could be administered.
  - Discuss possible problems encountered, in particular if the substrate doping is rather high.
- What function has a "buried layer" in bipolar technology?
- Draw a cross section of two integrated **p-MOSFET** transistor! Denote in the drawing the materials and the purpose of the decisive layers! Include typical lateral and vertical dimensions! Give key requirements for the dielectrics!
- Give a schematic drawing of a two-level metallization; make a list of the essential process steps and enumerate the materials used in each step.
- What exactly is a field oxide needed for?
- What is the difference between **MOS** and **CMOS**? 'Compare a **MOS** and **CMOS** inverter for this.
- Draw a schematic cross-section of two complementary **MOS** transistors next to each other. Indicate the major difference to **MOS**.

Cross Section of 16 Mbit DRAM and 64 Mbit DRAM

Illustration

Below a large **SEM** micrograph showing the cross-section of an (early) **16 Mbit DRAM**.

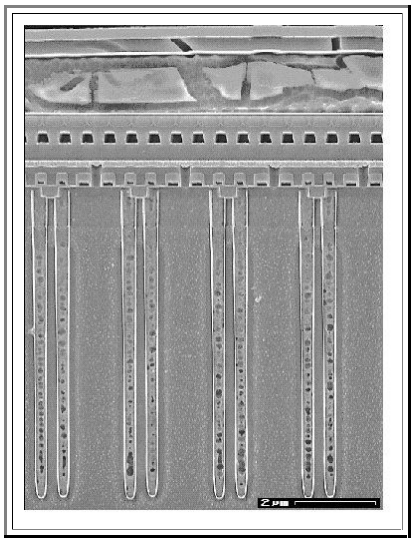




Some explanations:

- The two deep "trenches" (they are really holes) contain the capacitors. There dielectric (with ca. **7 nm** far too thin to be visible) is "**ONO**", a triple layer of Oxide - Nitride - Oxide.
- The trench os lined with poly-**Si** as a first electrode and as the second electrode.
- To the left and right two transistor gates are visible. The sources of both transistors is the (poly-**Si**) electrode lining the trench and the diffused areas being contacted by the "Polizid bitline". "Polizid" means a double layer of poly-**Si** and **$MoSi_2$** Molybdenum-silicide.
- The "poly-**Si** wordline" runs perpendicular to the picture and connects the gates of the transistors
- The "**BPSG**" layers denote **SiO_2** doped with **B** and **P** that serves as insulating dielectric. It is essentially a glass.
- Parallel to the word lines are **Ti/TiN/AlSiCu** lines. They contact the wordlines every once in a while to decrease the ohmic resistance. The consist of a layer sequence: **Ti**, **TiN**, and **Al** doped with about **0,5%** of **Si** and **Cu**.
- On top of this first metal layer is another one running across the picture.
- The metals are insulated by the intermetal dielectric composed of plasma-oxide (**PO**) that contains spin-on-glass (**SOG**) in the interstices.

Below the successor of the **16 Mbit DRAM**, the **64 Mbit DRAM** from a development stage around about **1996**.

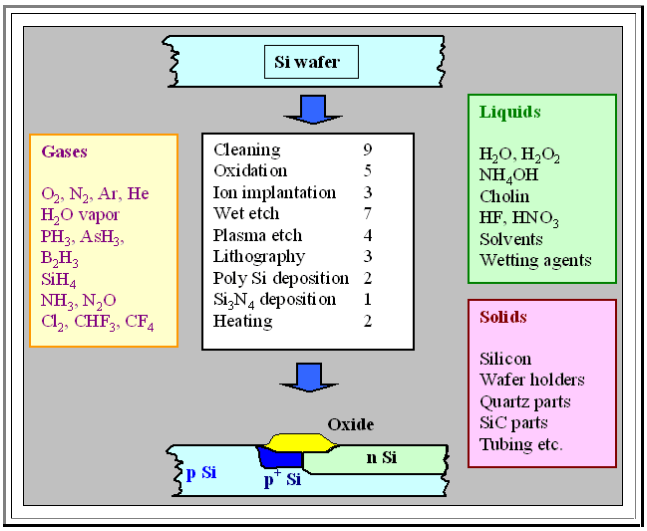


The structure is essentially the same, but all layers have been planarized.

Processes and Materials I

Illustration

Here is the list of processes and materials needed for the **16 Mbit DRAM** (an about **1999**) in a graphic way.



Note that any material that comes in contact with the wafer or with materials that will come into contact with the wafer, is an **electronic material** - utmost care has to be taken in selecting the right stuff!

- Wafer holders or tweezer, e.g., can totally ruin a wafer by leaving minute amounts (far below the detection limit) of heavy metals (most notably **Fe**) on the wafer if they are unsuitable (Never, really never, touch the wafer with a metal tweezer!).
- Gas pipes may corrode internally if made from the wrong metal and thus contaminate the gas flowing through it with traces of impurities - your factory then will only produce garbage.

Note also that some of the most dangerous inorganic chemical are used!

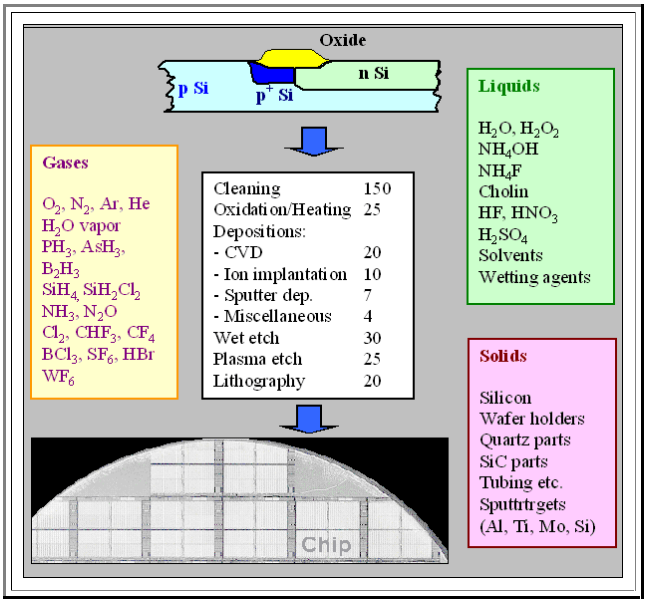
- ***HF*** (hydrofluoric acid) will cause heavy tissue and especially bone damage already by its vapors - you do not even have to touch it to get severly damaged.
- ***PH_3*** (**phosphine**) and ***AsH_3*** (**arsine**) are among the most toxic gases known to mankind; minute amounts are deadly (**PH_3**, in fact, was used as a poison gas in world war I).

To continue, use the link

Processes and Materials II

Illustration

- Here is the rest of the processes
  - A few more materials are needed, especially solids in the form of "sputter targets".
  - A great total of about **450** process- and control steps are needed.

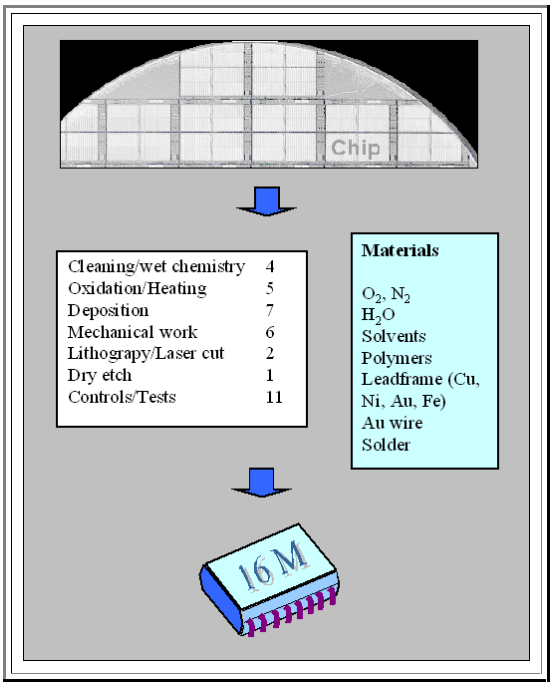


- We don't have a chip yet - we only have unpackaged chips on a wafer.
  - Next, the wafer has to be cut and the chips that work (this needs a measurement) are packaged.
  - Packaging, although not needing processes at very small dimensions, is not simple either.
  - **To continue, use the link**

Processes and Materials III

- Packaging is a completed process in its own right
  - It needs very special materials - even the lowly black plastic that dominates the appearance of chips is a sophisticated material!
  - It also includes extensive testing of the chip.

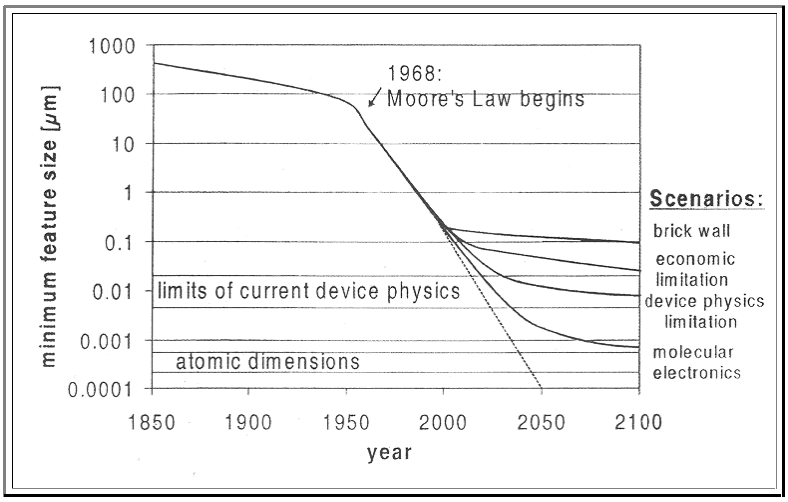
Illustration



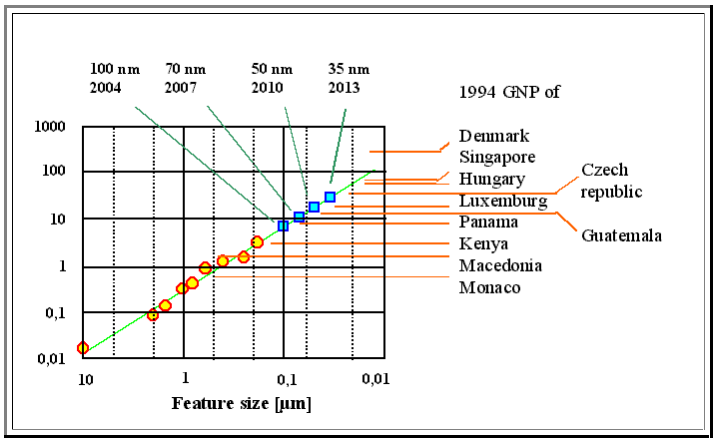
Moore's Law: Break Down Scenarios

Illustration

- Here is an illustration of **Moore's Law** with several scenarios for possible extrapolations.



- The dashed line shows the linear extrapolation. It is sure to break down - at the very last when feature size come down to the level of atoms
  - The "molecular electronics" scenario thus must be seen as the best possible case and would keep us going for another **20 - 30** years.
  - On the other extreme, the "Brick wall" scenario assumes that technology simply hits an insurmountable barrier right now. In other words, feature sizes will not come below about **0,1 µm** ever. The end then would be near.
  - In between are scenarios where feature size reduction is either limited by the costs (in other word, you can make it, but it is to expensive to sell) and device physic limitations. After all, a conventional **MOS** transistor needs at least a few doping atoms in its source and drain region and since the density of doping atoms must be considerably smaller than the density of **Si** atoms, you cannot make the transistor components arbitrarily small.
- What "economic limitations" mean is illustrated in the next picture

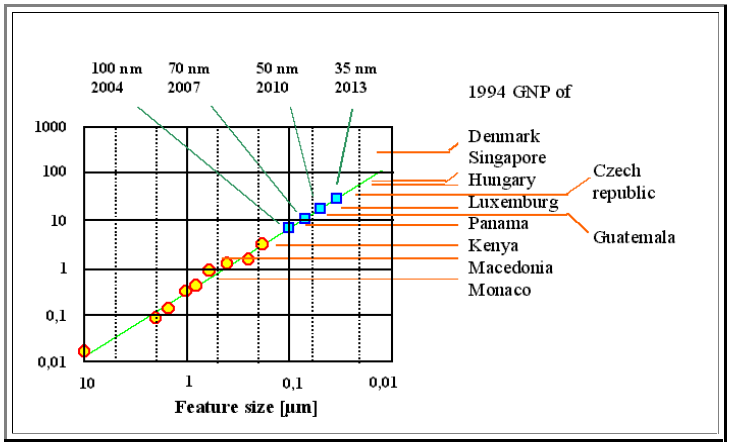


  - Here the costs of chip development is plotted (circles) and from the extrapolation of the resulting rather straight line predicted for the coming years (squares) and compared to the gross national product (***GNP***) of sizeable countries
  - The problem will be to recover these costs at roughly the same prize per chip. It means that you have to sell at lot more chips at each generation. Assuming that the capacity in memory chips increases fourfold every three years, the market for bits then has to grow much faster than about **60%** a year!

Costs of Chip Production

Illustration

This graph shows the development of the costs for developing a new chip generation (including the investments needed for building the factory) in comparison to the gross national product (GNP) of some nations.



There is an exponential increase in cost!

- While a **1 Mbit DRAM** (**1,0 µm** technology, about **1990** vintage) could be had for as little as **$ 2.000.000.-**, the **1 Gbit DRAM** (**0,1** µm technology, available in about **2003**) will run up a bill of about **$ 7.000.000.000.-**
- Compare this to the gross national product of some countries and you will see that "globalization" is more than a catch word for high-tech products!

Her the latest news that fits right in with the table above

- From **Oct. 2002**

Intel Opens $2B New Mexico Fab
Online staff -- 10/23/2002 **Semiconductor International**
Electronic News

Intel Corp. today announced the opening of a **$2 billion dollar** expansion to its manufacturing facility in Rio Rancho, N.M., adding 200,000 square feet of clean room space.

Designated Fab 11X, the plant will produce microprocessors on 300mm wafers using 0.13-micron process technology and will transition to 90nm process technology in 2003, Intel said.

"As computing and communications devices converge, the need for increasingly complex components with more capabilities will grow," said Paul Otellini, Intel's president and COO. "This facility will help us meet that growing demand. The combination of the 300mm wafers and 90-nanometer process technology will also reduce the costs of manufacturing, increase productivity and improve the availability of the world's most advanced semiconductor products."

In addition, Intel said that from an environmental perspective, water and chemical use will be more efficient at the plant, claiming that when compared to a 200mm facility, Fab 11X will produce 48 percent less volatile organic compound emissions, use 42 percent less ultra pure water and will use about 40 percent less energy.

- From **Oct. 2003**

Semiconductor International, Oct. 2003

It seems that hardly a day passes without a new announcement about the cost or economics of 300 mm wafer fabs: "$7.0 billion dollars estimated revenue to support a 300mm fab,"1 "**Leading-edge fab costs soar to $4 billion**,"2 "Prices for 300 mm wafers remain too high, says analyst,"3 "Revamping 200mm fabs will have a significant cost advantage over new 300mm facilities for some products for some time."4

- From **July 2005**

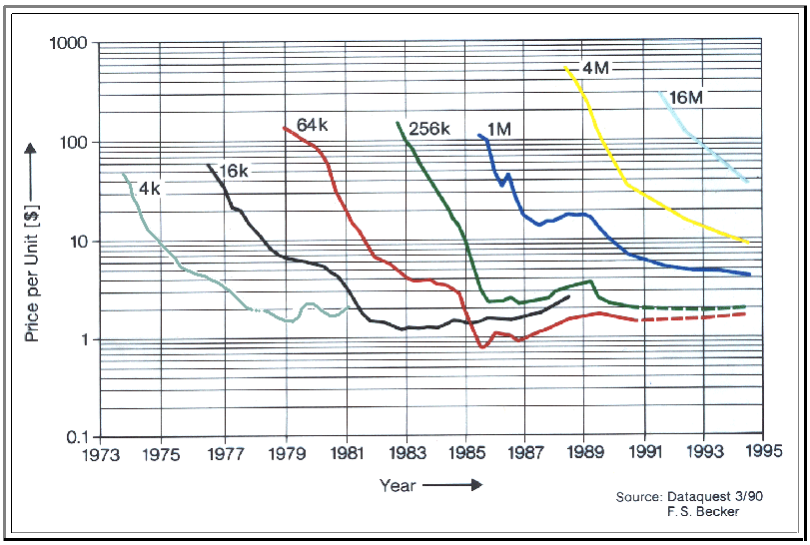
Solid State Technology, Aug. 2005

Intel Corp., Santa Clara, CA, announced on July 25 that it will build its sixth and newest 300mm/45nm fab at the site of its Chandler, AZ operation. Construction will begin immediately on the **$3.0 billion**, 1.0 million sq. ft Fab 32 facility, with production slated for 2H07. Intel operates four 300mm fabs in the US and Ireland, with a fifth expected to begin operations later this year -- a $2 billion conversion of an existing 200mm fab, also in Chandler.

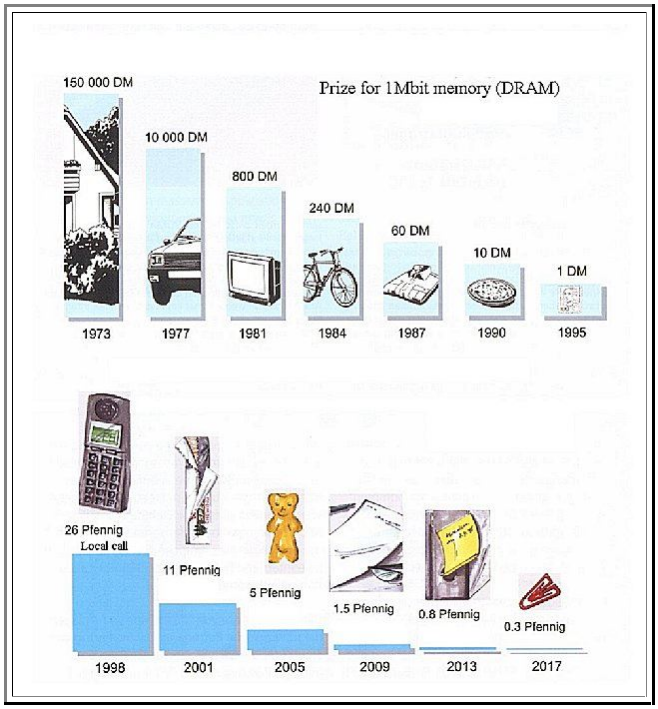
Prize Decay for Memory Chips

Illustration

- Shown are the average prizes for **DRAMs** as a function of time for each generation



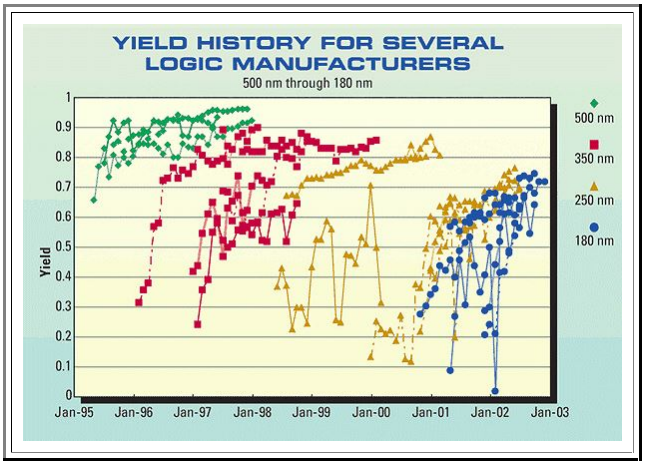
- Note that the scale is logarithmic!
  - While you may get around **$100** for chip for a short time, you will be down to a few **$** within **3 - 4** years!
  - There are influences beyond your control. In bad years the Dollar exchange rate may kill you (chips are always sold in **$$**). If you are lucky, trade restrictions (by the Americans in **87/88**) make the commodity scarce and prizes go actually up (see the **1 Mbit**; very unusual!).
- What this prize decay really means is very graphically illustrated below where the cost of **1 Mbit** of 1 Mbit memory is expressed in every day items



Yield Development

Illustration

- In the backbone part, there is a hypothetical yield over time curve, and some far-reaching conclusions are drawn form it.
  - What do *real* yield curves look like? Well, few people know, because that is one of the bigger company secrets.
  - However, in **2003**, Leachman and Berglund, working for *Sematech* did a study and release some real (anonymous) data. Here is the interesting curve:

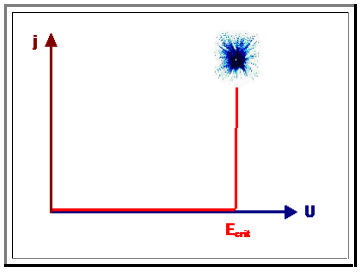


- The yield may go up somewhat faster than on the hypothetical curve, but several things are quit clear:
  - Companies start production with yields as low as **(10 - 20) %**: In other words: Most of the chips produced go in the garbage bin!
  - There is a lot frustration potential: For no recognizable reason your yield suddenly collapses substantially! (If the reason would have been recognizable, it would not have happened for more than a day or two and then not shown up on the monthly data given above!)
  - It's rather difficult to have yields in excess of **80 %** - and it is getting worse.

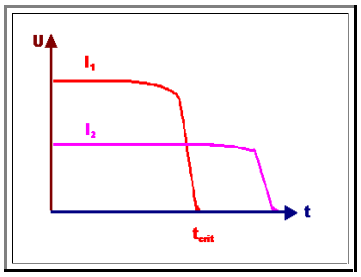
Electrical Breakdown and Failure of Dielectrics

Advanced

- As you know, the **first law of Materials science** is "Everything can be broken". Dielectrics are no exception to this rule. If you increase the voltage applied to a capacitor, eventually you will produce a big bang and a lot of smoke - the dielectric material inside the capacitor will have experienced "**electrical breakdown**" or electrical break-through, an irreversible and practically always destructive sudden flow of current.
  - The critical parameter is the field strength ***E*** in the dielectric. If it is too large, breakdown occurs. The (***DC***) current vs. field strength characteristic of a dielectric therefore may look look this:



  - After reaching E_{crit}, a sudden flow of current may, within very short times (10^{-8} **s**) completely destroys the dielectric to a smoking hot mass of undefinable structure.
  - Unfortunately, E_{crit} is *not* a well defined material property, it depends on many parameters, the most notable (besides the basic material itself) being the production process, the thickness, the temperature, the internal structure (defects and the like), the age, the environment where it is used (especially humidity) and the time it experienced field stress.
- In the cases where time plays an essential role, the expression "**failure**" is used. Here we have a dielectric being used at nominal field strength well below its breakdown field-strength for some time (usually many years) when it more or less suddenly "goes up in smoke". Obviously the breakdown field strength decreases with operating time - we observe a failure of the material.
  - In this case the breakdown may not be explosive; but a leakage current may develop which grows over time until a sudden increase leads to total failure of the dielectric.
  - The effect can be most easily tested or simulated, by impressing a constant (*small*) current in the dielectric and monitoring the voltage needed as a function of time. A typical voltage time curve may look like this:



  - A typical result is that breakdown of a "good" dielectric occurs after - very roughly - **1 C** of charge has been passed.
- The following table gives a rough idea of critical field strengths for certain dielectric materials

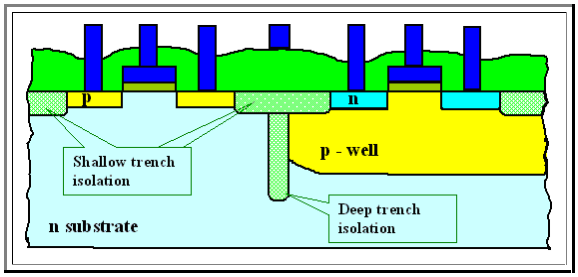
Material	**Critical Field Strength [kV/cm]**
Oil	200
Glass, ceramics	200...400
Mica	200...700
Oiled paper	1800
Polymers	50...900
SiO_2 in **ICs**	> 10 000

- The last examples serves to remind you that **field strength** is something *totally different from voltage*! Lets look at typical data from an integrated memory circuit, a so- called ***DRAM***, short for **Dynamic Random Access Memory**. It contains a capacitor as the central storage device (no charge **= 1**; charge **= 0**). This capacitor has the following typical values:
  - *Capacity* ≈ **30 fF** (femtofarad)
    Dielectric: ***ONO***, short for three layers composed of Oxide (**SiO_2**), Nitride (**Si_3N_4**) and Oxide again - together about **8 nm** thick!
    Voltage: **5 V**, and consequently
    Field strength **E = 5/8 V/nm ≈ $6 \cdot 10^6$ V/cm**.
  - This is *far above the critical field strength* for practically all *bulk* materials! We see very graphically that high field strength and voltage have nothing to do with each other. We also see for the first time that materials in the form of a *thin film* may have properties quite different from their bulk behavior - fortunately they are usually much "better".
- Last, lets just note in passing, that electrical breakdown is *not* limited to insulators proper. Devices made from "*bad* " conductors - i.e. semiconductors or ionic conductors - may contain regions completely depleted of mobile carriers - space charge regions at junctions are one example.
  - These insulating regions can only take so much field strength before they break down, and this may severely limit their usage in products

Box Isolation Technique

Advanced

- Advanced devices nowadays use a box isolation technique. However, the accepted name for that is "**shallow trench isolation**" or ***STI***.
  - If the trench is "*deep*" you may still call it *STI*, or possibly DTI, but this abbreviation may also mean "double trench isolation" - just get used to the fact that the semiconductor industry moves to fast to have time for standardizing those things and to force people to stick to it.
- How is it done? And why do we need it? Let's just look at this issue very briefly.
- First, we note that all the problems with **LOCOS** already mentioned in the backbone get rapidly worse as dimensions get smaller. On top of that, new problems develop, and the process flow becomes increasingly complicated (and expensive).
  - Many problems would disappear if you just would etch a suitable "hole", i.e. a trench in your substrate wherever you need isolation, and then fill it with oxide.
  - Or **CMOS** structure from before then would look like this:

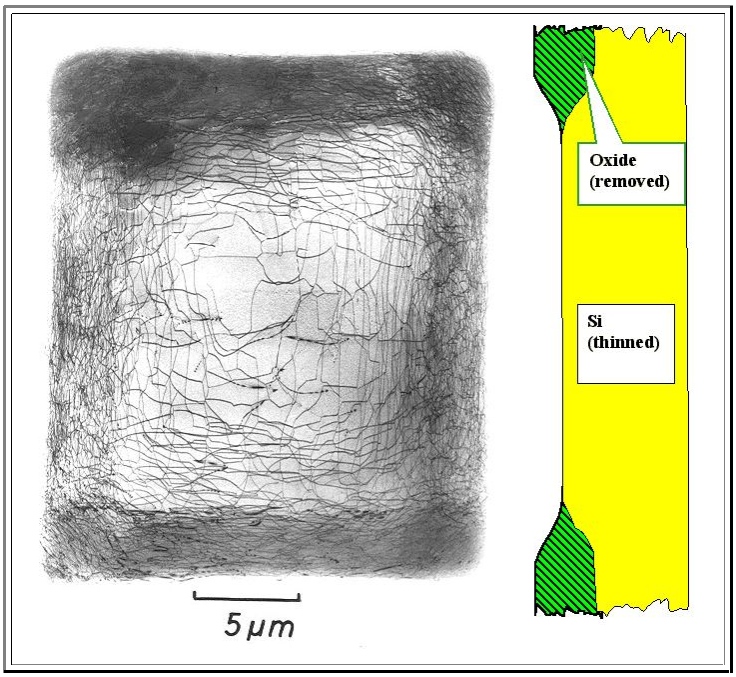


  - Of course, nobody would superimpose a shallow and a deep trench isolation as shown (you just would have the deep one and move the transistors closer together), but the picture illustrates the point nicely.
  - So why wasn't it done long since?
- Because it is neither easy to etch the required trenches, to fill them with (high-quality) oxide, and to planarize the surface.
  - The latter point is the key: Whatever process you use to *fill* the trenches with oxide, after your oxide deposition you have oxide everywhere, and you must take it off again wherever you don't want it, i.e. outside the trenches
  - If you wonder why you *fill* the trenches with oxide, and why you don't use simple *thermal* oxidation anymore, you missed some the essentials! Figure it out yourself; it is enough to consider what would happen if you start thermally oxidizing your **Si** after you etched the deep trench.

TEM of Oxide Edge Dislocations

Advanced

- Here is another picture of oxide edge dislocations, this time showing the complete area of the field oxide

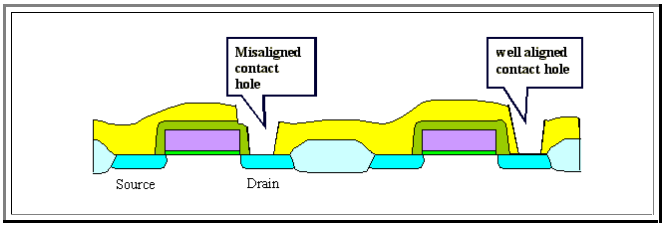


  - The graphics on the right indicate the cross section through the specimen (the oxide was removed during preparation). The areas of **Si** formerly covered by oxide are now so thin that they are totally transparent to the electron beam - the dislocated areas seem to float in empty space.
- You may wonder how pictures like this are taken. The link leads to a chapter from the "Defects in Crystals" Hyperscript explaining how transmission electron microscopy (**TEM**) can be used to image defects in crystals.
- However it works, this picture is special. In any **TEM** image, the electron beam is passed through the specimen, and pictures with good resolution can only be taken if practically all electrons emerge from the specimen backside with the same energy they had before entering the specimen
  - We know that electron beams do not travel very far in solids (the electron beam inside your monitor does not only not get out, after all, but deposits its energy into a rather thin layer of luminescent material); if they are supposed not to even slow down, the specimen must be very thin - say **0,5 µm** at the most for a conventional **TEM** with a beam energy of **(100 - 150) kV**.
  - This would mean that the specimen must be thinner than the structure it is supposed to contain - only the surface near parts of it would show.
  - This problem was overcome around **1976** by using the high-voltage **TEM** (**HVTEM**) of the Max-Planck-Institute for Metal Physics in Stuttgart - a monster that commanded a beam energy of **650 kV** which was enough to look through samples of **(2 - 4) µm** thickness. The picture shown above (and some of the other ones in the Hyperscript) were taken with this machine and therefore show the full richness of the defect structures.
- Very few **HVTEMs** are in use - they are too expensive (the newest on in Stuttgart costs about **$1{,}5 \cdot 10^7$ DM**), so pictures like the one above are rare.

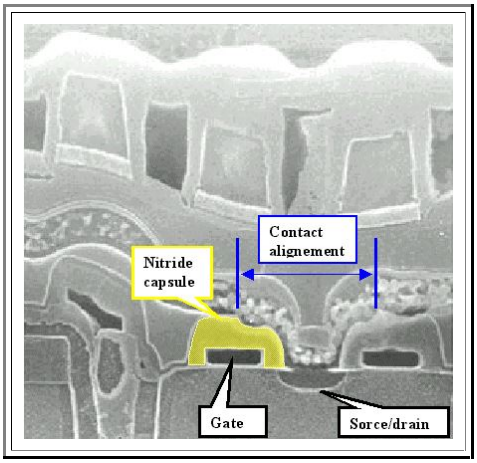
FOBIC Process

Advanced

- Consider making a contact hole through the **CVD** oxide that was deposited over the finished transistors.
  - Since the source/drain area are as small as possible, you have to make not only a small hole *but you must align it precisely* relative to the gate stack of the transistors.
  - Small misalignements would produce a short circuit between the gate electrode and the **Al** that will be put in the contact hole eventually.
- Now cover the whole gate stack with **Si_3N_4** that will resist whatever etching procedure you use to remove the oxide in the future contact hole. You gain a lot in your "**process window**" for the contact hole - misalignements don't matter as much any more as shown below.



- That looks pretty good - but there is a prize to pay:
  - First of all - how do we make the nitride encapsulation? And don't forget; **Si_3N_4** must never come in contact with **Si** - you always need a thin layer of buffer oxide underneath (not shown in the picture). There is quite a bit of added process complexity!
  - The topography gets worse. The aspect ratio of the contact hole - the relation between depth and diameter - increases and with it the problems of filling it with **Al**.
  - Still, starting around **1987**, **FOBIC** was used and helped to get the next generation onto the market.
- The picture below is a cross section through a **16 Mbit DRAM** memory cell shown before. It shows one of the contacts to a transistor (which is connected on the other side to a trench capacitor). The **FOBIC** structure has been outlined; it is clearly visible.

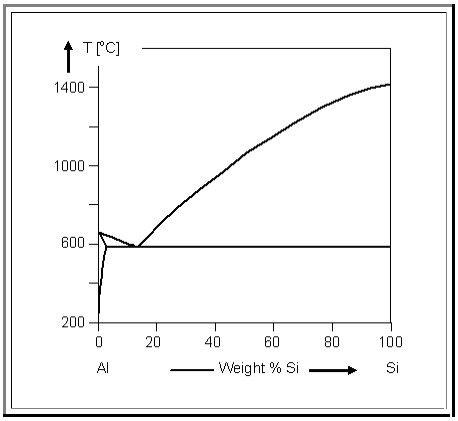


Spiking and Epitaxial Si in Contact Holes

Advanced

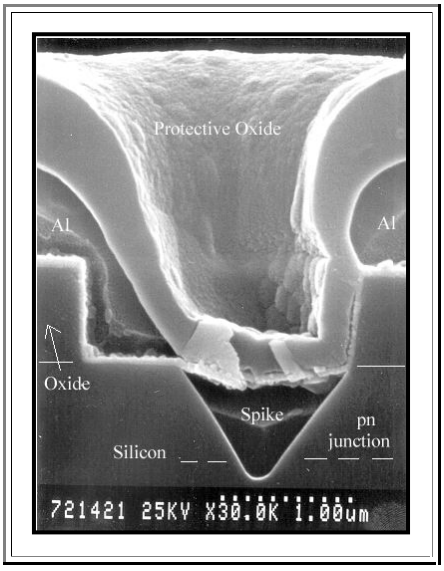
Spiking

- When we make a contact to the structures in the Si, for example to source and drain regions, we first cover everything with an insulator - **SiO_2** - and then make contact holes in the proper places. Use the link to refresh your memory.
  - Next we cover everything with a metal - if we stay simple we just use **Al**. We have a lot of other processes after that, and we will have to heat up the wafer to some extent for doing whatever needs to be done.
- What happens if we bring **Si** in contact with **Al** (or any other metal, for that matter) and heat it up to some extent?



- Well - consult the phase diagram, and it will tell you what you should expect at whatever temperature you chose. Here is the **Al - Si** phase diagram.
- What it tells us it that somewhat below **600 °C** we will have an eutectic.
- In other words, above the eutectic temperature, "things" will melt. Consult the chapter about phase diagrams if you don't know exactly what "things" means.
- In yet other words, after **Al** deposition, we must not raise the temperature above the eutectic temperature ever In order to stay on the safe side, we must even keep it below about **500 °C**
- This is not so good, but something else is worse: The solubility of **Si** in **Al** around **500 °C** is finite - about **2 %** one would estimate - , while at room temperature it is practically nil.

- So what happens if you heat up **Al** in contact to **Si**? The **Al** will try to incorporate some **Si**; it will sort of "suck it up" from the **Si** substrate.

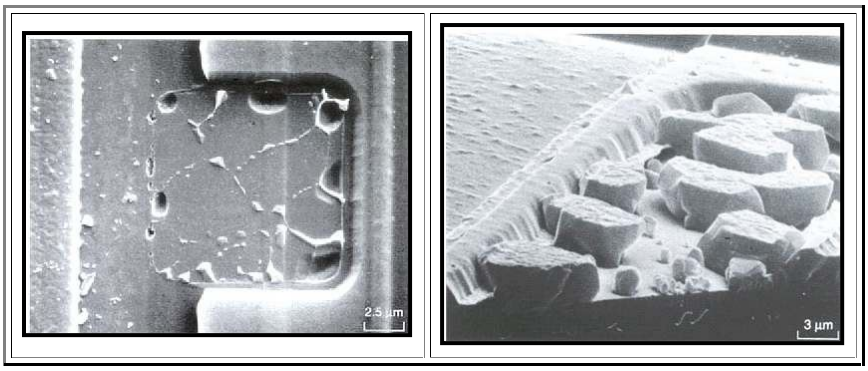


- In more scientific terms we talk about **Si** atoms diffusing into the **Al** (and **Al** atoms diffusing into the **Si**); so temperature dependent quantities like diffusion coefficients are involved.
- If we now have some of the **Si** in the **Al**, it must be missing somewhere else; obviously right below the **Al-Si** interface we must expect some "missing" **Si**.
- If we are lucky, the "missing" **Si** is uniformly distributed, i.e. the whole surface of the **Si** moved down a bit. However, Murphy's law (What can go wrong will go wrong) applies, and on occasion all the **Si** moving into the **Al** comes from one rather localized spot at the interface - we get a "**spike**". This is shown very drastically in the picture.
- At the same time, **Al** will diffuse in the room left by the **Si** - our spike is filled with **Al** and we have a short-circuited **pn**-junction!

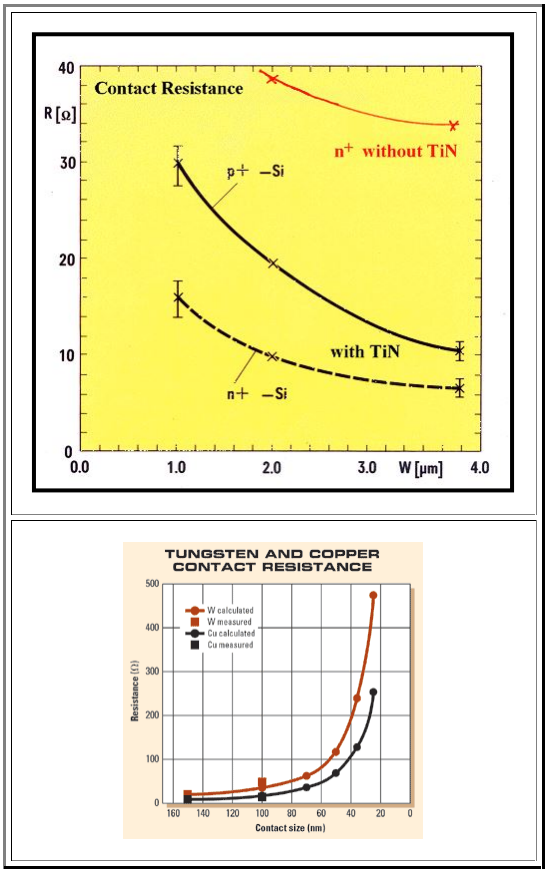
- This ain't so good. Remember: One spike / short circuit in just one of the > **50** million or so contact holes will kill the whole **IC**. We must fix that problem.
  - Easy, you say. (Do you see the obvious solution?). We do not use pure **Al**, but **Al** already containing some **Si**, so it does not have to "steal" **Si** from the substrate to meet its solubility needs at higher temperatures.
  - Right. That is exactly what we will do. We use **Al** alloyed with **0.5%** - **1 %** of **Si**. No more spikes will form, and as the process engineer in charge you can sleep well again at night.
- Really? Yes, you will. For a few years at least. But then you slumber will become unruly again, because as dimensions shrink, you run into new problem.

Si Precipitation in the Contact Hole

- You use **Al(0.7 % Si)** as your contact material. The dictate of the phase diagram with regard to the Aluminium needs of **Si** at your highest temperature are met; no spiking will occur.
  - How about the needs of the **Al** at room temperature? They are not met, because **Si** solubility at room temperature is negligible.
  - The phase diagram now dictates that **Si** precipitates should form. This will need some nucleation and may be kinetically difficult in the bulk of the **Al**, but at the **Al-Si** interface we already have **Si** and nucleation is easy.
- In essence, we must expect the reverse of the "sucking **Si** into **Al**" process to take place. The Al will "spit out" its surplus **Si** and deposit it right at the interface
  - **Si** precipitates will grow *epitaxially* (there is no reason why not) on the **Si** below the **Al**. We must expect to find some **Si** islands on top of the **Si**



- That is exactly what we see in the pictures above.
  - In the left hand picture we have a nice balance of a few shallow spikes and some **Si** precipitates following the contours of the grain boundaries in the **Al** (which has been etched off). This is not surprising, because precipitation is a diffusion process, after all, and diffusion along grain boundaries is faster then in the bulk.
  - The right hand picture shows rather large precipitates, almost taken up all the space there is in the contact hole.
- The question, of course, is: Does it matter? We still have some **Al** in contact to our **Si** substrate, and than we have **Al** in contact to the **Si** precipitates and the **Si** precipitates in contact to the **Si** substrate. Plenty of possibilities to run a current through the contact.
  - Right, but **Al** in **Si** is an acceptor, and we must expect the **Si** precipitates to be saturated with **Al** and thus heavily **p**-doped.

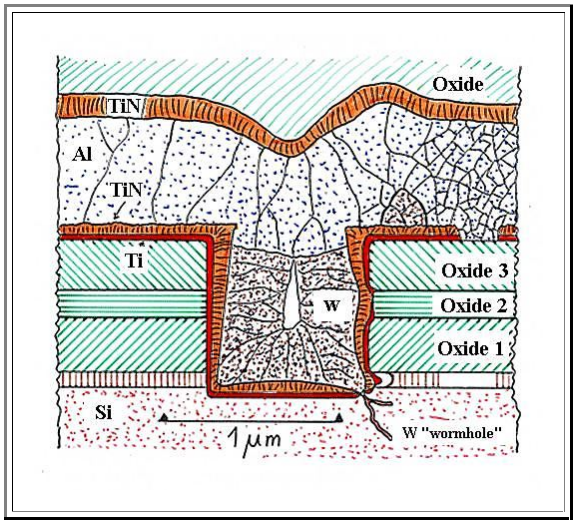


- No big problem as long as the **Si** below the contact is **p**-doped, too, but a big problem for **n**-doped **Si** and small contacts.
- In this case part of the contact area is now a **pn**-junction, blocking current flow in one direction.
- It is clear that the **contact resistance** always increases with decreasing contact area, but if we have **p**-doped **Si** precipitates, we have only part of the geometrical contact area for the contact to **n**-type **Si**.
- What we get as a function of the contact hole size for the **n**-doped case is shown in the figure in red.
- The message is clear: Contact resistance will be too large at some point. So what do you do?

- Without the **Si**, you get spikes, with the **Si** you get precipitates and a contact resistance that is too large.

- There is no choice anymore: You need a new material; in this case a diffusion barrier between the **Si** and the **Al**.

Introduce, e.g., a thin layer of **Ti/TiN** between the **Si** and the **Al**, and your contact resistance problems are solved, as seen in the figure. If you also throw in contact hole filling with **W-CVD**, you now have a structure like the one below, that also shows a few things that can go wrong on the right hand side of the contact hole:



- Of course, now you have to worry about how to make and structure this layer. How to measure how thick it is, and if it has the required properties (assuming that you know that).
- You must worry if the new diffusion barrier interferes with the reliability of the metallization (after all, electromigration in the metal is one of the major causes of premature device failures), and if can keep the additional processes cheap.

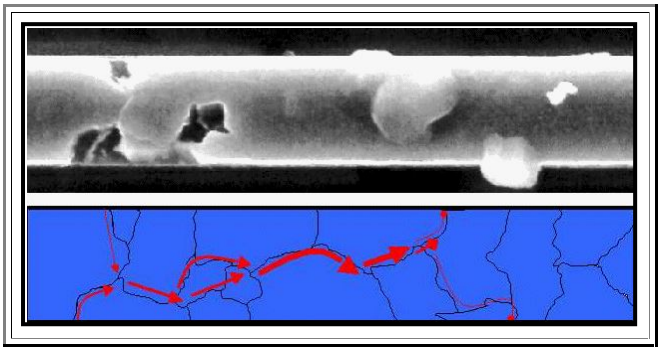
In other words. Introducing a new material is a long and cumbersome process.

- And a frustrating one. Nobody tells you, that **TiN** is it! You (meaning you and your team) find that out yourself. And in that process you will test many materials, most of which will not be the right ones.
- Out they go - and with them many hours of your time and a lot of (other peoples) money.

Electromigration

Advanced

- The picture below shows an **Al** circuit line of an **IC**; it is about **1 µm** across.
  - It has been "*stressed*" for some time by running a **DC** current through it at temperatures and current densities somewhat higher than what it would encounter in normal operation. The minus pole was on the left. The current density may have been in the**10^5 A/cm^2** region.
  - This is a hell of a current density - several orders of magnitude larger than what you would feel comfortable with in normal wires.
- The conductor line obviously has suffered from this treatment. In its left part - close to the minus pole - holes or pits have formed, and in its right part - close to the plus pole - we have extrusions or hillocks.
  - Obviously, substantial amounts of **Al** have migrated to the right: leaving back holes and generation hillocks. and
  - Considering that a huge number of electrons was driven from left to tight by the current source, it almost looks as if those electrons have pushed some **Al**-atoms along.
- This is indeed what has happened. The phenomena of **electromigration** is, in general terms, a drift in the diffusion currents induced by the "*electron wind*"
  - Remember that there is always some self-diffusion, but without any driving forces, the net diffusion current in one direction is always cancelled by the exact opposite current in the other direction.
  - Exactly how those electrons push atoms about, is still a bit obscure, but the effect is clearly there, and appreciable in finite time spans if the current densities are extremely large.



- You thus would expect that electromigration scales with the self-diffusion coefficient, and thus is more severe in materials with a low melting point.
  - Yes, but classical self-diffusion is a bulk property. In our poly-crystalline **Al**, we also have diffusion along grain boundaries, which is generally much faster than bulk diffusion.
  - A more detailed analysis of the picture above (which, incidentally, are from the thesis work of Dipl.-Ing. Wedemeyer in **1998**) that includes the grain boundaries clearly demonstrates that grain boundaries indeed provide the main pathway for the movement of the Al atoms.
- At the holes, the cross-section of the **Al** wire is reduced, i.e. its local resistance ***R*** is larger, and the locally dissipated power $L = I^2 \cdot R$at a constant current *I* goes up.
  - The wire heats up at the holes, electromigration becomes more severe, holes get bigger, and soon you have a complete interruption of the current path and thus the end of your **IC**.
- We have a problem. As dimensions shrink by a factor of κ, the cross-sectional area of the **Al** lines tends to shrink by κ^2, but the current only goes down linearly with κ.
  - With shrinking dimensions current densities thus tend to go up, and electromigration in pure **Al** (or **Al** with some **Si** to prevent spiking) would have made integrated circuits with a life time of **10** years simply impossible in the **80**ties of the last century.
  - You must do something. Avoiding grain boundaries roughly parallel to the wire helps, but how are going to do this? Well, there are some tricks which help to create the "bamboo structure" you want, but this would not be sufficient.

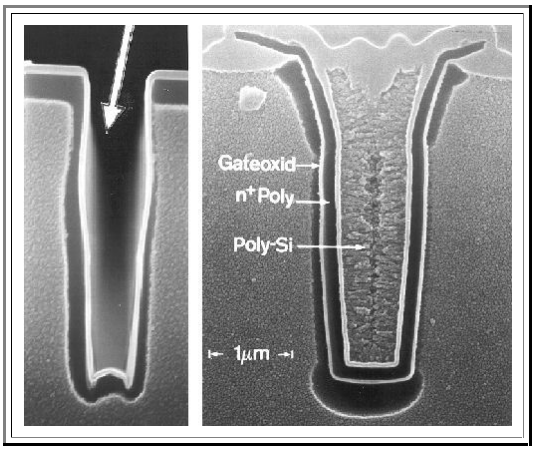
  So you do what you *always* do in this kind of situation: You add some dirt to your **Al** and hope for the best.
- Maybe, the right element "poisons" the grain boundaries, making grain boundary diffusion for **Al** atoms more difficult. Or maybe.....
  - Having a "theory" or just a hunch to work from helps; but not too much. Doing is what really helps. And yes, if you add about **0.5 %** of **Cu**, electromigration becomes less severe. This is also true for other elements, but **Cu** is what is mostly used.
  - However, if only a tiny fraction of that **Cu** ever diffuse into the **Si**, it will kill your device. Also, your etching process for **Al** may suddenly not work anymore. And your **Al** metallization now might be more sensitive to corrosion. And the adhesion of the next layer is now different . And....

- You get the drift: There will always be a lot of demanding work out there for highly qualified people like you!

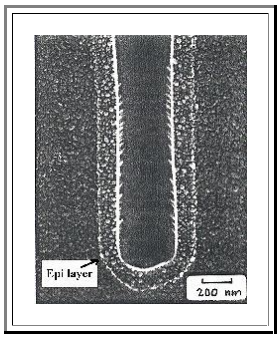
Doping Trench Walls

Advanced

- You have a trench (i.e. in reality a hole) with a diameter of about **1 µm** and a depth of **5 µm**. The **Si** around it serves as one of the capacitor electrodes and must be doped with something to a depth of about **0,3 µm** so it is highly conductive.
  - This task came up as an unprecedented challenge to process engineers in the context of the **4 Mbit DRAM** generation around **1985**. How do you do it?
- First you try what you have - ion implantation - and hope for the best. Since the ion beam does not impinge at right angels on the specimen, but somewhat off (say about **10^{o}**) for various reasons, there is hope.
  - Indeed, it works - see below:



- With a special etchant the doped region is etched off, leaving a dark rim around the trench (which was not very perfectly etched at this point in time).
  - The arrow in the left-hand picture marks the major direction of the ion beam, and the left-hand side of the trench is nicely doped, as one would expect. The right hand picture shows a fully processed trench capacitor for measurements of the electrical characteristics.
  - But some implantation also took place on the unfavorable side - there is a certain spread in the beam angle, some ions are reflected or scattered off their original direction - whatever happens, some ions hit the wall. If you turn the wafer during implantation, it might be possible after all, to obtain decent homogeneous doping.
  - But in the end, ion implantation was not used. Too expensive, too dependent on the precise trench geometry (which changes all the time), simply not good enough for prolonged mass production.
- OK, so you try epitaxy. Simply deposit the doped layer into the trench as you want to have it.
  - Allright - it works. Some problems are encounteres around the opening of the trench, but nothing that couldn't be solved eventually. The picture shows part of a trench with the doped epi layer; this looks good.



  - But we don't use it for production! The reason is simple: Too expensive, too hot - and we can do it better in a different way.
- How? Well, let's deposit a "**spin-on glass**" (***SOG***)containing the dopant, hope that it will completely fill the trenches, and then diffuse the dopant out of the **SOG** into the **Si**.
  - Spin-on glasses are used a lot, but do they fill up fine holes? The only way to find out is trying it.
  - To cut a long story short - it worked. But not reliably enough; didn't make it to production.
- What finally made it to production is a somewhat similar approach, except that the **SOG** was replaced by an **A**s doped **TEOS**-oxide.
  - Not an extremely nice process (the coating of the **CVD** tube is **As** contaminated and highly toxic), but without many unknowns and reproducibly.
- Is there something even better?

- Yes, there is a simple solution to the trench doping problem: *Don't do it*. Oxidize the trench walls and deposit a layer of highly doped poly-**Si** as the capacitor electrode. Simple. Elegant.
- Except that now you have to produce the very thin capacitor dielectric on poly-**Si**. Thermal oxidation of poly-**Si** will not be good enough - now you need **ONO**. The result was already shown in an illustration module

Dry Etching - Some Special Issues

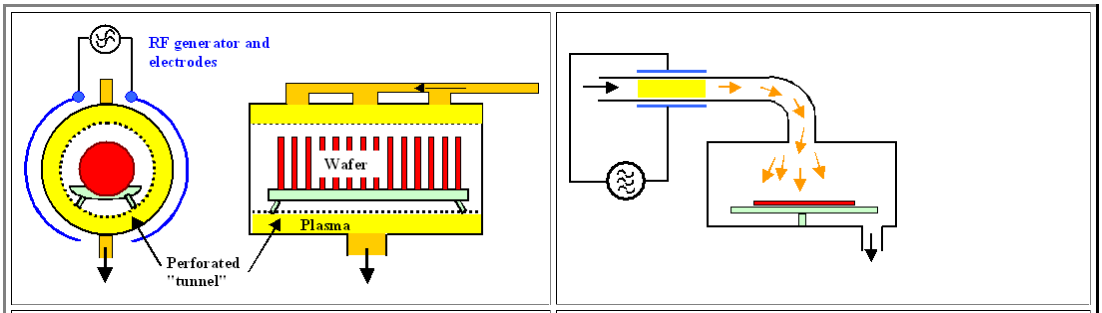
Advanced

Basic Dry Etching Modes

- There are four basic modes of dry etching
  - **1.** *Etching in a gas* (**Chemical dry etching**, ***CDE***). Rarely used, but possible. As an example, with (**HF + N_2 + H_2O**) vapor, some **CVD** oxides can be etched about **10** times faster than thermal oxide - a selectivity not achievable with wet etching. Like all chemical etches it is isotropic.
  - **2.** *Chemical etching in a plasma*. Whatever is produced in the plasma then must *diffuse* to the wafer, i.e. the physical component of large kinetic energies is absent. Like all chemical etches it is isotropic. This is a process widely used for hard-to-wet-etch layers, in particular photo resist (which is simply "burned off" in an **O** -plasma). This was the first dry etch process to hit production.
  - **3.** *Ion beam etching* (called ***RIBE*** for "**reactive ion beam etching**"). The etching process with high energy ions extracted form some plasma source is purely physical/mechanical. **RIBE** has all kinds of problems and is rarely used in its undiluted form.
  - **4.** *Chemical-physical etching* called ***RIE*** for "**reactive ion etching**". Here we mix everything from above and this is were it becomes complicated but powerful. This is what we use for all critical etch processes

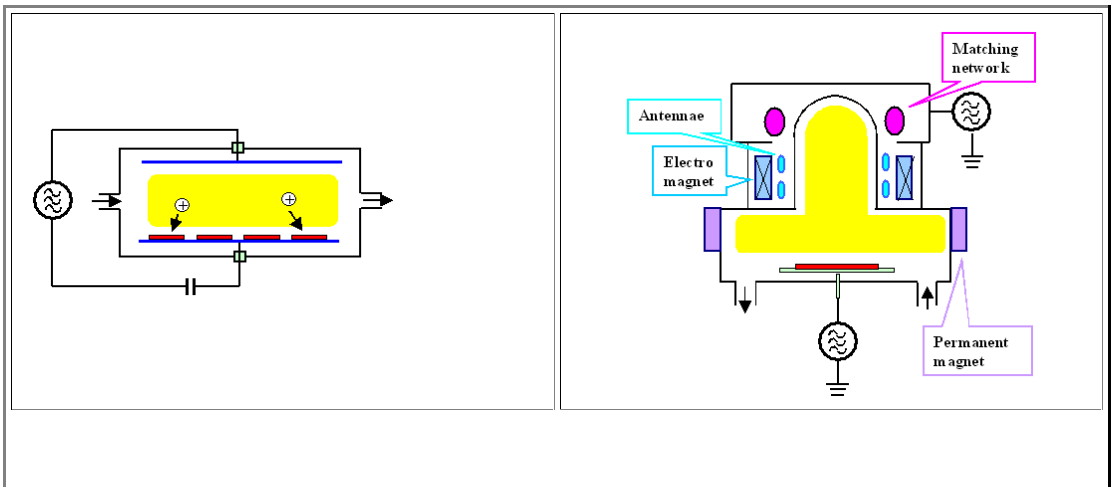
Basic Reactor Types

- Here are a few basic reactor types. You always go for batch processing, i.e. for etching several wafers (ideally **25 - 100**) at once, if it is possible (which ever more often it is not). We skip the reactors for **CDE** and **RIBE** because these etch modes are not so important.
  - *Chemical etching in a plasma*. There are two basic types of reactors: The **barrel reactor** and the "*down stream*" reactor. In both types the plasma does not extend to the wafer and the etching species must diffuse to the wafer.



Basic set-up of a barrel reactor - the work horse for resist removal.	Basic set up of down stream reactor

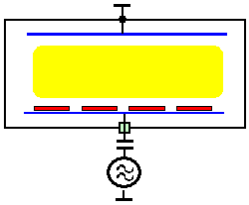
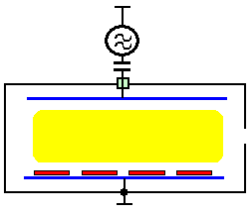
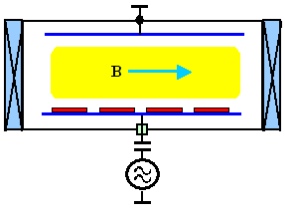
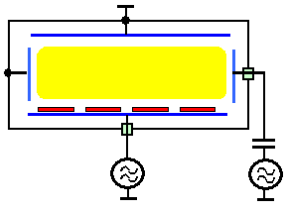
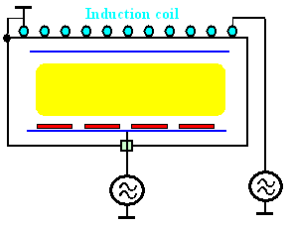
- The plasma in both cases might be produced by high frequency excitation, or even with microwave induced discharge. The pressure is relatively high, around **100 Pa**.
  - Besides photo resist etchimng in an **O**-plasma, **Si**, **SiO_2** and **Si_3N_4** might also be etched (isotropically) if some **F** - bearing gas is used.
- For reactive ion etching, some kind of **parallel plate reactor** or some kind of plasma-source reactor is used. Designs might look like this:



A basic parallel plate reactor - quite similar to the basic set-up for sputter deposition	A rather complicated plasma reactor, just to give an idea of where plasma etching is going

.

Lets just look at etching in parallel plate reactors for some relevant examples.

Mode	Basic configuration	Advantages	Disadvantages
RIE **Reactive ion etching**		* Large anisotropy * Good structure transfer	* Low selectivity * Low etch rate * Surface damage
Anodically coupled plasma etching		* High selectivity * High etch rate * Low surface damage	* Tendency to underetch mask (bad anisotropy)
MRIE **Magnetically enhanced RIE**		* Large anisotropy * High etch rate * Reduced surface damage * Etching relatively independent of loading	* Homogeneity is a problem
TRIE **Triode reactive ion etching**		* Large anisotropy * Increases process optimization potential with generators	* Increased costs and complexity
TCP **Transmission coupled plasma etching or Inductively coupled plasma etching**		* Etching relatively independent of loading * Very large etch rates because of high plasma density	* Constructive problems (no ferromagnetic steels can be used)

Basic Etching in One Reactor Type

- Well, you get the idea. And if you do not understand exactly how this works (consider the variations in what is grounded and if their is a capacitor), don't worry: It *is* quite involved. Lets just look at the first two examples:
- Here is ***RIE***, the first example given in the table.
  - In the case of pure **RIE**, the high frequency voltage is *coupled capacitively* to the lower electrode that carries the wafers. The upper electrode is grounded, and together with the chamber it has a much bigger surface then the lower electrode.
  - This will lead to a *negative charging* of the lower electrode (the reasons for that are tricky and due to the fact that no **DC** currents can flow, and that the mobility of the negative electrons and positively charged ions is very different).
  - The positive ions thus are strongly accelerated before the hit the substrates to be etched; typically they have energies > 100 eV.
  - If we keep the pressure low enough, there is hardly any scattering of the ions, and the impinge nearly vertically on the substrates.
- Now lets look at the second case, *anodically coupled plasma etching*.
  - It looks rather similar to the first case, we just changed the connection of the **HF** generator and the ground.
  - This simply reduces the negative charge on the lower electrode, and thus the energy of the ions hitting the substrates.
  - In addition, we will keep the pressure rather high, making sure that the ions accelerated in the direction of the substrate have many collision with gas molecules and thus bombard the substrate from all directions. This also reduces their energy, and etching is less "physical" and more "chemical". We gain in selectivity and loose anisotropy.
- But now we will stop. While you should marvel at the complexity of plasma etching, you should not despair. It was people like you - and not the Noble prize winners in physics - who made it work and continue to improve the technique.

Some Issues from Advanced Lithography

Advanced

General

- A somewhat better equation than the one in the backbone for the resolution limit or minimal feature size d_{min} of an optical system is

$$d_{min} \approx \frac{k \cdot \lambda}{NA}$$

  - λ is the wave length. The parameter ***k*** lumps together the effects of, e.g., photoresist response, or reticle properties.
  - The numerical aperture ***NA*** can be defined as $NA = n \cdot \sin\Phi$ with ***n*** = refractive index of the medium above the photo resist, and Φ = largest angle of converging rays hitting the resist at a "point".
- If we want the ultimate in resolution, we have to work at all three parameters

Wavelength

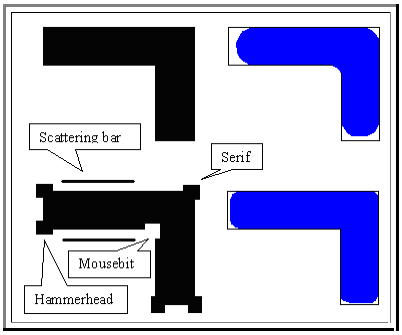
- The visible range of wavelengths extends from about **780 nm** (red) to **380 nm** (violet). Obviously we need to go to even smaller wavelengths in the ultraviolet part of the spectrum if we want to make structures in the **100 nm** region. Obvious, so where is the problem? Well, there are two major problems with this approach.
- *First*, we need a *powerful* and fairly *monochromatic* illumination source, and *second* we need materials to make an extremely good lens from.
- Let's look at the illumination source issue first:
  - A *powerful* light source we need because we cannot afford to wait forever before an exposure is finished. The maximum exposure time should be below a second or so, and you simply need intense light for that.
  - *Monochromatic* light we need, because we cannot possible built a supreme lens for many wavelengths (there are things like chromatic aberration and so on). Taking a small part of the spectrum out of some blackbody radiation (the spectrum emitted by something hot like a light bulb), however, leaves very little intensity.
- The solution lies in going for an intense line in the emission spectrum of some element - mercury (**Hg**) in this case.
  - In the **80**ties, the so-called **G**-line at **436 nm** was used (coming from a high-pressure **Hg** discharge lamp). Next came the **I**-line at **365 nm**, and then a **250 nm** line.
  - But that was already pushing the **Hg** lamp to its limits, and it was soon replaced by so-called **DUV** (for deep ultraviolet) **excimer lasers**.
  - Excimer lasers are based on rather strange materials: Compounds of noble gases like **KrF**, or **ArF**. Rather unstable stuff, but emitting at **248 nm** (**KrF**) or **193 nm** (**ArF**). With the **KrF** system, dimensions down to **130 nm** have been realized, but this is already pushing it quite a bit.
  - The **ArF** excimer laser has been used from about **2003**, so it is still in its infancy. It is expected to cover the "**65 nm** node", and possibly also the **45 nm** node.
  - That will be the end. After that, the age of "**EUV**" (extreme ultraviolet) might start, at a wavelength around **12 nm** (its really rather soft **X**-rays). There is no way of having a lens anymore, "optical processing" must then be done with mirrors.
- If we now look at the *lens* issue, we first should realize that high-aperture lenses are generally difficult to make. But the overwhelming issue is to find suitable materials that have a sufficiently large index of refraction at the wavelength considered.
  - We not yetr dealt with this issue - the frequency dependence of the dielectric "constant" ϵ_r or of the index of refraction $n = (\epsilon_r)^{1/2}$, but you can check the following links to get a first impression
    - General remarks to the frequency dependence of ϵ and ***n***.
    - Dielectrics and optics.
- Illumination source and lens materials are not the only problems encountered by switching to a smaller wavelength. Of course, there are many others, too.
  - To mention just one: The "pellicles", the thin foils protecting the mask, will turn dark in intense **UV** illumination. Not good, so let's take a better material. Easy fix, but do you know a better material? No? Too bad - since nobody else does either, you missed your change of getting rich quickly.

Numerical Aperture

- In air, ***NA*** obviously than has a maximum value of **1**. The best lenses built so far have a ***NA*** of about **0.8**; but **0.9** is already aimed for
  - Keep in mind that what you gain in resolution by increasing ***NA***, you loose in the depth of focus. Large ***NA*** lenses thus only make sense in the context of rather perfect **planarization**.
- Nevertheless, increasing ***NA*** even more helps, and there is - in principle - a simple way of doing it: Replace the air between your lens and the wafer with something that has an appreciable index of refraction, e.g. oil.
  - "Oil immersion objective lenses" have been used for about a century in conventional optical microscopes; in this way the numerical aperture and thus resolution can be increased in a rather simple way by up to **40%**.
  - But this is far easier said than done. Just consider that the name "stepper" comes from the fact, that you *step* the wafer (rather rapidly) below the lens. How do you keep you oil in place? And how will the wafer respond to be covered with oil?
  - Well, let's not use oil, let's use high-purity water (***n* = 1.437** at **193 nm**), but that only solves some of many problems and creates some new one (your **CaF_2** lens, for example, will dissolve in water).
  - Nevertheless, "**liquid immersion lithography**" will most likely be the next big fashion in lithography, with the potential to keep microelectronics alive well into the next decade (i.e. after **2010**).

Recticles and Resists

- What is left is to make the parameter ***k*** as small as possible, i.e. to pay some attention to *reticles* and *resist*, or, more general, to resolution enhancing techniques.
  - There is quite a potential here, "historically" parameter ***k*** has decreased steadily form about **0.8** in the **1980**s to **0.4** today.
  - While optimizing the resist is critical, it does not introduce new principles, and we will not cover it here.
  - That leaves the reticle and the way it is illuminated. There is quite a bit that can be done, but you must pay the prize of sharp increases in complexity.



  - The proper catchwords giving some idea to what is meant are:
    - Off-axis illumination
    - Optical proximity correction (***OPC***)
    - Phase shift masks (***PSM***)
  - For the latter two cases the general idea is to have a structure on the reticle that is different from what you want to have projected into the resist on the wafer. If, for example, a sharp corner is "smeared out" to a roundish image, than make the corner look different. The figure gives a rough idea what that means
  - In phase shift masks you add structures that do not only manipulate the amplitude of the light transmitted through the mask, but also the phase.
  - In this way you can produce constructive or destructive interference in the image plane in places where that is helpful to sharpen the image.
- Of course, all these additional features on the mask must first be computed (not easy), than made (very difficult), and finally tested (exceedingly difficult).
  - Testing your mask is essential, that any mistake in the mask will automatically be transferred to the chip and, remember Murphy's law, more likely than not kill the chip.
- In the grand total a set of masks will quickly cost you up to **2.000.000 €**. You must a sell a hell of a lot of chips (at a profit) just to recover that cost
  - For customized chips, that are not made by the untold millions, its simply not possible to pay that prize.
  - This drives a large-scale effort to find some better solutions. For mor details (and for the source of some of the data here), refer to "Materials today" from Feb. **2005**.

Multiple Choice Test zu

6.1.1 Si Oxide and LOCOS Process - Summary

Start Multiple Choice

Multiple Choice Test zu

6.2.1 CVD for Poly-Silicon, Silicon Nitride and Miscellaneous Materials

Start Multiple Choice

Multiple Choice Test zu

6.3.1 Physical Processes for Layer Depositions - Summary

Start Multiple Choice

Multiple Choice Test zu

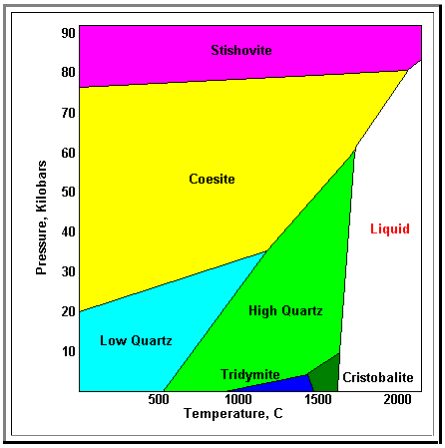
6.6.1 Materials and Processes for Si Technology - Summary

Start Multiple Choice

Forms of SiO_2

Illustration

- Here is a **phase diagram** of **quartz**; unfortuantely I do not know the source
  - What we have at room temperature and standard pressure is "low quartz" or α - quartz

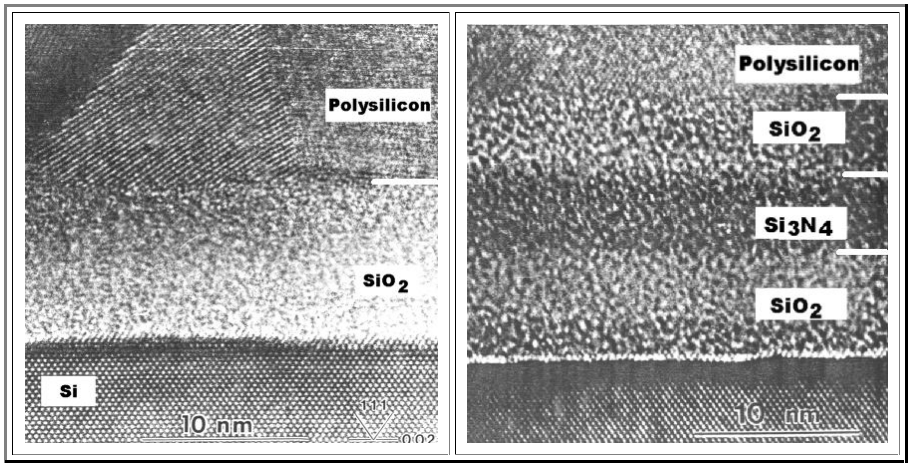


HRTEM of Gateoxide and ONO

Illustration

Here are two ***HRTEM*** pictures of "gate" dielectrics form around **1990** from the Siemens Laboratories in Munich. It is actually not the gate dielectric in these pictures, but the dielectric in the capacitor of a ***DRAM*** memory cell. The properties of this oxide are just as critical as under the gate of a transistor.

- The one on the left show the "conventional" single layer of **SiO_2** that was grown by thermal oxidation.
- The picture on the right shows "***ONO***". Here the first oxide is grown by thermal oxidation of the substrate wafer, the nitride is deposited by **CVD**, and the top oxide layer is produced by oxidizing the nitride.



Furnaces

Illustration

- Furnaces for thermal oxidation, but also simply for annealing or for **CDV** processes are complicated and expensive pieces of equipments.
  - While horizontal furnaces dominated the scene for **150 mm** wafers or smaller, with **200 mm** wafers a switch to vertical furnaces took place
  - Below two pictures showing a horizontal and a vertical furncace for **200 mm** or **300 mm** wafer, respectively. Of course, you don't see much; nevertheless, they are big pieces of equipment.

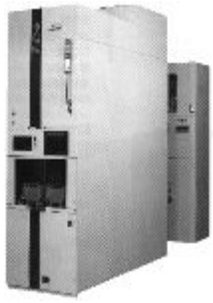
	
Horizontal oxidation furnace; three tubes	ASM A412 300mm twin vertical furnace system for high temperature atmospheric pressure oxidation and LPCVD processing of polysilicon

- Here is a comparison between horizontal and vertical fucnaces, taken straight form the homepage of a major furncae manufactorer.

Comparison Vertical vs. Horizontal Furnaces

Koyo Thermo Systems manufactures numerous versions of vertical furnaces and horizontal furnaces with full automation or manual loading. Smaller versions for pilot and laboratory applications are available.
We are frequently asked for the differences of vertical furnaces and horizontal furnace and for a justification of the higher price of vertical systems. Therefore we worked out the following table sheet. We hope that this overview, this comparison sheet will help you to make a proper planning.
We compare in this sheet a horizontal furnace with a vertical furnace for mass production (VF5300) and with our small vertical furnace VF1000, which is a good alternative to horizontal systems for small companies as well as for research institutes.

Feature	3- or 4-Tube Horizontal Furnace e.g. Model 206	Vertical Furnace for Production e.g. VF5300	Vertical furnace for RD e.g. VF1000

			
Heat up speed	slower	faster	faster
Quartz boat loading time	slower	faster	faster
Cool down speed	equal	equal	equal
Max. temperature using KLL's advanced LGO heaters	1100°C	1250°C	1250°C
Temperature uniformity	lower	higher	higher
Temperature interference between the tubes	1 - 2 °C	none	none
Oxygen concentration tube center (end open)	16%	0,1%	0,1%
Oxygen concentration tube end (end open)	high	500ppm	500ppm
Oxygen concentration tube center (end closed)	0,1%	300ppm	300ppm
Oxygen concentration tube end (end closed)	10 - 30 ppm	10ppm	10ppm
Air tight process chamber (atmospheric process)	no	yes	yes
HCl leak free	no	yes	yes
Cross contamination	possible	not possible	not possible
Process independence	not completely	yes	yes
Particle data	worse	very good	better
Flexibility: Run mixed diameter of wafers in one run	possible	not possible	not possible
Flexibility: Run different diameter of wafers run to run	possible	not possible	possible
Flexibility: Range of wafer diameters that can be processed	3" - 6" (8")	4" - 300mm	3" - 300mm
Stock wafer cassettes	no	yes	no
Automation level	lower	very high	higher
Thickness uniformity wet oxidation 10nm, 8" wafer	no data	± 0.9 %	± 0.9 %
Thickness uniformity dry oxidation 20nm, 8" wafer	± 2.4 %	± 1.2 %	± 1.2 %
Thickness uniformity poly-Silicon 400nm, 8" wafer	± 2.0 %	± 1.0 %	± 1.0 %
Thickness uniformity Nitride100nm, 8" wafer	± 2.5 %	± 1.5 %	± 1.5 %
Capacity	150 wafer	100 - 150 wafer	25 wafers
Power consumption	high	lower	very low

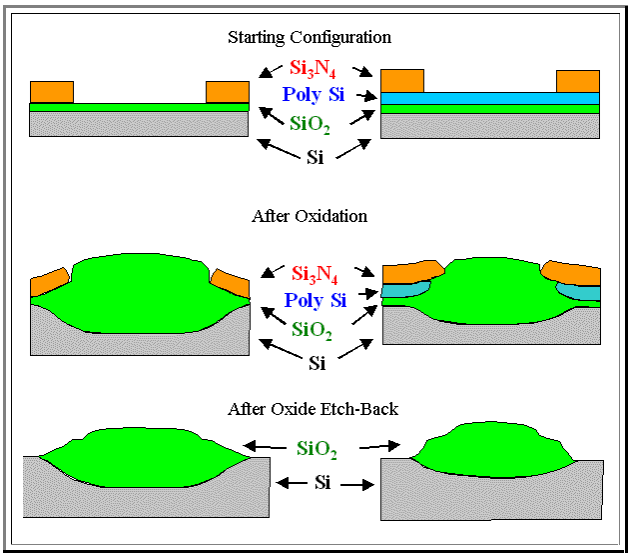
Maintenance independence	no	yes	yes
Maintenance work, necessary	higher	lower	very low
Footprint / tube	2.6 - 3.4 m^2 (partially cleanroom)	3.0 m^2 (grey room)	1.5 m^2 (grey room)
Price	low	high	low

Nowadays, mass production of semiconductor chips happens with silicon wafers with 200mm or 300mm diameter. Vertical furnaces are used almost exclusively. Only in older factories, which still use smaller wafer diameters, horizontal furnaces are still common. For wafer diameters until 150 mm the performance of such systems is in many cases still good enough to fulfil the customer requests. However, the advantages of vertical systems are already evident for this wafer size.
The result of this situation on the oven market was, that almost all large furnace manufacturers stopped the further development of horizontal furnaces. Development work is done today almost only for vertical systems. Therefore vertical furnaces are superior to horizontal ones not only for physical reasons, but also because they are the more modern production tool. Their performance is much higher.
A main issue is furnace automation. Automation for horizontal furnaces means mainly the installation of an elevator system for the loading of the boat on the cantilever arm. The loading area is open to the clean room. Vertical furnaces however are closed system with clean room class 1 inside. The loading happens fully automatically from the cassette by advanced robot systems.
Other technical advantages of the vertical furnaces are the better gas tight sealing of the furnace tube, as well as several options, availabe only on vertical furnaces like improved temperature uniformity by boat rotation or nitrogen load lock chamber.

LOCOS with Sacrificial Poly-Silicon

Illustration

- Here is a comparison of the **LOCOS** process with and without some sacrificial poly-**Si:**

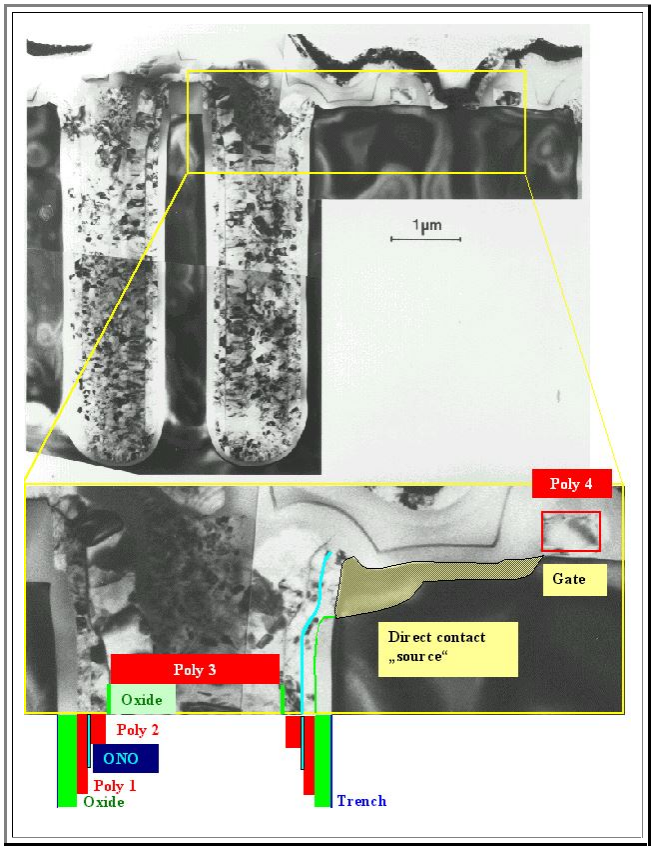


  - Provided, some of the oxide is removed by an "etch-back" process, the lateral extension can be kept somewhat smaller than in the conventional case.
- While this makes things quite complicated, the final versions of the **LOCOS** process were even more complicated.

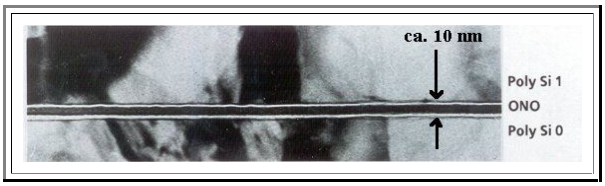
Poly-Silicon

Illustration

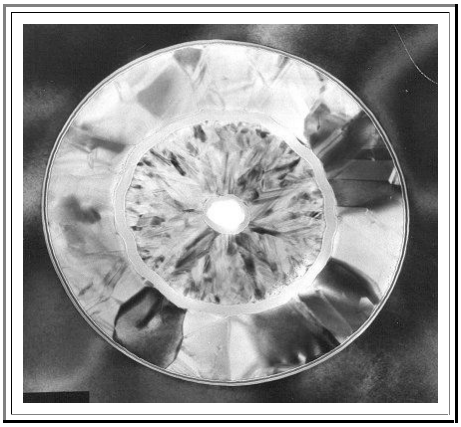
- Here is a large size rendering of the **TEM** picture of a **16 Mbit DRAM** memory cell. A drawing of this cross section can be found in the link.



- .Many details are not visible in this low magnification picture. Below we see the ONO layer between the two poly Si layers at high magnification
  - The two oxide layers appear white; the nitride layer dark



- Next, we look at a trench capacitor "*from above*" and not in *cross section*.
  - Shown is the "simple" capacitor from the **4 Mbit DRAM** generation. **ONO** us used as dielectric; and the substrate **Si** served as one electrode.
  - Again the poly electrode was oxidized for insulation and the trench filled with poly. The smaller (and dendritic) grain structure indicates that a large deposition rate (at somewhat higher pressure) was used; and a little hole remained unfilled in this case.



Requirements for Chip Metallization

Illustration

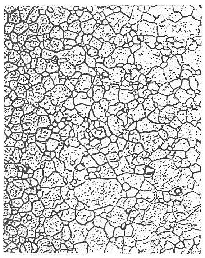
- The metal lines connecting transistors or other components on a **Si** chip must meet many, partially conflicting, requirements. Below is a list, including some materials that do *not* meet the particular requirement very well.
  - Can you guess the winner?

Desired Property	**Materials not meeting requirement**
Very good conductivity	All but **Ag**, **Cu**
High eutectic temperature with **Si** (**> 800 °C** would be good)	**Au, Pd, Al, Mg**
Low diffusivity in **Si**	**Cu, Ni, Li**
Low oxidation rate; stable oxide	Refr. Metals, **Mg, Fe, Cu, Ag**
High melting point	**Al, Mg, Cu**
Minimal interaction with **Si** substrate	**Pt, Pd, Rh, V, Ni , Mo, Cr** (form silicides easily)
Minimal interaction with poly **Si**	Same as above
No interaction with **SiO_2**	**Hf, Zr, Ti, Ta, Nb, V. Mg, Al**
But must stick well to **SiO_2**	?
Must also comply with other substrates, e.g. **TiN**	? (see example for **Al** below)
Chemical stability, especially in **HF** environments	**Fe, Co, Ni, Cu, Mg, Al**
Easy structuring	**Pt, Pd, Ni, Co, Au**
Electromigration resistant	**Al, Cu**
.... and many more,...	

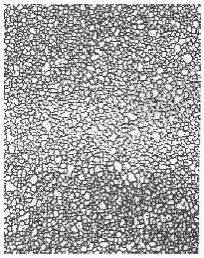
- The winner is: *Aluminum* (with **<1%** of **Si** and **Cu** added).
  - **Al**, in fact, is pretty bad - but all others are worse!
  - Presently (**2001**) a switch to **Cu** takes place (the better conductivity is definitely needed). The industry will pay several **10^9** Dollars to develop the new material technology and change the production facilities.

Al Grain Structure on Different Substrates

- Around the late eighties, the necessity came up to use a **diffusion barrier** between the **Al** - metallization and the **Si** substrate becasue the rection of **Al** with the **Si** in contact holes with cross sections **< 1 µm^2** became a problem. One material of choice was **TiN**, another one **Ta**.
  - The grain structure of the **Al** layer (and with it other properties, e.g. the electromigration resistance, depends significantly on the substrate).
  - Below you can see the representative pictures (identical scale) that illustrate this point.



Aluminium layer on top of a **Ta** layer



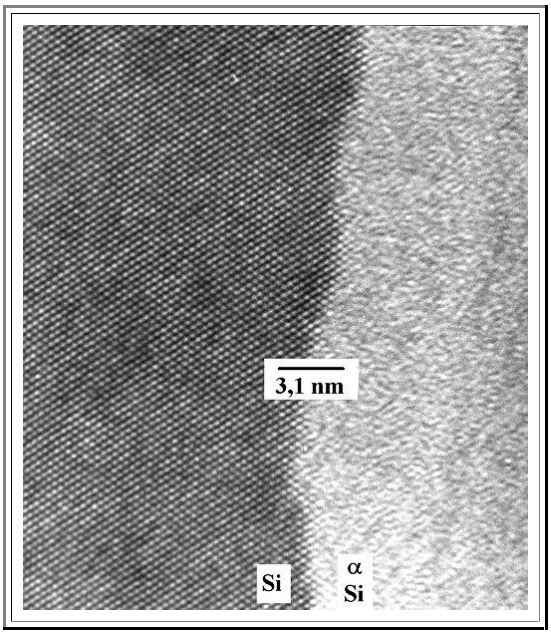
Aluminium layer on top of a **TiN** layer

- Close examination revealed that the substrate influences:
  1. Grain size.
  2. Grain size distribution.
  3. Texture.
  4. Degree of **Si** precipitation.
  5. Macroscopic stress.
  6. Microscopic stress.
- All of these properteis may influence the performance ot the **Al** conductor - and this gives you an idea of what it means to introduce a new material into a fine-tuned product.

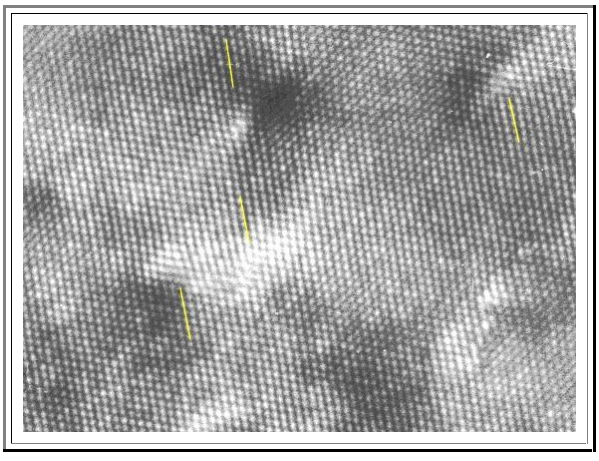
Amorphization after Ion Implantation

Illustration

- The picture below shows the boundary between crystalline **Si**, and **Si** that has been rendered *amorphous* by an ion implantation (the ion beam came from the right). The picture is slightly remarkable because it was the first high resolution transmission electron microscopy (**HRTEM**) picture ever taken from damage after ion implantation.
  - There is indeed an amorphous **Si** (α-**Si**) layer. (Think a minute how you could ascertain that without transmission electron microscopy).
  - There is a pretty abrupt, if somewhat wavy boundary between the amorphous and the crystalline **Si**.
  - There seems to be little disturbance in the **Si** lattice - it looks pretty perfect.



- Quick glances at **HRTEM** pictures may be deceiving, however.
  - A lot of point defects may be contained in the lattice - they would not clearly show in this picture
  - Looking a bit more closely at some greater depth (to the left of the above picture), a high density of dislocations is found. An example is shown below; the ending lattice planes are indicated with yellow lines.
  - Note that not all dislocations will show up in this kind of imaging mode.



Loosing Large Amounts of Money with Wet Chemistry

Illustration

- Usually, the big disasters in chip production are kept quite confidential - you don't want to admit that you have problems, and you don't want to help your competitors to avoid these problems.
  - However, as time goes by, confidentiality is no longer necessary. But people have other problems now, so still nothing is published.
  - Well, here is the exception: The short story of a major disaster in the early production of the **1 Mbit DRAM** in the new Siemens factory in Regensburg, as it happened back in about **1985**.
- Managerial wisdom had decided not to trust the research and development team with the development of the **1 Mbit DRAM** (Siemens (and Phillips) then were about **1** year behind the leading Japanese and had started a race to catch up), but to take a license from Toshiba, the top memory producer then.
  - Not that we liked it. But the new factory was dutifully converted to the Toshiba process - all the equipment, all materials, everything whatsoever, was identical to what Toshiba had and did - with good yields of functioning devices.
  - But the German factory produced exclusively junk - the memory chips didn't work. Nobody, including the Toshiba engineers, had the faintest idea why.
  - This went on for almost **6** months - at losses of about **5 - 10** million marks a month. Then the problem was found and solved. First empirically, then by understanding what happened.
- Since it was evident that something *we* did must be different from what *Toshiba* did, the search focussed on the few differences that were unavoidable for some reason or other.
  - The culprit that was finally identified, was an extremely simple chemical: **H_2O_2**; used in the mixtures for cleaning the wafers.
  - In contrast to practically all other chemicals, the **H_2O_2** was not bought in Japan from the source Toshiba used, but from a German company because it simply would not have survived months of traveling aboard a ship. **H_2O_2** always decays into water and **O_2** in the course of weeks, and since it is a slightly dangerous chemical, the airlines refused to transport it.
  - So there was no choice but to buy it in Germany - and of course the German **H_2O_2** was carefully checked for cleanliness (it was actually cleaner than the Japanese stuff).
- Somebody finally convinced an airline (Alitalia) to fly in a barrel of the Japanese stuff - and a miracle happened and good chips were produced using it.
- What has happened? Nobody knew, but who cares if it works? Well the research oriented guys do care, and in due course the mystery was unraveled.
  - As it turned out, **H_2O_2** always contains some *stabilizer*, and this is neither displayed on the label nor do the producers tell you what it is. The stabilizer is needed to keep the remaining traces of metal ions, that are still present - even in ultrapure **H_2O_2** - complexed (i.e. surrounded by the stabilizer molecules); in particular **Fe^{++}**.
  - And this complexation is necessary because "naked" metal ions would catalyze the decay of **H_2O_2** into water and oxygen - the "shelf life" of your chemical would be very short without a stabilizer.
  - Now Siemens, as most other western producers, used some variant of the classical "**RCA**" cleaning procedure which is always acidic, i.e. it works at **pH** values **<< 7**.
  - The Japanese, however, had invented a new alkaline cleaning procedure, relying heavily on "Choline", a simple organic leach, i.e. they worked at pH values **>> 7**.
- As it turned out, the stabilizer in the Japanese **H_2O_2** worked in an alkaline environment, while the German stabilizer did not.
  - This was purely accidental, neither the Japanese, nor the Germans, nor anybody else, had ever to worry (or even knew) about the stability of **H_2O_2** stabilizers.
- As a consequence, whenever an alkaline cleaning was carried out with the German **H_2O_2**, Fe was no longer complexed and some of it was deposited on the **Si** substrates.
  - This must be expected to happen, because **Si** is less noble than **Fe** with respect to its electrochemical potential. We are talking tiny amounts of deposited **Fe** here, far less of what is still contained in ultra-hyperpure chemicals.
  - The iron deposited in this way would diffuse into the **Si** as soon as it was heated. This did not do much damage, and that was why every measurable parameter always looked quite good during processing - only at the end the chips started to deteriorate.
  - The reason for this was that at every heating cycle, **Fe** was diffusing around a bit more, until eventually small precipitates formed (needle-like **$FeSi_2$**)
- And these precipitates killed some gate and capacitor oxides - and since it needs only *one* dead transistor (out of about **1,5** million) to kill a chip, the yield was practically non-existent.

- There is just no way you can anticipate that. And the detective work in this case was complicated because the effect (dead transistor) was not traceable to the reason (incorporation of iron), because first measurable deviations from expected behavior occurred many process steps *after* the original cause.
- Many process lines could tell similar stories. From what one hears or suspects, one rule that might be good to know evolves: The really big disasters in chip manufacture are more likely to have their roots in humble wet chemistry than in the sophisticated processes everyone talks about.

Casimir Effect

Advanced

- Imagine some space filled with "absolute" vacuum. It contains nothing whatsoever.
  - Is that possible? Technically no. We don't have the equipment (vacuum pumps and so on) to really suck out the last molecule or atom from inside some container, but conceptually you should have no problem to imagine an absolute vacuum.
- However, if you can imagine an absolute vacuum, you didn't pay enough attention to your quantum mechanics. There you learned about the Heisenberg uncertainty relations in the form

$$\Delta x \cdot \Delta p_x \geq h$$

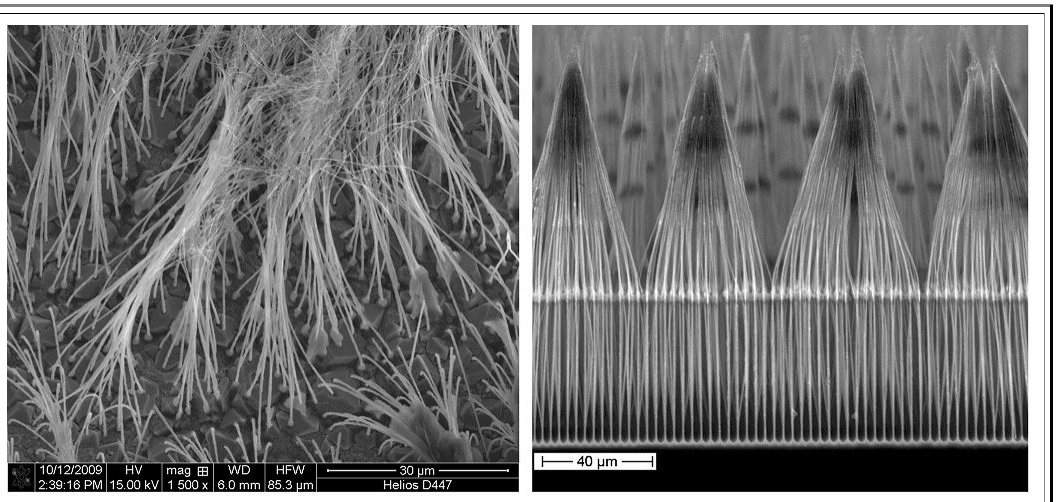
$$\Delta E \cdot \Delta t \geq h$$

- The second one is the interesting one. here. It says that for small times Δt you may violate energy conservation by ΔE.
  - That means that even in absolute vacuum, some elementary particle may come into existence (needing some energy ΔE for its generation) if it has a short enough live.
  - This is not a joke but an accepted fact of quantum electrodynamics: even an absolute vacuum is teeming with so-called virtual particles that come into existence for a short time span and the disappear again. For some reasons of conserving charge etc, virtual particles always appear as particle - anti-particle pair. While they "live" they have some wave function with some **exp(i*kr*)** term and thus also some wavelength. The same thing, off course, also happens in air
- In the words of an article from Alexander E. Braun, Senior Editor in the ***Semiconductor International, 1/19/2010***:
  - "*Predicted in 1948 by Dutch physicist Hendrik B.G. Casimir, the Casimir Effect results from the fact that space is filled with vacuum fluctuations, virtual particle-antiparticle pairs that continually form out of nothing then quickly vanish. This effect is observable between two metallic parallel plates placed 100 nm apart. The gap between the plates restricts the range of wavelengths possible for these virtual particles. With fewer of them present within that space this results in a lower energy density between the two plates than is present in open space. This creates negative energy and pressure that pulls the plates together; the narrower the gap, the more restricted virtual particles' wavelength, the more negative the energy and pressure, the stronger the attractive force. At nanoscale, it becomes the dominant force between uncharged conductors. Depending on surface geometry and other factors, at separations of 10 nm the Casimir Effect can produce the equivalent of 1 atmosphere of pressure (101.3 kPa)*."
- In other words: the Casimir effect is real and a problem (possibly also an asset) to ***NEMS*** = Nano-EMS
  - The rest of the article delves into details: the problems caused to **NEMS**, possible ways to avoid them, and what else one might encounter at these dimensions.

Stiction

Advanced

- We all know some kind of stiction that we also could call "wet hair effect".
  - The picture below shows what your hair looks like after it became wet. The surface tension of the liquid "glues" the hair fibers together.



Vertical and horizontal sticking of cantilevers

- The "hairs" here, however, are **Si** nanowires; far smaller than an actual hair. Their making involves some wet chemistry and at some point you pull the whole arrangement out of some liquid.
  - The **Si** nanowires stick together as shown - exactly like hair.
  - *However,* in contrast to hair, they do not come apart again as soon as they are dry. The "sticking forces" (properly called adhesion or secondary bonding) acting on the large specific surface cannot be overcome by the "spring forces" resulting from the elastic bending of the little volume of the nanowire.
  - Maybe Mother Nature saw to it that hairs are thin (the thinner the better for elasticity, thermal insulation, ...) but not too thin - just right for what they are being utilized for.

Exercise 7.1-1

Class Exercises and Quick Questions to

7.1 MEMS - Products and Developments

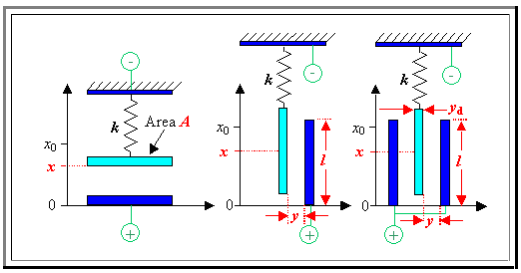
We had the following class exercises:

- How large is the resistance of a **1 pF** capacitor at **10 GHZ**?
- Consider that the layer of whatever it is that forms the cantilever in the picture above would be under tensile stress in its top part (maybe the cantilever consists of two different materials stacked on top of each other). What would happen?
- Calculate Δ**R/R** for a rectangular piece of material with length ***l***, width ***w***, thickness ***t*** and specific resistivity ρ that is strained by ϵ in ***l***-direction.
- Give examples of **MEMS** products, their working principles and raneg of applications.
- Describe the working principle of a **DLP** beamer. Consider strengths and problems.
- Describe the working prinicple of a **MEMS** gyro. Provide a rough sketch of a possible implementation.
- Describe possibilites for inducing and detecting mechanical movment in a **MEMS** device.
- Compare a gyro or acceleration sensor operated around resonance or at lower than resoance frequencies. Give curves of amplitude and damping as a function of frequncy and discuss the role of damping.
- Give some pricnciples for making actors and compare the relative merits of the approach.

Exercise 7.1-2 Capacitors and Forces

Forces in Capacitive Structures

- Let's look at the three different situations where we can produce forces electrostatically by using capacitive structures
  - We consider one fixed plate with either a given area ***A*** or a dimension ***h*** · ***l*** (***h*** would be the height in the ***z***-direction in the drawing below), and a moveable plate of identical dimension to keep things easy. We also have a spring with a spring konstant ***k*** to keep the moveable plate in force equilibrium. The position x_0 is the equilibrium position for zero voltage. The terminals show schematically how voltage would be applied.



  - We consider *only* movements in the ***x***-direction as shown for the three configurations given (where the plates have moved some distance ***x*** from their equilibrium position (zero voltage at ***x*** = **0**).
  - The third configuration embodies one element of a comb structure that we have encountered a few times already. We assume that there is a potential difference***U*** between the plates of the capacitors.

- Here are the questions:
  - **1.** Find the proper relations for the forces pulling at the moveable plates for all three configurations.
  - **2.** Compare the relative strength of the first and third configuration (you may assume that $y \approx y_d$)
  - **3.** Discuss the pros and cons of the two configurations for driving an actual actuator.

- **Hint:** Consider the work ***W*** = ***E*** needed to move a plate and remember that Force $F = -dW/dx$

Solution

Exercise 7.2-1

Class Exercises and Quick Questions to

7.2 MEMS - Processes and Specialties

- Here are some quick questions
  - What is "stiction" and what role does it play in **MEMS** technology?

Exercise 7.3-1

All Quick Questions to

7. Si MEMS

Products and Developments

- How large is the resistance of a **1 pF** capacitor at **10 GHZ**?
- Consider that the layer of whatever it is that forms the cantilever in the picture above would be under tensile stress in its top part (maybe the cantilever consists of two different materials stacked on top of each other). What would happen?
- Calculate Δ**R/R** for a rectangular piece of material with length ***l***, width ***w***, thickness ***t*** and specific resistivity ρ that is strained by ϵ in ***l***-direction.
- Give examples of **MEMS** products, their working principles and raneg of applications.
- Describe the working principle of a **DLP** beamer. Consider strengths and problems.
- Describe the working prinicple of a **MEMS** gyro. Provide a rough sketch of a possible implementation.
- Describe possibilites for inducing and detecting mechanical movment in a **MEMS** device.
- Compare a gyro or acceleration sensor operated around resonance or at lower than resoance frequencies. Give curves of amplitude and damping as a function of frequncy and discuss the role of damping.
- Give some pricnciples for making actors and compare the relative merits of the approach.

Processes and Specialities

- What is "stiction" and what role does it play in **MEMS** technology?

Solution to Exercise 7.1-1

Class Exercises and Quick Questions to

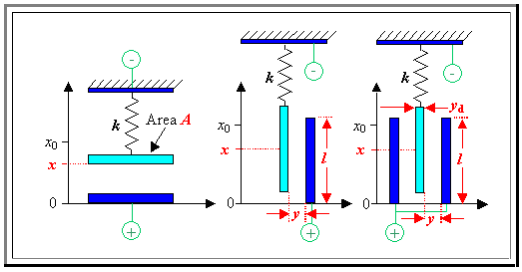
7.1 MEMS - Products and Developments

- No solutions at present

Solution to Exercise 7.1-2

Forces in Capacitive Structures

Here is the drawing once more to avoid jumping back and forth:



1. Find the proper relations for the forces pulling at the moveable plates for all three configurations.

- The total energy ***E*** for all three configurations is simply given by

$$E = \int U \cdot I \cdot dt + \int k \cdot (x - x_0) \cdot dx$$

- With the simple relations $C = Q/U$, $I = dQ/dt$ and therefore $U \cdot I \cdot dt = (Q/C) \cdot (dQ/dt) \cdot dt = Q/C \cdot dQ$, we obtain

$$E = \frac{Q^2}{2C} + \frac{k \cdot (x - x_0)^2}{2}$$

$$= \frac{C \cdot U^2}{2} + \frac{k \cdot (x - x_0)^2}{2}$$

$$F = -\frac{U^2}{2} \cdot \frac{dC}{dx} - k \cdot (x - x_0)$$

- Without integration limits we cannot get proper signs (one energy term must decrease if the other one increases because we have energy conservation) - but that is not important here since we know that the spring force and the capacitive force must have opposite signs, and we are only interested in the capacitive force F_C.

For the capacity ***C*** and the force $F_C = \frac{1}{2}U^2 \cdot (dC/dx)$ we obtain:

- **1.** Configuration:

$$C_1 = \frac{\epsilon_0 \cdot A}{x}$$

$$F_1 = \frac{U^2}{2} \cdot \frac{\epsilon_0 \cdot A}{x^2}$$

- **2.** Configuration:
  x* is the (easy to calculate) plate overlap for zero voltage. But since it disappears upon differentiation, we do not need to spell it

$$C_2 = \frac{\epsilon_0 \cdot (x^* - x) \cdot h}{y}$$

$$F_2 = \frac{U^2}{2} \cdot \frac{\epsilon_0 \cdot h}{y}$$

out.

- **3.** Configuration:
  C_3 is simply given by **2** C_2 in parallel

$$C_3 = 2C_2 = \frac{2\epsilon_0 \cdot (3l/2 - x) \cdot h}{y}$$

$$F_3 = 2F_3 = U^2 \cdot \frac{\epsilon_0 \cdot h}{y}$$

2. Compare the relative strength of the first and third configuration.

- If we simply take the relation F_1/F_3 for equal distances between the plates (i.e. $x = y$), we obtain

$$\frac{F_1}{F_3} = \frac{A}{y \cdot h}$$

- Considering that $y \cdot h << A$ for typical structures, configuration **1** can transmit much more force than the other ones for about identical size.

3. Discuss the pro and cons of the two configurations for driving an actual actuator.

In configuration **1** the force decreases with the square of the distance between the plates; in the extreme case of zero distance the plates would stick together *forever* (in reality a fuse will blow).

- The design rule is obvious: Use with extreme care!

In configuration **3** the force is independent on the position, which makes the design reasy. However, the force is relatively small.

- The consequences are obvious too: This is the preferred configuration, but you need to employ many combs to achieve sufficient force.

MEMS Market

Illustration

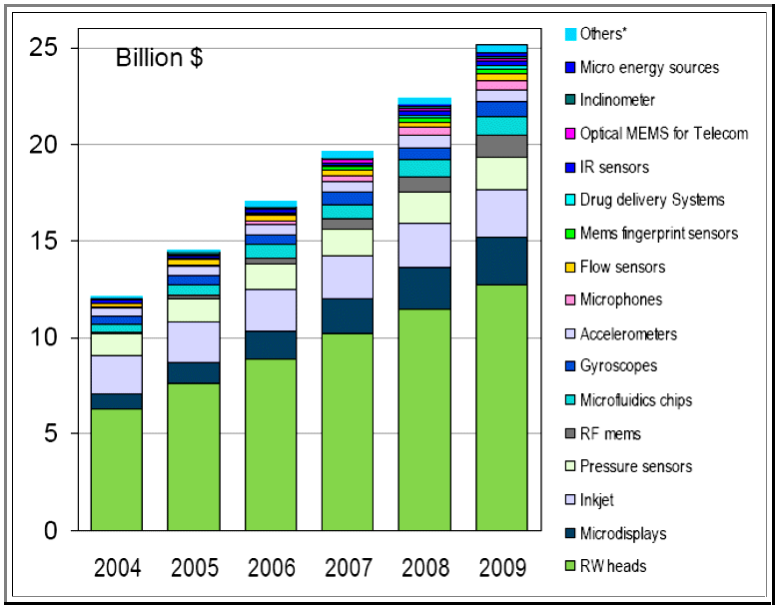
- It is surprisingly difficult to find reliable numbers about the **MEMS** market in the Net. What one finds quite easily, however, are announcements of detailed studies about the **MEMS** market that you can buy for a price of several **1 000 $**.
  - That makes it clear that some kind of **MEMS** "gold-rush" is happening right now. If you know where the nuggets are, you won't tell anybody but go for it yourself. If you can't do that, you try to sell your treasure map.
- Nevertheless, here are a few numbers that seem to be trustworthy:
- The worldwide **MEMS** *systems* market is expected to reach **$ 95 · 10^9** by **2010**.
  - If we look at the *device* market in **2005**, we have the following numbers (partially calculated from other data):

Region	Gross 10^9 $	%
USA	2,6	53
Japan	1,1	20
Europe	0.815	16
Singapore	0.45	11
Σ	4.22	100

- Besides the market for the **MEMS** devices, there is a market for all the things you need to make **MEMS** devices.
  - You need machinery or equipment for the **front-end** and **back-end** process, and you need materials. Front-end means more or less everything going on in the clean room, back-end describes essentially the packaging of the raw chips - an often quite sophisticated process in **MEMS** technology.

Area	Gross 10^6 $	Share
Front-end	?	53 %
Back-end	?	31 %
Materials	≈ 270	

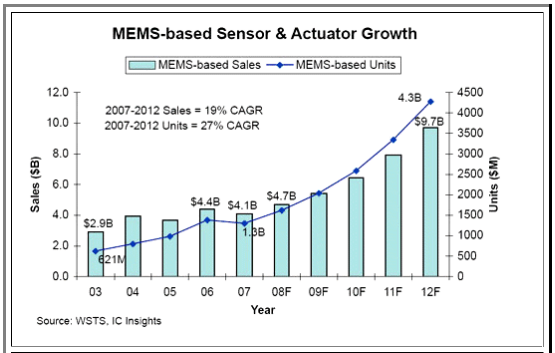
  - No numbers for the machinery, but **270 Mio $** is really money, too; and the gross for the equipment is certainly much higher.
- Here is an overview from the "Silicon Technology I" script of **Prof. Wagner**, belonging to the Master Study course.



The discrepancy to the numbers above comes from unclear definitions: Are "RW" (reading and writing) heads included in the numbers above? Are we dicussing devices or systems? Are we discussing just the product market or also the "making **MEMS**" market?

- Whatever. Two things are clear enough: There is a lot of money in **MEMS**, and it is a rapidly growing market.

The "Semicondcutor International" in March **2008** agrees: Below are some data concerning the **MEMS** sensor and actuator market.



From **IC Insights Inc.** (Scottsdale, Ariz.) as published in the **Semiconductor International Newsbreak**. March **26th 2008**

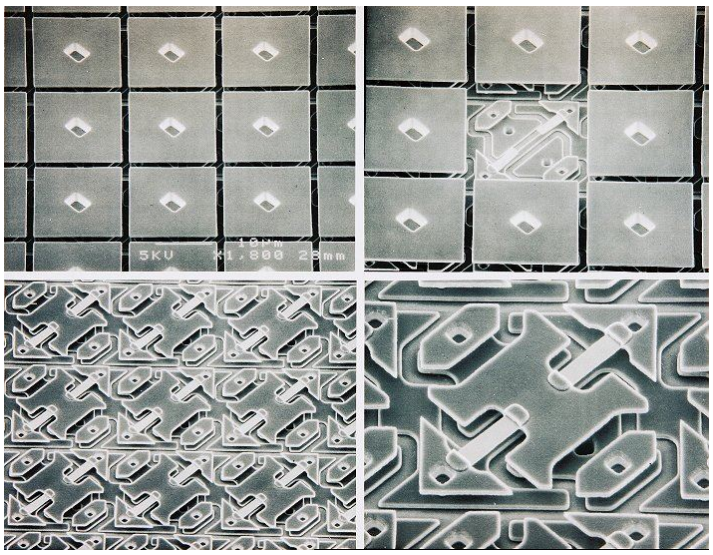
- Note that between "Units" and "Sales" we have roughly a factor of **3**. There might be a mistake here: Units probably should not have the $ sign - it's just how many devices are made. If this is true, we see that a **MEMS** chip sells for about **3 $**.

The message is loud and clear: The future belongs to **MEMS** (and **Solar**, and ...) - and that may be part of your future as materials scientist, too.

The Texas Instrument DLP Chip

Illustration

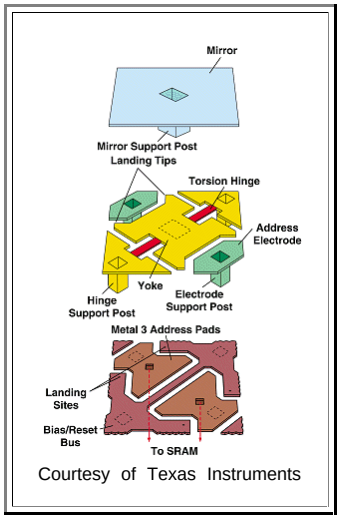
- The pictures below show some of the **> 900 000** to more then **2 000 000** micromirrors of a DLP chip and what is below.



Courtesy of Texas Instruments

  - What you see are the "hinges" of individual mirrors. What you don't see is that below the hinges is an array of transistors etc. that allow to address an individual mirror to "do its thing".
- It looks pretty complicated because it is pretty complex. But still rather coarse if you look at the size of the smallest entities - about **1 µm** or larger, so there is room for much improvement in years to come.
- How does it work? The basic principle is shown below

- The **DLP** chip's micromirrors are mounted on tiny hinges that enable them to swivel around a defined axis in two defined positions called "ON" and "OFF".
  - Any mirror can be addressed individually and the time for being in an ON or OFF position is individually adjustable.
  - The system is driven electrostatically at some **kHz** swivel frequency; a ***SRAM*** memory cell contains the pixel information for the long time - tens of milliseconds! - the pixel is held at some intensity value.
  - The light from the lamp is reflected by any mirror either into the the optical system that projects it on the screen or wall (ON position), or into a light sink (OFF position). The intensity you see on the screen for every pixel depends on the ratio of ON to OFF times for the mirror processing that pixel.



Courtesy of Texas Instruments

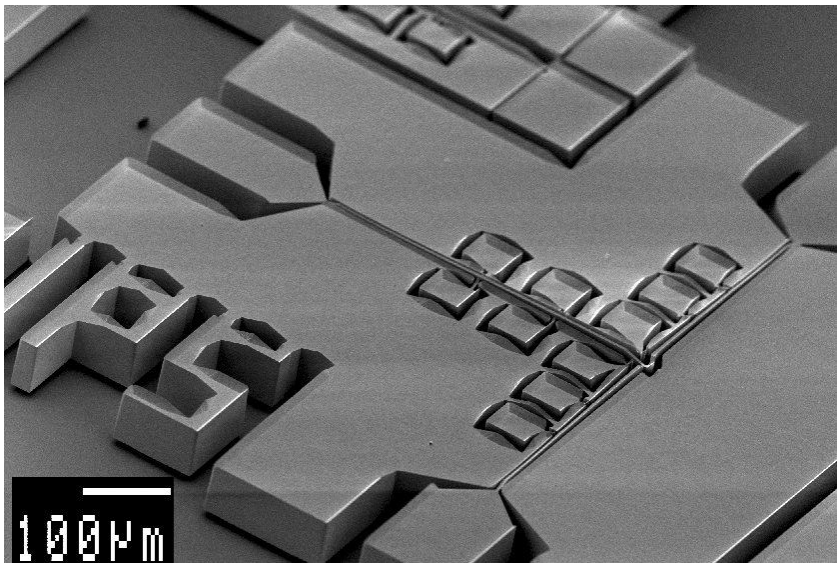
- Now we can form a grey scale image. How about color?
  - Simple. Illuminate through a revolving filter wheel with red-green-blue in succession, and have you individual mirror project the right red-green-blue intensity synchronously.
  - Of course, you loose intensity. So have three **DLP** chips, one for each primary color, and superimpose the thee images. But that makes the beamer more expensive.
  - OK - now we are talking product diversity and the **R&D** that goes with constant product improvement. How about a beamer for a big cinema screen? What kind of **DLP** chip would that need?

You get the idea. Digital light processing may still have a long and very rewarding (Dollar wise) career ahead. TI already made more than **10^9 $** with it by now (2007). Or it may not - because something better may come up.

RF MEMS

Illustration

- ***RF*** actually means **Radio frequency**, i.e. at best **MHz**, but it is now understood to mean "high frequency" (sometimes abbreviated ***HF*** in German) in the sense of ***GHz*** frequencies.
  - In this frequency range you can forget most of what you learned about running **AC** current through wires. You actually wire, if you go up the frequency range, first becomes a coaxial cable (**MHz**), than a wave guide (**GHz = 10^9 Hz**), i.e. a hollow tube, and finally an optical wave guide, i.e. a glass fibre or simply "nothing" (= air) for optical frequencies at ≈ **300 THz = 300 · 10^{12} Hz**.
  - Capacitors and inductors as individual elements start to loose their meaning in the upper **GHz** range (ever seen a capacitor for light?) and signal generation, propagation and detection becomes a special science and engineering discipline with special components and products
- Presently, we have Microwaves in use everywhere - from your microwave oven (operating roughly around **10 GHz**) via all kinds of microwave sensors e.g. in your car, to serious ***RADAR***; actually an abbreviation for **RA**dio **D**etection **A**nd **R**anging.
  - We also have "optics" from the far **IR** to the **UV**. The only frequency range not used for some application is the lower **THz** region, but it is a major research area at present.
  - In any case, if you want to do something useful in "**RF**", you need some devices. Most trivial, you simply need switches to turn some transmission on and off (without simply unplugging the power supply) or to route it form here to there. Simply interrupting a piece of wire won't necessarily do the job because that may just produce is a capacitor with very little resistance at high frequencies.
- If we now look at the connection between **RF** technology and **MEMS**, we find, e.g., the following statements in the Net:
  - **RF** microelectromechanical structures (**MEMS**) are replacing conventional microwave devices in various wireless transceiver applications, offering the advantages of improved isolation, lower power dissipation, and reduced cost, size, and weight. Emerging **RF MEMS** devices include switches, high **Q** capacitors and inductors, couplers and power dividers, filters, resonant structures, etc.
  - Microelectromechanical Systems (**MEMS**) applications in **RF** and microwave electronics are on the verge of revolutionizing wireless communications. In particular, **RF MEMS** promises to endow wireless handsets, base stations and satellites with the key properties of low-power consumption and reconfigurability, which in turn will enable superior functionality and performance. In this course, a comprehensive exposition of the state-of-the-art in **MEMS** technology applied to **RF** devices, circuits and systems is given.
    The topics to be presented include: **RF MEMS** fabrication technology, **MEMS** Actuators, Passive devices (Transmission Lines, Capacitors, Inductors, Switches, Varactors, Resonators), Circuits (Filters, Oscillators, Phase Shifters, Couplers), Systems (Transceivers, etc.)
    The IEEE Boston Section Techsite
    http://www.ieeeboston.org/edu/2006spring/mems.htm
  - **MEMS** technology can be used to implement high quality switches, varactors (variable reactors), inductors, resonators, filters and phase shifters. Among the broad range of applications the **MEMS** technology gives a unique possibility to implement micromechanical resonatores and filters with high performance regarding selectivity and **Q**-factors. When combining these mechanical structures with microelectronics, central parts in wireless systems, RF systems (Radio Frequency systems) can be implemented.
    Examples can be various types of oscillators, **VCOs** (Voltage Controlled Oscillators), mixers and sharp filters. The **MEMS** structures can thereby replace traditional costly and large off-chip discrete components by making possible integrated solutions that can be batch processed. Vibrating **MEMS** resonators and filters that have been implemented so far are based on mechanical vibrations in lateral or vertical directions on Silicon wafers. Different types of beams, comb structures and disks can be used.
    From **RF MEMS** at the Department of Informatics, University of Oslo.
    http://heim.ifi.uio.no/~oddvar/rfmems.htm
- Subtracting the hyperbole still leaves us with some possible applications that we may use in the not too distant future.
  - The picture shows an actual example of a "**RF switch**" and the picture caption gives an idea of why **RF** technology is special.



Source and copyright: KTH - Royal Institute of Technology, Sveden

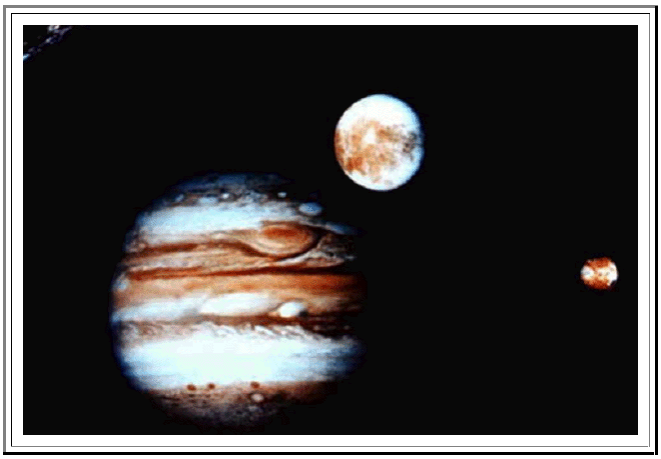
http://www.ee.kth.se/php/index.php?action=research&cmd=showproject&id=23

2006-2007: Micromachined **3D**-transmission line embedded, mechanically-multistable single-pole-single-throw (**SPST**) and single-pole-double-throw (**SPDT**) switches with very low intrusive **RF** design (low reflections, low insertion loss), since the complete switch mechanism is fully embedded in the signal line of a coplanar waveguide. The switches are mechanically multi-stable, feature active opening, and are fabricated in a single photolithography mask process

Two Planets

Advanced

Two planets meet on some social function. The first one says: "Long no see. How you're doing?"



"Not so good", says the second one. "I have a severe case of Homo Sapiens".

"Oh my God" replies the first one, "I feel sorry for you. I had this myself some eons back and it was thoroughly disgusting. But don't worry. It never lasts very long".

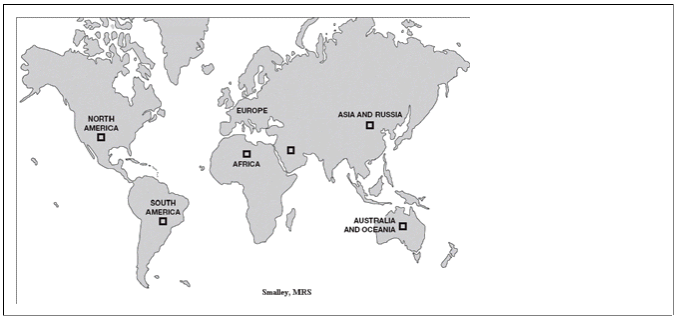
Frequently Asked Questions Concerning Solar Cells

Area Needed for Producing Lots of Energy by Solar Cells

Advanced

- Do we have enough **area** that we can cover with solar cells to produce the energy we need?
  - Let's look at *you* - an average **EU** citizen. Here are your basic data:
    - You need roughly **150 kJ/a** ≈ **40 000 kWhr/a** primary energy.
    - Your direct electricity only (secondary energy) needs are around **2 000 kWh/a**.
    - Your direct plus indirect electricity only energy needs are around **6 000 kWh/a**.
    - Your **1 m²** solar module with **η = 15 %** solar cells produces in Germany **150 kWh/a**.
    - There are roughly **500** million of you; including **82** million Germans.
  - Obviously *you* need.
    - **267 m²** for *your* total primary energy needs and **132 000 km²** for all of you Europeans (**21 900 km²** for the Germans)
    - **40 m²** for *your* total electricity needs, and **20 000 km²** for all of you Europeans (**3 280 km²** for the Germans)
    - **13 m²** if *you* don't want to pay electricity bills anymore, and **6 500 km²** for all of you Europeans (**1 070 km²** for the Germans)
  - How much available area is there?
    - The **EU** (Germany) covers about **4.2 · 10⁶ km²** (**357 · 10³ km²**). The total needs of all of you are thus **3.14 %** of the area of the **EU** or **21 900 km²** for the Germans,
    - The German "Autobahnen" alone have a length of **12 000 km** or an area of roughly **12 000 km · 25 m = 300 km²**.
    - All the other (paved) roads have a length of about **500 000 km** or an area of **500 000 km · 7 m = 3 500 km²**. The railroads add up to **43,457 km · 5 m = 218 km²**. All traffic space together (without airports, etc.) thus consumes **> 4 000 km²** or **> 1 %** of the German area.
    - How much average rooftop area do you have? Count the roof on your home; but also on the places you work and hang out. Let's guess it's around **5 m²** - **25 m²**, with a total of **410 km² - 2 050 km²** in Germany.
  - All the German area presently build over with something thus amounts to **(5 000 - 7 000) km²** in our crude guess, only a quarter or a third of what we would need.

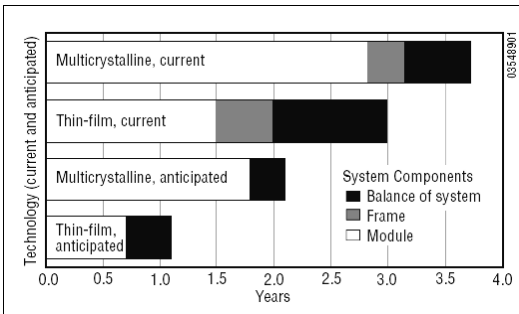
- *Conclusion*: Obviously it is a good idea to also use wind power, solar heating and so on, and above all, to *reduce consumption*, because we don't want to cover even more of the ground by solar panels than we have covered already by concrete and asphalt.
  - But the numbers also suggest that we can produce a lot of electricity by simply covering rooftops with solar cells. Even the low number of **400 km²** would produce about half of your secondary *direct* electricity needs, i.e. would substitute for power plants with three times the *primary* power capacity.
  - "**Bio fuel**" that is grown and burned in regular power plants for producing electricity is *not* a good idea in this context. Photo synthesis, while remarkable for its adaptability and so on, runs at far lower efficiency than a solar cell. Comparing the energy that one can harvest from a given area by bio fuel and by solar cells shows that solar cells "harvest" orders of magnitude more solar energy per **m²** than "e**nergy plants**". Bio fuel only comes into its own if used as energy storage medium or if produced as a by-product of producing food, for example.
  - Right now, we Germans import a lot of our energy needs in the form of oil, ga,s and coal. We could just as well import solar energy. Our total area needs add up to a square with **150 km** side length. You could hide an area of this size in the Sahara desert with little chance of someone finding it by accident.
- Here's a world map showing the total area (= squares) needed in some deserts to supply mankind with all the energy it reasonably needs.



From the article of Nobel prize winner **Richard Smalley** in the MRS Bulletin **30, 2005**

Energy Pay Back Time

- What is the energy **pay-back time** for solar cells? To ask more pointedly: Does a solar cell during its life time (let's say **20** years) produce the energy that it took to make it?
  - The **energy pay back time** thus is simply the time a solar cell has to "work" to generate just that amount of energy that it took to make it. Of course, any regular power plant has an energy pay back time, too
  - How large is the energy pay back time for any energy "producing" contraption? What do you count for a standard coal-fired plant, for example? For sure the energy needed to make its components and to assemble the whole contraption. You should count the energy for digging out the coal and for transporting it, of course. Do you also count the energy needed to tear it down one day and to redo the damage done by the emissions? Should you also count the energy needed to keep the humans alive that run the plant (you must feed them on occasion and transport them to the plant)?
  - You realize that the concept of an energy pay back time is not so easy as it appears on a first glance. If you consider on top of this that the energy needed to build a solar cell also keeps people in business, it becomes even more complicated and questionable.
- Nevertheless, there are useful definitions that allow comparison of different energy producers and based on this one can come with numbers. There is a considerable spread in those number, depending on who produced it. You can bet that supporters of solar energy come up with smaller numbers for the pay-back time of solar cells than the supporters of nuclear power, and vice verse.
  - In any case, the energy pay back time will be several years at the most
- The picture shows some recent data from an unimpeachable source not likely to favor solar energy (George W. Bush ~~is~~ was the Boss!).



From the ***DOE*** (Department of Energy) of the USA government

The Costs of Energy

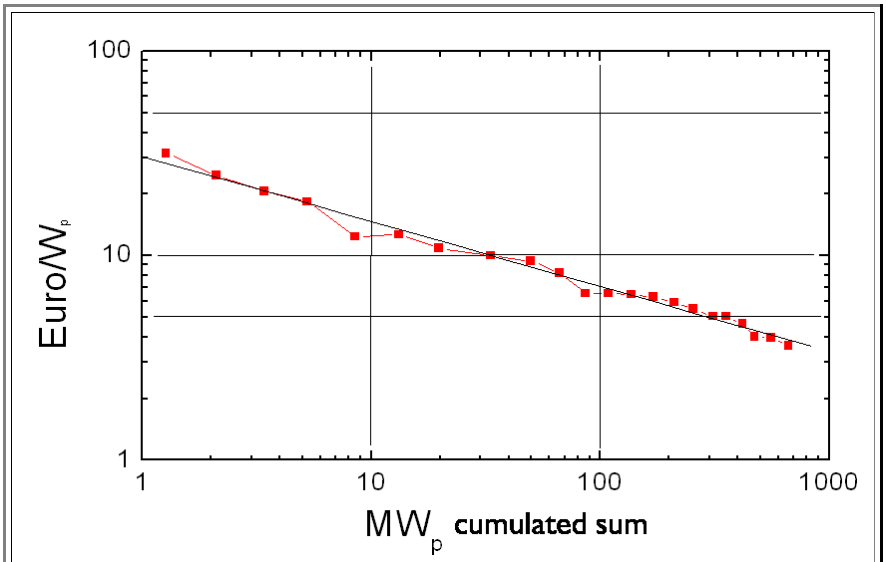
- So we have enough land and solar cells would produce net energy - far more than it took to make them. But can we **afford** it?
  - Right now, in Germany we pay something like **0.10 €** for **1 kWh** of electrical energy delivered to our homes, and something like **1.50 €** for a liter of Gasoline, which amounts to about **10 kWh**. Solar electricity costs are much higher, it is said. If prices were to double and triple, we simply couldn't afford it and would be reduced to poverty.
- Well - yes and no! If prices were to double but consumption would be halved, we have no problem since we can maintain the present quality of life at half the energy consumption if we just try!
  - Of course, halving consumption will need a lot of money for investments, which we cannot *afford*. In particular, we have to build a lot of new and better buildings - heating-wise. But this would be a big boost to the construction business, which would create lots of jobs and lots money.... .

- National economy is just not that easy. Right now (end of **2007**) we are told by the experts that the world faces an unavoidable major financial crisis because the Americans overspend, meaning that they spend money they don't have but borrowed from the Chinese and others. The only way to avoid this crisis, we learn, is if the Americans keep overspending and go even more into debt. There must be reasons why the experts on this are almost always wrong in their predictions. It simply boils down to the fact that money, to a small part, consists of silly pictures printed on paper and to a larger part of bits stored in some computer memory.
- After the war, a lot of Germany was destroyed, nobody had money, and we certainly could no more afford to rebuild Germany then than we can afford to rebuild the energy infrastructure now. However, people then didn't know this and the economic miracle took place. Obviously, all we have to do, is to bomb a large percentage of our buildings (in particular the power plants) ourselves (the Americans are busy in this respect elsewhere, anyway) and rebuild them in energy-smart ways.

1st Conclusion: There is simply no real money problem except for the one in our heads.

Even better, the costs of solar power is coming down exactly on schedule. Costs of technology always come down with experience *and* with mass production, and this always follows an exponential decay curve in the beginning, called the **learning curve**.

- Here is the solar cell learning curve. It it continues like that, we will have "**grid parity**" in **7 - 12** years, i.e. the solar **kWh** costs exactly whatever you will pay for conventional (electrical) power.



Learning curve of solar cell energy costs

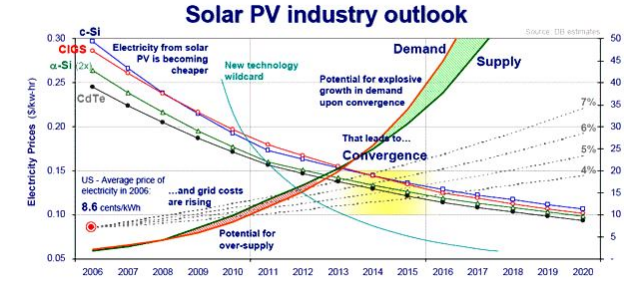
2nd Conclusion: Whatever solar energy money problems there might be, they are small (the problems, not the amount of money needed) and pale in comparison to the no-solar-energy problems looming in the background!

- The war in Iraq so far has run up a direct bill (money out of congress) of about **450 000 000 000 $** (about **$3 \cdot 10^{11}$ €**) and simply shows that large amounts of money are not the problem, when it comes to securing energy (the wrong way, in this case).

So What's Going to Happen?

How the hell should I know? But one thing is certain: if some people out there believe that there is serious money in solar energy, they will go into it. If that means saving the world as a side effect - so be it.

- Here is a scenario put together by a guy from "Deutsche Bank". It contains nothing new to Materials Scientists and Engineers who keep in touch of what is going on, but ... see above.



Scenario for the photo voltaic (***PV***) Industry according to Deutsche Bank research (Spring 2008)

- Even in the most pessimistic scenario, solar energy will simply be cheaper than what you pay to you utility company in **2015**. Then there is a potential for "explosive" demand, meaning that a decent amount of money will be involved - couple thousand billions a year or so.
  - Be sure to buy stock from the right companies at the right time!

Efficiency Limits

- There is a wide-spread feeling in the (published) public opinion that if on would spend a lot of money on solar cell **R&D**, or even better would have spend a lot of money already in the past, we would have much better solar cells now. "Better" in this context seems to be often confused with "higher efficiency".
  - After all, converting just a little more than **10 %** of the energy contained in the sun light into electrical energy is a pitifully small efficiency η; even old-fashioned coal fired power plants have efficiencies around **η = 30 %**.
  - So, give those tinkers more money, and the will (or already would have) come up with the **η = 30 %** solar cell!
- **No!** No matter how many tinkers slave away in their garages for all that time, they will not beat the iron-clad limit for efficiency that comes straight out of semiconductor theory. The first law of engineering science obtains:
  - *There is nothing more practical than a good theory*.
  - There are strict limits for efficiencies. You will *never* be able to make a **Si** solar cell with **η > 25 %**, so don't even try it. We will probably also never be able to make millions of *very cheap* **Si** solar cells with **η = 24 %**; but if you give me lots of money, I could try.
- While we have and will be struggling to increase the efficiency of present-day **Si** solar cells *somewhat* (**0.5 %**, lets say) every year (and thanks for the money!), the real problem is not in semiconductor physic.
  - The only game in town with respect to solar cells is: Make a lot *and* make 'em cheap!
  - It's not physics, it is Materials Science *and* Engineering!

Night

- Half of the time the sun isn't shining and solar cells are not producing any energy. How about that?
  - Now we have a real problem. While we can hope that somewhere in Europe the wind is always blowing, the same is certainly not true for sunshine. A Europe-wide power grid would not help in this case since it's nighti n all of Europe at about the same time.
  - An **AC** power grid actually never helps very much for that task, because the losses in transmission of power via high-voltage cables are too large after a couple of **1000 km** or so. We can thus neither transport "wind energy" from Sicily to Denmark at some reasonable efficiency via **AC** electricity, nor direct solar energy. However, with a high-voltage **DC** grid we have fewer losses and eneryg transport over larger distances (several **1.000 km**) now might be envisioned. Look for "Desertec" in the net to get some idea of what is in the making.
- This is a real and big problem - coming up as soon as solar energy has a sizeable chunk of the energy market. Right now, the fluctuations of the solar energy input into the grid are easily compensated by the other power providers. At night, some gas turbines somewhere run a bit more energetically, and that is all it takes.
  - What we clearly need is some way of storing energy. Unfortunately, there is presently no good solution for energy storage.
  - If you look at the energy density of various energy storage technologies, you realize that any mechanical storage by lifting masses ("$\boldsymbol{E = m \cdot g \cdot h}$") is extremely inefficient. You must lift your liter of gasoline to a height of about **360 km** to have the same potential energy you get as thermal energy by simply burning it.
  - This is not to say that ***pumped-storage hydroelectricity*** is useless, just that it takes large amounts of space to install sizeable capacities.

- The article in the link has a lot to say about the energy storage problem; the author (a Noble prize winner) is cautiously optimistic about this issue.
  - If and how the problem can be solved remains to be seen, the only save bet is that it will provide important and interesting work for scores of materials scientists.

Modules

- There really should be a substantial module here. Sorry, but I never got around to it.
  - Go at it yourself. Perhaps as a seminar topic?

Advanced

CELLO

Advanced

CELLO is short for Solar CELl LOcal characterization. It allows to determine any parameter of solar cell at any given "Point" (= pixels) of a solar cell. It is rather sophisticated. What you do is:

1. Hold the solar cell at some fixed condition with respect to global illumination and either voltage or current as first input.. You then get a defined voltage or current respectively, as an "output" that can be deduced straight from the global characteristics
2. Scan a focussed Laser beam across it, adding a second input. The added illumination at any pixel will "disturb" the solar cell a tiny little bit and the voltage ***V*** or current ***I*** output will be disturbed a little bit by **d*V*** or **d*I*** too. Measure how much the is for every pixel.
3. Use more than 1 Laser (up to 4) with different colors simultaneously. Modulate the intensity with several frequencies at once. Measure the phases of the output signals, too. (Do impedance spectroscopy, in other words)
4. You now get **d*V*'s** or **d*I*'s** sorted by color and frequency (do a FFT to the signals)
5. Produce a complete dynamical (amplitudes and phases) mathematical model for the solar cell. It contains all the desired parameters as unknowns.
6. With at least as much output data as unknowns, solve the equations for the desired parameters and display in some color scheme

- Not simple. And rather demanding wih respect to signal-to-noise, scanning speed, and software. But immensely powerful
- Link to a detailed description of the early system
  Link to a powerpoint presentaion of full system

Series and Parallel Connection

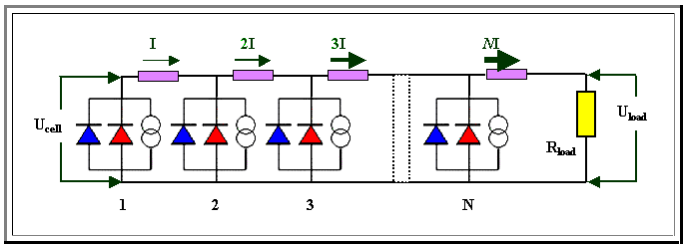
Advanced

First Basic Question

We take a number ***N*** of solar cells; assuminng that all oft them are perfectly identical, to keep things simple at first. We also assume that there are no shunts in those solar cells, but that they have some series resistance R_{SE}.

- The first basic question now is : How can you deliver maximum power into some load resistor R_{load} - by switching the ***N*** solar cells in series or parallel?

Let's look at switching in parallel first.



- Every individual solar cell delivers a current ***I***, theses currents add up, and the current flowing through the load resistor R_{load} of the solar cell No. ***k*** under short circuit conditions is $I_{load}(k) = k \cdot I$
- The corresponding total voltage drop in all ***N*** series resistor each with resistivity R_{SE} is U_{SE} given by

$$U_{SE} = I \cdot (R_{SE} + 2R_{SE} + 3R_{SE} + ... + NR_{SE})$$

$$= I \cdot R_{SE} \cdot \tfrac{1}{2}N \cdot N \qquad \text{(remember young Gauss?)}$$

$$= \frac{I \cdot R_{SE} \cdot N^2}{2}$$

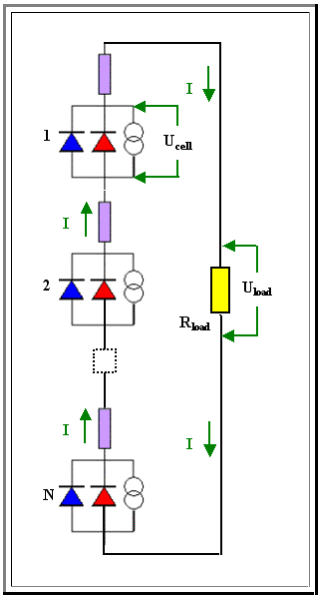
The product of maximum output voltage times current give some fictive power $P_{parallel}$ akin to just multiplying I_{SC} and U_{OC} of the "ideal" solar cell. The actual power $P_{load,\ p}$deposited into the load resistor depends, of course, on the value R_{load}; it will be maximal for some specific value of R_{load} but always smaller than $P_{parallel}$.

$$P_{parallel} = N \cdot I \cdot (U_{cell} - \tfrac{1}{2} N^2 \cdot I \cdot R_{SE})$$

$$= N \cdot I \cdot U_{cell} - \tfrac{1}{2}N^3 \cdot I^2 \cdot R_{SE}$$

The total losses due to series resistors thus are at least $P_{loss,\ p} = \tfrac{1}{2}N^3 \cdot I^2 \cdot R_{SE}$

Now let's look at a series connection.



- The current flowing through the load resistor is ***I***
- The voltage U_{load} is ***N*** times the cell voltage minus the voltage drop in the internal resistance, i.e. minus $N \cdot R_{SE} \cdot I$; we have $U_{load} = N \cdot (U_{cell} - I \cdot R_{Se})$
- The power P_{series} deposited in the load resistor is such

$$P_{series} = I \cdot N\,(U_{cell} - I \cdot R_{SE})$$

$$= I \cdot N \cdot U_{cell} - I^2 \cdot N \cdot R_{SE}$$

- Your total losses due to series resistors thus are $P_{loss,\ s} = N \cdot I^2 \cdot R_{SE}$

The relation between the losses of a solar cell array in in parallel or in series thus are

$$\frac{P_{loss,\,p}}{P_{loss,\,s}} = \frac{N^3 \cdot I^2 \cdot R_{SE}}{2 \cdot N \cdot I^2 \cdot R_{SE}} = \frac{N^2}{2}$$

- In words: You *must* switch your solar cells in series! You loose far too much power if you have them in parallel

Second Basic Question

We take a number ***N*** of solar cells; assuminng that all oft them *except for one* are perfectly identical, to keep things not too simple anymore.

- We consider two extrem cases for both variants, series and parallel connection of ***N* – 1** identical good cells and one bad one:

  1. The bad cell has a very large series resistance and no noticeable shunt resistance
  2. The bad cell has a regular series resistance and a very small shunt resistance.

Let's see what will happen in the parallel world. All we have to do is to image that the series reistor of cell No. ***N*** is very large, or that the whole cell is just a small resistor in parallel to the load. What will happen?

- In the *first* case you simply loose most of your voltage at the high series resistance of the defective cell. The power output will go down dramatically, or simply proportionally to the added series resistance. It will also matter which one of the ***N*** cells is the defective one.
- In the *second* case you just add another parasitic load resistor to the intended load, funneling off some of the current. Again, the power output will go down dramatically, proportional to $\mathbf{1}/R_{SH}$.

Let's see what will happen in the serial world.

- In the *first* case you simply loose most of your voltage at the high series resistance of the defective cell, and it doesn't matter which cell it is. You can consider the surplus resistance as a parasitic addition to the load resistor. The power output will go down dramatically, again proportionally to the added series resistance.
- In the *second* case you just loose the power not generated by the defective cell. The current of all the other cell passes through the (small) shunt resistance without much of a voltage drop.

Again, we have compelling reasons to go with series connections. We also see that we must watch out for high series resistance - it does more damage on the whole than the occasional shunt.

The most important insight, however, is that one bad cell in an array of, let's say **100** cells, may bring down the total efficieny of the module - that's essentially what we are discussing - quite disproportionally, i.e. not by just **1 %**, but by far more.

Third Basic Question

We take a number ***N*** of solar cells; assuminng that *all* oft them are perfectly identical, but that *one* of them is completely shaded because a big leaf or a splash ot birds shit comßpletely covers it surface.

- In other words: We take out the constant current generator of one cell.
  What happens?

Ans so on. By now you get it.

Moduel technology is tricky and needs attantion!

Basic *IU*-Characteristics of Solar Cells

Advanced

Foreword

- This module is at once the solution of an exercise and a detailed discussion of the ***UI***-characteristics of a theoretical and practical *ideal* solar cell.
  - If you are not already familiar with the diode equation and what exponential terms can do, you will profit very much by going through it in detail.
  - For easy readability parts of the data given in the exercise will be repeated her in a somewhat modified form.

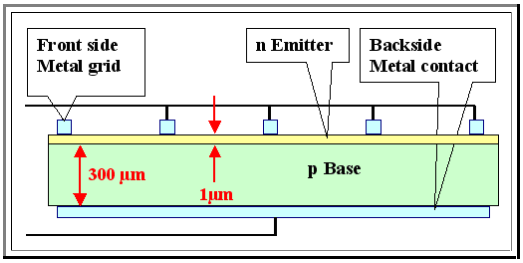
Starting Point

- The diode equation with generation and recombination in the space charge region part describes a **pn**-junction made from a semiconductor like **Si** that has well defined properties, for example the doping concentration and the diffusion length / life time. It thus also describes an "ideal" solar cell. We will now try to see what we can do with this equation with respect to solar cells.
  - We use the abbreviations j_1 and j_2 for the current densities in the brackets and $d(U)$ as the "abbreviation " for the equation shown below that gives us the width of a space charge region in a **pn**-junction.
  - For the current density $j(U)$ and the (less important) space charge region width $d(U)$ we have

$$j(U) = \underbrace{\left(\frac{e \cdot L \cdot n_i^2}{\tau \cdot N_A} + \frac{e \cdot L \cdot n_i^2}{\tau \cdot N_D}\right)}_{j_1} \cdot \left(\exp\frac{eU}{kT} - 1\right) + \underbrace{\left(\frac{e \cdot n_i \cdot d(U)}{\tau}\right)}_{j_2} \cdot \left(\exp\frac{eU}{2kT} - 1\right) - j_{Ph}$$

$$d(U) = \left(\frac{2 \cdot \epsilon_{Si} \cdot \epsilon_0 \cdot (\Delta E_F - e \cdot U_{ex})}{e^2 \cdot N_D}\right)^{1/2}$$

- In evaluating this equations for solar cells as schematically given in the drawing, we must first find numerical values for important parameters (all others have their usual meaning and values). We have:



  - ***L*** = **diffusion length** = $(D\tau)^{1/2}$ average distance a minority carrier travels between its birth by a generation event (mostly caused by light in a "working" solar cell) and its death by recombination. A suitable value for good bulk **Si** is $L = 100\ \mu m$.
  - ***D*** is the diffusion coefficient and τ the (minority carrier) life time. A good enough value for the life time going with a diffusion length of **100 µm** is $\tau = 1$ ms.
  - n_i is the intrinsic carrier concentration. It increases exponentially with temperature ***T***. A good value for **Si** at room temperature (**RT**) is $n_i(RT) = 10^{10}\ cm^{-3}$.
  - N_A and N_D are the acceptor and donor concentrations in the **p**-part called **base**, the usually several **100 µm** thick part of a bulk **Si** solar cell, and the **n**- part, called **emitter**, the thin "layer" on top of the solar cell, respectively. The base is lightly doped (otherwise the diffusion length suffers) whereas the emitter is heavily doped (good conductivity is important). $N_A = 10^{16}\ cm^{-3}$ and $N_D = 10^{19}\ cm^{-3}$ are good round numbers for the purpose here.
  - We could also calculate the width of the space charge region with the equation given but take it as $d(U) = 1\ \mu m$ as a first simple approximation.
  - Now we consider a *real* but still "ideal" solar cell under "standard" illumination. This gives us the following (simplified) second set of numbers:
    - Area of the **Si** bulk solar cell = $100\ cm^2$. It's actually more like **200** cm^2 in **2008** but let's stay with easy

numbers.

- Photo current density $j_{Ph} = 30\ mA/cm^2$ at "***AM 1.5***" the condition of maximum solar intensity on earth ("High noon in the tropics")
- The photo current here is thus $j_{Ph} = 3\ A$.

The *first question* was:

1a: Using only the first term in the bracket for j_1 as a sufficient approximation, give an equation for the relation of j_2/j_1 and some numbers for these current densities.

1b: Does the result imply that you can neglect one of the j_i terms in the equation above in the *forward* direction? How about the *reverse* direction?

For the numerical values of j_1 and j_2 we obtain

$$j_1 = \frac{e \cdot L \cdot n_i^{\,2}}{\tau \cdot N_{dot}} = \frac{1.6 \cdot 10^{-19} \cdot 10^{-2} \cdot 10^{20}}{10^{-3} \cdot 10^{16}} \frac{C}{s \cdot cm^2} = 1.6 \cdot 10^{-14}\ A/cm^2$$

$$j_2 = \frac{e \cdot n_i \cdot d(U)}{\tau} = \frac{1.6 \cdot 10^{-19} \cdot 10^{10} \cdot 10^{-4}}{10^{-3}} \frac{C}{s \cdot cm^2} = 1.6 \cdot 10^{-10}\ A/cm^2$$

For the relation of j_1 to j_2 (and using N_{Dot} instead of $N_{A,\,D}$) we obtain

$$\frac{j_1}{j_2} = \frac{\dfrac{e \cdot (n_i)^2 \cdot L}{\tau \cdot N_{Dot}}}{\dfrac{e \cdot n_i \cdot d(U)}{\tau}} = \frac{n_i \cdot L}{N_{Dot} \cdot d(U)}$$

Inserting the numbers from above yields once more

$$\frac{j_1}{j_2} = \frac{10^{10} \cdot 10^{-2}}{10^{16} \cdot 10^{-4}} = 10^{-4}$$

Now we can address the second part of the first question. Since j_2 is so much larger than j_1, can we simply neglect the j_i terms in the diode equation? We should know the answers from before:

For biasing in the reverse direction, we have $j_{rev} \approx j_1 + j_2 \approx j_2$ and we can indeed neglect j_1.

For the forward direction - which is the one of interest to us - we have approximately

$$j_{for} \approx j_1 \cdot \exp\frac{eU}{kT} + j_2 \cdot \exp\frac{eU}{2 \cdot kT}$$

No, we cannot neglect the j_1 term "just so", we also have to consider what the exponential terms will do. That will become very clear as soon as we look at the third part of the question.

The prelude to the second question was: If we now *measure* the actual ***UI*** characteristics of a good *real* solar cell and fit the curve obtained to our equation from above, we find values for the current densities j_1 and j_2 like

- $j_1 = 10^{-9}$ A/cm^2.
- $j_2 = 10^{-7}$ A/cm^2

The *second question* was:

2a: Do the measured values of j_1 and j_2 and their relation meet your expectations based on your results from *question 1*?

2b: If not, what could be reasons for the discrepancy?

Let's compare what we have in a table:

	Calculated	Measured
j_1	$1.6 \cdot 10^{-14}$ A/cm^2	10^{-9} A/cm^2
j_2	$1.6 \cdot 10^{-10}$ A/cm^2	10^{-7} A/cm^2
$\frac{j_1}{j_2}$	10^4	10^2

- We neglected the second term for j_1, which we will now call j_1^E for the time being; i.e. the reverse current flowing from the heavily doped thin **n**-emitter into the lightly doped **p**-base. If we would naively calculate j^E_1, we would get an even smaller value than what we already have for j_1^B because the doping concentration N_D of the emitter is larger than N_A of the base and appears in the *denominator* of the equation for j_1^E.
- However, we would commit a grave error in doing this because the diode "master" equation from above is only valid for one-dimensional junctions in "infinitely" long semiconductors, meaning that the semiconductor must extend at least a few diffusion lengths in both directions as seen from the junction. This is clearly not the case here.
- More advanced theory teaches us that in the case of "thin" semiconductors we have to replace the diffusion length L by the thickness d of the layer.. This makes sense because the diffusion length came into the equation as the dimension over which carriers are collected that could make it across the junction.
- If we use this insight, however, j_1^E gets even smaller because L is found in the *nominator* of the equation.

A first but *wrong* conclusion could be the discrepancy between theory and experiment cannot come from the emitter part of the reverse current.

- However, we forgot the life time τ, which we find in the *denominator* of the equation, and now we must take into account that heavy doping simply "kills" the life time, i.e. makes it very small. The diffusion length L gets smaller, too, but the combined effect is that $L/\tau \propto \tau^{½}$ so heavy doping always increases the j_1 part coming from the heavily doped region, and this increase can be substantial

As a first insight we note that a heavily doped thin emitter can indeed lead to a substantial increase in j_1. But there are more reasons for this.

- At the most elementary level of deriving j_1 we simply equated it with diffusion length L times the generation rate G; and G was equal to the recombination rate $R = n_{min}/\tau$. An increased j_1 thus demands for an increase in generation -we simply need more charges to have larger currents.
- In not-so-perfect **Si** we might have generation of carriers at grain boundaries or at the huge surfaces in excess of what just thermal generation can produce in a perfect lattice. To be sure, the recombination rate R must still be equal to G in equilibrium, but j_1 will go up with increasing generation anyway.

We see that there are several reason why we have the discrepancy. We simply must accept that the experimental j_1 and j_2 values are essentially **fitting parameters** of something called "solar cell" that do not fall within the range of a simple theory but still allow to describe the solar cell by the "simple" ideal theory if one accepts these empirical "fitting parameters" instead of the theoretical constructs in the equation.

Now to the *third question*: Given the measured j_i values from above and the j_{Ph} value given, we now can consider the short circuit current I_{SC} and the open circuit voltage U_{OC}

Question 3:

- **3a:** What do you get for I_{SC}? Does it depend on j_i and j_2? If not, what determines its value?
- **3b:** What can you say about the open circuit voltage U_{OC}?

The *first* part is easy: I_{SC} is what we get for $U = 0$ and that is simply $-j_{Ph}$.

- We have used $-j_{Ph}$ as a constant in the master equation, it thus does not depend on the values of j_1 or j_2 or on the variables determining their numerical values.
- This is not generally correct, of course. For example, if the diffusion length L increases, more carriers generated by light deep in the volume of the solar cell can reach the junction and $|j_{Ph}|$ should increase with L.
- However, we have assumed *good* solar cells along, and this means that practically all carriers generated by light end up in the photo current. This simply implies that for diffusion lengths good enough not much can be gained anymore by increasing L. In other word, if the longest distance between a generation event and the next contact is **20 µm**, it just doesn't matter much if your carriers could go **200 µm** or **500 µm**.

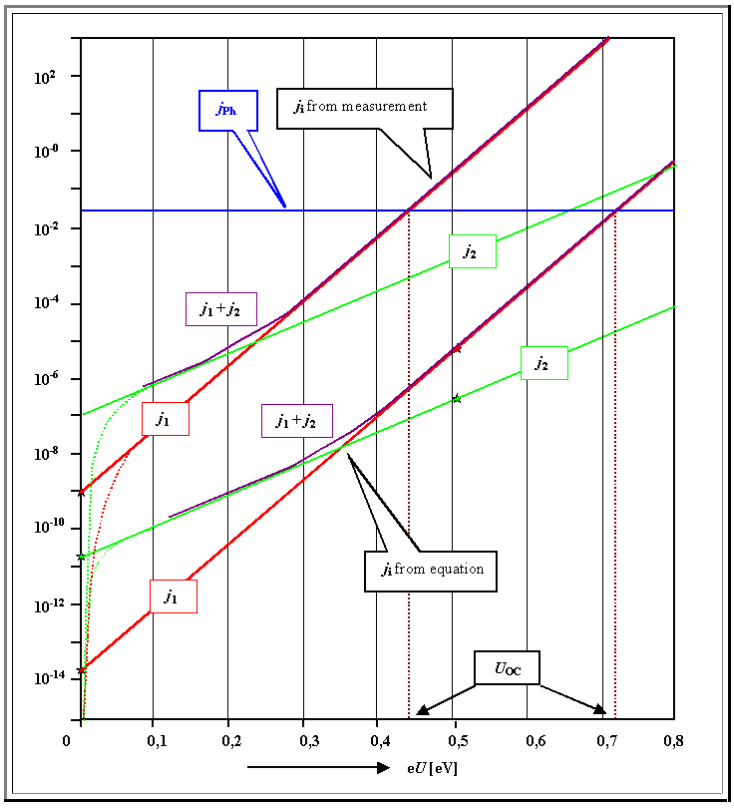
The *second* part is tough: If we try to solve the master equation for U_{OC}, i.e. setting $j = 0$, we realize that it can't be done.

- There is no analytical expression for U_{OC} that we can gain from the master equation. Short of going numerical, we need to consider other ways of gaining some insight, including approximations.
- One way is to go for a graphical solution of the problem. We actually have done that already, but probably not recognized what we can learn for solar cells from this. All we have to do is to draw the master equation in a **log *j* - e*U*** plot. This is actually a very good exercise and you should do - at least look at the solution - and learn how it's done.

Exercise 8_1_6

log *j* - e*U*

- The result looks like this:



From looking at the graph we can learn a lot of things

1. The "**-1**" term in the master equation is only noticeable for currents **<≈ j_2**.

- We can safely neglect it for solar cells as long as we have a photo currents in just the **µA/ cm²** region, i.e **1 000** times smaller than the maximum photo current density on earth.

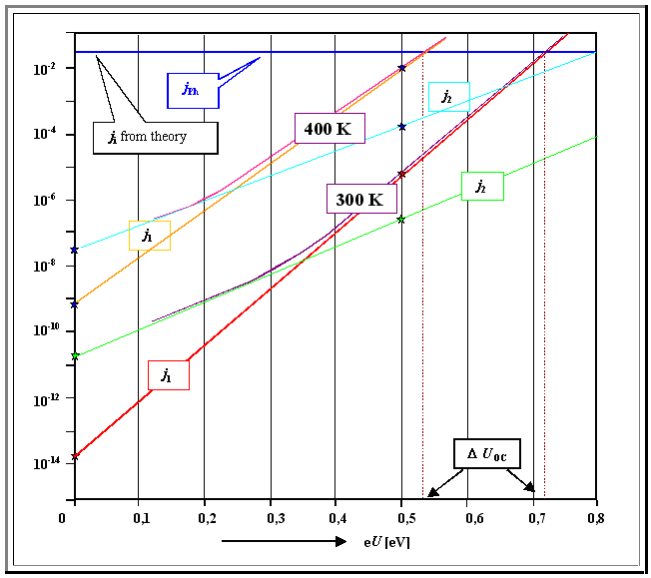
2. The open circuit voltage **U_{OC}** depends *only* on **j_1** for reasonable photo currents. Even so **j_2** is much larger, the exponential term going with **j_1** always "wins" for voltages above **0.3 V - 0.4 V**.

- If we neglect the **j_2** term in the master equation, we can solve it for **U_{OC}** and obtain for **eU_{OC}** as measured in **eV**:

$$eU_{OC} = kT \cdot \ln \frac{j_{Ph} - j_1}{j_1} \approx kT \cdot \ln \frac{j_{Ph}}{j_1}$$

- This means that **j_1** is the decisive term for **U_{OC}**, one of the prime properties of a solar cell.

Now we look at the temperature dependence of **U_{OC}**. From the solution of the exercise we take only one curve here:



- Increasing the temperature has the following effects:
  1. The slope of both exponentials decreases. This would lead to a higher **U_{OC}**.
  2. The **j_i** increase exponentially because their defining equations contain the intrinsic carrier density **n_i**, which increases exponentially with **T**.
  3. The total effect is a decrease of **U_{OC}** with **T**

We can see that also in the equation for **U_{OC}**. Inserting **$j_1 = c_1 \cdot n_i^2 = j_1' \cdot \exp(-E_g/kT)$** yields

$$eU_{OC} = E_g + kT \cdot \ln \frac{j_{Ph}}{j_1'} = E_g - kT \cdot \ln \frac{j_1'}{j_{Ph}}$$

- It first looks like we *add* something to **eU_{OC}** with increasing **T**; increasing **eU_{OC}**. However, it is important to realize that **$j_1' >> j_{Ph}$** - even if that is counterintuitive - and **$\ln(j_{Ph}/ j_1')$** thus is a negative number! That the open circuit voltage indeed decreases is better seen in the final formulation where we *subtract* a positive number from **E_g**.

If you made it to this point, you learned quite a bit about basic solar cell characteristics. You also learned another thing:

Don't rely on "feeling" if exponentials are involved!

Joules, Watts and Kilowatthours

When we talk about **energy**, we often mean really **power** = energy / time. Let's first see what the basic units are, how they can be expressed and how they are connected. The links take you to the modules of the Hyperscript "Einführung in die Materialwissenschaft I", where more details can be found

Units for Energy, Power, and Electricity					
Quantity	**Name**	**Symbol**	**Units**		**Relations**
			In basic units	In secondary units	
Energy	Joule	J	$m^2\ kg\ s^{-2}$	N · m	1 kWh = 3.6 · 10^6 J
				W · s	1 kcal = 4,1868 kJ = 1,163 kWh
					1 to ***SKE*** = 8 140 kWh
Power	Watt	W	J/s	$m^2\ kg\ s^{-3}$	1 kW = 3.15 · 10^{10} J/yr 1 kW = 3.6 · 10^6 J/hr 1 kW = 10^6 J/s
				V · A	
Electric potential, voltage	Volt	V	W/A	$m^2\ kg\ s^{-3}\ A^{-1}$	
Time	Second	s	s	s	1 yr = 1 a = 3.15 · 10^7 s 1 d = 86 400 s

There are a few points to note:

- The time unit is the second or "**s**". Then we have the minute = "**m**", the hour = "**h**" and the day = "**d**" as officially (allowed) **SI** units. There is no official abbreviation for year; here we use either "**yr**" or "**a**" (for annum).
- Terms like **energy consumption** do not mean (of course!) that energy is really *consumed*, but that it is just transformed (to heat in the end). There is such a thing as the energy conservation law, after all!
- In case of doubt, when just the term "energy" is used, what is meant is often *energy per year* - formally this is *power*. However, it is not always sensible to refer to formal "power" as power: If your gas tank contains ***x*** liters of gasoline, it contains some equivalent of pure *energy*. If your *consumption* of gasoline is **2 000 l** *per year*, you talk about the energy *per year* that you have consumed (about **20 000 kWh**)! You could express that as an *average* consumption of **2 000/3,15 · 10^7 l/s = 6,5 · 10^{-5} l/s** in **SI** units - but that doesn't make much sense.
- There is **primary energy** and **secondary energy**. If your personal consumption of electrical energy is, for example, **2 000 kWh** (*per year*; a good number for an average German), then this is what your power company charges your for. In this case you are charged for your use of *secondary* (electrical) energy. The *primary* energy needed for producing that amount of *secondary* energy (per year!), is roughly *three times* larger, because the ***efficiency*** of our fuel-burning power plants is around **35 %**. This means that in terms of the energy contained in coal, oil or gas - whatever was burnt to produce your personal **2 000 kWh** (*per year*), you actually consumed around **6 000 kWh** (*per year*).
- Your **direct** consumption of these secondary **2 000 kWh**, that you paid for directly, is not all the electrical energy that you personally consumed (*per year*; last time!). You also consumed electrical energy in the form of light at the place you work, products you buy (that somebody made somewhere, using electrical energy), and so on. Your *indirect* (secondary) electrical energy consumption is probably much higher than your *direct* consumption - about a factor of **2,5**.
- Besides electrical energy, *you*, personally, consumed energy in the form of heat (including hot water for showers), gasoline (for driving your car and for riding a bus and so on) eating and drinking (somebody had to ride a tractor to harvest the hops, transport your beer, and so on); *your* grand total of primary energy consumption is around **50 000 kWh** (per year) - if you are a German or **EU** citizen. If you are an American, it's **2,5** times more; it's far less if you are, e.g., from Ghana or Peru

We have now defined the *quality* of energy and power; next let's get a feeling for the **quantities** involved.

We will not care about *precise* numbers. If humankind right now consumes **13 TW**, **12,8 TW** or already **14,783 TW**, is just as irrelevant in this context, as the question if the **50 W** of power produced by the slave mentioned below includes his sleeping time, or only the time he actually works.

Typical examples for energy and power		
Example		**Formula**
Energy *E*	Potential energy E_{pot} if *I* climb ***h* = 1.000 m** (For *you* it might be a little less)	$E_{pot} = m \cdot g \cdot h$ $= 100\ kg \cdot 9{,}81\ m/s^2 \cdot 1000\ m$ $= 9{,}81 \cdot 10^5\ J$ $= 0{,}273\ kWh$
1 kWh corresponds to:		367 t lifted to 1 m 9,5 l water à 10°C bring to boil 2 km - 10 km car driving
1 kWh is **stored** in:		Large (85 Ah) truck batterie 0,1 l gasoline / Diesel 0,25 kg dry wood 0,12 m^3 natural gas 0,28 m^3 H_2 7,3 t H_2O in a reservoir with 50 m height difference
Power *P*	What *I* can sustain for ≈ **1 hr** on a bicycle	**175 W**
	Hard working **slave** on average	**50 W**
Power consumed by a **light**.		≈ **50 W**
Power consumed by a **toaster**.		≈ **1 kW**
Energy consumed by a toaster in **3** minutes		**50 Whr**
Time your **slave** has to work to power your toaster for **3 min**.		**1 hr**
Power consumed by your **car**		≈ **20 kW - 100 kW**
Power consumed by a (small) **jet engine**		≈ **1 000 kW**
Output of "standard" **power plant**.		≈ **1 000 000 kW = 1 GW**
Global secondary *electrical* power demand.		**1 000 GW = 10^{12} W = 1 TW** 1 "**Terawatt**" = 1 trillion Watt (USA) = 1 Billion Watt (Europe)
Total *primary* global **power** produced (**2001**) Oil: 4,66 TW Gas 2,89 TW Coal 2,98 Nuclear 0,92 TW Rest: 1,81 TW		**13 TW** USA (303 Mill. people) **3,2 TW** Rest (6.331 Mill. people) **9,8 TW**
Total **energy** produced (= consumed) *per year* in **2001** (**EJ = Exa Joule = 10^{18} J**)		**13 TW · 3,15 · 10^7 s = 4 · 10^{20} J/yr = 400 EJ/yr**
Total **energy** produced (= consumed) **per year *and capita*** USA: **350 GJ/yr = 97 200 kWh/yr** Australia: **240 GJ/yr = 66 600 kWh/yr** Japan; ***EU***; S.-Korea: **150 GJ/yr = 41 700 kWh/yr** Brazil: **40 GJ/yr = 11 100 kWh/yr** China: **30 GJ/yr = 8 330 kWh/yr** India: **15 GJ/yr = 4 170 kWh/yr**		**60 GJ/yr · capita = 16 660 kWh/yr · capita**
Time it takes if we built **1** renewable standard **1 GW** power plant per day on Terra to replace all "carbon emission" power plants		**11 TW / 1 GW = 11 000 days** **= 30 years**

Total **energy** needed (*per year*) in **2050**	Who knows? Extrapolation: All like **EU**:	**33 TW**
Additional power plant building rate needed to account for the expected increase.	**22 TW / 1 GW · 50 yr = 1.2 power plants / day**	
Present (**2007**) rate of **1 GW** solar cell power plants built per day	**$3 \cdot 10^{-4}$ power plants / day**	
Conclusion:	*You*, the young student, are in trouble! (*I*, the old professor, will get by)	

- What you find in this table you may call the "The Terawatt Challenge", and it has been called this way. Read more about it in the article Future Global Energy Prosperity: The Terawatt Challenge of Nobel prize winner **Richard Smalley** published in the **MRS** Bulletin **30, 2005**, and in the article "Powering the Planet" of **Nathan S. Lewis** in the **MRS** Bulletin **32, 2007**

- This looks like somebody should do something. Right. This somebody is *you* - and *you*, over there in the **USA**, too!

- While the shear magnitude of the numbers may induce a feeling of hopelessness, fixing the problem within the next **50** years or so is *not* impossible. Let's look at the bright side:
  - *Building power plants:* Build one big **1 GW** power plant *per day*?? Impossible! Wrong - that's exactly what we have been doing for many years! The energy - time curve in the past had the same slope we used for the extrapolation, roughly **1 GW/day**. So it can be done, and if we really want to do it (and pay the higher price!) we can do it with renewable energy plants, too, in the not-so-distant future.
  - *Reducing energy consumption:* You (and I) don't really need **50 000 kWh/year** or even more to experience a high quality of life. The average quality of life in the **USA** is certainly not higher than in the **EU**, but energy consumption is more than **2** times larger. In **1970** *I*, personally, had a pretty good life, too - but consumed far less energy than I do now. So let's reduce energy consumption without compromising the quality of life - it can be done!
  - *Renewable energy is too expensive:* Bullshit! It is nominally more expensive than *my*, the old Professor's, present **kWh** price - yes! But I'm not paying the full bill; it does not include, for example, the cost of climate change or the destruction of the environment, the costs of the wars for resources etc. -Either *you*, the young student, will have to pay for this later, or all of us pay somewhat more soon.
  - *Slave labor:* A slave could give you **50 W · 24 · 365 hr = 438 kWh/year** if you worked him really hard (your female slave may give you things not always measured in **W**). You European thus command **114** slaves working all the time for you; even more if you let them go to the bath room on occasion. You have this much power at your fingertips *only* because you have access to technology. Think a moment about this! It is the *only* reason why *you*, personally, are doing so well in modern society! In good old-fashioned society, only one out of **114** or more could command that much *power*, so chances are **> 114 : 1** that *he* was your Lord and *you* one of his slaves / serfs / indebted servants, or whatever you like to call it.
  - *Exponential growth:* Nobody has a feeling for exponential growth - you must sit down and calculate. OK. Here is the exercise:
    (Look at least at the solution!)

Exercise b8_1_1

Exponential growth

Exercise Basic 8.1-1

Exponential Growth

The output of the solar cell industry in **2006 - 2008** grew by **40 %** per year. Let's assume that all solar cells installed in **2007** produced a total energy of **0.1 GW /year**.

Calculate (and plot) the installed power as a function of time up to **2050** for growth rates of **20 %**, **30 %**, **40 %**, and **50 %**

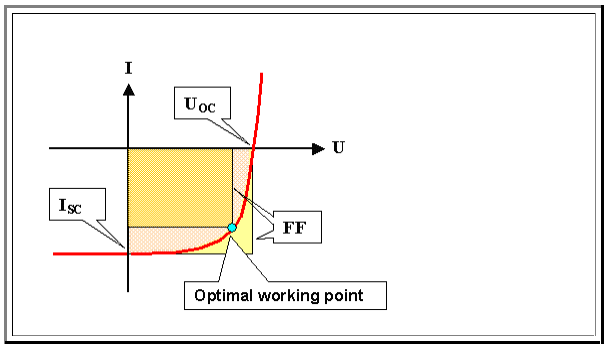
- What is the proper equation?
- What follows from the results with respect to the world-wide power scenario as described in the link??
- Plot the demand for **Si**, assuming that a standard (**1000 x 1000 x 0.1**) **mm^3 Si** solar cell generates **10 W** on average. Will there be enough **Si**? How do the amounts of **Si** needed compare to other essential raw materials?

Solution

Exercise 8.1-2

Optimal Working Point of Solar Cells

The figure shows the ***IV***-characteristics of a real solar cell.



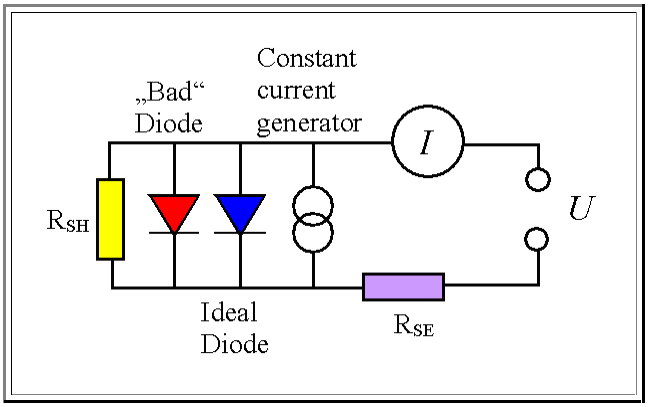
- Derive the optimal working point by simply constructing the ***I*** · ***U*** = Power curve graphically
- Discuss the result with respect to solar cell parameters
- What kind of load resistor would you need for this solar cell? What are the implications ?

Link to the solution

Exercise 8.1-3

IV Characteristics of Real Solar Cells

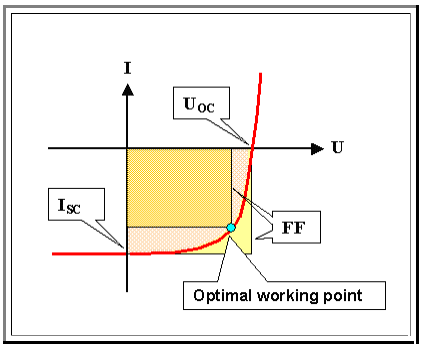
Lets consider a solar cell as described in the backbone, with a built in **series resistance** R_{SE} and a **shunt resistance** R_{SH}



- We have the equivalent circuit diagram as shown.
- The shunt resistance takes into account that the huge area of the **pn**-junction of a solar cell might have weak points (locally, e.g. at the edge) which short-circuits the junction somewhat. These defects are summarily described by a *shunt resistor*.
- The constant current source mimics the current generated in the junction by light. It simply defines a current value I_{Ph} (not to be mixed up with the terminal current I) that is given by the light and added (with a negative sign) to the junction current, i.e. $I_{junct} = I_{diode}(U) - I_{Ph}$. The photo current I_{Ph} thus simply moves the total characteristics of the diode downwards on the current scale.

Take the following schematic curve of the ***I-U***-characteristics as a reference and for the definition of the following terms:

- The **fill factor** is the relation between the area of the large yellow rectangle to the more orange area centered at the **optimal working point.**



Discuss qualitatively the influence of the *two resistors* (and, as a more minor point, the *idealiy factor* ***n***) on the ***IV*** characteristics with particular respect to:

- The **open-circuit voltage** U_{OC}.
- The **short-circuit current** I_{SC}.
- The **reverse dark current** if the solar cell is bisased in the reverse direction.
- The **fill factor** ***FF*** (the degree of "rectangularism" of the characteristics).
- The **efficiency** η which is proportional to U_{OC}, I_{SC}, and ***FF***, i.e.

$$\eta = \text{const} \cdot U_{OC} \cdot I_{SC} \cdot FF$$

Link to the solution

Exercise 8.1-1

Quick Questions to

8.1 Solar Cells - General Concerns

Here are some quick questions

- Give some rough numbers (with some reasoning whereever applicable), always per $\mathbf{m^2}$, for
  - Maximum solar power.
  - Maximum and practical efficiency of "standard" **Si** solar cells.
  - Average power for "standard" **Si** solar cell.
  - Average energy harvest of "standard" **Si** solar cells per year.
- Compare indirect and direct semiconductors with respect to light absorption at the "band edge", i.e. for light energies around bandgap energy. What follows for solar cells?
- What is your first priority with respect to the coupling of light and semiconductor when you want to make a solar cell with a good efficiency?
- Draw the current density (***j***) - voltage (***U***) characteristics of a **pn**-junction in the dark and under illumination in the interesting part of the ***j-U*** plot. Point out the important points of the illuminated diagram and give some approximate numbers for typical intense sun light.
- The typical ***j-U*** equation for a **Si pn**-junction is

$$j = j_1 \cdot \left(\exp \frac{eU}{kT} - 1 \right) + j_2 \left(\exp \frac{eU}{2kT} - 1 \right) - j_{Ph}$$

  Discuss the origin of the ***j*** terms. Compare (qualitatively) the magnitude of j_1 and j_2. What kind of properties of **Si** influence the value of j_{Ph}?
- Why is "dirty" **Si** not good for solar cells? *Hint*: Follow the fate of a photon-generated carrier.
- Draw the ***j-U*** curve of an illuminated decent solar cell. Draw (qualitatively) the power curve into this diagram. Discuss the curve shortly with respect to real power applications
- Your electrical energy bill shows that you, personally, consumed **2 000 kWh** electrical energy per year in your home. How man square meters of solar cells (roughly) would you need on your roof to supply this much energy?
- Give the eqivalent circuit diagram of a realistic **Si** solar cell. Discuss the components with the aid of schematic **IV**-characteristics.
- Define the fill factor of a solar cell and discuss its dependence on solar cell parameters.

Exercise 8.1-5

IV Characteristics of Real Solar Cells

We take the diode equation, including generation and recombination in the space charge region part, as given. We will now try to see what we can do with this equation with respect to solar cells. We have

$$j = \underbrace{\left(\frac{e \cdot L \cdot n_i^2}{\tau \cdot N_A} + \frac{e \cdot L \cdot n_i^2}{\tau \cdot N_D} \right)}_{(\ j_1\ )} \cdot \left(\exp \frac{eU}{kT} - 1 \right) + \underbrace{\left(\frac{e \cdot n_i \cdot d(U)}{\tau} \right)}_{(\ j_2\ )} \cdot \left(\exp \frac{eU}{2kT} - 1 \right) - j_{Ph}$$

- We must first look at the important parameters (all others have their usual meaning) and find numerical values:
  - L = **diffusion length** = $(D\tau)^{1/2}$ average distance a minority carrier travels between its birth by a generation event (mostly by light in a "working" solar cell) and its death by recombination. A good values for bulk Si that we take for simple calculation is $L = 100\ \mu m$
  - D is the diffusion coefficient and τ the (minority carrier) life time. A good enough value for the life time going with a diffusion length of **100 µm** is $\tau = 1\ ms$
  - n_i is the intrinsic carrier concentration. It increases exponentially with temperature T. A good values for **Si** at room temperature (**RT**) is $n_i(RT) = 10^{10}\ cm^{-3}$.
  - N_A and N_D are the acceptor and donor concentrations in the **p**-part (called **base**; the usually several **100 µm** thick part of a bulk **Si** solar cell) and the **n**- part (called **emitter**, the thin "layer" on top) of the solar cell. The base is lightly doped (otherwise the diffusion length suffers) whereas the emitter is heavily doped (good conductivity is important). $N_A = 10^{16}\ cm^{-3}$ and $N_D = 10^{18}\ cm^{-3}$ are good round numbers for the purpose here.
  - The width of the space charge region we take as $d(U) = 1\ \mu m$
- Now we consider a *real* good solar cell under "standard" illumination. This gives us the following (simplified) second set of numbers:
  - Area of the **Si** bulk solar cell = **100 cm²**. It's actually more like **200 cm²** in **2008** but let's stay with easy numbers.
  - Photo current density $j_{Ph} = 30\ mA/cm^2$ for a very good solar cell, less for a not-so-good one.
  - The photo current here is thus $j_{Ph} = 3\ A$.

Question 1:

- **1a:** Using only the first term in the bracket for j_1 as a sufficient approximation, give an equation for the relation of j_2/j_1 and some numbers for these current densities
- **1b:** Does the result imply that you can neglect one of the j_i terms in the equation above in the *forward* direction? How about the *reverse* direction?

If we now *measure* the actual ***UI*** characteristics of a good *real* solar cell and fit the curve obtained to our equation from above, we find values for the current densities j_1 and j_2 like

- $j_1 = 10^{-9}\ A/cm^2$.
- $j_2 = 10^{-7}\ A/cm^2$

Question 2:

- **2a:** Do these values and their relation meet your expectations based on your results from **question 1**?
- **2b:** If not, what could be reasons for the discrepancy?

Given the measured j_i values from above and the j_{Ph} value given, we now can consider the short circuit current I_{SC} and the open circuit voltage U_{OC}

Question 3:

- **3a:** What do you get for I_{SC}? Does it depend on j_1 and j_2? If not, what determines its value?
- **3b:** What can you say about the open circuit voltage U_{OC}?

Solution

Exercise 8.1-6

Constructing Quantitative Logarithmic *IV* Characteristics

Question 1. Construct rather quantitatively the logarithmic ***IV*** characteristics (= **log *j* - e*U*** plot) of two solar cells with the j_1 and j_2 values as given in the table. Here are a few hints:

- Draw first the straight lines for the two exponential terms resulting from the master equation (i.e. omit the "**-1**" term) into a **log *j* - e*U*** plot. Note that for room temperature **k*T* = 0.025 eV** or **exp(e*U*/k*T*) = exp(40 · e*U*)**. Use the given numbers for the various **$j_i = j(U = 0V)$** and calculate, for example, **$j_i(U = 0.5\ V)$** to get a second point.
- Correct by "hand" for the "**-1**" term (justify your reasoning) and add "by hand" the two resulting curves to the *full characteristic*.
- Repeat the procedure for a temperature of **400 K**. Note, however, that all j_i contain the intrinsic carrier density n_i and that changing the temperature changes n_i accordingly.

	Calculated	Measured
j_1	$1.6 \cdot 10^{-14}$ A/cm^2	10^{-9} A/cm^2
j_2	$1.6 \cdot 10^{-10}$ A/cm^2	10^{-7} A/cm^2

Question 2: Determine the open circuit voltage U_{OC} for room temperature and for **400 K** and discuss your finding.

Note: U_{OC} is the voltage where the (positive) forward current in the dark is exactly equal to the magnitude of the photo current I_{PH}

Solution

Exercise 8.2-1

Quick Questions to

8.2 Making Bulk Si Solar Cells

Here are some quick questions

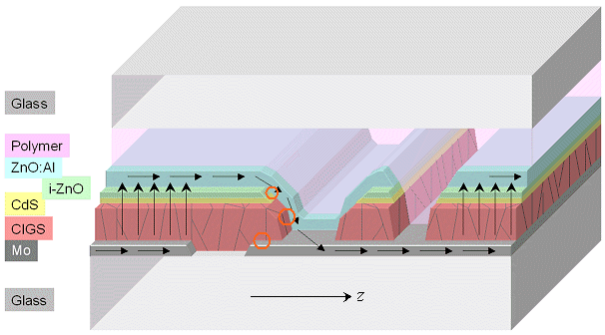
- Discuss the basic requirements for mass production of solar cells including technical constraints resulting from economical boundary conditions
- Describe the essential production steps of a **mc-Si** solar cell. Start with suitable poly-**Si** and discuss essential problems encountered with the solutions. Use schematic drawings.
- Describe "anti reflection" technologies.
- Compare screen-printing for the deposition of the metallic grid on a **Si** solar cell to other layer deposition methods. *Hint:* Consider that a good solar cell may deliver **5 A** at **0.5 V** and consider how that converts into thickness requirements of the metal grid layer. The *specific* resistivity ρof a decent metal is about **2 μΩcm**, the resistance ***R*** for a cross sectional area of ***A*** **cm^2** and a length ***l*** is $R = \rho \cdot l / A$.

Exercise 8.3-1

Making In-Situ Series Connections

Consider the basic structure of a thin film module about **(1.000× 1.000) mm^2** in size. made "in one piece" and consisting of individual solar cells **(10× 1.000) mm^2** that are all switched in series in-situ during the production process.

- The final structure seen in cross-section at the position of the series connection should look like this:



Courtesy of Dr. Powalla; ZWS Stuttgart,

1. Suggest a way to make this module. You don't have to you describe how layers are deposited; all that counts is the interconnect structure. You may also forget about the polymer and glass top layer; which are trivial.

- **Hint** : You should consider making three "cuts" at the right time a the right place

2. Assuming that there is some tolerance for the alignments of whatever has to be aligned, discuss pro and cons of the structure above with respect to the nominal short circuit because of the **Mo** overlap between the two solar cells.

Solution

Exercise 8.3-2

Quick Questions to

8.3 Making Thin Film Solar Cells

Here are some quick questions

- List basic requirements that semiconductors must meet if they are to be used for *thin film* solar cells.
- Give some examples of existing thin film solar cell technologies and list their strengths and weaknesses.
- What is the logic behind "concentrator cells"? Discuss the basic principle and necessities concerning application.
- Give a schematic cross section through a **CIGS** solar cell, indicating the major layers and the in-situ series connection.

Exercise 8.4-1

All Quick Questions to

8. Solar Cells

General Concerns

- Give some rough numbers (with some reasoning whereever applicable), always per $\mathbf{m^2}$, for
  - Maximum solar power.
  - Maximum and practical efficiency of "standard" **Si** solar cells.
  - Average power for "standard" **Si** solar cell.
  - Average energy harvest of "standard" **Si** solar cells per year.
- Compare indirect and direct semiconductors with respect to light absorption at the "band edge", i.e. for light energies around bandgap energy. What follows for solar cells?
- What is your first priority with respect to the coupling of light and semiconductor when you want to make a solar cell with a good efficiency?
- Draw the current density (***j***) - voltage (***U***) characteristics of a **pn**-junction in the dark and under illumination in the interesting part of the ***j-U*** plot. Point out the important points of the illuminated diagram and give some approximate numbers for typical intense sun light.
- The typical ***j-U*** equation for a **Si pn**-junction is

$$j = j_1 \cdot \left(\exp \frac{eU}{kT} - 1 \right) + j_2 \left(\exp \frac{eU}{2kT} - 1 \right) - j_{Ph}$$

  Discuss the origin of the ***j*** terms. Compare (qualitatively) the magnitude of j_1 and j_2. What kind of properties of **Si** influence the value of j_{Ph}?
- Why is "dirty" **Si** not good for solar cells? *Hint*: Follow the fate of a photon-generated carrier.
- Draw the ***j-U*** curve of an illuminated decent solar cell. Draw (qualitatively) the power curve into this diagram. Discuss the curve shortly with respect to real power applications
- Your electrical energy bill shows that you, personally, consumed **2 000 kWh** electrical energy per year in your home. How man square meters of solar cells (roughly) would you need on your roof to supply this much energy?
- Give the eqivalent circuit diagram of a realistic **Si** solar cell. Discuss the components with the aid of schematic **IV**-characteristics.
- Define the fill factor of a solar cell and discuss its dependence on solar cell parameters.

Making Bulk Si Solar Cells

- Discuss the basic requirements for mass production of solar cells including technical constraints resulting from economical boundary conditions
- Describe the essential production steps of a **mc-Si** solar cell. Start with suitable poly-**Si** and discuss essential problems encountered with the solutions. Use schematic drawings.
- Describe "anti reflection" technologies.
- Compare screen-printing for the deposition of the metallic grid on a **Si** solar cell to other layer deposition methods. *Hint:* Consider that a good solar cell may deliver **5 A** at **0.5 V** and consider how that converts into thickness requirements of the metal grid layer. The *specific* resistivity ρof a decent metal is about **2 μΩcm**, the resistance ***R*** for a cross sectional area of ***A*** $\mathbf{cm^2}$ and a length ***l*** is $R = \rho \cdot l / A$.

Making Thin Film Solar Cells

- List basic requirements that semiconductors must meet if they are to be used for *thin film* solar cells.
- Give some examples of existing thin film solar cell technologies and list their strengths and weaknesses.
- What is the logic behind "concentrator cells"? Discuss the basic principle and necessities concerning application.
- Give a schematic cross section through a **CIGS** solar cell, indicating the major layers and the in-situ series connection.

Solution to Exercise 8.1-1

Exponential Growth

Illustration

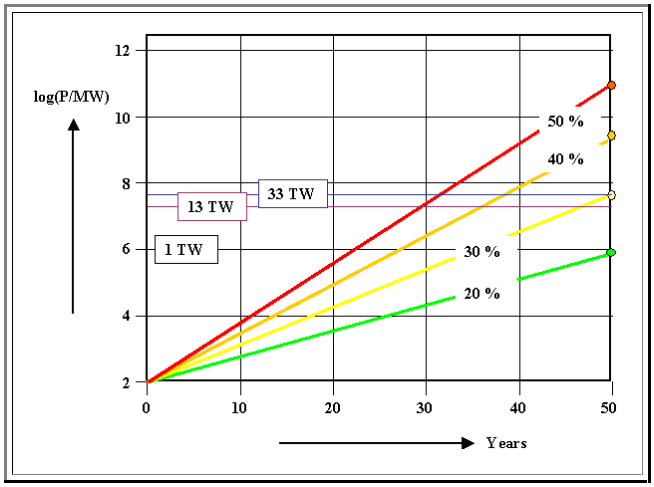
1. The output of the solar cell industry in **2006 - 2008** grew by **40 %** per year. Let's assume that all solar cells installed in **2007** produced a total energy of **0.1 GW /year**. Calculate (and plot) the installed power as a function of time up to **2050** for growth rates of **α' = 20 %**, **30 %**, **40 %**, and **50 %**. What is the proper equation?

- The general equation is $P(t) = P_0 \cdot \exp(\alpha \cdot t)$ and we know $P(t = 0\ a) = 100$ MW and $P(t = 1a) = 100\ \text{MW} + (\alpha'/100) \cdot 100\ \text{MW}$; α' is the given growth rate in %
- It follows that

$$
\begin{aligned}
P(t = 1\text{a}) &= 100\ \text{MW} \cdot \exp(\alpha \cdot 1\text{a} \cdot \text{a}^{-1}) \\
&= 100\ \text{MW} + (\alpha'/100) \cdot 100\ \text{MW} \\
\alpha &= \ln(1 + \alpha'/100)\ \text{a}^{-1} \\
&= (0.182;\ 0.262;\ 0.336;\ 0.405)\ \text{a}^{-1} \\
&\quad \text{for growth rates of} \\
&\quad 20\%\ ;\ 30\%;\ 40\%;\ 50\%
\end{aligned}
$$

2. Calculate (and plot) the installed power as a function of time up to **2050** for growth rates of **20 %**, **30 %**, **40 %**, and **50 %**.

- That's easy and we do it, of course, in a **log** $P(t)$ plot. What we get looks like this:



3. What follows form the results with respect to the world-wide power scenario as described in the link??

- It follows that with the present growth rate of **40 %** all of the world's energy demands can be produced by solar cells in **35 - 38** years - be it the present **13 TW** or the predicted **33 TW**
- That looks like a "Milchmädchenrechnung" (i.e. very naive), because that's what it is. If we can sustain a growth rate of **40 %** for **30 - 40** years remains to be seen. It's unlikely, but not impossible. The semiconductor industry, for example, sustained a growth rate of about **30 %** by now for more than **35** years, and no end is in sight.

4. Plot the demand for **Si**, assuming that a standard (**1000 x 1000 x 0.1**) **mm^3 Si** solar cell generates **10 W** on average. Will there be enough **Si**? How do the amounts of **Si** needed compare to other essential raw materials?

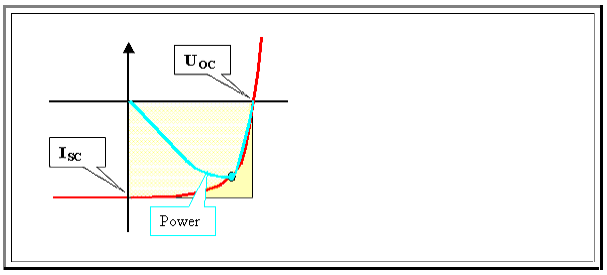
- The volume is **10^5 mm^3 = 100 cm^2**. With a density of **2.33 g/cm^3** we have **23.2 g/W**.
- The present (**2007**) production of (solar grade) **Si** per year is roughly **20.000 to = $2 \cdot 10^{10}$ g**; corresponding to **862 MW**. If we want to produce **1 TW,** we would need **$23.2 \cdot 10^{13}$ g = $23.2 \cdot 10^7$ to** of **Si**.
- That looks like a lot of **Si**. Yes, but look at the present world production of:
  - **Iron / Steel:** $\approx$ **$780 \cdot 10^6$ to**.
  - **Coal:** $\approx$ **$5\,000 \cdot 10^6$ to**.
  - **Al** $\approx$ **$22 \cdot 10^6$ to**.
- So a few million tons of **Si** is definitely within present day capabilities

Solution to Exercise 8.1-2

Optimal Working Point of Solar Cells

The figure shows the ***IV***-characteristics of a real solar cell. Derive the optimal working point by simply constructing the ***I · U =*** Power curve graphically

Illustration



- This is easy. Power is $P = U \cdot I$; as long as the current is relatively constant, the power curve is a straight line from the origin.
- At U_{OC} the power is zero; for voltages a bit smaller it will go up sharply. All in all the power curve must look like the blue curve with the point of maximum efficiency somewhat below U_{OC}

- What kind of load resistor would you need for this solar cell? What are the implications ?

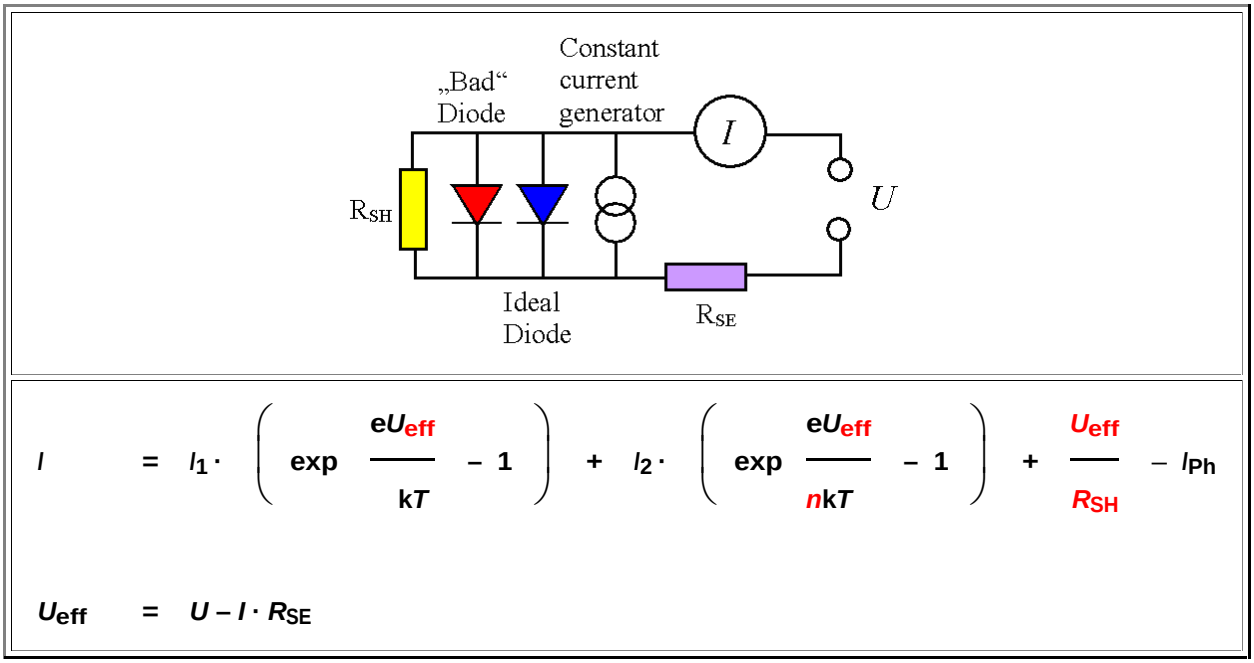
- In full sun light a **(15 × 15) cm^2** solar cell delivers something like **5 A** at **0.5 V**, calling for a load resistor with R_{load} **= 100 m**Ω. It the current drops by a factor of **10** because of clouds, the load resistor must increase **10** fold, otherwise the voltage drops below the optimum value .
- This means that not only do we have very small (and difficult to handle) load resistors, we also need an active load management if maximum power is to be harvested from solar cells with necessarily strongly changing output.

Solution to Exercise 8.1-3

Characteristics of Real Solar Cells

Illustration

This is the starting point for doing the exercise:



$$I = I_1 \cdot \left(\exp \frac{eU_{eff}}{kT} - 1 \right) + I_2 \cdot \left(\exp \frac{eU_{eff}}{nkT} - 1 \right) + \frac{U_{eff}}{R_{SH}} - I_{Ph}$$

$$U_{eff} = U - I \cdot R_{SE}$$

Discuss *qualitatively* the influence of the *two resistors* (and, as a more minor point, the *idealiy factor* ***n***) on the ***IV*** characteristics.

We will get to this, but here we will actully discuss the questions first *quantitatively*. As input parameters we need j_1, j_2 and the ideality factor n, which we take as (see also exercise 8.1-5)

- $j_1 = 10^{-9}$ A/cm^2.
- $j_2 = 10^{-7}$ A/cm^2.
- $n = 2$.

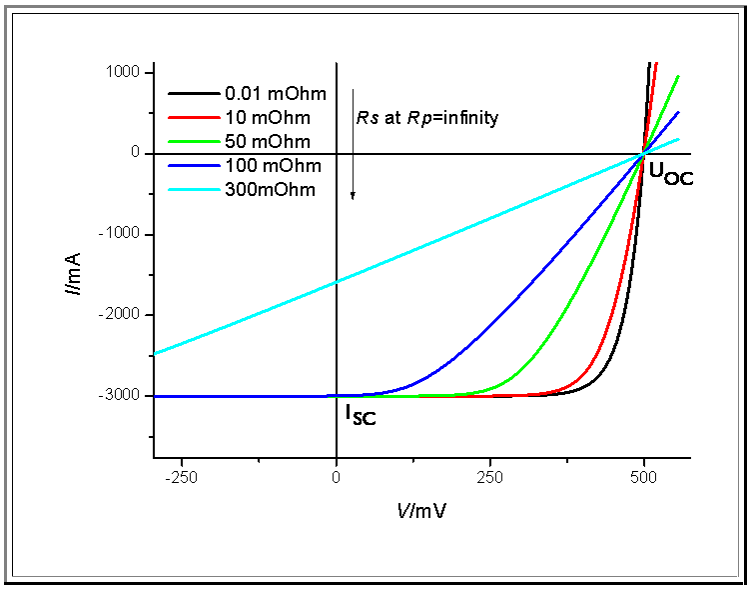
All we have to do is to solve the the equation from above numerically for various values of the parameters:

- Series resistance R_{SE}.
- Shunt resistance R_{SH}.
- Diode ideality factor n.
- Pre-exponential factors I_1 and I_2.
- And possibly the whole thing as a function of temperature T.

This is a big program, but it is not too difficult to see some major points.

Series resistance R_{SE}

Here is a plot of the ***IV*** characteristics of a *typical* solar cell with **5** different series resistances R_{SE}.



- Everything else has been kept "ideal". This means that the shunt resistance R_{SH} is very large ("infinity"), the ideality factor of the second diode is $n = 2$, and the two pre-exponential factors are $I_1 = 0{,}1\ \mu A$ and $I_2 = 10\ \mu A$. The photo current is **3 A**.

Even without looking a the (numerically) calculated figure, we can deduce *qualitatively* a few facts from our basic equation above, aw asked in the exercise.

- For $I = 0$ **A**, we have $U_{eff} = U$. That means that *all* ***IU***-characteristics *must* run through U_{OC}, no matter what kind of serial resistance we might have.
- For large negative ***U*** (*reverse direction*) , the current ***I*** is simply constant. We loose a part of the applied voltage in the serial resistance, but that does not effect the current. The characteristics in the **3rd** quadrant thus does not depend on R_{SE} if $|U|$ is large enough.
- For large positive ***U*** (*forward direction*), the diode by itself will admit large currents for voltages above about **0.5 V**, i.e. the *diode resistance* becomes very low. The ***IU***-characteristics then *must* be dominated by R_{SE}; it will simply turn into an ohmic *straight line* with a slope given by $1/R_{SE}$
- In the fourth quadrant for voltages below U_{OC} some of the voltage drops a the series resistor. The magnitude of the current thus can only be lower than in the case without a series resistor.

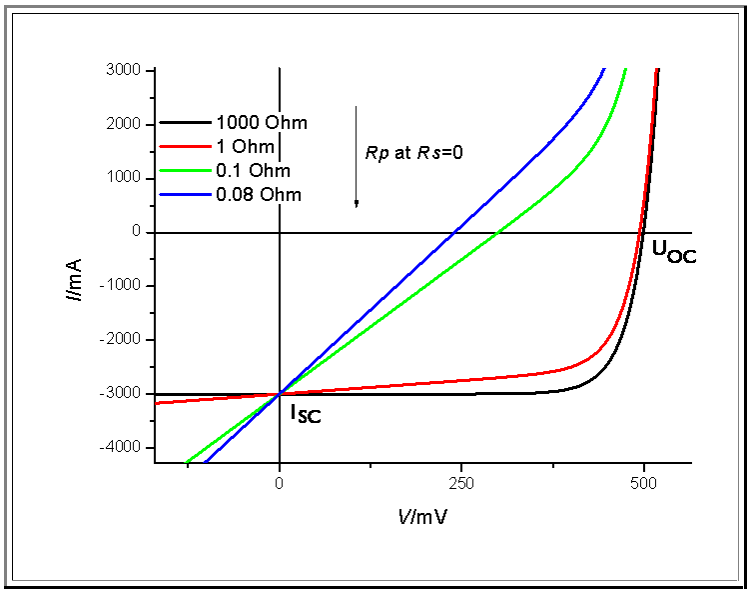
All of this is exactly what the calculated figure shows- taking into account the "Ohmic" straight line we could have derived most of the graph above without any quantitative calculations

We now can draw several conclusions:

- **1.** The efficiency η is proportional to the product of $I_{Sc} \cdot U_{OC} \cdot FF$. As long as R_{SE} is not too large (e.g. $R_{SE} < 100\ m\Omega$ for the example given), series resistances primarily decrease the fill factor ***FF*** and thus reduce the efficiency η.
- **2.** While in normal "electrical" life, "*milliohms*" hardly count, a few **m**Ω serial resistance are enough to make your solar cell measurable worse.
- **3.** Given the specific resistivity of good metals of $\rho \approx 2\ \mu\Omega cm$, a **Cu** wire of **1 cm** length and **1 mm²** cross section has a resistance of $R = 2\ m\Omega$. The cross sectional area of the grid metallization on a solar cell is **< 1 mm²**, which means *we have a real and unavoidable problem* with series resistances of real solar cells!

Shunt resistance R_{SH}

Here is a plot of the ***IV*** characteristics of a *typical* solar cell with **4** different shunt resistances R_{SH}.



- Everything else has been kept "ideal". This means that the series resistance R_{SE} is now close to zero.

Again, without looking a the (numerically) calculated figure, we can easily deduce *qualitatively* what is going to happen.

- $U_{eff} = U$ is always true. For $U_{eff} = U = 0$ V *all* characteristics must run through I_{SC} since the term U_{eff}/R_{SH} is zero.
- Otherwise, for any voltage U in *reverse* and *forward* direction we have a current $I_{SH} = U_{eff}/R_{SH}$ that must be added to the diode current and thus shifts the total current *upwards* (towards *larger* values (**-1** is larger than **-2**!) and thus decreases its *magnitude* in the fourth quadrant by just U_{eff}/R_{SH}. The flat part of the ideal characteristic thus turns into a straight line with slope $1/R_{SH}$
- In the fourth quadrant, where it counts, we will loose voltage *and* fill factor and thus severely reduce the efficiency η.
- We also have a reverse current increasing linearly with the reverse voltage - very bad in a module!

All of this is exactly what the calculated figure shows- taking into account the "Ohmic" straight line centered at I_{SC} we could have derived most of the graph above without any quantitative calculations

We now can draw several conclusions:

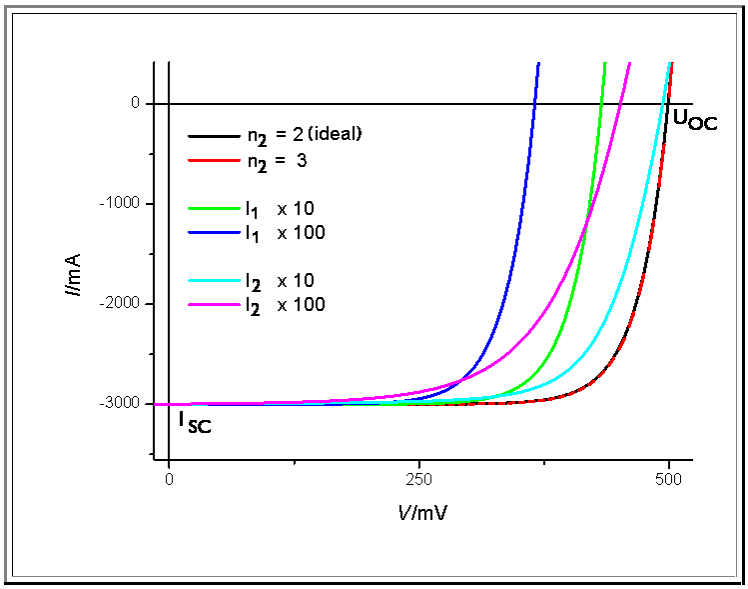
- **1.** The efficiency η is proportional to the product of $I_{Sc} \cdot U_{OC} \cdot FF$. As long as R_{SH} is not too small (**>**≈ **1**Ω) for the example given, shunts are not too bad. Real short circuits **<** ≈ **1**Ω, however, are disastrous for the efficiency.
- **2.** Since the **pn**-junction is very large and extends all the way out to the *edge* of the solar cell, we must expect that local short circuits happen. The rather difficult question coming up now is how a few *local* short circuits affect the *global* solar cell.

We have now answered the exercise questions.

- However, we will go on and discuss a few more points.

Ideality factor *n*

Here is a plot of the *IU* characteristics of a *typical* solar cell with deviations from ideality expressed in the ideality factor n and the pre-exponential factors I_1 and I_2



- Everything else has been kept simple - no shunt or series resistors. This means that the shunt resistance R_{SH} is very large ("infinity"), R_{SE} is zero. The ideality factor of the second diode is $n_2 = 2$ or $n = 3$ (the ideality factor of the first diode is always $n_1 = 1$ by definition), and the two pre-exponential factors are I_1 **= 0,1 µA** and I_2 **= 10 µA** as starting values once more, but also **10** times and **100** times that number. The photo current is **3 A**.

Without looking a the (numerically) calculated figure, we *cannot* easily deduce what is going to happen.

- Well, looking a the figure, we see that changing the ideality factor of the second diode from $n = 2$ to $n = 3$ does not produce a noticeable change in the characteristics. The simple reason for this is that in reverse direction the exponentials in the $I(U)$ equation don't matter, and that in the forward characteristics the ideal diode always "wins" except for small positive voltages.
- However, the relation between the two diodes is also influenced by the pre-exponential factors. Divided by the cell area, they were abbreviations for the following current densities:

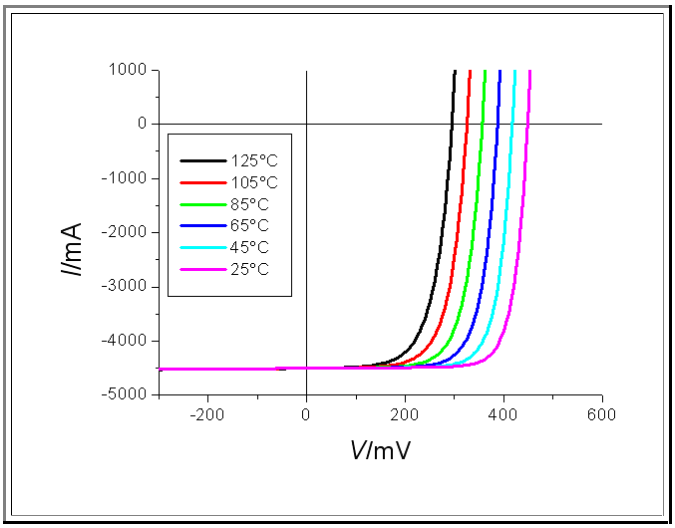
$$j_1 = \left(\frac{e \cdot L \cdot n_i^2}{\tau \cdot N_A} + \frac{e \cdot L \cdot n_i^2}{\tau \cdot N_D} \right)$$

$$j_2 = \left(\frac{e \cdot n_i \cdot d(U)}{\tau} \right)$$

- Why did we pick j_2 so much larger than j_1? We have, in fact, already discussed the relation $j_2 \, / \, j_1$ for **pn**-junctions, even so you probably forgot it all, and found that $j_2 >> j_1$ is unavoidable for **Si** and other semiconductors with bandgaps $<\approx$ **1 eV**. This is why the values chosen for the pre-exponential factors and given above have the relation they have.
- On the other hand, both factors are functions of variables like the diffusion length L or the recombination time τ, i.e. of crystal perfection; of the doping N_{Dop}; and of the temperature T (via the intrinsic carrier concentration n_i).
- I_1 and I_2 are thus variables up to a point, and we want them as small as possible because the currents they cause diminish the photo current and the open circuit voltage. The figure shows that clearly. Increasing I_1 or I_2 substantially, decreases U_{OC} and, for the case of I_2, also the fill factor FF.
- However, they should not be too small, either. If they would be zero, we would just have a constant photo current and no voltage ever builds up. The values chosen are rather optimal, that's why we called them "ideal".

Temperature *T*

- Our basic equation on top contains the temperature explicitly in the two exponentials and implicitly in the two pre-exponential factors j_1 and j_2.
  - The two equations right above for these two factors contain $n_i(T)$, the intrinsic carrier density, which grows exponentially with increasing temperature.
  - On top of that, the lifetime τ might be temperature dependent as well as the series and parallel resistors, but we will neglect that here.
- So what it the total effect of temperature? This is shown below for negligible resistances and an ideality factor $n = 2$.



- What we have is quite clear: As long as I_{Ph} does not depend on temperature (e.g. because we have a very good solar cell where all photo generated carriers are turned into photo current), the influence of the temperature comes from the exponents of our basic equation and from the (exponentially; via n_i)) temperature dependent j_1 and j_2
  - The major effect is that the open circuit voltage decreases a lot (which is bad).
  - Taking into account that I_{Ph} might be somewhat temperature dependent too (via the temperature dependence of the diffusion length, for example), that the series and shunt resistors most likely will be temperature dependent like most everything else, the situation can become quite complicated.
  - However, the total effect is practically always that the efficiency comes down quite a bit with increasing temperature - high temperatures are bad for solar cells!
  - This gives at least some comfort to cold and sun-deprived areas like Schleswig-Holstein. We may not have as much sun as the people in Spain or Sicily, but we don't have to worry as much about keeping our solar cells cool!

Final Conclusion

- We now can draw some conclusions:

- If you want to understand solar cells at the most fundamental non-trivial (= University) level, you better make damn sure that you understand the basic equation above and *all* its connotations by heart!
  - That's not as difficult as it may appear! It's all in "Introduction to Materials Science II" - see chapter 6! That's why *we* spend so much time on the **pn**-junction *and* its finer details, which are not usually covered in standard undergraduate text books.

Solution to Exercise 8.1-5

IV Characteristics of Real Solar Cells

The solution to the questions asked is so basic and so important that it commands its own advanced module.

Illustration

Solution to Exercise 8.1-6

Constructing Quantitative Logarithmic *IV* Characteristics

Illustration

- First we get a few important relations and numbers.
  - If $\mathbf{1/kT = 40\ eV^{-1}}$ at **300 K**, we have $\mathbf{1/kT = 40 \cdot 300/400 = 30\ eV^{-1}}$ at **400 K**
  - The current densities j_1 and j_2 can always be written as

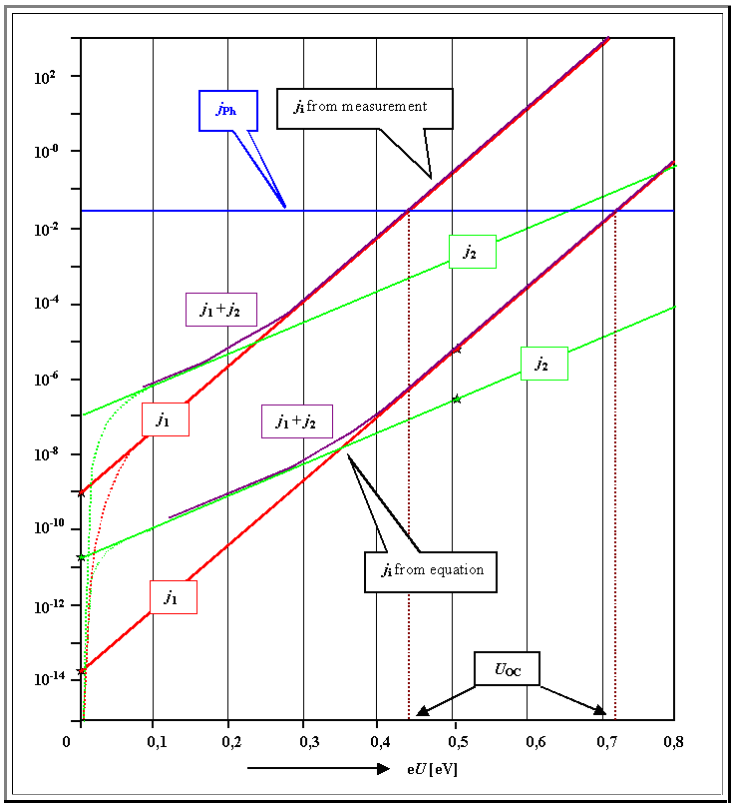
$j_1 = c_1 \cdot n_i^{\,2} = j_1' \cdot \exp(-E_g/kT)$	$j_1' = j_1 \cdot \exp(E_g/kT)$
$j_2 = c_2 \cdot n_i = j_2' \cdot \exp(-E_g/2kT)$	$j_2' = j_2 \cdot \exp(E_g/2kT)$

- This gives us the following numbers:

	Theory		Measured	
	j_1	j_2	j_1	j_2
Calculated j_i'	$2.06 \cdot 10^{5}$ A/cm^2	$5.74 \cdot 10^{-1}$ A/cm^2	$1.29 \cdot 10^{10}$ A/cm^2	$3.58 \cdot 10^{2}$ A/cm^2
	***T* = 300 K**			
Starting values j_i U = 0 V	$1.6 \cdot 10^{-14}$ A/cm^2	$1.6 \cdot 10^{-10}$ A/cm^2	10^{-9} A/cm^2	10^{-7} A/cm^2
Calculated j_i U = 0.5	$7.76 \cdot 10^{-6}$ A/cm^2	$3.52 \cdot 10^{-7}$ A/cm^2	0.46 A/cm^2	$2.2 \cdot 10^{-3}$ A/cm^2
	***T* = 400 K**			
Starting values j_i U = 0 V	$9.60 \cdot 10^{-10}$ A/cm^2	$3.92 \cdot 10^{-8}$ A/cm^2	$6.01 \cdot 10^{-5}$ A/cm^2	$2{,}44 \cdot 10^{-5}$ A/cm^2
Calculated j_i U = 0,5 V	$9.67 \cdot 10^{-3}$ A/cm^2	$1{,}5 \cdot 10^{-4}$ A/cm^2		

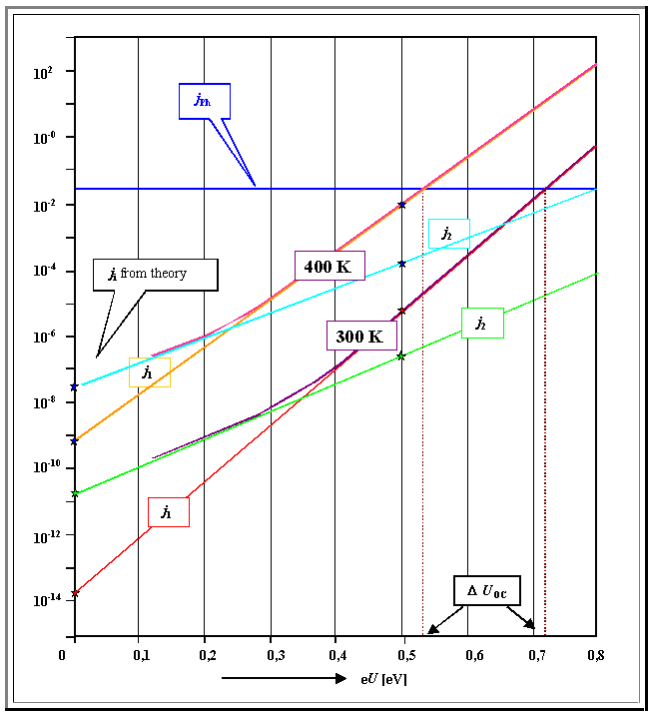
- Now to the questions:
  Question 1. Construct rather quantitatively the logarithmic ***IV*** characteristics (= **log *j* - e*U*** plot) of two solar cells with the j_2 and j_2 values as given in the table.
  Question 2: Determine the open circuit voltage U_{OC} for room temperature and for **400 K** and discuss your finding.

- Constructing the graph is easy now; here is the result:



- We note that the **"-1"** term can be neglected as soon as we have current density values about **10** times larger then the starting values, i.e. below $U \approx 0.1$ V. At lower values this term dominates the characteristics by forcing the currents to zero, i.e. to $-\infty$ in a **log** plot, but that is of no interest here.
- The addition of both curves only introduces a slight "rounding" at the intersection point.
- The open circuit voltage follows from the intersection of the $j(U)$ curves with a straight line at $j = -j_{Ph}$. It is immediately clear that only the j_1 part is of interest here.

The effect of temperature is shown in a separate graph and only for the "theoretical" set of the j_{Ph}:



- While the decreasing slope of the curves would increase U_{OC}, the large increase in the starting value of j_1 has a much stronger effect and causes a substantial decrease of U_{OC} with temperature.

Solution to Exercise 8.3-1

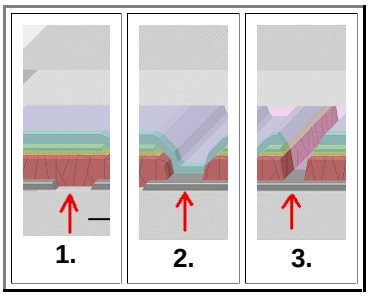
Making In-Situ Series Connections

Illustration

1. Consider the basic structure of a thin film module about **(1.000× 1.000) mm^2** in size. made "in one piece" and consisting of individual solar cells **(10× 1.000) mm^2** that are all switched in series in-situ during the production process.
Suggest a way to make this module. You don't have to you describe how layers are deposited; all that counts is the interconnect structure. You may also forget about the polymer and glass top layer; which are trivial.
Hint : You should consider making three "cuts" at the right time a the right place.

Thius is easy because one can deduce hat needs to be done right form the picture

- **1.** Deposit the **Mo** contact metal and structure it as shown.
- **2.** Deposit the **CIGS** layer plus the **CdSe** and **i-ZnO** and structure as shown.
- **2.** Deposit the**ZnO:Al** layer and structure as shown by cutting through all layers down to the **Mo** but not through the **Mo**.



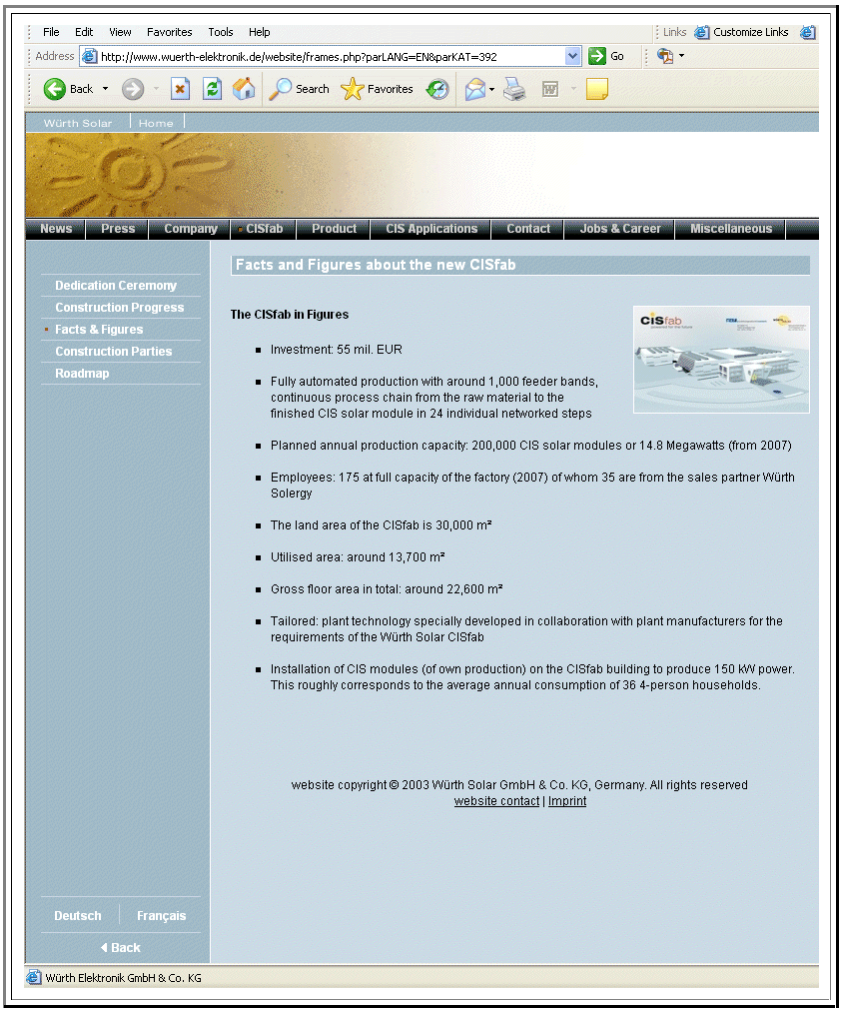
2. Assuming that there is some tolerance for the alignments of whatever has to be aligned, discuss pro and cons of the structure above with respect to the nominal short circuit because of the **Mo** overlap between the two solar cells.

- Well - do it!

CIGS Production by Würth Solar

Illustration

- **CIGS** solar cells got a big boost when one of the classical entrepreneurs from the heydays of the German ***economic miracle*** took an interest in the topic.
  - Reinhold Würth, started as an apprentice in the hardware store (specialized on screws) of his father in Künzelsau (in Suebia). He took over, **19** years old, in **1953** when his father died. Then he turned the small local company into the "Würth group", nowadays present in **86** countries and grossing **$8.5 \cdot 10^9$ Euro** in **2007**
  - **1994** his daughter took over the company he founded - giving **61** year old Reinhold Würth the opportunity to pursue a few hobbies like collecting art in a major way (he owns a few museums for his **> 10.000** substantial art pieces), teaching at various universities (and collecting honorary degrees in the process), riding his Harley-Davidson whenever he doesn't pilot his plane (he has a professional licence and his **CIGS** factory is next to his private airport in Schwäbisch-Hall), or furthering the cause of alternative energy.
- **1999** he founded Würth Solar, being fascinated with photovoltaics and in particular with **CIGS**. Being a successful business man, his goal was to get the **CIGS** technology into industrial large-scale production
  - Cooperations with the "Zentrum für Sonnenenergie- und Wasserstoffforschung Baden-Württemberg" (***ZSW***) lead to a pilot factory in Marbach (next to the town I grew up in and birthplace of Friedrich Schiller) and to the first major **CIGS** factory built in **2006** in Schwäbisch Hall for **$55 \cdot 10^6$ Euro**.
  - Here is their Internet page with some data about the factory.



Quasi Particles

Advanced

How to Imagine a Quasi Particle

- Protons, neutrons and electrons are particles. Nobody has a problem with that because most everybody just imagines them as some little ball that can exist by itself even in the absolute vacuum of space. While the "little ball" part of that imagination is faulty, the "can exist by itself" is correct.
  - Now let's look at **photons**. Definitely a particle, but the "little ball" picture is now completely off. A photon can exist be itself but must keep moving.
- If we move one step away from the "little ball" image we can perceive all those particles in a more abstract way as more or less localized carriers of fixed and immutable *properties* like rest mass, charge, spin, and some others that we don't need to worry about.
  - Then we have another set of properties tied to what those particle "*do*", expressed foremost in energy and momentum - quantities that are usually coupled by some dispersion relation and containing other parameters like wavelengths.
  - The important point is that all of the above is covered by quantum theory and that means that all properties are usually quantized and that interactions between those particles must take into account the Pauli principle. This leads to the Fermi-Dirac or Bose-Einstein distributions if more than one particle needs to be "filled" into a system with defined energy levels.
- Now we make another leap of imagination and consider any entity that can be perceived as a carrier of defined properties in the sense alluded to above. This entity then could be either a "real" particle or a *quasi-particle*. Discounting the many elementary particles of high energy physics that are not stable, whatever we are discussing now must have protons, neutrons, electrons and photons as building blocks. So let's see what kind of "quasi" particles that we have already encountered we can compose with these basic ingredients:
  - *Atoms* obviously. We can easily perceive atoms as real particles in their own right; nothing "quasi" about them. So we don't have to dwell on this.
  - *Molecules* and *crystals*. Now we have a problem. We can certainly "see" a **Cl_2** molecule as a "real" particle, but not a **DNA** molecule consisting of some **100.000** atoms or a **300 mm Si** wafer. But this is not a real problem, it is just the old tiring story that as things get larger, it is more convenient to switch from particle-oriented quantum mechanics to good old classical mechanics. We don't think of big things as particles anymore (but still as "mass points" for certain questions!).
  - *Holes*, behaving for all purposes like a positively charged electron. Now we have a "real" quasi particle in the sense that you can't take it out of the crystal in which it dwells and look at it. You cannot, to belabor this point somewhat more, make a fine point to a crystal and by applying a high field strength extract a *hole* beam which you focus and scan a cross on a specimen, running a scanning *hole* microscope. With electrons this is everyday technology. Holes in "reality" are still collectives of electrons that follow certain rules.
  - *Phonons* or quantized vibrations of a crystal. The term "elastic waves" is actually better than the term "vibrations", because that's what phonons are: Waves running through the crystal with amplitudes resulting from a local *elastic* deformation - the atoms are somewhat off their equilibrium position. Being a wave, it has a wavelength and thus momentum and some energy; always quantized, of course. So what is the difference to a *photon*? Only that the term "$\boldsymbol{A}$" in the basic wave equation $\boldsymbol{A(\underline{r},t) = A_0 \cdot \exp(\underline{kr} - \omega t)}$ has a different unit (distance instead of electrical field strength). If a photon is a particle in vacuum, so is a phonon inside a crystal!
- Holes and phonons are thus *quasi particles* that can only exist inside crystals (or matter). They describe some collective behavior of electrons and atoms, respectively, in a simple and consistent way.
  - It should come as no surprise that we will find more "collectives" that behave in a fashion with defined properties that we can ascribe to a suitable quasi-particle. Let's first look at list of what we have and then discuss those quasi particles very briefly

The Quasi Particle Zoo

Here is a list of crystal-dwelling quasi particles:

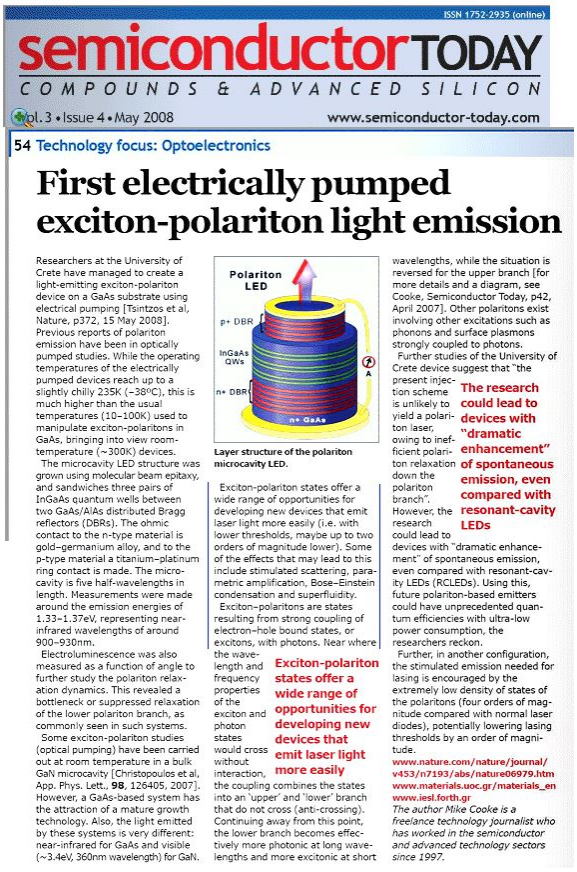
Quasi particle	Constituents	Remarks
Phonons	Crystal atoms	large momentum little energy Smallest wave length = lattice constants
Holes	Electrons in Valence band	"Define" semiconductors
Polarons	e^- + phonon	Essential to organic semiconductors
Plasmons	Electron collective	Determines optical properties of nanocrystals
Excitons	Bound electron hole pair	Produce the light in **GaP LED's**
Polaritons	Photon + phonon	Pretty strange - but those things do exist.
	Photon + electron	
	Photon + exciton	
Magnons	Crystal magnetic moments (= spins)	The "phonons" of spin waves
..	.....	.....
Cooper pairs	2 electrons coupled by phonons	The quasi-particle responsible for **superconductivity**

Now we would need a closer look (= large part of a advanced solid state physics lecture). I'm not doing that here but only give a minimal glimpse at:

Plasmons: Imagine a nanoparticle of **Au**, for example. There is only a small number of atoms and accordingly a small number of free electrons. Now imagine an (oscillating) electrical field acting on that particle - e.g. a photon "coming in".

- If a photon homes in on a large piece of **Au** the electrons in surface near regions feel a force, move a bit and by doing this screen the electrical field - it will not penetrate in the interior of the metal.
- However, if the particle is small enough, all electrons feel the same force, and all electrons behave as one, as an ensemble called "plasma" for reasons easy to guess. What this ensemble of electrons can do is, of course, subject to quantum mechanics, i.e. we must expect that there is some quantization of the energy. The *plasmon* then is nothing but the quantized excitation states of an electron collective - the "plasma". Excitations are more or less restricted to longitudinal oscillations of the plasma. In other words, our electrons in the nanocrystal swing from "left" to "right" in unison at some quantized "Eigenfrequencies" and then can be described just as well as a (standing wave like) plasmon.
- Quantization of energy means quantization of frequencies and that means that only "light" with the right frequency can excite a plasmon in our **Au** nanoparticles. Only this light then will be absorbed, transferring its energy to the plasmon.
- This is easy to show: Put some of your **Au** nanoparticles in otherwise fully transparent glass - it will now look colored because your nanoparticles, depending on their size, take out some wavelengths since light with wavelengths just right to excite plasmons will be absorbed. But you don't have to do this experiment, just look at stained glass windows in medieval churches. The beautiful dark red you see there is obtained in exactly this way! Our forbearers actually knew how to get **Au** nanoparticles in Glass. They just had no idea why and how it worked
- What you can imagine for **Au**, you can imagine for everything else, of course.

- Finally, just to show that I'm not making this up, the beginning of a collection of articles etc. where quasi particles come up as being useful.
- First we have **polaritons**



ISSN 1752-2935 (online)
semiconductorTODAY
COMPOUNDS & ADVANCED SILICON
Vol. 3 • Issue 4 • May 2008 www.semiconductor-today.com

First electrically pumped exciton-polariton light emission

Researchers at the University of Crete have managed to create a light-emitting exciton-polariton device on a GaAs substrate using electrical pumping [Tsintzos et al, Nature, p372, 15 May 2008]. Previous reports of polariton emission have been in optically pumped studies. While the operating temperatures of the electrically pumped devices reach up to a slightly chilly 235K (–38°C), this is much higher than the usual temperatures (10–100K) used to manipulate exciton-polaritons in GaAs, bringing into view room-temperature (~300K) devices.

The microcavity LED structure was grown using molecular beam epitaxy, and sandwiches three pairs of InGaAs quantum wells between two GaAs/AlAs distributed Bragg reflectors (DBRs). The ohmic contact to the n-type material is gold–germanium alloy, and to the p-type material a titanium–platinum ring contact is made. The microcavity is five half-wavelengths in length. Measurements were made around the emission energies of 1.33–1.37eV, representing near-infrared wavelengths of around 900–930nm.

Electroluminescence was also measured as a function of angle to further study the polariton relaxation dynamics. This revealed a bottleneck or supressed relaxation of the lower polariton branch, as commonly seen in such systems.

Some exciton-polariton studies (optical pumping) have been carried out at room temperature in a bulk GaN microcavity [Christopoulos et al, App. Phys. Lett., **98**, 126405, 2007]. However, a GaAs-based system has the attraction of a mature growth technology. Also, the light emitted by these systems is very different: near-infrared for GaAs and visible (~3.4eV, 360nm wavelength) for GaN.

Layer structure of the polariton microcavity LED.

Exciton-polariton states offer a wide range of opportunities for developing new devices that emit laser light more easily (i.e. with lower thresholds, maybe up to two orders of magnitude lower). Some of the effects that may lead to this include stimulated scattering, parametric amplification, Bose–Einstein condensation and superfluidity.

Exciton–polaritons are states resulting from strong coupling of electron–hole bound states, or excitons, with photons. Near where the wavelength and frequency properties of the exciton and photon states would cross without interaction, the coupling combines the states into an 'upper' and 'lower' branch that do not cross (anti-crossing). Continuing away from this point, the lower branch becomes effectively more photonic at long wavelengths and more excitonic at short wavelengths, while the situation is reversed for the upper branch [for more details and a diagram, see Cooke, Semiconductor Today, p42, April 2007]. Other polaritons exist involving other excitations such as phonons and surface plasmons strongly coupled to photons.

Exciton-polariton states offer a wide range of opportunities for developing new devices that emit laser light more easily

Further studies of the University of Crete device suggest that "the present injection scheme is unlikely to yield a polariton laser, owing to inefficient polariton relaxation down the polariton branch". However, the research could lead to devices with "dramatic enhancement" of spontaneous emission, even compared with resonant-cavity LEDs (RCLEDs). Using this, future polariton-based emitters could have unprecedented quantum efficiencies with ultra-low power consumption, the researchers reckon.

The research could lead to devices with "dramatic enhancement" of spontaneous emission, even compared with resonant-cavity LEDs

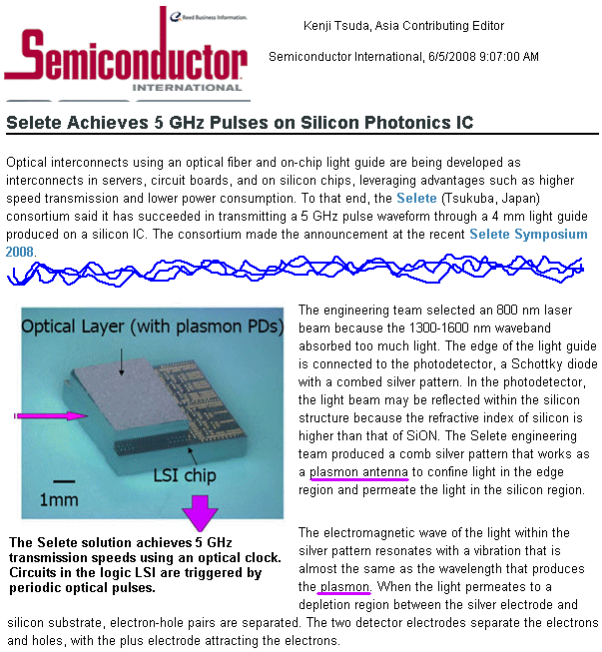
Further, in another configuration, the stimulated emission needed for lasing is encouraged by the extremely low density of states of the polaritons (four orders of magnitude compared with normal laser diodes), potentially lowering lasing thresholds by an order of magnitude.

www.nature.com/nature/journal/v453/n7193/abs/nature06979.htm
www.materials.uoc.gr/materials_en
www.iesl.forth.gr

The author Mike Cooke is a freelance technology journalist who has worked in the semiconductor and advanced technology sectors since 1997.

- It can't get much more exotic than that: putting exciton-polaritons to use! The blue lines indicate a part of the article where you will find a lot of the vocabulary used here in this context.

- Now let's look at **plasmons**:



Semiconductor INTERNATIONAL

Kenji Tsuda, Asia Contributing Editor

Semiconductor International, 6/5/2008 9:07:00 AM

Selete Achieves 5 GHz Pulses on Silicon Photonics IC

Optical interconnects using an optical fiber and on-chip light guide are being developed as interconnects in servers, circuit boards, and on silicon chips, leveraging advantages such as higher speed transmission and lower power consumption. To that end, the Selete (Tsukuba, Japan) consortium said it has succeeded in transmitting a 5 GHz pulse waveform through a 4 mm light guide produced on a silicon IC. The consortium made the announcement at the recent Selete Symposium 2008.

The Selete solution achieves 5 GHz transmission speeds using an optical clock. Circuits in the logic LSI are triggered by periodic optical pulses.

The engineering team selected an 800 nm laser beam because the 1300-1600 nm waveband absorbed too much light. The edge of the light guide is connected to the photodetector, a Schottky diode with a combed silver pattern. In the photodetector, the light beam may be reflected within the silicon structure because the refractive index of silicon is higher than that of SiON. The Selete engineering team produced a comb silver pattern that works as a plasmon antenna to confine light in the edge region and permeate the light in the silicon region.

The electromagnetic wave of the light within the silver pattern resonates with a vibration that is almost the same as the wavelength that produces the plasmon. When the light permeates to a depletion region between the silver electrode and silicon substrate, electron-hole pairs are separated. The two detector electrodes separate the electrons and holes, with the plus electrode attracting the electrons.

Recombination Channels

Advanced

High Injection Approximations for Recombination Rates

- Since optoelectronic devices usually are made to produce *plenty* of light, the deviation of the carrier concentrations from equilibrium in the recombination zonemust be large to obtain large recombination rates and thus light
  - If we write the concentrations as $n_{e,h} = n_{e,h}(equ) + \Delta n_{e,h}$, we now may use *the simplest possible approximation* called **high injection approximation**:

$$\Delta n_{e,h} >> n_{min}(equ)$$

  - i.e. the minority carrier concentration is far *above* equilibrium.
- The surplus carriers contained in $\Delta n_{e,h}$ are always *injected* into the volume under consideration (called **recombination zone** or **recombination volume**), usually by forward currents across a junction. They always must come in equal numbers, i.e. in pairs to maintain charge neutrality; otherwise large electrical fields would be generated that would restore neutrality. We thus have

$$\Delta n_e = \Delta n_h$$

  - The recombination volume usually is the space charge region of a junction or an other volume designed to have *low carrier concentrations* in equilibrium, cf. the picture in the backbone. Since the equilibrium concentration of both carrier types in the **SCR** is automatically very low, we may easily reach the high injection case.
- The surplus concentration of carriers decays with a characteristic lifetime τ_{hi}. For the recombination rate R we have in analogy to "normal" recombination more close to equilibrium:

$$R = \frac{\Delta n}{\tau_{hi}}$$

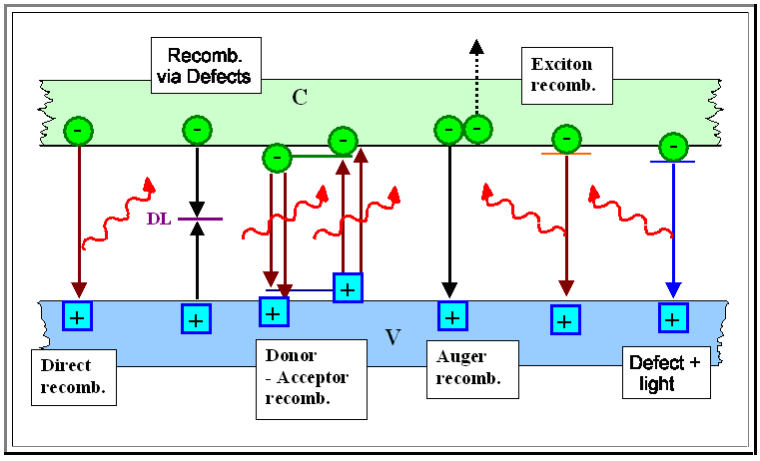
  - The only difference is that the high-injection life time τ_{hi} can be quite different from the equilibrium minority carrier life time. If τ_i is the (high-injection) life time of recombination channel No. i, τ_{hi} is given by

$$\frac{1}{\tau_{hi}} = \sum_i \frac{1}{\tau_i}$$

  - The important thing for optimizing **LED's** and Laser is that the τ_i are not constants but depend on the degree of injection as we will see.

Some Specifics of Recombination Channels

- Let's repeat the picture from the backbone to have a listing of the more important recombination channels



- Now let's look at the more mportant recombination channels and their dependence on the injection ratio, i.e. the carrier concentration

The **band-band recombination channel** is easy to understand:

- A large number of electrons and holes finds themselves in some volume of a semiconductor at concentrations far above equilibrium. They are running around in a random matter and every now and then a hole and an electron get real close on their ***perambulations*** and recombine. The probability for this to happen is clearly proportional to the concentration $n_{e,h}$ electrons and holes which as we have postulated above is Δn for both carrier types.
- the recombination rate $R_{b\text{-}b}$ have for the band-band recombination channel is thus given by

$$R_{b\text{-}b} = B_{b\text{-}b} \cdot n^2$$

- The proportionality constant ***B*** is occasionally also called a **recombination coefficient**.

If we now look at the **recombination channel via defects** (also called "**deep level**" recombination because the defect must have a energy level deep in the bandgap).

- The story was that an electron on its random migration might encounter a defect, e.g. an impurity atom with an energy level somewhere in the bandgap which it occupies and now is trapped and mellow (low in energy, at least for some time). A hole, somewhat later, also finds the impurity atom plus the electron unable to run away, and happily recombines with the electron. In other words: a girl, wandering around at random finds an irresistible café and sits down for a while. A boy, coming accidentally by the café, seeing the girl trapped there and in a mellow mood, knows what to do... This also means that *no light is produced*
- We obtain a rather simple relation for the recombination rate R_{defect}

$$R_{defect} = B_{dl} \cdot n_{dl} \cdot n$$

- With B_{dl} = recombination coefficient for this case. We have a proportionality to the density of defects and the density of carriers, giving the rate that an electron (or hole) is trapped by a defect. The rest, recombination with the opposite carrier, happens "automatically"

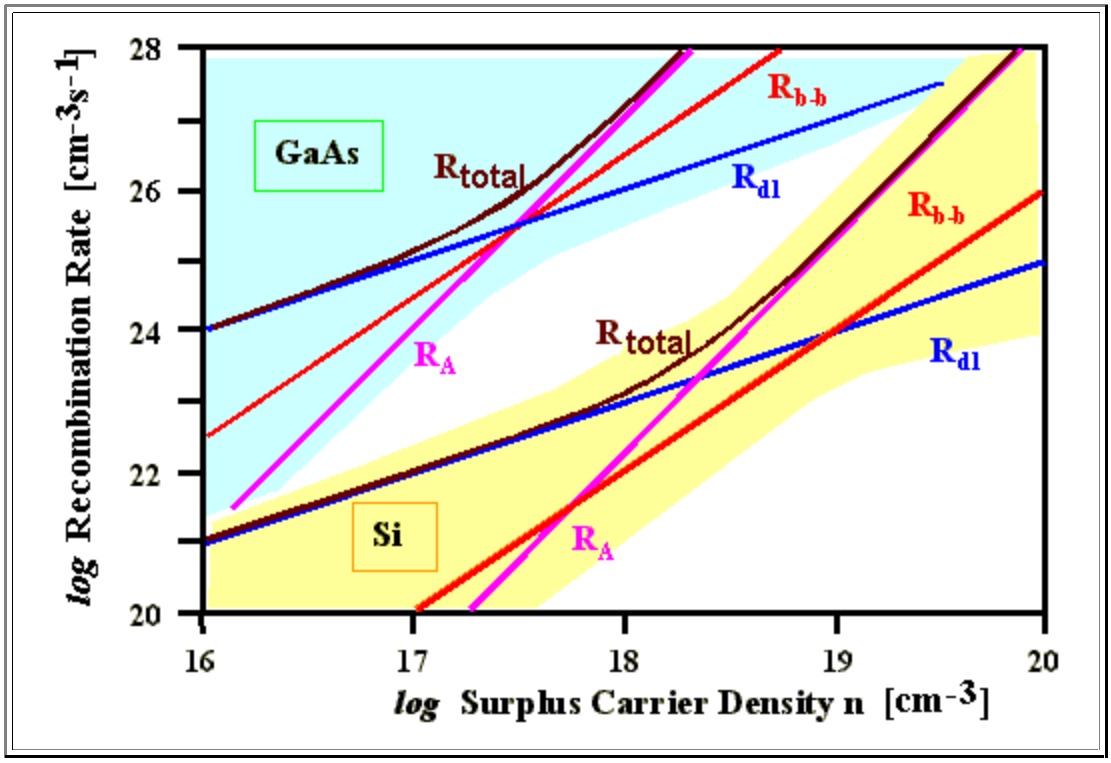
If we omit the recombination from donor and acceptor energy levels (which is quite similar to band-band recombination anyway) and the "exotic" recombination via excitons, we only have to consider "**Auger recombination**".

- In this case the energy of the recombination event is transferred to another electron in the conduction band, which then looses its surplus energy by "thermalization", i.e. by transferring it to the phonons of the lattice. This means that *no light is produced*.
- It also means that we have two have two electrons and one hole at the same place at the same time or that the Auger recombination rate R_A is given by

$$R_A = B_A \cdot n^3$$

Taken everything together we see that we have recombination rates for the major recombination channels that depend on the first, second and third power of the carrier concentration ***n***

If we plot the total recombination rate as a function of carrier density (in a double log plot) with the proper proportionality constants (wherever they come from) and some assumption for defect densities for **GaAs** and **Si**, we obtain the following highly informative picture.



- The recombination rate in **Si** is generally far lower than in **GaAs** and for carrier concentrations **<≈ 10^{18} cm^{-3}** dominated by defect recombination.
- In both cases we can increase the relative strength of radiative band-band recombination by decreasing defect recombination, i.e. by making the semiconductor more pure and perfect and by not allowing too large carrier concentrations.
- Here we have the explanation for the fact that very good solar cells, having by definition very low defect recombination rates, will show measurable luminescence if a large carrier concentration is introduced via an intensive flash of light. This effect is presently (**2008**) exploited for the characterization of solar cells.

Cleaving Semiconductors for (Blue) Lasers Diodes

Advanced

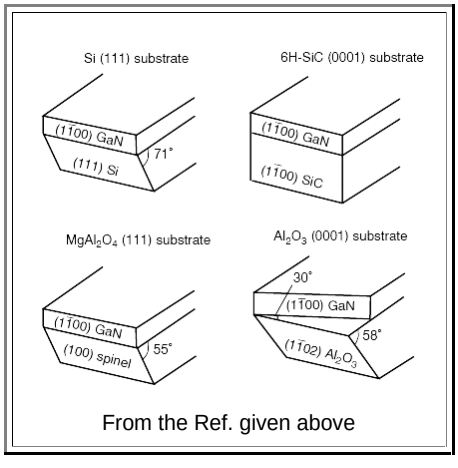
- The module is based on the paper of Kuramta et al: in FUJITSU Sci. Tech. J. **342** (1998). p.191. The authors provide data about **cleavage planes in semiconductors**.
  - The preferred cleavage planes of a semiconductor are not as clear-cut as it seems. As everybody know who just once dropped a **{100}** Si wafer knows, the fracture plans are the **{110}** planes; a large part of the literature, however, including the paper given above, insists that it should be **{111}**.
  - Be that as it may, here we take the planes given in the article.

Material	Crystal structure	Effective lattice mismatch (%)	Difference in thermal expansion coefficient ($\times 10^{-6}$)	Cleavage	Stability
Si	Diamond	20.1	-2.0	(111)	Good
GaAs	Zinc blende	25.3	0.4	(110)	Fair
GaP	Zinc blende	20.7	-0.9	(110)	Fair
MgO	Rock salt	-6.5	4.9	(100)	Fair
MnO	Rock salt	-1.4		(100)	Bad
CoO	Rock salt	-5.4		(100)	Bad
NiO	Rock salt	-7.6		(100)	Bad
$MgAl_2O_4$	Spinel	-10.3	1.9	(100)	Good
$NdGaO_3$	Perovskite	-1.2	1.9		Fair
ZnO	Wurtzite	2.0	-2.7	(1-100) (11-20) (0001)	Fair
6H-SiC	ZnS 6H	-3.4	-1.4	(1-100) (11-20) (0001)	Good
$LiAlO_2$	ß-$NaFeO_2$	1.7	1.7	(001)	Fair
$LiGaO_2$	ß-$NaFeO_2$	-0.1	1.9	(010)	Fair
Al_2O_3	Corundum	-13.8	1.9	(1-102)	Good
$LiNbO_3$	Ilmenite	-6.7	9.9	(1-102)	Bad
$LiTaO_3$	Ilmenite	-6.8	10.6	(1-102)	Fair

From the Ref. given above

Comparison always to **GaN**

- **GaN** as a semiconductor technology material only comes as a thin layer on a substrate other than **GaN** since there simply are no usable **GaN** single crystals
  - As we know, if we grow thin layers with different lattices, we have to watch out for misfit dislocations. It is important to look for substrates with a lattice constant as similar as possible to that of the thin layer to be grown. The table above shows the lattice mismatch of prospective substrates to **GaN** and thus gives a guideline.
- If we want to make a Laser diode form the thin film, we have a few more requirements besides "just" avoiding misfit dislocations as best as we can:
  - The substrate should have a high electrical and thermal conductivity. The first property would make it easier to supply the large current densities we need to operate a Laser diode, the second to remove efficiently the heat generated during operation.
  - The whole stack of substrate and layers should cleave nicely on a well-defined and very flat plane because the two relevant surfaces obtained by cleavage will serve as the mirrors of the Faby-Perot resonator we need for a Laser. Now look at the possible cleavage relations:



From the Ref. given above

- Summing up: There is no ideal substrate - you have to find the optimal compromise once more if you wan to make the blue Laser diode.

Exercise 9.1-1

Quick Questions to

9.1 - Optoelectronics - General Concerns

Here are some quick questions

- Name some optoelectronic devices. List their strengths and weaknesses in comparison to competing products / technologies
- Identify and discuss specific properties of some optoelectronic materials.
- Why is the index of refraction an important property of optoelectronic semiconductors? How is it defined and what kind of numbers can you give?
- Compare the operation of a **CMOS** processor and an optoelectronic device in terms of power. What follows for some material properties?
- Why do we still use light bulbs or fluorescent light for general lighting? What are the prime conditions that optoelectronics has to meet in order to impact the lighting market?
- List and briefly discuss the advantages and disadvantages of semiconductor Lasers
- Describe some recombination mechanisms, how they impact optoelectronic devices and what can be done to optimize recombination.
- What are **OLED's**? For what kinds of products are they of prime importance and why?

Exercise 9.2-1

Quick Questions to

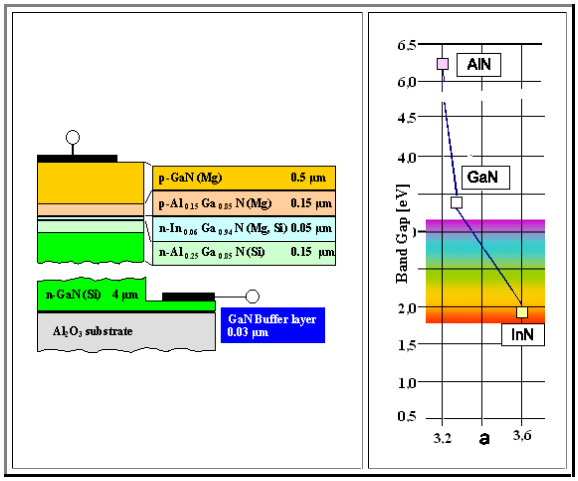
9.2 - Optoelectronics - Important Principles and Technologies

Here are some quick questions

- Describe some measures necessary of you want to produce a high-efficiency **LED**. Use hand drawings to illustrate (at least) three major points.
- How can you best define a recombination volume?
- Draw the band diagram of an **Np** junction in equilibrium with the bandgap of the **N**-type semiconductor about twice the size of the **p**-type material. Discuss choice you make if necessary.
- Discuss advantages an problems of hetero junctions for light emitting devices
- Describe the working principle of **MBE** and what it means in terms of realization.
- Describe the mechanisms of *fundamental absorption* (**FA**) and *stimulated emission* (**SE**) with the aid of a band diagram. What kind of relation between the rates R_{Se} and R_{FA}, i.e. the number of events per second (and **cm**^{2}) must you have if amplification of light is to take place?
- What is the meaning of "inversion" in the context of a semiconductor Laser?
- What is "pumping" in the context of a Laser and why is a semiconductor very well suited for efficient "pumping"?
- How can you turn a light amplifier into a Laser? What does it mean technically for processing your semiconductor?

Here is a somewhat "longer" question:

The picture shows a schematic "to scale" drawing of a simple blue Laser diode and the relevant material parameters in the "master" diagram.



- What kind of approximate band gap energy can you assign to the various layers? Draw a schematic band diagram of this kind of situation.
- Which layer is the light-producing one?
- What is the function of the three relatively thin layers in the center region?
- What could be the function of the two thicker pure **GaN** layers?
- Why is the whole structure on an Al_2O_3 substrate (= sapphire) and what is the *electrical* problem encountered?
- What might be the problem necessitating a "buffer" layer between the Al_2O_3 substrate and the stack of functional layers?
- Where are the mirrors necessary for a Laser?
- What is obviously used for **n**- or **p**-doping? Make a guess as to why the very thin central layer contains **Mg** *and* **Si** as dopants

Exercise 9.3-1

All Quick Questions to

9. Optoelectronics

General Concerns

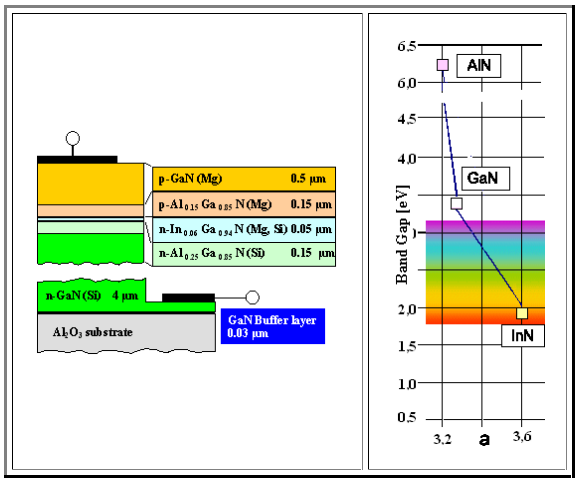
- Name some optoelectronic devices. List their strengths and weaknesses in comparison to competing products / technologies
- Identify and discuss specific properties of some optoelectronic materials.
- Why is the index of refraction an important property of optoelectronic semiconductors? How is it defined and what kind of numbers can you give?
- Compare the operation of a **CMOS** processor and an optoelectronic device in terms of power. What follows for some material properties?
- Why do we still use light bulbs or fluorescent light for general lighting? What are the prime conditions that optoelectronics has to meet in order to impact the lighting market?
- List and briefly discuss the advantages and disadvantages of semiconductor Lasers
- Describe some recombination mechanisms, how they impact optoelectronic devices and what can be done to optimize recombination.
- What are **OLED's**? For what kinds of products are they of prime importance and why?

Important Principles and Technologies

- Describe some measures necessary of you want to produce a high-efficiency **LED**. Use hand drawings to illustrate (at least) three major points.
- How can you best define a recombination volume?
- Draw the band diagram of an **Np** junction in equilibrium with the bandgap of the **N**-type semiconductor about twice the size of the **p**-type material. Discuss choice you make if necessary.
- Discuss advantages an problems of hetero junctions for light emitting devices
- Describe the working principle of **MBE** and what it means in terms of realization.
- Describe the mechanisms of *fundamental absorption* (**FA**) and *stimulated emission* (**SE**) with the aid of a band diagram. What kind of relation between the rates R_{Se} and R_{FA}, i.e. the number of events per second (and $\mathbf{cm^2}$) must you have if amplification of light is to take place?
- What is the meaning of "inversion" in the context of a semiconductor Laser?
- What is "pumping" in the context of a Laser and why is a semiconductor very well suited for efficient "pumping"?
- How can you turn a light amplifier into a Laser? What does it mean technically for processing your semiconductor?

Here is a somewhat "longer" question:

The picture shows a schematic "to scale" drawing of a simple blue Laser diode and the relevant material parameters in the "master" diagram.



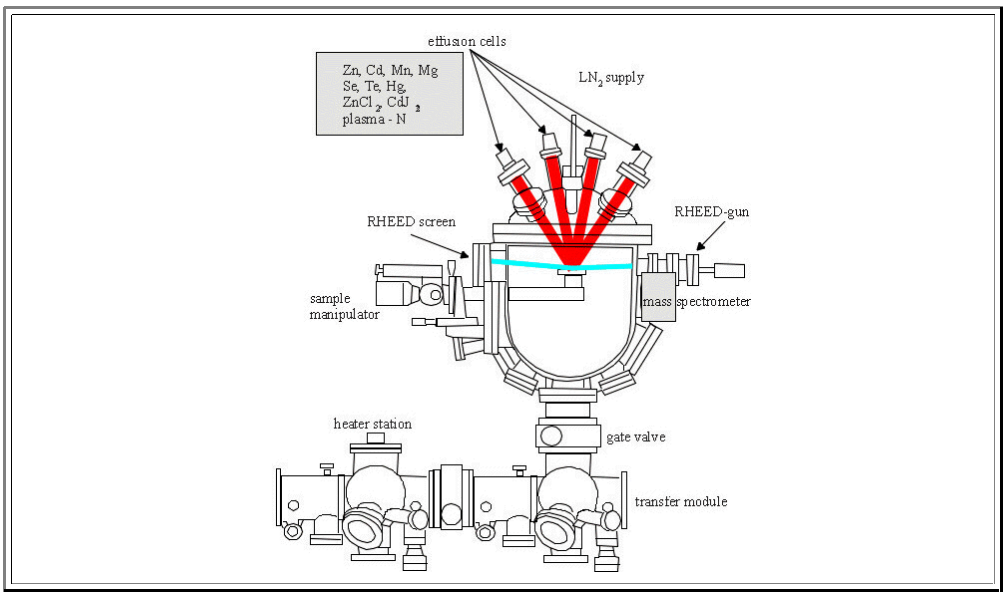
- What kind of approximate band gap energy can you assign to the various layers? Draw a schematic band diagram of this kind of situation.
- Which layer is the light-producing one?
- What is the function of the three relatively thin layers in the center region?
- What could be the function of the two thicker pure **GaN** layers?
- Why is the whole structure on an $\mathbf{Al_2O_3}$ substrate (= sapphire) and what is the *electrical* problem encountered?
- What might be the problem necessitating a "buffer" layer between the $\mathbf{Al_2O_3}$ substrate and the stack of functional layers?

- Where are the mirrors necessary for a Laser?
- What is obviously used for **n**- or **p**-doping? Make a guess as to why the very thin central layer contains **Mg** *and* **Si** as dopants

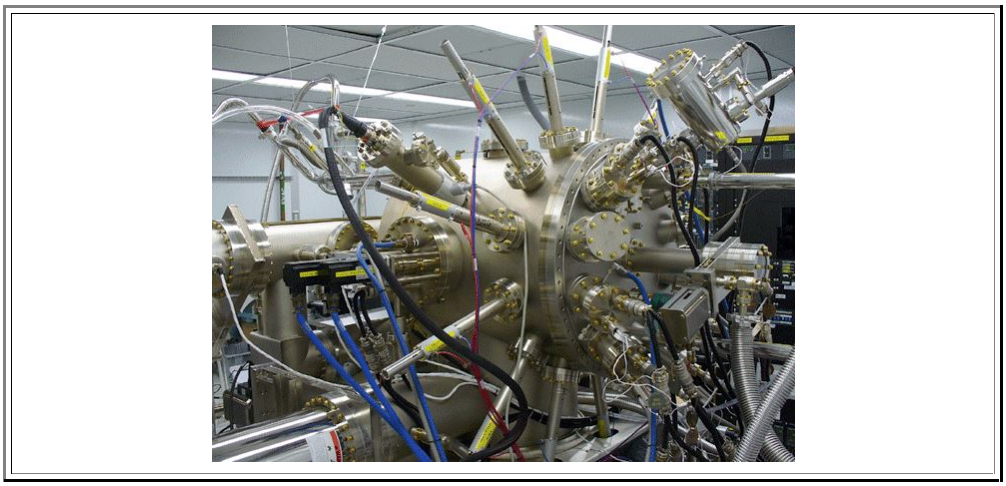
MBE

Illustration

- Here are two pictures that give an idea of what **MBE** means in reality.
  - First, just a schematic drawing showing the necessary parts.



  - Second, the real thing. The picture shows part of a **MBE** machine at the University of Delaware, used for making world-record multi-junction solar cells.



  - Same thing from the University Tokyo:

- While it doesn't look particularly cheap, it is still possible to use MBE for low-cost mass production.