1. Introduction

1.1 Scope of the Course

1.1.1 General Remarks

1.2. Introduction to the Course

1.2.1 Motivation for the Content

1. Introduction

1.1 Scope of the Course

1.1.1 General Remarks

Some Important Links

- For a detailed *table of contents* use the link
- The *organization, timetable, etc.* for the running term could be found in the link. Now you have to look elswhere.
- If you like to read *prefaces* - tough luck; there is none
- For book recommendations: Consult the *list of books*

What is Special About this Course

- The lecture course "Semiconductor Technology" has a somewhat *special status* for two reasons:
  - **1.** Nothing comparable has been done before in Germany - because this is one of the very first lecture courses taught within a "Bachelor" study course, i.e. without "advanced" parts.
  - **2.** It has a *special format* for the exercise part - we will conduct it as seminar.
  - **3.** In the fall term of **2009** it has become part of the "teaching module" **Semiconductor Technology and Nano Electronics**
  - **4.** I stopped teaching it in 2014. In 2019 it's still taught, however, by my predecessor Prof. Adelung.

1.2. Introduction to the Course

1.2.1 Motivation for the Content

General Remarks

- The course will first take a look at **products** containing semiconductor technology. Analyzing this products leaves us with **components** and finally with materials and processes for semiconductor technology.
  - *Products* we define simply as something you and I *do* buy or at least *could* buy. We also include services in this category.
  - *Components* are whatever one finds inside a product, e.g. "**Chips**", but also **light emitting diodes** (***LED***'s) or **liquid crystal displays** (***LCD***'s)
  - At least some components of our products of interest are made from **semiconductors**. What we want to learn in this course then is simply
    - What Semiconductors do we use?
    - How do we make the component we want?
- These questions go deeper then it may appear on first sight. Let's look at two examples:
- So we use **Si** for our component. But just saying **Si** is not good enough.
  - Do we use single crystalline **Si**, poly crystalline **Si** or amorphous **Si**? Or perhaps nanocrystalline **Si** with some amorphous regions?
  - If we use single crystalline **Si**, do we go for Czochralski-grown (***CZ***) or **float zone** (***FZ***) crystals, or possibly an *epitaxial layer*?
  - OK - we take the **CZ** wafer. What doping type would you like? **p** - or **n**-type? All right, we take the **n**-type, well done - thank you very much.
  - Would you prefer **P**-doping or **As**-doping? Or may we recommend today's special: **Sb**-doping? And what kind of interstitial oxygen concentration may I offer to you? We have a large selection for every taste.
  - You get the drift. And as in any good restaurant, you will "taste" the difference. What you get as a component depends on your detailed specification.
- No let's make a solar cell. From **Si** or from something else?
  - In fact, we make (and you and I can buy it) from all kinds of **Si** mentioned above, from **GaAs**, from **$CuInSe_2$**, from **CdTe**, from **TiO_2** and from a growing number of other semiconductors and combinations of different semiconductors,

---

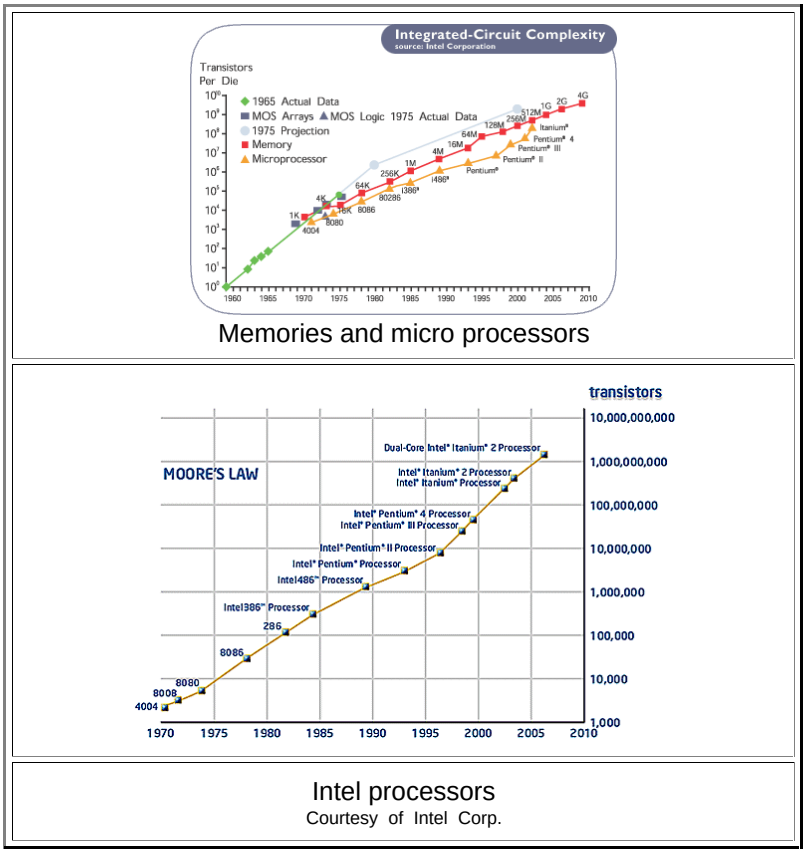
- Why, oh why are we doing this? It seems to make life so complicated. Can't we decide on the best material and process for solar cells and be done with it?
  - Well, being the boss of a large solar cell company, you actually must make this decision - you can not possible run a multitude of factories, each with its own materials and processes. **You** must make a choice for one, or maybe just two basic product lines.
  - The same is true for your competitor. If his choice is different from yours, time and in particular the market will tell, which one of you guys made the *better* choice.
- In other words: if we look at products, we do not just look at technical topics, we actually look at economical issues! **Money**, not Nobel prizes is the decisive factor in the end!

Products, Components and Materials

- As a human being, you encounter all kinds of products and services all the time - and you rarely think about what is hidden behind the obvious. You pick up your (cell)phone, dial a number or press a button, and expect that within seconds you will be able to talk to the person of your choice - whoever and wherever that person might be.
  - If a regular human being gives the "behind the obvious" any thoughts, he or she will probably conclude, in the words of **Dave Barry**, "that cell phones are operated by **magic**".
  - As a (budding) Materials Scientist and Engineer, you know better. Behind the obvious is semiconductor technology. Not exclusively, and not always, but "immer öfter" (ever more often).
- Note that not all that long ago (for elderly professors) - in the **1950** ties - the number of semiconductor products was exactly zero.
  - Now we have such a large diversity of products, components and semiconductors all around us that we can hardly do more than scratch the surface in this course.

Markets and Growth

- The very first semiconductor products intentionally made (i.e. based on understanding what is going on) hit the market in the late fifties / early sixties of the **20**th century - in the form of "transistors", a word not used for a transistor per se, but for a **transistor radio**.
  - This was an unbelievable big *product* achievement because, for the first time in human history, it enabled everybody, not just skilled muscians, to make loud noises in public. Of course, the transistor radio was an instant success.
  - A portable, battery-run "transistor" contained about **20** (**Ge**) transistors, already some progress in comparison to your big and heavy home radio, that may have contained about **10** vacuum tubes as active elements.
- Putting several transistors on *one* piece of **Si**, i.e. making an **integrated circuit** (***IC***), was the next big (double) step in technological development; it consisted of switching to **Si** as base material and in finding ways for integration.
  - Since then semiconductor technology is an unprecedented success story - it is now (**2007**) arguably the world-wide biggest industry with respect to product penetration.
- The key word in this respect is "**Moore's law**", simply stating that any quantitative measure of progress in **IC** technology grows exponentially "forever" with growth rates in the **30 %** range.
  - Typical measures are, for example, the number of transistors on one chip, almost the same as the number of bits one can store in one memory chip. What that looks like is shown in the figure below - note the logarithmic scale



Memories and micro processors

Intel processors

Courtesy of Intel Corp.

- The implications of exponential growth for by now more than **40**years are staggering - use the link if you want to have just a flavor of this. We will just look at two points in this context:
  - The world has changed in a major way in the last **20** years or so because of semiconductor technology. Think about this yourself. *Hint:* Consider what hides behind catch words like "Internet", "electronic warfare", "Resonance tomography", "globalization", "energy crisis", ..., in technical, social and political terms.
  - Exponential growth *never* continues forever. In fact, since about **1985**, serious people believed (or actually tried to prove) that it will be over soon. When it will be over, meaning that growth rates come down to "normal", all kinds of problems might occur - witness the bursting of the (stock market) "Internet bubble" in **2000** or the bursting of the USA real estate bubble right now (Aug. **2007**) - all caused by the sudden end of exponential growth. If we are lucky, we will have a "soft landing" in the **IC** business; if we are even luckier, the foreseeable slack in the **IC** business will be compensated for by growth in other areas of semiconductor technology, e.g. **solar cells**.
- Where does this leave *you*? If the **30%** per year growth rate peters off, will there be jobs? Is it sensible to learn about semiconductor technology now when it soon will be "over"?
  - For an answer, look at the German car industry. Seen with semiconductor industry eyes, technical progress in making cars during the last **40** years or so was close to zero - compared to semiconductor products. A factor of two in total performance progress (top speed, gas consumption, ..) in these **40** years is already seen as enormous achievement. Compare with memory chips: **1978: 16 kbit**, **2007: 16 Gbit**; improvement factor 10^6.

- Yes - but:. The car industry is still the largest industrial branch in Germany with lots of jobs....

2. Semiconductor Materials and Products

2.1 General Chemistry and Structure

2.1.1 General Considerations and Elemental Semiconductors

2.1.2 Compound Semiconductors

2.1.3 Some Relevant Properties and Products

2.1.4 Summary to: 2.1 General Chemistry and Structure

2.2 Silicon

2.2.1 Silicon and Microelectronics

2.2.2 Other Uses of Silicon

2.2.3 Summary to: 2.2. Silicon

2.3 III-V Semiconductors

2.3.1 III-V Semiconductors and Optoelectronics

2.3.2 Other III-V Products

2.3.3 Summary to: 2.3. III-V Semiconductors

2.4 Other Semiconductors and Products

2.4.1 Germanium and SiC

2.4.2 II-VI Semiconductors

2.4.4 Organic Semiconductors

2.4.5 Summary to: 2.4: Other Semiconductors and Related Products

2.5 Exotic Semiconductors, Processes and Products

2.5.1 Porous Semiconductors and Product Ideas

2.5.2 Sensors

2.6 Summary

2.6.1 Summary to: 2. Semiconductor Materials and Products

2. Semiconductor Materials and Products

2.1 General Chemistry and Structure

2.1.1 General Considerations and Elemental Semiconductors

Some General Statements

- All semiconductors come in a certain structural form. We will look at the possibilities by just comparing extremes:
- *Large* in three dimensions or *Small* in at least one dimension?
  - *Large:* Can you see *it* and only it (i.e. not its substrate) with the naked eye? Touch it, hold it, break it?
    Examples:
    - Single crystals of e.g. **Si** (**300 mm** diameter, **1 m** long!), **GaAs**, **InP**, **GaP**, **SiC**, ...
    - **Si** wafers (up to **450 mm** diameter, about **1 mm** thick), **GaAs**, etc.
  - *Small:*, at least in one dimension. Can you see it with the naked eye? But, maybe, not hold it?
    Examples:
    - Small in *one* dimension: All **thin films**; always on a substrate. This contains most **optoelectronic** and all "thin film" **solar cell** uses. Many usable semiconductors exist *only* as thin films!
    - Small in two dimensions: **Micro**- or **Nanowires**. A big research field right now (2007) and in use as connections on **IC**s.
    - Small in three dimensions: **Micro**- or **Nanoparticles**. Semiconducting micro- or nanoparticles are not yet part of products coming out of semiconductor technology (unless you count an integrated transistor as a microparticle which we do not), but a big research item, e.g. for "Nano" solar cells
- *Crystalline* or *Amorphous*?
  - So you need a (thin film) semiconductor on a really large area - for a **flat panel display**, for example. Or a solar cell module made from "one piece". There is no single crystal big enough for that and it would be prohibitively expensive to make on (provided you could).
  - You then must try live with a poly-crystalline thin film or, maybe with an amorphous one. Be prepared to spend some **10.000** man-hours in getting it to work.
    Class Exercise: *Ponder the history of "**LCD**" flat panel displays.*
- Here are some alternatives:

Mono Crystalline	*or*	**Poly Crystalline**
Contains dislocations and other defects *or* is (almost) perfect?		Large Grains *or* Small Grains?
Electronic parameters are adjustable *or* Fermi level pinning is observed?		Grain boundaries problematic *or* tolerable?

- By now you get the drift: This may turn out to be quite complicated. Thank God, there are some specialists who have to know all this stuff; the rest of us can forget about it and just be good consumers.
  - Right. Those specialists, by the way, are called **Materials Scientists and Engineers**. Sorry. But it will be up to you (and a few others) to **save the world** - your world.
  - **Class Exercise:** *Why does the world need saving? How shall it be done?*

Elemental Semiconductors

There isn't much. All we need to do is to look at a rather small part of the periodic table:

- Here is a part of the complete periodic table accessible by the link. Semiconductor are outlined a reddish background and big letters. The redder, the better!

IIA	The Rest		IIIA	IVA	VA	VIA	VIIA	VIII
								He
Be			B	C	N	O	F	Ne
Mg			Al	Si	P	S	Cl	Ar
Ca			Ga	Ge	As	Se	Br	Kr
Sr			In	Sn	Sb	Te	J	Xe
Ba			Tl	Pb	Bi	Po	At	Rn

What we have, in (subjective) order of importance, are

- ## Silicon (Si)
  It's so obviously of top importance that we are not going to say anything more to it at this point.
- ## Germanium (Ge)
  A true fine semiconductor. Good single crystal can be grown, doping etc. is easily possible, but the band gap is a bit too small for most applications. Far worse: There is no "good" Germanium Oxide (**GeO_2**)
  The first semiconductor put to commercial use in the early sixties - and then phased out almost completely.
  In the last few years **Ge** experiences a kind of "come back"; we take that as a reason to start an "advanced" page at some point.
- ## Selenium (Se)
  An often overlooked semiconductor. Historically of some interest, and in particular because it made "Xeroxing", i.e. photo copying possible. We take that as a reason to start an "advanced" page at some point.
- ## Diamond (C); **metastable fcc form**
  There are some technical uses (besides the obvious non-technical ones in (hetero) human relations, but nothing we have to concern ourselves with at present.
- ## Tin (Sn); α **- Sn (below 13 °C)**
  Forget it!
- ## Boron (B)
  Forget it!
- ## Phosphorous (P);
  Forget it!
- ## Arsenic (As)
  Forget it!

Are we going for ***Crystalline*** or ***Amorphous***?

- To make it short: In case of doubt we use crystals, preferably single crystals, preferably "perfect" single crystals.
  Class Exercise: *Why?*
- Applications on large areas, however, use amorphous thin films or poly crystals for obvious *(???)* reasons.

Chemical Element - Technical Semiconductor

We finally must concern ourselves a little with what exactly it is we are looking for when we consider semiconductor technology. To a small extend, we have already discussed some points of interest above.

- To make the issue clear, consider that you can buy a **kg** of e.g. "Silicon" from a chemical company like Merck. What you will get is a bottle full of a greyish powder, which will be of no use whatsoever for semiconductor technology. We are obviously looking for more than just the **element**.
- Let's look at some material parameters that are of interest to us when we want to make a product or component by doing some semiconductor technology. Here we just list key words (hoping that they will strike a chord). In the next sub-chapter we will take a closer look:

Properties	**Remarks**
Crystal	
Crystal structure	**fcc**, **bcc**, **hex**,..
Lattice constant	*Will turn out to be very important*
General structure	single-, poly-crystal, thin layer, ..
Defect densities	dislocation density, impurity concentration, ..
Defect properties	Formation-, migration enthalpies of point defects, ..
Unit weight [mol], Density [g/cm^3]	
Mechanical properties	Yield strength (as **f(*T*)**), fracture strength, surface energies, ...
Electronic Properties	
Band gap **[eV]**	*Gives n_i(**T**)*
Type	direct, indirect, dispersion function
Effective mass of electrons and holes **[m*/m]**	*Important, but beyond the scope here.*
N_{eff} in ***C*** and ***V***; n_i	*Needed for calculating **n**(**T**)*
Mobility (undoped)	*Very important for speed*
Lifetime; Diffusion coefficient of electrons and holes; Diffusion length	*Appear in any equation!*
Mechanism of luminescence	*Important, but beyond the scope here.*
Deep levels of impurities and defects	*Important if you can't be perfect*
Dielectric properties	
Dielectric constant	*Appears in most equations!*
Break through field strength	*Obviously important*
Specific intrinsic resistance	*Not so important*
Electron affinity	
Thermal Properties	
Therm. expansion coefficient	*Very important in many cases*
Therm. conductivity, Specific heat, Melting point	*of some interest, important in power applications*

Economical / Ecological Properties	
Supply / Price	*Potentially important; depends on product.*
Poisonous / polluting; directly or indirectly	***Si** you can eat; **GaAs** is poisonous!*

Class Exercise: *Supply examples for critical parameter - component couplings*

2.1.2 Compound Semiconductors

Some General Points

- The elemental semiconductors give us the all-important **Si** and, far less important, **Ge**. on the outer fringe there is **Se**. Forget about the rest.
  - Anything else coming up in semiconductor technology, by definition, must be a **compound semiconductor**. In fact, the periodic table provides for untold compounds or "Verbindungen" and among the solid ones are a lot of semiconductors.
  - How can we tell? Either by having experimental proof that there is a band gap with about **0.5 eV** - **5 eV** (those are of course arbitrary numbers) or by reliable band structure calculations for the compound in question. Both approaches might be tricky, and there are most likely compounds out there that have not yet been recognized as semiconductors. We have looked at some known semiconductors before - here is the link.
  - We have also discussed that the vast majority is of no interest to semiconductor technology at present.
    Class Exercise: *What makes a semiconductor interesting for technology?*
  - What is left (at present) is shown in the table:

Type	Examples	Remarks
Group **IV** and compounds	**Si**, **Ge** (**C**); **SiGe**, **SiC**	Only **Si**is extremely important, rest so-so
III-V compounds **Al**, **Ga**, **In**- **N**, **P** , **As**, **Sb**.	**GaAs**, **GaP**, **GaN**, **InP**, ...; $In_xGa_{1-x}As_yP_{1-y}$, ...	Some combinations are more important than others
II-VI compounds **Zn**, **Cd**, (**Be**, **Mg**) - **O**, **S**, **Se**, **Te**	**CdSe**, **CdTe**, **ZnO**, **ZnSe**	Much research, few products
Metal oxides	**TiO_2**, **ZnO**	Investigated for solar cells
Chalkogenides **A_xB_y(S, Se, Te)$_2$**	**CdTe**, **$CuInSe_2$** (**"CIS"**), **$Cu_xGa_{1-x}In_yS_ySe_{1-y}$ "CIGS"**;	All in production for solar cells
Others	Organic semiconductors	Used for small flat panel displays and as lighr source ("***OLED***")

- So we have quite a number of semiconductors that we can buy *right now* as part of a product or component. And before *you* are going to earn serious money as Materials Scientist and Engineer, there will be more, for almost sure.
- Once more we come to the crucial point: A semiconductor material in the form of a powder in a bottle (what you get when you buy it as "chemical") is almost always useless for us (exception: **TiO_2**). We must have two sets of additional properties that ensure that we can use the materials for "*semiconductor technology*"
  - There must be a market for the component or product (it must be better or cheaper than the competition). **Ge**, as an example, did not really meet this point for the last **45** years or so.
  - It must have the properties needed (e.g. be very perfect single crystal) *and* processes must exist to turn it into the component or product envisioned. This point, if enlarged upon, will turn into a long list of requirements.
- Both point together provide for a threshold that most semiconductors cannot pass. It is the (demanding) job of Materials Scientists and Engineers like *you* to make more semiconductors ***pliable*** for technical uses.

2.1.3 Some Relevant Properties and Products

Some General Topics

- Let's look at some general properties that come up for *pretty much* all technical semiconductors.
  - *From now on qualifiers like "pretty much" or "almost all" will be omitted. We simply assume from now on that there might be exceptions to "hard" rules that are given in what follows.*
  - What do we have? Let's see:
- **Bandgap**. This is, after all, what defines a semiconductor.
  - We need to know the value in **eV** and the type: direct - indirect. What are we looking for? How can we get what we need?
    Class Exercise: *How would you like your bandgap, Sir?*
- **Doping**. Ideally, you would like to be able to adjust the carrier density for both types - electrons and holes - within several orders of magnitude. Ideally, this is easy: Introduce the right concentration of defects that produce shallow levels at the band edges.
  - In reality, this may be already the end of many promising semiconducting materials. Dirty words like **"Fermi level pinning"** may come up in this context.
    Class Exercise: *What does it mean to dope a semiconductor in reality?*
- **Structure and Shape**. We already looked at this to some extent. Keyword are crystallinity, size, thin film or bulk, lattice type and constant, ..
  - **Class Exercise:** *Come up with 2 - 3 examples where product requirements transfer to shape / structure requirements.*
- **Money**. Can you afford to make it? Can you still afford to make it if your present product is hugely successful? Can you make it cheaper?
  - **Class Exercise:** *Can you still afford it if your present product is hugely successful? - What could that mean?*

Looking at Products

- Let's look at a list of semiconductor products and components and see if we can make out what specific semiconductor property is essential for the function and the commercial success of these *products*
  - **Integrated circuit** (***IC***),
    e.g. memories or microprocessors.
  - **Solar cell**
  - **Liquid crystal display** (**LCD**)
    Where is semiconductor technology involved?
  - **OLED** displays.
  - **Micro electronic and mechanical systems** (***MEMS***)
  - **Light emitting diodes** (**LED**)
  - **(Diode) Lasers**
  - **Thermoelectric devices**
  - **Sensors**
  - **Weird stuff**.
  - **Class Exercise:** *Provide examples (and criteria) for each entry.*
- And those are just *direct* products.
  - Direct semicondictor products are often not useful by themselves (who needs a chip?) but as part of end products like computers, cars, TV's, any modern machine from a toothbrush via a house-sized printing press to small-town sized power plants,...
  - Then we have the machines that one needs to make semiconductor products: Crystal growers, furnaces, plasma etchers, ion implanters "steppers", ...
- In short: Semiconductors and semiconductor technology are behind a good part of the world-wide industry. The industrial sector described above is by far the largest in the world as we know it now, accounting for **> 10^{12} €** turnover over year.
  - **40** years back, when I was your age, semiconductors and products containing semiconductors accounted for next to nothing!

2.1.4 Summary to: 2.1 General Chemistry and Structure

- Structure and size matter!
  - Mostly we need single crystals, as perfect (and as large) as possible
  - Either in bulk, or thin films
  - If thin film, substrates matter.
- For some applications (solar cell , **LCD**, ...) polycrystalline or amorphous semiconductors are used.
  - "**CIGS**" or **CdTe** for solar cells.
  - Amorphous or poly-**Si** for **LCD** transistor matrix.

Typical Si wafer:	**300 mm**, **850 µm** thick, perfect single crystal
Solar cell: Si	• Single crystalline, bulk. • Poly crystalline, large grain, bulk. • Polycrystalline, micro grain, "thick" film • Polycrystalline, nano grain, thin film. • Amorphous (plus **H**), thin film

- Important *elemental* semiconductors are **Si** and marginally **Ge**.
  - Forget **Se**, **C**, **P**, **As** and **B**.
- Compound semiconductors are important.
  - Group **IV** and compounds: **SiGe**, **SiC**.

  III-V compounds (**Al**, **Ga**, **In**) - (**N**, **P**, **As**, **Sb**). Important G**aAs**, **$Ga_xAl_{1-x}As$**, **GaP**, **InP**, ..

  Chalkogenides A_xB_y(S, Se, Te)$_2$. Important "**CIGS**" = **$CuIn_xGa_{1-x}Se_2$**.

  "**Newcomers**" like organic semiconductors, Metal oxides (e.g. **TiO_2**).
- Properties matter! Some properties are rather independent of the structure (= defects), others can be structure sensitive

Some important **Properties**	**Remarks**
Lattice type, lattice constant	**Structure independent**
Melting point, diffusion constants	
Bandgap type and energy	
Dielectric constant	
Thermal expansion coefficient	
Doping range	**Structure dependent**
Transport of electron / holes (mobility, life time, diffusion length, ..	
Unwanted levels in bandgap	

- What counts in the end are products that sell and make a profit!
  - Besides the direct semiconductor products, there are also products that contain semiconductors (**PC's**, Cars, TV's, any modern machine,...) and products that are needed to make semiconductor products (crystal growers, ovens, plasma etchers, ion implanters, ..).

Integrated circuits, Solar cells, Liquid crystal displays, Micro electronic and mechanical systems, Light emitting diodes, (Diode) Lasers, Sensors, ...

Exercise 2.1-1

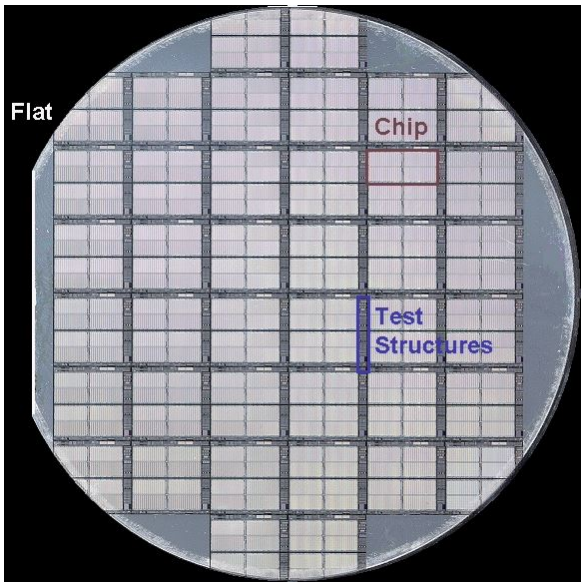
All Class Exercises to 2.1

2.2 Silicon

2.2.1 Silicon and Microelectronics

What is Microelectronics?

If you don't have some idea about what "microelectronics" means, you might be studying the wrong subject. So let's be brief: **Microelectronic** means that



Wafer (**150 mm** diameter) and chips
(**16 Mbit *DRAM*** memories from about **1990**)

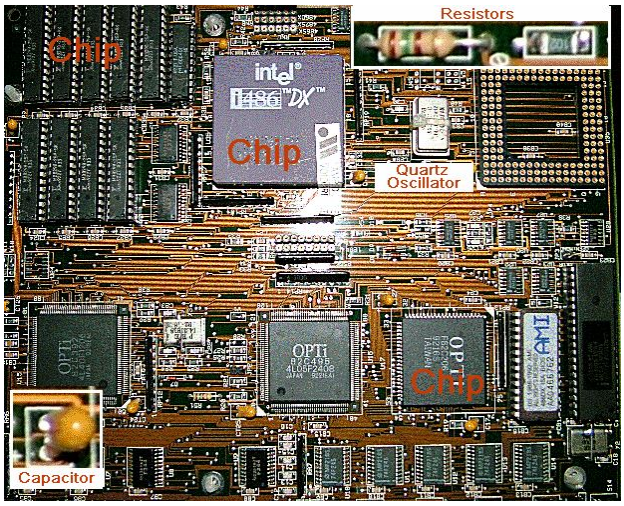
- Many basic electronic components are integrated into an **integrated circuit**, an "***IC***"
- The components of choice are mostly **transistors**, sometimes capacitors, rarely resistors, and almost never inductors.
- Integrated transistors come in two variants:

  - ***MOS* transistors**
  - **Bipolar transistors**

  MOS transistors are the most popular ones, and **MOS technology** (especially in the variant ***CMOS*** technology) is the name of the major game in semiconductor technology. You know all there is to know at present about these transistor types because you learned it before - see the links from above.
- Usually, integration means that **> 1.000.000.000** components are made in and on one piece of **Si**, about **1 cm^2** in size, called a **chip**. The transistors and other components are first isolated from each other and are then linked by conductors in such a way that a circuit or *system* results. The system could be a memory, a microprocessor, or whatever.
- Chips are made as part of a **Si wafer**, typically **150 mm - 300 mm** in diameter. Obviously you want as many chips as possible on one wafer.
- Next, single chips are cut out and "packaged". As a result you get the typical "**IC**" - e.g. a little black "brick" with many little "legs" by which it is connected (via soldering) to the circuit board.
- **Class Exercise:**

  - *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?*
  - *Where will it end?*

Here is a picture of an (old) printed circuit board with plenty of chips and other components:



Circuit board (about**1995**) with *chips* and other stuff

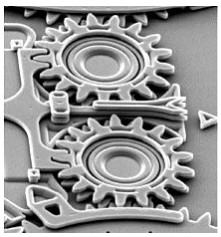
2.2.2 Other Uses of Silicon

The Obvious

- As a budding Materials Scientist and Engineer, you must have heard or read about at least two other major **Si** products. If not, do the following: i) Start reading a real newspaper and ii) Read the Science and Technology part. What you definitely should be able to come up with are.
- **Solar Cells**.
  - What do you know about solar cells? Quite a bit, actually - provided you remembered what you have learned already.
  - Let's recapitulate a few essentials you should know:
    - Maximum efficiency η and how it relates to the band gap.
    - The energy density given by the sun and how much power we can generate at high noon per $\mathbf{m^2}$.
    - The fact that we need a (**pn-**) junction to collect *minority* carriers.
    - The fact that the diffusion length ***L*** plays a major role, and that this has to do with **Si** being an indirect semiconductor
    - The ***I-U*** characteristics and how it is calculated.
    - That the only decisive parameter in the solar cell business is money.

Solar Cell

- "**MEMS**", i.e. microelectronic and micro-mechanic (and micro-optics and micro-fluidic and...) *systems*.
  - What do you know about **MEMS**? Probably not all that much from what you have learned so far.
  - **Class Exercise:** *What do you know about MEMS?*
  - To get some idea of what is going on in your immediate neighborhood in Itzehoe, check this link!

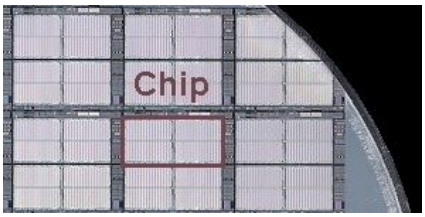


MEMS device
Courtesy of Sandia National Laboratories, SUMMiTTM Technologies, "www.mems.sandia.gov"

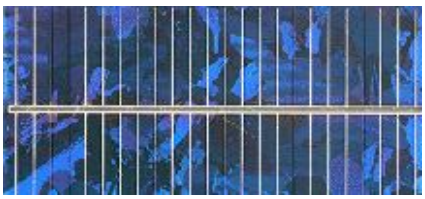
- We will come to these devices or components in due course. Meanwhile you can activate the link and look ahead a bit.
- Now ask yourself: **Class Exercise:** *Are there any other uses of Si you know off (or can find quickly)?*
  - Only after you pondered the questions above for some time, you should activate this link

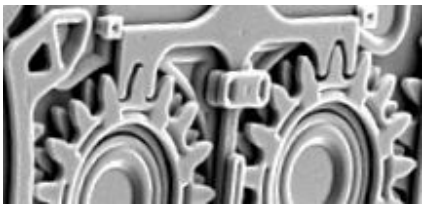
2.2.3 Summary to: 2.2. Silicon

- Silicon, and *only* **Si**, enables integrated circuits of amazing complexity, with billions of transistor on one chip
  - Two kinds of integrated transistors exist.
    - **MOS** - the absolute majority
    - bipolar - if speed counts
  - Wafers diameter are up to **300 mm** (2007), smallest (lateral) structures on a wafer are in or below the **100 nm** range.
  - Integrated circuits are packaged chips with some connections to the outside world



- Besides integrated circuits, Si is increasingly used for other semiconductor products:
  - Solar cells based on **Si** consume more Si than IC's, and demand rapidly increasing Si production. The key point of Si solar cell technology is to have high efficiencies η at low prices.
  - Microelectronic and micro-mechanic (and micro-optics and micro-fluidic and...) = **MEMS** systems find increasing uses for many tasks.





Exercise 2.2-1

All Class Exercises to 2.2

2.3 III-V Semiconductors

2.3.1 III-V Semiconductors and Optoelectronics

The Need for III-V's

- Next in importance to the elemental semiconductor **Si**, we have the **III-V compound semiconductors** obtained by combining group **III** elements (essentially **Al**, **Ga**, **In**) with group **V** elements (essentially **N**, **P** , **As**, **Sb**). This gives us **12** possible combinations; the most important ones are probably **GaAs**, **InP GaP** and **GaN**.
  - All of these **III-V** combinations crystallize either in the diamond lattice like **Si** or **Ge**, often called "**Zinc blende**" or **ZnS** structure (the term "**sphalerite** structure" is used, too), or in an hexagonal lattice known as "**wurtzite**". For your ***edification*** both structures are shown and explained in the link.
- What can **III-V**'s do that **Si** cannot do? This is an absolutely essential question for an *engineer*.
  - In your **engineer mode** (as opposed to your scientist mode) you think *exclusively* in terms of **applications** and **products**.
  - In a good first approximation, using a new material for an existing product is *only* sensible if it makes the product *better* or *cheaper* (or both). Looking at just "raw" **Si** single crystals, no other semiconductor comes even remotely close with respect to prize / performance. There are simply *no* large practically defect-free cheap wafers of other semiconductors!
  - So there must be a very compelling reason to use **III-V**'s for application that **Si** just can't hack. Obviously, this is **optoelectronics** for starters.
  - Obviously, because by now you know that **Si** has an *indirect* band gap and that means it will not **emit light**. If we want to produce **light emitting diodes** (***LED's***), we simply cannot use **Si**.
- This brings us right to the most important set of **III-V** properties: Size *and* nature of the band gap:
  - **Class Exercise:** *What would we like to have here?*
  - Let's look at what we really have - if we like it or not:

Properties	Si	GaAs	InP	GaP	GaN	$In_{0,53}Ga_{0,47}As$
Band gap [eV]	1,12	1,42	1,35	2,26	3.39	0,75
Type	Indirect	Direct	Direct	Indirect	Direct	Direct
Lattice	fcc	fcc	fcc	fcc	hex	fcc

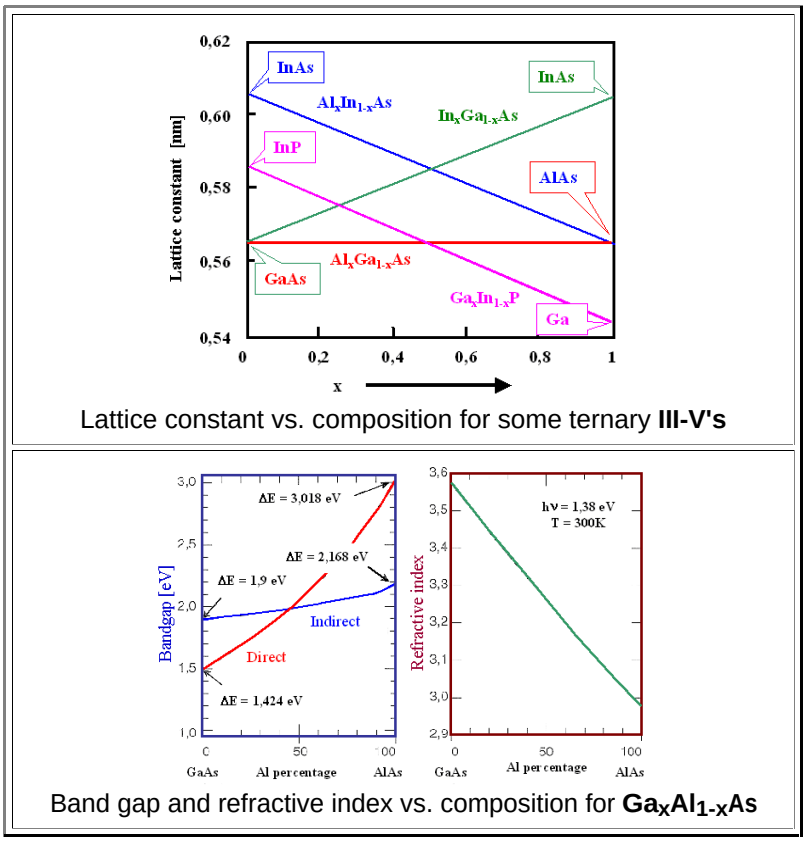
- Some questions should come up:
  - Where does that leave us with **optics** - what kind of light can we get out of these compound semiconductors? The next figure will provide the answer for this question.
  - How about **GaP**? It has an indirect band gap but is still used for making **LED**'s (just believe it)?
    Yes - there are tricks to get an indirect semiconductor to emit light. For some materials they work, for others they don't. How it is done is beyond our ken at present; things like "excitons" (one of the many quasiparticles of solid state quantum theory) "quantum wires" or "quantum dots" will be invoked. This link moves you on to a suitable module of a graduate course if you are curious.
  - Does the last column imply that we can also have mixed cases? Yes - but only for thin films
  - Is it technically important if we have wurtzite or sphalerite; how about the lattice constants? Yes and yes - this is even *extremely* important.

Ternary and Quaternary III-V's

- We have **GaAs** and **GaP** - what keeps us from *mixing*, forming for example $GaAs_xP_{1-x}$?
  - Nothing, of course - provided that the (**ternary**) phase diagram provides for such a compound.
  - We can do even better by producing a **quaternary III-V** compound with the structure $III_yIII_{1-y}V_xV_{1-x}$.
- Now we have a large research program: Find out what can be done for all kinds of combinations of **III's**, **V's**, ***x*'s** and ***y*'s**.
- Fortunately we can make a few educated guesses of what might happen; and we do that for *ternary compounds* (**III-**V_xV_{1-x} or $III_xIII_{1-v}V$) to keep it easy.
  - **Lattice constant**: As long as the lattice *type* doesn't change, the lattice constant most likely will just smoothly change from one extreme value - e.g. **GaAS** - to the other - e.g. **GaP**.

- **Band gap**: If we have no choice but guessing, your best guess would be exactly as above: The band gap probably changes from one extreme value to the other one - somehow
- **Direct - indirect band gap**: That's where guessing ends - except that for very small or very large ***x***'s we probably get whatever the pure material will have.
- **Index of refraction**: Well - we will have a smooth change, most likely.

OK - we are now ready for a few diagrams

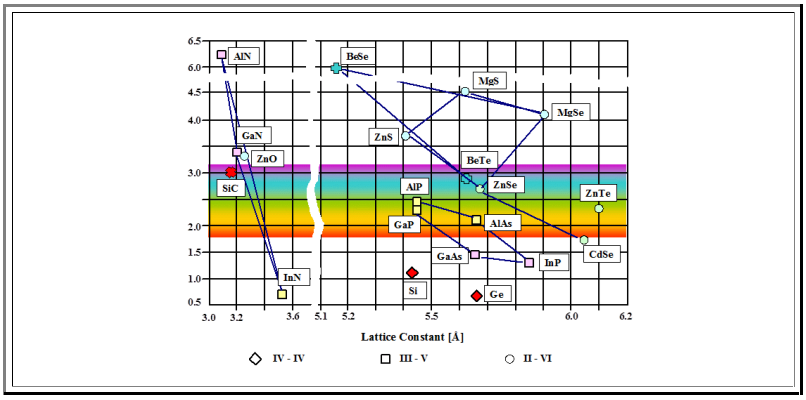


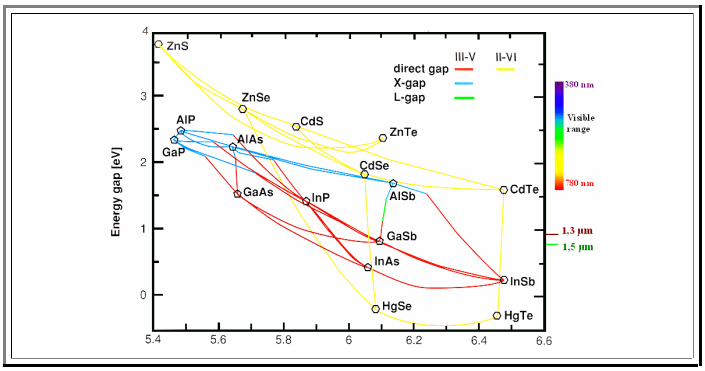
Lattice constant vs. composition for some ternary **III-V's**

Band gap and refractive index vs. composition for $Ga_xAl_{1-x}As$

- So it is pretty much as expected. Pretty smooth changes of lattice constants and index of refraction with composition, but not strictly linear with composition. Notice that the lattice constant hardly changes going from **GaAs** to **AlAs**. This will have tremendous technical consequences.
- The band gap case shows two curves - direct and indirect band gap for all compositions. Remember that in wave vector (=***k***-space you might have all kinds of band gaps and only the smallest one left after "adding up" is what we call "the" band gap. In the picture above, the "direct" band gap "wins" for ***x*** **<** ≈ **0,5**; if you go more in the **AlAs** direction, it will be indirect.

We are now ready for a first glimpse at the semiconductor "**Master Graph**":





- We see a large number of semiconductors in a band gap - lattice constant plot. We have two plots because either one is a bit restricted. The spectrum of light is schematically superimposed, to give some idea about the relation of the band gap energy to light color.
  We also see a whole range of values for **InN**. The lower one is probably the better one, as it turned out more recently. That tells us that even the most basic property of a semiconductor material is not always ewasy to assess.
- Lines connecting two semiconductors indicate that some mixture of the two is possible, and how the lattice constant and the band gap will behave. For a **HgTe - HgSe** mixture, for example, the band gap would decrease coming from both ends at first. However, the upper diagram does not have all possible lines drawn in.
- Look at the **GaAs** - **AsAl** case in the lower diagram. You see that the lattice constant does not change very much and that the band gap changes from direct (=red line) to indirect (=blue line) as soon as you get **AsAl** "heavy". What we also see it that the common red **LED** is *not* made from **GaAs** but from e.g. $\mathbf{Ga_{0.7}Al_{0.3}As}$.

It looks like we have a great selection of semiconductors and their mixtures to chose from.

- ***Alas***! Choosing is one thing, *making* the semiconductor of your choice is something else. That's were *semiconductor technology* comes in.
- In reality, only a few of the many semiconductors shown in the Master Graph could be tamed to perform by now. That's good news because it leaves something to do for *you*.

Optoelectronics - A Few Products

Optoelectronics is a formidable and strongly growing field with many facets. Here we can only look at a few products in some arbitrary selection

Light emitting diodes
Where do we find **LED**'s now, and what is going to happen to the field? There is a long story to this question, we can only look at some major points here.

- We have **LED**'s for all the little lights dotting every product that needs electricity - ***TV'***s, dashboards in cars, coffee makers, ... the list is rather long. There are two major and some minor technical requirement:

  1. Must be extremely cheap (**< 1 €** per **LED**); otherwise no mass market.
  2. Must last for many years because you cannot change or replace it.
  3. Should come in all colors (including infrared (***IR***) for remote controls)
  4. Should have low power consumption.
  5. Should have defined angular dependence of emittance.

- The first two points are *musts*; the next three may be ***negotiable***. For informations about the state of our coffee machine or **TV** we don't really need all colors. In fact, blue **LED**'s are a rather young achievement; they necessitate the mastering of **GaN**, which happens long after red **LED**'s were already ***ubiquitous***. Low power consumption is nice, but for the low intensity **LED**'s it doesn't matter all that much. For some applications you want to see that your machine is on from all angles - your **LED** thus should emit in all space angles; for some other uses you want it more directed.
- Then we have **LED**'s replacing regular light bulbs, or at least the light bulb in your flash light. In other words, they are competing against long established technology and must be cheaper or better. What are the requirements?
  Class Exercise: *Answer the question and compare your answer to the list above.*
  Class Exercise: *What is the state of the art?*
- Next we have special **LED**'s, e.g. for *infrared* light. Here is a link illustrating unexpected uses of **IR LED**'s right at the Institute of Materials Science in Kiel

So what do we need in terms of semiconductors and *optoelectronics* technology? Let's start a list.

- Band gaps in direct semiconductors with energies fitting all the photons or $\mathbf{h \cdot \nu}$ wanted - **0.5 eV - 4 eV** would be fine, for example
- High efficiencies of operation, i.e. Power in (=$\boldsymbol{U \cdot I}$) / light power out should be close to **1** meaning **100 %** efficiency.
- Absolute light power should be large ("**100** Watt light bulb")
- White light should be possible.
- (Product) Lifetime is a concern.

BNow let's look at more optoelectronic products. You do that:

Class Exercise: ***Amend*** *and discuss the list given so far.*

2.3.2 Other III-V Products

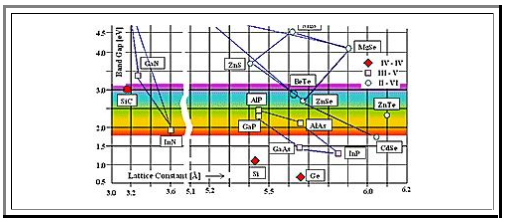
- Besides Optoelectronics, **III-V's** (mostly **GaAs**) are used for:
  - High frequency devices
  - Sensors
- We will not go into details here. A collection of short highlights will be presented later

2.3.3 Summary to: 2.3. III-V Semiconductors

- III-V semiconducrors combine the group III elements Al, Ga, In) with the group V elements N, P , As, Sb; giving **12** possible combinations.
  - The most important ones are probably **GaAs**, **InP GaP** and **GaN**
  - Band gap energies and types vary; lattices are zincblende / sphalerite (= **fcc**) or wurtzite (= **hex**).

Properties	Si	GaAs	InP	GaP	GaN	$In_{0,53}Ga_{0,47}As$
Band gap [eV]	1,12	1,42	1,35	2,26	3.39	0,75
Type	Indirect	Direct	Direct	Indirect	Direct	Direct
Lattice	fcc	fcc	fcc	fcc	hex	fcc

- Ternary and quaternary ($III_xIII_{1-x}V_yV_{1-y}$) compounds are relatively easy to make.
  - Properties like band gap, lattice constant, refractive index then adjustable to some extent.
  - Main materials for optoelectronic products. Some high-speed and sensor applications.
  - "Master diagram" = bandgap vs. lattice constant is of elementary importance for semiconductor technology.



Exercise 2.3-1

All Class Exercises to 2.3

2.4 Other Semiconductors and Products

2.4.1 Germanium and SiC

Germanium

What do we need to know about **Ge**?

- There is only one thing to be aware of: **Ge** was the material for the very first transistors in the 60ties but has not been used for many years (with a few marginal exceptions) until 2000 and later. It is, however, experiencing a sort of "come back" now - but in a tricky way.

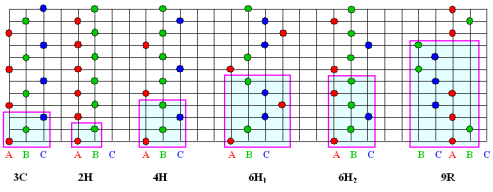
Just a short list of why **Ge** is of interest again:

- You can alloy it with **Si** - up to a point - and thus change the band gap, carrier mobilities and the lattice constant in potentially beneficial ways.
- You can use **Ge** single crystals as a substrates for growing certain layers better than on other substrates. This is enticing, e.g., for **GaAs** based space solar cells.
- You can use **Ge** single crystals for any uses where a low band gap is beneficial (detectors, sensors).

Silicon Carbide

What do we need to know about **SiC**?

- That it has quite interesting basic properties but is very difficult to produce as nearly perfect single crystal. The reason for that is that it comes in many different lattice types - the word is **polytypes** - meaning that there are many different stacking sequences of the **Si-C** base in a basically hexagonal lattice
- We will use a couple of pictures from a different Hyperscript at this point that illustrate what happens.



Polytypes of **SiC**

- All we do is to stack the building unit - **SiC** - in more tricky variants than the **fcc** and **hex** structure shown under "**3C**" and "**2H**"; the nomenclature used for **SiC**. Look up the original page for details, but you don't have to know this.
- What you should know is that **SiC** actually exists in all those polytypes (there are many more) and that given crystals may even be mixes of several polytypes.
- It is not easy at all to grow a good single **SiC** crystal in a defined polytype; unfortunately the properties depend on the polytype:

		4H-SiC	6H-SiC	15R-SiC	3C-SiC
Band Gap [eV]		3.265	3.023 3.03	2.986	2.390
Lattice Constant [Å]	**a**	3.08 3.073	3.08	3.08	4.36
	c	10.05	15.12	37.70	-
Effective Mass [m_c]	m_e	0.37	0.69	0.53 - 0.28	0.68 - 0.25
	m_h	0.94	0.92	-	-
Mobility (@ 300K) [cm^2/Vs]	μ_e	500	300	400	900
	μ_h	50	50	-	20

Thermal conductivity (RT) [W/cm · K]	3.0 - 3.8	3.0 - 3.8		

- You get the drift: **SiC** would be great for certain uses (high power, high speed, ...), but there are hardly any real products out there - despite major (usually military inspired) efforts.
  - We thus will not go into **SiC** much more but wait and see. Just one little innovation - a cheap way of making good single crystals of one polytype - would generate a new technology and a new market.
  - As a last use of **SiC** it should be mentioned that it often serves as substrate for producing thin layers of **GaN**..

2.4.2 II-VI Semiconductors

If we look back at our Semiconductors Master graph; we see a number of **II-VI's** drawn in - **ZnO**, **ZnS**, **ZnSe**, or **ZnTe**, for example. Generally, we are talking combinations of group **II** elements:

- **Zn**, **Cd**, **Be**, **Mg**,

and group **VI** elements:

- **O**, **S**, **Se**, **Te**

to name the most important.

- So we can play the same game again that we played with the **III-V's**?
- Yes and no. Yes - look at the Master graph and you see it. No - because *here* we are not interested in playing games but in products. Presently (**2007**), there are no products worth our attention. That does not mean that there aren't any, only that they are either "trivial" like resistors with a negative ***T***-dependence or simple sensors, or very special.
- In **2010** the situation has changed a bit - activate this link to get a glimpse of what is in store concerning **II-VI** technology.
- However, you should also be interested in the *science* of oxide semiconductors. The "Nano Electronics" part of this lecture course will deal with this.

Besides the **II-VI** compounds, there are also some **III-VI** semiconductors that do not yet play any role at all in technology, but who knows what we will see in some years.

- If you wonder how such a combination can form a crystal - you won't be able to form the usual ***fcc*** or ***hcp*** lattice if you think about it - you are doing fine. These compound semiconductors have a very special crystal structure, more to that in the link.

2.4.3 $CuInSe_2$ and other Chalcogenides

- There are a lot of "**Chalcogenides**", meaning compounds with "Chalcogens", i.e. **S**, **Se**, and **Te** as major elements (**O**, in the same **IIa** group, forms "oxides").
  - The general recipe is to form a **IB - IIIA - VI** compound. In group **IB** we have essentially **Cu**, **Ag**, **Au**; in group **IIIA** we find **B**, **Al**, **Ga**, **In**.
  - That allows us, for example, to produce $CuInS_2$ or $CuInSe_2$ ("**CIS**"), but also $Cu_3In_5Se_9$, $Cu_2In_4Se_7$ - look up the link for many more.
  - Like before, we can "mix", e.g. produce $CuIn_xGa_{1-x}S_ySe_{2-y}$ and so on. In case of doubt we call the whole family "**CIGS**".
  - It certainly looks like there is plenty of work left for *you*, but "**CIS**" or "**CIGS**" solar cells are actually on the market.
  - Moreover, there are "simple" chalcogenides like **CdTe**, which are on the market for solar cells but not even contained in the long list in the link from above.
- We obviously have a big success story here. We will look at some of this later in more detail.

2.4.4 Organic Semiconductors

- This is were the action is - in **2007**. **Organic semiconductors** are hot topics in **R&D**, and first products in the form of **OLED**'s are on the market. **RFID**'s may or may not follow soon.
  - Materials Science and technology for organic conductors and semiconductors is far from being well understood and there are major technological challenges, too. To give just one example: Oxygen, quite ***ubiquitous*** in air, is deadly for organic semiconductor devices. How can you keep a (cheap) device absolutely airtight for **20** years or so?
- But first things first: What exactly are organic semiconductors?
  - There is no simple answer. Essentially you need two ingredients: Some organic molecule with a **conjugated** carbon-carbon chain. This means that there is a succession of "single bond - double bond", i.e. **–C=C–C=C–C=C–C=C–** with all kinds of stuff on the one remaining free valence of any **C** atom. There also must be some "doping" because the conjugated backbone chain of the polymer molecule is (surprisingly?!) not conductive or semi-conductive.
  - **Doping** is written in quotation marks because it has nothing to do with what we have learned about doping in **Si** - except that you add some impurities to your semiconductor.
- We will come back to this topic later (if there is time). Meanwhile you may activate the folllwing links:
  - Basics about semiconducting polymers
  - The Peierls instability: Why conjugated C-chains are *not* conductive - contrary to expectation!

2.4.5 Summary to: 2.4: Other Semiconductors and Related Products

- Germanium (**Ge**) and **SiC**
  - Germanium was almost "useless" but is experiencing some comeback now (**2007**) in conjunction with **Si** technology.
  - **SiC** is very difficult to obtain as a good single crystal (many polytypes) but has some desirable properties for high speed or high power devices
- **II-VI** semiconductors are objects of heavy research but hardly used for products at present.
  - The only used material is **CdTe** for solar cells that are actually on the market. We might see, maybe, **ZnO** being used for **LED's** in the future.
- "Chalcogenides", meaning compounds with "Chalcogens", i.e. **S**, **Se**, and **Te** as major elements, are often semiconductors
  - Oxygen, in the same **IIa** group, forms "oxides"!
  - The most prominent representative of chalcogenides (besides **CdTe**) is "**CIS**" (**$CuInSe_2$**) or better "**CIGS**" (**$CuIn_xGa_{1-x}Se_2$**) used for solar cells and actually on the market.
- **Organic semiconductors**. A relatively recent addition to the club, organic semiconductors seem to have a bright future at least in optoelectronics
  - **OLED's** are on the market, in particular as part of a flat panel display; the first **OLED** based TV screen has been announced for **2008**.
  - The big problem of **OLED's** is their sensitivity to oxygen.

Exercise 2.4-1
Some quick questions to 2.4

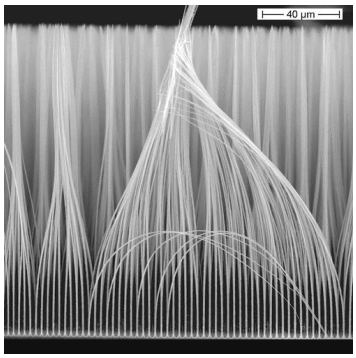
2.5 Exotic Semiconductors, Processes and Products

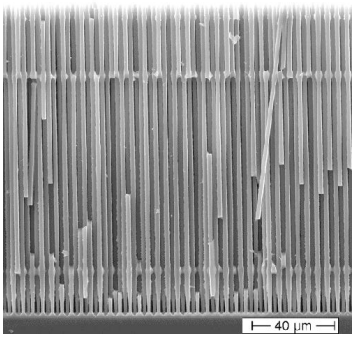
2.5.1 Porous Semiconductors and Product Ideas

Microporous Silicon

- Imagine a piece of perfect single crystalline **Si** that you have turned into a sponge by drilling holes into it that meander around like, well, just like in a sponge.
  - Now imagine that the diameter of your holes is only a few **nm**, and that the average distance between the holes is also just a few **nm**. Now ask yourself: Where is my periodic potential that I need in order to evolve a band structure? How many atoms do I need to be lined up in some periodic arrangement before I can talk about a periodic potential? Two are probably not enough, but **200** might do.
  - Tricky question. Let's simplify this a bit by considering a **quantum wire** - a **Si** crystal arbitrarily long but with a very small diameter. As long as the diameter is a few **10 nm**, nothing happens. You have a nice semiconductor, just a bit on the small side. Now decrease - in your mind - the diameter to just a few **nm**. You will now encounter "**quantum wire**" effects. With decreasing diameter the bandgap (perpendicular to the wire length) seems to increase and finally you just get a bunch of discrete energy levels - because you are loosing your periodic potential.
  - Now look back at your sponge. Between the pore, you have some quantum wire like pieces of **Si**. You must expect that the **Si** sponge behaves different from solid **Si**.
- As it turned out in **1991**, a Si sponge on a **nm** scale is extremely easy to make - all you need is a simple electrochemical cell with **Si** as the anode through which you run some current at the right conditions.
  - Your **Si** sponge actually falls into a new class of materials called "**metamaterials**"; man-made things with properties not encountered in the constituents. For reasons deeply routed in ancient chemistry, all materials with pores in the size range below **10 nm** must be called *micro*porous and not *nano*porous (s would be proper).by sone codifies convention. I know it makes not sense, but it comes from chemistry, for God's sake..
  - The properties of microporous **Si** are just amazing. To give just two:

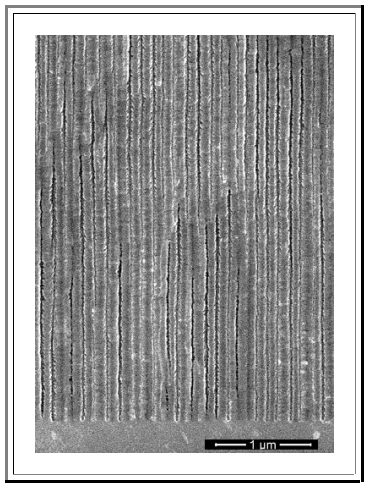
    1. It behaves like a direct semiconductor with a band gap of **1.5 eV** or so (depends on porosity), showing strong luminescence.
    2. If you put oxygen-rich stuff in the pores (e.g. **$KCLO_4$**) you have produced a high explosive with three times the bang (as measured in **kJ/kg**) than **TNT**
- Beside microporous **Si**, we have also mesoporous (**10 nm - 50 nm**) and macroporous (**> 50 nm**) **Si**; many other semiconductors can also be turned porous.
  - Porous semiconductors are objects of active research. Many possible uses have been proposed, none is on the market right now.
  - The picture below shows **Si nanowires** (actually microrwires but nowadays we call all that "nano" because it's more sexy); they were made via pore etching. The structure has been optimized for an extremely hot new application: Anodes in **Li ion batteries** with an 11-fold capacity increase relative to the state-of-the-art. If you want to know more, use the link.





Si "nanowires" made via electrochemical pore etching.

- Here is another possibly "hot" application:
  - Porous Si with a pore geometry as shown on the right has a very small **thermal conductivity**. The reason is quite simply that the wavelength of the phonons transporting the thermal energy does not fit anymore in the small space between the pores.
  - Electrons can still squeeze through, however. What that means is that the ratio of electrical conductivity to thermal conductivity increases by orders of magnitude if the pore geometry is just right.
  - This is exactly what one needs (besides some other stuff) for making efficient thermoelectric generators.
- Efficient (and cheap) thermoelectric generators are a "hot" topic right now (2010) because if we (=Materials Scientists and Engineers) can make them, "**energy harvesting**" from all kinds of hot surfaces - car exhausts for example - will be big business.



2.5.2 Sensors

I don't need to go into this here because it will be dealt with in the "Nano Electronics" part of this lecture course.

2.6 Summary

2.6.1 Summary to: 2. Semiconductor Materials and Products

- Structure and size matter!
  - Mostly we need single crystals, as perfect (and as large) as possible
  - Either in bulk, or thin films
  - If thin film, substrates matter.
- For some applications (solar cell , **LCD**, ...) polycrystalline or amorphous semiconductors are used
  - "**CIGS**" or **CdTe** for solar cells.
  - Amorphous or poly-**Si** for **LCD** transistor matrix
- Important *elemental* semiconductors are **Si** and marginally **Ge**
  - Forget **Se**, **C**, **P**, **As** and b
- Compound semiconductors are important
  - Group **IV** and compounds: SiGe, SiC

    III-V compounds (Al, Ga, In) - (N, P , As, Sb). Important GaAs, $Ga_xAl_{1-x}As$, GaP, InP, ..

    Chalkogenides $A_xB_y(S, Se, Te)_2$. Important "CIGS" = $CuIn_xGa_{1-x}Se_2$

    "**Newcomers**" like organic semiconductors, Metal oxides (e.g. TiO_2)
- Properties matter! Some properties are rather independent of the structure (= defects), others can be structure sensitive

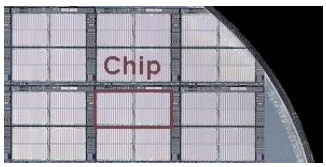
Typical Si wafer:	**300 mm** diameter, **850 µm** thick, perfect single crystal
Solar cell: Si	• Single crystalline, bulk. • Poly crystalline, large grain, bulk. • Polycrystalline, micro grain, "thick" film • Polycrystalline, nano grain, thin film. • Amorphous (plus **H**), thin film

Some important **Properties**	**Remarks**
Lattice type, lattice constant	**Structure independent**
Melting point, diffusion constants	
Bandgap type and energy	
Dielectric constant	
Thermal expansion coefficient	
Doping range	**Structure dependent**
Transport of electron / holes (mobility, life time, diffusion length, ..	
Unwanted levels in bandgap	

- What counts in the end are products that sell and make a profit!
  - Besides the direct semiconductor products, there are also products that contain semiconductors (PC's, Cars, TV's, any modern machine,...) and products that are needed to make semiconductor products (crystal growers, ovens, ion implanters, ..).

Integrated circuits, Solar cells, Liquid crystal displays, Micro electronic and mechanical systems, Light emitting diodes, (Diode) Lasers, Sensors, ...

- Silicon, and *only* **Si**, enables integrated circuits of amazing complexity, with billions of transistor on one chip
  - Two kinds of integrated transistors exist.
    - **MOS** - the absolute majority
    - bipolar - if speed counts
  - Wafers diameter are up to **300 mm** (2007), smallest (lateral) structures on a wafer are in or below the **100 nm** range.
  - Integrated circuits are packaged chips with some connections to the outside world

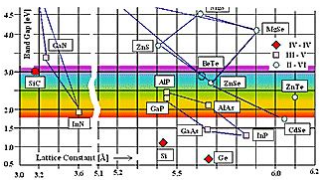


- Besides integrated circuits, Si is increasingly used for other semiconductor products:
  - Solar cells based on **Si** consume more Si than IC's, and demand rapidly increasing Si production. The key point of Si solar cell technology is to have high efficiencies η at low prices.
  - Microelectronic and micro-mechanic (and micro-optics and micro-fluidic and...) = **MEMS** systems find increasing uses for many tasks.

- III-V semiconducrors combine the group III elements Al, Ga, In) with the group V elements N, P , As, Sb; giving **12** possible combinations.
  - The most important ones are probably **GaAs**, **InP** **GaP** and **GaN**
  - Band gap energies and types vyr,; lattice are wurtzite or zincblende (= fcc) and sphalerite (= hex)

Properties	**Si**	**GaAs**	**InP**	**GaP**	**GaN**	**$In_{0,53}Ga_{0,47}As$**
Band gap [eV]	1,12	1,42	1,35	2,26	3.39	0,75
Type	Indirect	Direct	Direct	Indirect	Direct	Direct
Lattice	fcc	fcc	fcc	fcc	hex	fcc

- Ternary and quaternary ($III_xIII_{1-x}V_yV_{1-y}$) compounds are relatively easy to make.
  - Properties like band gap, lattice constant, refractive index then adjustable to some extent.
  - Main materials for optoelectronic products. Some high-speed and sensor applications.
  - "Master diagram" = bandgap vs. lattice constant: of elementary importance for semiconductor technology.



- Germanium (**Ge**) and **SiC**
  - Germanium was almost "useless" but is experiencing some comeback now (2007) in conjunction with **Si** technology.
  - **SiC** is very difficult to obtain as a good single crystal (many polytypes) but has some desirable properties for high speed or high power devices
- **II-VI** semiconductors are objects of heavy research but hardly used for products at present.
  - The "hot" contenders **CdTe** used for solar cells and actually on the market, and, maybe **ZnO** in the near future.
- "Chalcogenides", meaning compounds with "Chalcogens", i.e. **S**, **Se**, and **Te** as major elements are often semiconductors
  - Oxygen, in the same **IIa** group, forms "oxides"!
  - The most prominent representative of chalcogenides is "**CIS**" (**$CuInSe_2$**) or better "**CIGS**" (**$CuIn_xGa_{1-x}Se_2$**) used for solar cells and actually on the market.
- Organic semiconductors. A relatively recent addition to the club, organic semiconductors seem to have a bright future in optoelectronics
  - **OLED**'s are on the market, in particular as part of a flat panel display; the first **OLED** based **TV** screen has been announced for **2008**.
  - The big problem of **OLED's** is their sensitivity to oxygen.

Exercise 2.6-1
All Questions to 2.

3. Thin Films

3.1 General

3.1.1 General Remarks and Some Definitions

3.1.2 Applications of Thin Films

3.1.3 Summary to: 3.1 Thin Films - General

3.2 Mechanical Properties

3.2.1 Geometry and Topology

3.2.2 Adhesion

3.2.3 Stress and Strain

3.2.4 Summary to: 3.2 Mechanical Properties

3.3 Nucleation and Growth

3.3.1 In the Beginning

3.3.2 Nucleation and Growth Modes

3.3.3 Summary to: 3.3 Nucleation and Growth

3.4 Thin Films: Structure, Interfaces and Some Properties

3.4.1 Single Crystalline Thin Films

3.4.2 Thin Film Structure

3.4.3 Special Properties of Thin Films

3.4.4 Summary to: 3.4 Structure, Interface and Some Properties

3.5 Properties, Measurements and Characterization

3.5.1 More Properties

3.5.2 Important Techniques

3.5.3 Local Properties

3.5.4 Summary to: 3.5 Properties, Measurements and Characterization

3.6 Summary

3.6.1 Summary to: 3. Thin Films

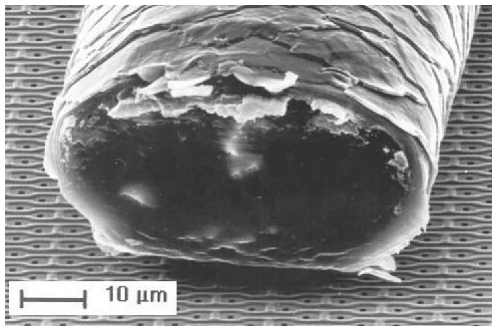
3. Thin Films

3.1 General

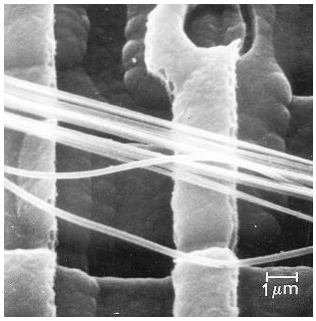
3.1.1 General Remarks and Some Definitions

The Meaning of "Thin"

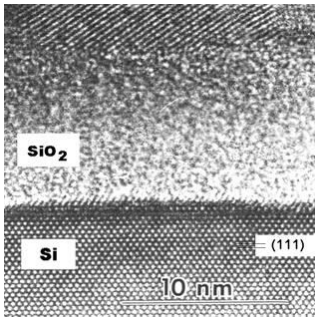
- The expressions "Thin films" and "Semiconductor technology" are almost synonyms. It is true, there is some semiconductor technology that does not need thin films but not much comes to mind right away.
  - There is, however, quite a bit of thin film technology outside of semiconductor technology, e.g., in **optics**. In fact, thin film technology is far older than semiconductor technology. In ancient times, for example, people already knew how to beat gold into a thin film (**< 1 µm** thickness) with hammers and knew how use this "**gold leaf**" for coating all kinds of stuff.
  - When you wax your car, or paint a wall, you are actually applying a thin film - *or are you*?
- How thin does a *thin* film have to be to fall under the notion of "**Thin Film**"? The "Ohring", for example doesn't tell you. The "Smith" doesn't give you a number either but offers the following working definition:
  - **Thin film technology** involves deposition of individual molecules or *atoms.*
    Thick film technology involves deposition of *particles*.
  - Painting thus is *thick film* technology, and evaporation is thin film technology. Good enough, but what about beating gold with hammers to sheets with a thickness of **5 nm**? Or depositing **100 µm** of **Ag** or **Cr** on a metal galvanically? - atom by atom, to be sure.
- All in all, there is no natural distinction between "thick" and "thin", it always has to be practical. In what we look at here, we consider in a first approximation thin films to by typically thinner then **1 µm**, or if needs be a **few µm**. Let's get an idea of what that means:
  - One of the smallest things we still can *see* and touch is a **human hair**. They come fine and coarse, but a typical thickness value is **(30 - 50) µm** as you can see below.



Human hair on top of a **4Mbit DRAM** (around **1990**)



Spider "silk" on top of memory chip.
One strand consists of several rather fine strings.

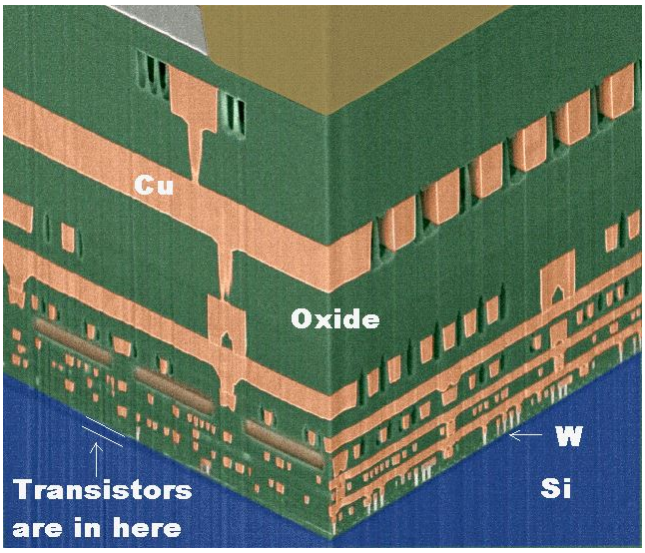


Thin layer of oxide on **Si** (with poly-**Si**on top);
atomic resolution. About **10 nm** thick

  - We we can also see the strands of a spider web. Yes - but only because what we see is a bundle of many strings with diameters **> 1µm**. An individual string would be rather invisible to us humans.
  - If we look at a **Si** wafer covered by **10 nm** of oxide we see - nothing whatsoever. It looks exactly like a **Si** wafer with no oxide; at best there might be just a hint of some greyish-brown ***hue***.
- So the word "*see*" is of interest here. If *you* can *see* it - it may not be "thin". The number to remember is the wavelength of light: **<≈ 1 µm**
  - That simply means that if its's smaller than roughly **1 µm** in all dimensions, you can't *see* it anymore.

- This is not quite true, of course, if something is only **< 1 µm** in one dimension, i.e. if we have a thin film. We all have seen **interference colors** from *thin* oil films on water; the thickness then is only some fraction of the wavelength or well below **1 µm**. Actually, the color you see from an otherwise completely colorless and transparent thin film is directly coupled to its thickness - we have a first method to actually **measure** the thickness of a thin film
- **Class Exercise:** *How was that? Interference causes the color of a thin film and betrays its thickness?* Use the link for a reminder.

OK - so if we want to "see" more than interference effects, we must use **electron microscopes**. That's a simple but costly first conclusion. Let's buy a **scanning electron microscope** or ***SEM*** (take out **€ 200.000** - **€ 400.000** from your savings account) and look at the cross-section of a modern **IC**, a cut-open chip:



Cross-section of modern chip; colors are artificial.
Picture courtesy IBM

- The first thing to note in the picture above is that **IBM** did not provide a scale. For a Materials Scientist this is not acceptable (=failing grade if *you* provide a picture without a scale). However, the size of the letter "**W**" (=tungsten) is about **1 µm** so you get the idea that there are a lot of (still) thin layers involved. There are actually far more thin films than meet the eye in this picture; just wait.

Anyway, you now have a first impression of the realation between thin films and semiconductor technology. Now let's look a bit more detailed on the meaning of "thin".

The Meaning of "Thin"

Again, we say layers are thin if their thickness $d_z < d_0$ with d_0**=5 µm** for example.

- This is a fine definition (and implicitly used a lot), but it is also arbitrary. Why **5 µm** and not **0.1 µm** or **10 µm**? Well, quite often, without thinking too much about it, d_z is scaled with other typical geometrical dimensions. If we look at a single transistor in a modern integrated circuit, its lateral dimensions are in the **1 µm** region, and we certainly would demand that d_0 must now be smaller than this if we consider *thin* films on *top* of the transistor.

From this example, we get a clue for a good alternative but qualitative definition that helps to keep our perception of thin films focussed:

- A film is *thin* if its thickness is in the same order of magnitude or smaller than some **intrinsic length scale** of the system we are considering. There is a surprisingly large number of such length scales; let's look at a few in the context of semiconductor technology:

Intrinsic length scale	Magnitudes	Remarks
Structural Scales		
Geometric dimensions $d_{x,\,y,\,z}$	Any; "Thin" if $d_z \ll d_{x,y}$	Trivial.
Changes in dimensions	$\Delta d \approx \epsilon \cdot d$	Thermal expansion; other stress / strain sources ϵ=strain
Grain size d_{grain}	**nm - cm**	Strong influence on mechanical and electrical properties

Other internal structural sizes (e.g. phases in multi-phase compounds).	From **nm to > 10 µm**	Important in proper context
Roughness amplitude.		
Interfacial layer thickness.		
Radii of curvature.		
(Average) distance between dislocations or other defects.		
Lattice constants a_0	**(0.3 - ...10) nm**	Ultimate limit. $d_z < a_0$ doesn't make sense
Wavelength Scales		
Wavelength of interacting radiation - Light (including **IR** and **UV**)	≈ **5 µm - 0.2 µm**	Determines what you "see"
- X-rays - Electron beams	**"≈" nm**	
Internal wavelengths λ - Electrons in conduction band. - Quasiparticles (phonons, excitons, plasmons, polarons, polaritons, Cooper pairs, ...	You don't have to understand that here.	What happens if $d_z >$ ≈ λ
Interaction Scales		
Absorption depths - Light - Electron beams	- **km** (glass fibers) - **nm** (metals) - **nm** - few **µm**	
Mean free paths' - Electron scattering	≈ **10 nm - 1 µm**	
Diffusion length of minorities	≈ **10 nm - 1.000 µm**	
Electrical Scales		
Space charge region width d_{SCR}	≈ **10 nm - 10 µm**	
Debye length d_{Debye}	**0.1 nm** (metal) - **m** (insulator)	
Scale of doping gradients	≈ **10 nm - 10 µm**	
Critical thickness for electrical break down	≈ **1 nm - 100 µm**	
Critical thickness for tunneling	< ≈**5 nm**	

Wow! Lots of scales - some you (should) know, some will be new. There are even more internal scales, but what we have is enough to get a feeling for:

1. "*Thin*" is indeed a relative measure.

- **2. Properties** of thin films might be quite different from that of the bulk material if that property is some expression of an internal length scale..

The Meaning of "Film"

- After we have defined (or confused) the meaning of "*thin*", we will now ponder the meaning of "*film*".
  - What we don't mean is the (thin) *film* of water on your wet windshield, nor do we mean the *layer* of dust on your furniture. While thin films of liquid might be legitimate objects of thin film semiconductor technology, and the avoidance of thin films of dust is in fact a major topic in semiconductor technology, in this lecture we concentrate on
    - *Solid* films: single crystalline, poly crystalline, amorphous; whatever.
    - *Adhesive* films: There is some *bonding at the interface*, i.e. the thin film does not easily disattach from its substrate.
- That's all. We might go a bit further and demand that the thin film has about the same thickness everywhere, and that it should be homogenous (same properties everywhere), that it should not contain holes or cracks, and so on. But this is either a matter of course or a legitimate special topic in thin film technology that needs to be treated on its own merits.

3.1.2 Applications of Thin Films

Applications Outside Semiconductor Technology

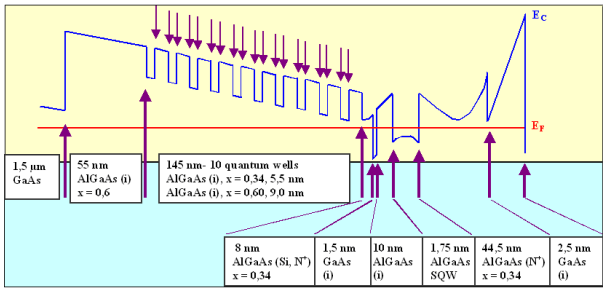
Let's first look at applications of thin films outside of semiconductor technology - so we know and then can forget it for the time being. What we have, very briefly and not exhaustively, is

Application Field	Examples
Optics	Antireflection coating; on lenses or solar cells, .. Reflection coatings for mirrors. Coatings to produce decorations (color, luster, ...), Interference filters. **CD's**, **DVD's** and upcoming **D's.** Waveguides. Photosenistive coating of "analog" film for old cameray
Chemistry	Diffusion barriers. Protection agains corrosion / oxidation. Sensors for liquid / gaseous chemicals.
Mechanics	"Hard" layers (e.g. on drill bits). Adhesion providers. Friction reduction.
Magnetics	"Hard" discs. Video / Audio tape. "***SQUIDS***"
Electricity (without semiconductors)	Insulating / conducting films; e.g. for resistors, capacitors. Piezoelectric devices

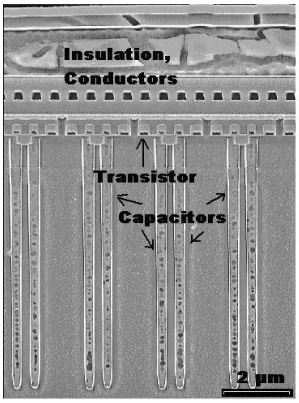
- You should know some of this stuff from experience (do your glasses have an antireflection treatment? an antiscratch layer?) or from your studies.
- For some other applications you may easily guess where thin films come in (remember the formula for the capacity of two plates with a dielectric in between? The thickness or better thinness of the dielectric does play a crucial role, after all).
- Some others may be totally unknown, but no matter: Thin films do play an important role in many branches of Materials Science and Engineering, and a lot of what we learn in this course can be directly transferred to those applications.

Illustrations of Applications in Semiconductor Technology

Let's just look at a few pictures of thin films in semiconductor technology to get a first flavor of what we are up to. Nothing more needs to be said.



The blue line shows the conduction band of a modern Semiconductor Laser. It is what you want for superior perfomance.
It follows from the sequence of **>25** thin films deposited on top of each other, starting from the **1,5 µm GaAs**, as shown.

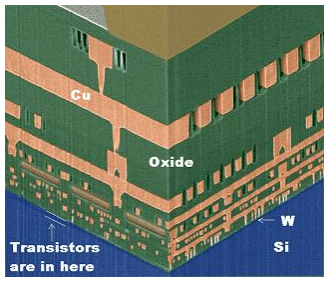


SEM cross section through a **64 Mbit DRAM** (**1996** or so).
The "holes" (idiotically called "trenches") contain the capacitor.

TEM picture of a cut through a "trench" containing
one capacitor and **6** thin layers .
"**ONO**" = Oxide-Nitride-Oxide layer sequence;
about **3 nm** per layer.

3.1.3 Summary to: 3.1 Thin Films - General

- Semiconductor technology is almost synonymous with thin film technology
  - A thin film is adhering to a substrate and (at least orginally) continous.
  - Thin films may still be found in the product or may have been "sacrificed" during the making of the product.
  - An **IC** is a study of thin films in and on the **Si** substrate.
  - The same is true for pretty much every semiconductor product.



- Thin always means "thin" relative to some intrinsic (internal) length scale. Examples are:
  - Structural length scales
  - Wavelength and interaction length scales
  - Transport parameter length scales
  - Electrical scales

• Dimensions $d_{x, y, z}$ • Grain size d_{grain} • Lattice constants a_0 • λ radiation (light, **IR**, **UV**) • Absorption depths • Mean free path lengths. • Diffusion length	• **SCR** width d_{SCR} • Debye length d_{Debye} • Critical thickness d_{crit} for electrical break down • Critical thickness d_{tu} for tunneling

- There are many thin film applications outside of semiconductor technolgy:
  - Optical, electrical, chemical, mechanical, magnetical technologies use thin films

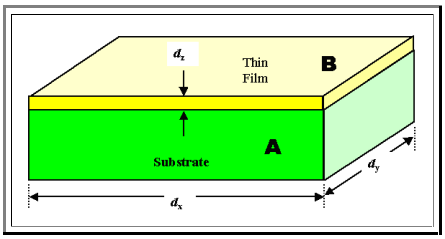
Exercise 3.1-1

All Questions to 3.1

3.2 Mechanical Properties

3.2.1 Geometry and Topology

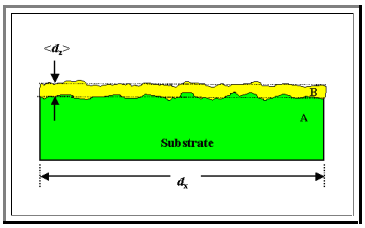
At this stage, when you think about a "thin film" you probably have this picture in mind:



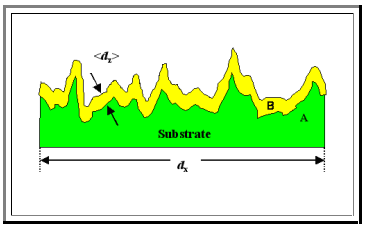
- Some solid Material **B**, with thickness d_z supposed to be "thin", on top of some substrate **A** having a lateral extension $d_{x,y} >> d_z$.

That is fine, but now let's look at some thin films with a more involved **geometry** or **topology** as the case might be:

In the first picture we have a more realistic situation, The surface of the substrate **A**, onto which we deposit our thin film **B** is **rough**. This is certainly realistic, because complete absence of roughness would mean atomically flat, which is not impossible but hard to imagine in a real world.



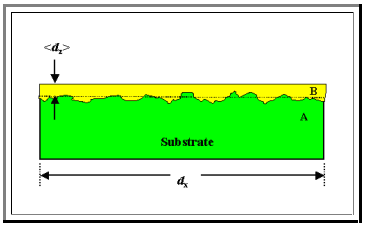
- Note that the roughness of the interface and the roughness of the thin layer surface may be correlated (as drawn), but this must not necessarily be so. The link provides an example for **$NiSi_2$** (thin) layers on a **Si** substrate where the interface roughness and the surface roughness is quite different.



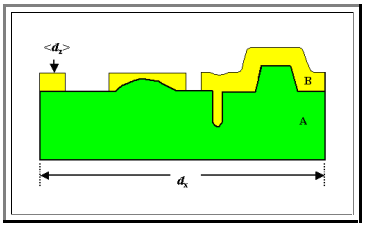
- Two questions come to mind

1. How do we define and measure roughness?
2. How rough will it be in typical situations?

You know the answer to the first question from your Lab classes:



- Measure the deviations z_i at regular intervals i from an ideally flat average surface (dotted line in the drawing; note that you may have two surfaces or interfaces with different roughness) and take the "**root mean square**" (***RMS***) for the **roughness *R***, i.e.



$$R = \left(\frac{1}{N} \sum_{i=1}^{N} z_i^2 \right)^{1/2}$$

- Usually you can't do that, so you measure from an arbitrary ideal surface *somewhere,*. Then simply subtract the *average*, i.e. use $(z_i - \langle z_i \rangle)^2$ in the sum. Simplified, we might also use an **average roughness R_a** according to

$$R_a = \frac{1}{N} \sum_{i=1}^{N} |z_i|$$

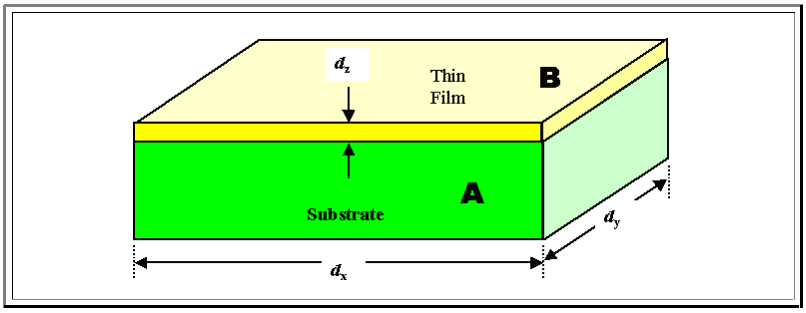
- What kind of roughness do we find (or want) in typical situations? There is a clear and simple answer: It depends! We will look at that whenever we encounter it in semiconductor technology.
- By the way, the **RMS** of the roughness may be much larger than the thickness of the thin film. The second picture shows just that - and it would be a good illustration for certain solar cells where d_z would even be considerably smaller!

- The third picture shows a case where there is some interface roughness, but the surface is quite smooth.
  - Well - why not? Realistically, however, the surface is quite smooth because you made it so - by a special process called "chemical mechanical polishing" (***CMP***).
  - Actually, what we see here in a highly stylized form is the *key process* for integrated circuits after, let's say, **1995**. If you look at the picture of the IBM chip shown before, you could see exactly that: A thin layer of oxide was deposited on a rough substrate (with **Cu** wires on it), but the surface at the oxide is perfectly flat.
- The last picture show some aspects of thin films in semiconductor technology.
  - Our thin film may be flat on top but accommodating "roughness" in the substrate (on the left), *conformally* following the substrate (and then completely cover "little" holes; on the right), and it may be structured, i.e. having defined "holes".

- All in all, just defining, describing and measuring "geometry" and "topology" (and ignoring exactly what the difference is between those two terms) can be a demanding task.
  - However, we will not waste time and brainpower to delve into this topic more deeply but will tackle it whenever it comes up in semiconductor technology.

3.2.2 Adhesion

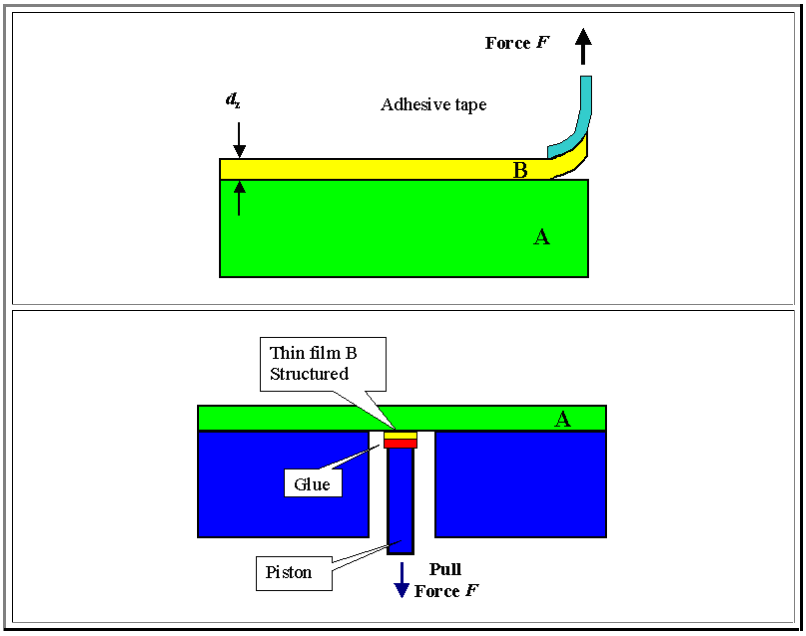
- Once more, what we want to have - idealized - is shown below. A thin film of a material **A** on top of material **B**. Of course, as we have seen, the interface (and thus the surface of material **B**) does not have to be perfectly flat but can be somewhat rough or even extemely rough.



- But *now* we look at the quite important parameter of **adhesion**. What does that mean?
  - It is obvious. Thin films **A** that do not *adhere* to their substrate **B** will not be of much use, since they will come off with little force.
  - We have a first problem. Let's say you want a thin film of **Teflon** on top of your semiconductor, or whatever. Maybe not *you* personally, but lots of other people would just love to have that. Unfortunately, as we all know, nothing *sticks* to Teflon. In other words: The **sticking coefficient** of Teflon molecules (whatever exactly that will be) to any substrate is close to zero.
  - There is indeed such a thing as a well-defined sticking coefficient of **B** on **A**, and we will come back to this.
- Adhesion is thus a relatively clear thing, easy to conceive, and what we worry about now is how to define and measure it in some quantitative way. This is relatively easy, too - in principle.
  - A direct measure of the amount of adhesion that we have for some interface is the work or energy we need to employ per **cm^2** to remove **B** from **A**. This is nothing but the good old concept of surface and interface energies, applied to thin films, i.e. interfaces between two different materials. The questions now coming up are:
    - Can we calculate the interfacial energy between **A** and **B**?
    - Can we measure the interfacial energy between **A** and **B**?
    - Is it always the same, or does it depend on, e.g., the thickness d_Z or the detailed structure of the interface between **A** and **B**?
- Let's look at the answers:
- *Can we calculate the interfacial energy between **A** and **B**?*
  - Not really. Or better put: That depends on many things. If we know the exact nature of the bonding between **A** and **B**, and if we can do all the calculations required, we will get relatively good results.
  - But even without calculations, "theory" can give us a lot of important inputs. The interfacial properties, as far as adhesion is concerned, come just as much from the bonds and their binding potentials in the interface between **A** and **B**, as most of the mechanical properties of pure **A** or pure **B**. Essentially the same rules apply.
- A few examples for this:
  - If you deposit a metal **B** on top of a metal **A** and you know from the phase diagram that these metals are completely ***miscible***, you know that you will have no adhesion problem. Complete miscibility, after all, necessitates that the bond strength between **A-A**, **B-B**, and **A-B** is not too different, so your **A-B** bonds in the interface should be just about as strong as those in **A** and **B**.
  - The same kind of thinking will tell you that **Si** on **Ge**, for example, or any **III-V** semiconductor on top of any other covalently bonded group **IV**, or **III-V** semiconductor should give good adhesion.
  - The same kind of thinking, however, will also teach you that if there is just *one* molecular layer of "dirt" in between **A** and **B**, you might be up shit's creek, as the saying goes, because you don't know how **A** and **B** bond to "dirt".
  - Everyday experience also tells you that if there is not a continuous layer of dirt in between **A** and **B**, but just some dirt particles (called e.g. dust), you simply will not get **A-B** bonding either. And now you even realize that just a few particles of less then **1 nm** in diameter or some roughness of the two layers before they make contact will already be enough to prevent bonding between **A** and **B** on an atomic scale. After all, if you put a macroscopic piece of metal **A** on top of a metal **B** the adhesion will always be zero.
  - Things of any kind, brought into what we think is "close contact" do practically never stick together, and we sure are glad for this!
  - That leads straight to a trivial, but important insight: Not counting some exotic techniques ("Wafer bonding"), we *never* produce a thin film by "putting" the film **B** on **A**.

Thin films are always *grown* on their substrate

- Now to the second question: *Can we measure the interfacial energy between **A** and **B**?*
  - Yes, of course - but... . Let's look at the "standard" techniques first. Essentially you try to delaminate your layer by "pulling" or "pushing". Measuring the force needed to do that gives you a quantitative number that you may be able to convert into the interfacial energy.
  - The simplest technique is to attach an adhesive tape ("Tesa") to your layer. If there is better adhesion at the interface tape- layer than at the interface layer-substrate, you might be able to pull off your layer as shown. Or you make a hole in the substrate and push up with a small piston, or you pull down the piston after glueing it to the thin film.



  - If you can structure your layer as shown in the second picture, you could glue some "piston" to it and try to pull the layer off
- If that looks pretty unconvincing - that's because it is! Imagine doing that to a **5 nm** film. You won't even be able to see if anything is doing.
  - Nevertheless, these techniques often work, and, of course, we can conceive of more sophisticated versions, too. But quite often you must admit defeat. You just can't measure the interfacial energy of your thin film easily or at all - in particular if their is very good adhesion, i.e. a high interfacial energy.
  - The good news in this case, of course, is that you don't care much then.
- We have one question left: *Is the interface energy always the same*, or does it depend on, e.g., the thickness d_z or the detailed structure of the interface between **A** and **B**?
  - What do you expect? Of course it does, at least to some extend. A soon as we look at epitaxial growth, we will see what might happen in interfaces and how it relates to its energy.

3.2.3 Stress and Strain

Thermal Stress and Strain

- You have made a **200 nm SiO_2** layer on you **100 mm x 100 mm Si** substrate (it is easier to consider it to be it square here) with some process at **1.000 K** and now you cool it down to room temperature (**=300 K**).
  - **SiO_2** on **Si** is one of the most frequently made thin layers. But you will find that you now have a tremendous stress in your **SiO_2** film.
  - The same is true for pretty much any thin film made at a temperature different from room temperature.
  - The question coming up now is obvious. *Why* stress, and *how much* stress?
- The "*why*" is easy. Upon cooling, substrate and thin film will shrink by some factor given by their **thermal expansion coefficient** α, a property we have encountered before and seen that it comes straight from the bonding potential.
  - The thermal expansion coefficient α was defined as

$$\alpha = \frac{l(T) - l_0}{l_0 \cdot T} = \frac{\epsilon_{therm}}{T}$$

  - l_0 is the length at the reference temperature, usually room temperature. Of course, α might be a function of the temperature; more generally we would define it as $\alpha(T)=d\epsilon_{therm}/dT$.
- It follows that two different materials with an α_A and an α_B, having the same length l_0 at some T_0, will differ in length at the temperature T by some $\Delta l(T)$ directly proportional to the **mismatch in the thermal expansion coefficient** and given by

$$\Delta l(T) = l_0 \cdot \Delta T \cdot (\alpha_A - \alpha_B)$$

  - Dividing $\Delta l(T)$ from above by l_0 obviously describes a strain ϵ_{TF}. The subscript "TF" stands for "Thin Film" since the strain (and the stress) will occur almost exclusively in the thin film, as explained below. We have

$$\epsilon_{TF} = \Delta T \cdot \Delta\alpha$$

- Let's suppose that we cover a **100 mm** Si wafer with some oxide ar **700 °C**. What happens when we cool down?
  - Fierst we need numerical values for α, which we can easily find - Wikipedia, for instance, gives us
    - $\alpha_{Si}=3 \cdot 10^{-6}\ K^{-1}$
    - $\alpha_{quartz}=0.6 \cdot 10^{-6}\ K^{-1}$

    More values are given in the link
  - Taking these numbers we get:
    - $\Delta l(T)_{Si}=100 \cdot 700 \cdot 3 \cdot 10^{-6}\ mm \cdot K \cdot K^{-1}=0.21\ mm$
    - $\Delta l(T)_{SiO2}=100 \cdot 700 \cdot 0.6 \cdot 10^{-6}\ mm \cdot K \cdot K^{-1}=0{,}042\ mm$
  - In words: The **Si** substrate would like to be **0,17 mm** smaller than the **SiO_2** layer after cooling down to room temperature.
  - However, since **SiO_2** adheres very strongly to **Si**, both materials must have the *same* lateral size and there is now a contest of strength. The Si tries to pull the **SiO_2** layer to a smaller size, the **SiO_2** layer pulls back, trying to keep the **Si** larger then it wants to be.
  - There can be no doubt about who will be the winner. Since there is far more **Si** than **SiO_2**, silicon "wins" and the **SiO_2** layer is simply forced to shrink more during cooling than it would like to.
  - The **SiO_2** thus is forced to the same lateral size as the **Si** substrate and builds up compressive stress that produces a "shrinkage" of just **0.2 mm** at room temperature. It will be under compressive stress, in other words.
- We can generalize to all thin films: A difference $\Delta\alpha$ of the thermal expansion coefficients between substrate and thin film material and a temperature difference ΔT relative to a stress few state produces a strain $\epsilon_{TF}=\Delta T \cdot \Delta\alpha$ in the thin film.
  - If that strain is purely elastic, it will lead to a stress σ_{TF} in the thin film given by

$$\sigma_{TF} = Y \cdot \epsilon_{TF} = Y \cdot \Delta T \cdot \Delta\alpha$$

- ***Y*** is, of course, ***Young's modulus*** ("Elastizitätskoeffizient").

Once more: why is the stress in the thin film? Well, you try to sit atop a huge steel bridge and stop its thermal expansion relative to yours by pulling hard. You are not going to win, neither is a thin film on a thick substrate. However, this is only true up to a point that we will discuss below.

- For our example we get with Y_{quartz}**=70 GPa**

$$\sigma_{quartz} = 70 \cdot 10^9 \cdot 0.2\ mm / 100\ mm = 140\ MPa$$

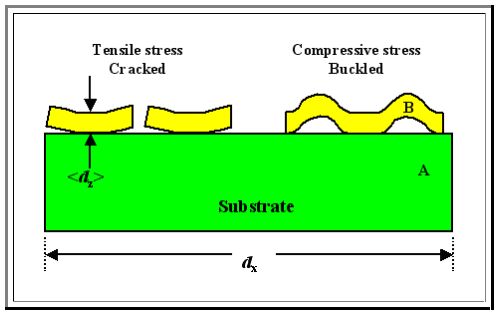
- This is a lot of stress! It means that you have to press on the tiny cross section of $(100 \cdot 2 \cdot 10^{-4})\ mm^2 = 2 \cdot 10^{-2}\ mm^2$ of our thin film with a force $F = 147 \cdot 10^6\ N/m^2 \cdot 2 \cdot 10^{-8}\ m^2 = 2.94\ N$.

Stress Relaxation Mechanisms

Now let's generalize some more: We know now that if we change the temperature of a substrate - thin film system, we generate stress. We also know now that we might generate a *lot* of stress.

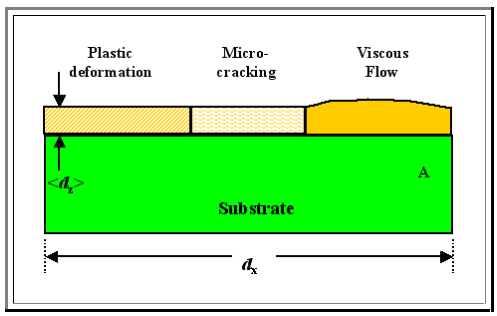
From before, we know that no material will be able to sustain arbitrarily large amounts of stress without something happening. We now ask: What is going to happen? How will the stress in a thin films be relieved, or, to use the common expression: what are the mechanisms of *stress relaxation*?

There are many "easy" answers *in general*. Unfortunately, the answer for your particular system may not be easy to obtain. Let's look at some special cases:



1. The adhesion is not very good.

- If your adhesion is not very good, your film may simply come "loose" - in total or in parts. Wherever it does not adhere to the substrate anymore, the stress can be completely relieved.
- The criterion, as always, is that if the energy gained by stress relieve (=½ $\sigma \cdot \epsilon$ per volume times thickness) is larger than the interface energy, it "pays" to rip off. Obviously, this scales with the film thickness.
- The problem is that the energy balance of a mix of partially coming off, partially still sticking, is not that easy to calculate. First you have to consider if you have tensile or compressive stress, and then what additional energy terms come into play
- **Fracture** or cracking in the tensile stress case, or "**buckling**" in the compressive stress case, also takes some energy. What you will get thus depends on many things. But whatever you get - *you probably won't like it!*

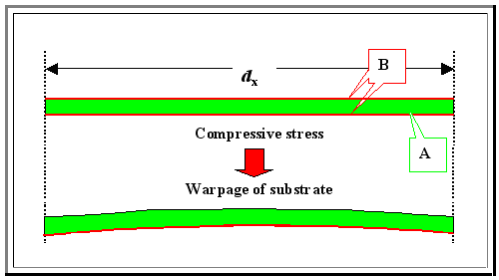


2. The adhesion is very good - the interface does not "give". This means that you will find a fully adherent thin film *and* still a lot of stress. But not necessarily the full stress you would calculate!

- There are some stress relieve mechanisms without the layer coming off that may become activated and reduce the stress in your thin film.
- Most simple, of course, is plastic deformation if your thin film is crystalline. As soon as you exceed the ***critical yield strength*** or shear stress, you will get plastic deformation. **Dislocations** will move; if too few are present, some will be generated.

- However, for plastic deformation to occur, at least one of your two materials must be ductile and not brittle. At this point it is important to note that perfectly brittle materials at room temperature (like **Si** or **Ge**) might be quite ductile at high temperatures. Note that some plastic deformation may also occur in the substrate.
- If your thin film is brittle at all temperatures, it might "microcrack", whatever that means. The link provides some insights into the some puzzles investigated right now in this context.
- In the case of our example from above, where we have produced *amorphous* **SiO_2**, plastic deformation by dislocation movement is impossible by definition. We might have plastic deformation by **viscous flow**, however.
- In fact, amorphous **SiO_2**, being a glass, is somewhat viscous at our chosen temperature of **1.000 K**; and some of the stress generated during cooling down will be relieved

Whatever happens, chances are that you will have some appreciable stress left after cooling down because the stress in the this layer could not be fully relaxed. We now give a quick glance of what might happen to the substrate in this case.



- Now me must look at little bit closer on how we made the thin layer. In the example from above it would have been thermal oxidation - we simply stick the **Si** wafer in a furnace at **1.000 K** with pure **O_2** inside. An **SiO_2** layer will grow in essentially the same way as a rust layer grows on iron. We look at this quite closely in the context of making **IC's**.
- The thing to realize here is that we do not just get an oxide layer on the wafer surface, we get exactly the same one on its backside. Whenever you make a thin layer, you always must ask yourself from now on: Will it grow just on one side or on both sides? What would be better?
- But now you have an oxide layer, being under considerable compressive stress, on both sides. Otherwise your square of **Si** looks perfect.
- Clearly, if you want to make an *integrated circuit*, you don't want or need an oxide layer on the backside. So you take the backside layer off - stick it in hydrofluoric acid (**HF**) for example, with the front side protected by some layer of wax you painted on.
- Big mistake! Now your whole substrate will be warped - you have just encountered the phenomenon of "**warpage**".

Warpage can be a serious problem in thin film technology. The reason for it is clear. Your thin film exerts a two-axial normal force on any pixel of your substrate, and your substrate will respond somehow.

- If it doesn't deform plastically (not possible for e.g. **Si**, ceramics, and so on) it will respond by elastic deformation. It may not respond a lot, but respond it will.
- For a (round) wafer with thickness d_A and Young's modulus Y_A, on top of which a thin film with thickness d_B has been deposited, that is under a stress σ_A, we can make an easy educated guess of what we would find for the **radius of curvature** ***R***. Remember: ***R*** is *small* if the wafer is heavily warped, and *large* if it is almost flat.
- What we must expect is
  - $R \propto 1/\sigma_A \cdot d_B$ because $\sigma_A \cdot d_B$ gives the force bending the wafer.
  - $R \propto Y_A$ because Y_A is a measure of how much the substrate will "give" for some force.
  - $R \propto d_A^2$ because the bigger d_A the less it will bend. That ***R*** is proportional to the square of d_A may not be obvious right away (the "Wolf and Tauber", by the way, give a wrong equation without the square), but a radius of curvature in this case is necessarily defined in two dimensions, so we must have $d_A \cdot d_A$.
  - ***R*** must not depend on $d_{x,y}$ because for a radius of curvature this dimension doesn't matter - at least as long as $d_{x,y} >> R$.
- What we will get by really deriving the radius of curvature ***R*** by going deep into elasticity theory is the "**Stoney**" **formula** (cf. e.g. "Ohring" p. 727).

$$R = \frac{1}{6\sigma_A \cdot d_B} \cdot \frac{Y_A \cdot d_A^2}{(1-\nu)}$$

- So, besides ν, ***Poisson's ratio*** coming in (as always when considering two-dimensional elasticity problems), with $1-\nu$ coming out to an almost negligible $\approx$ **0.7**, the only thing we missed is the factor **1/6**.

One thing we can learn already here. When we consider **IC** technology we will see that we make a lot of layers on the **Si** backside, where we don't need them. We will take those layers off - but at the exactly right moment in time, when we can afford to loose the "counterbalance" they offer in terms of stress and strain management.

Stress and Strain in Thin Films Not Coming From Temperature Differences

You do not necessarily need a temperature difference to generate stress in your thin film. Your method of depositing or making the thin film may already produce *intrinsic* stress and strain.

Since there are many different ways to produce thin films, we will not go into details here. We will just look into two extreme cases:

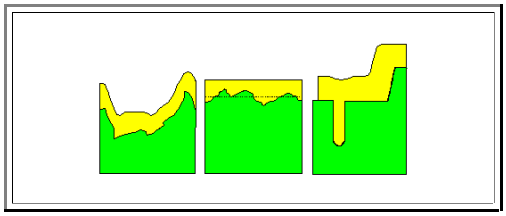
- **1.** Producing stress is an integral part of the film generating method. This is true, for example, if the thin film is produced by a chemical reaction with the substrate. The paradigmatical example is **Si** oxidation as outlined above, where a certain volume of **Si** is converted into a different volume of oxide (almost twice as large). This is just not possible without generating considerable stress.
- **2.** The stress produced is a function of many parameters. You may or may not understand the exact relationship between your deposition parameters and the stress produced, but you may now start to optimize. Note that you do not necessarily want stress-free layers. What you want is control!

The long and short of this paragraph is a simple message:

Stress management is an integral part of thin film technology

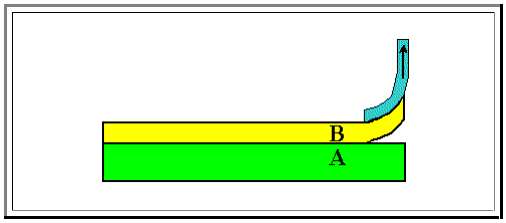
3.2.4 Summary to: 3.2 Mechanical Properties

- Thin films have other spatial properties besides their thickness, i.e. roughness
  - Interface roughness and surface roughness ***R*** defined by their "root mean square".



$$R = \left(\frac{1}{N} \sum_{i=1}^{N} z_i^2 \right)^{1/2}$$

- Useable thin films adhere to their substrate.
  - A direct measure of adhesion is the interfacial energy γ_{AB} between film **A** and substrate **B**.
  - The phase diagram provides some guideline. Complete miscibility = good adhesion, (eutectic)) decomposition =(?) low adhesion. Calculations of γ are difficult.
  - Full adhesion can only be obtained for films *grown* on a substrate. Adhesion energies can be measured.



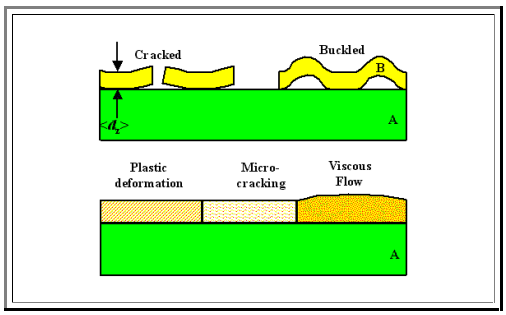
- Generally, there will be stress σ and strain ϵ in a thin film and its substrate.
  - A major source of strain is the difference of the thermal expansion coefficients α

Stress and strain in thin films can be large and problematic!

$$\epsilon_{TF} = \Delta T \cdot \Delta\alpha$$

$$\sigma_{TF} = Y \cdot \Delta T \cdot \Delta\alpha$$

- Stress in thin films may relax by many mechanisms; and this might be good or bad:
  - Cracking or buckling
  - Plastic deformation
  - Viscous flow
  - Diffusion
  - Bending of the whole system (Warpage)
  - Warpage can be a serious problem in semiconductor technology.



Exercise 3.2-1

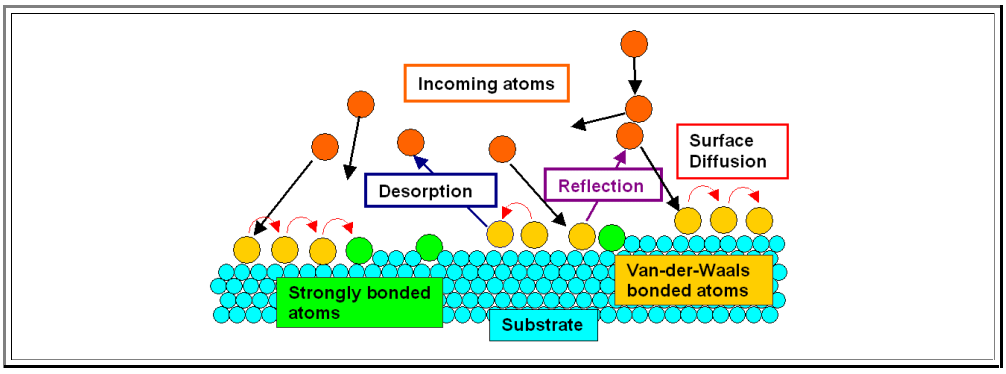
All Questions to 3.2

3.3 Nucleation and Growth

3.3.1 In the Beginning

First Condensation

- In the beginning of a thin layer we have a substrate with a "receptive" surface (being pretty clean at the minimum) and a source of atoms or molecules that are supposed to build up on the substrate.
- How do we build up a thin layer? Let's do it in a very simple manner. We put on our good glasses, the ones that allow us to see atoms. (*Hint*: you don't see with your eyes, you actually *see* with your brain. Eyes are just one source of input to the seeing center of the brain). Moreover, we do the most perfect experiment we can imagine, because here we will work with perfect materials in a perfect world - in our minds, where this is easy.
  - So we have a perfect substrate crystal with a perfect surface - no oxide on it, no other dirt, no whatever Also no **surface reconstruction** (if you want to find out what that means - activate this link). But perfect, in the sense of minimum free enthalpy, still entails some vacancies even in thermal equilibrium; being of a gentle disposition we also allow mild deviations from perfection as required by thermal equilibrium and allow a few surface steps and the occasional dislocation.
  - From some unnamed source we produce a constant stream of atoms (or molecules) that are supposed to form our thin film on the substrate. They have some average velocity and move more or less in the direction of the substrate.
- What happens when an incoming atom hits the surface of the substrate? Well, just about anything you care to imagine:
  - The atom may just be reflected like at solid wall and then runs away to infinity.
  - The atom may just be reflected like at solid wall but then hits an incoming atom and is redirected to the substrate.
  - The atom may loosely bond to one or two of the substrate atoms (by one or the other of the **secondary bonding** mechanisms, like "**van der Waals**"), or, to use surface science terminology, it will become **"physisorbed"**.
  - At finite temperatures, it may now jump to equivalent positions in its neighborhood, i.e. it diffuses in a **2**-dimensional random walk manner on the surface.
  - On occasion, while running around at random, it may just get **desorbed** again and flies off into the wide world out there.
  - But then it may also find a cozy place where it can *seriously* bond to more than just one (or two) substrate atoms, it will get **chemisorbed**; i.e. bonded by one of the strong bonding mechanisms.
  - This *chemisorption* will happen with higher probability if our so far loosely bonded or physisorbed atom is seriously cornered by the substrate atoms - at a **step** or a surface vacancy, for example



- It's all in the picture, and all of that (and more) is really happening.
  - Atoms, as the saying goes, are only human after all (or was it: humans are only atoms, after all?). They fool around if they can, and only get seriously stuck if they fall into "traps", or to put it more positively, if they find highly attractive situations.
- All in all, we can draw two major conclusions from this simple picture:
  1. Not all atoms reaching the substrate "get stuck". The percentage of "**B**" atoms (or molecules) remaining on the substrate **A** (the **sticking coefficient**") depends on many things, but the probably most important parameter is the binding energy between **A** and **B**.
  2. In the initial phase of layer growth - the **nucleation** phase - first clusters of **B** atoms (or molecules) most likely form at irregularities of the substrate, in particular at (atomic) steps.
- The question coming up now is how many steps (per $\mathbf{cm^2}$) we might have on a given substrate and how we can optimize the step density if that is what it takes.

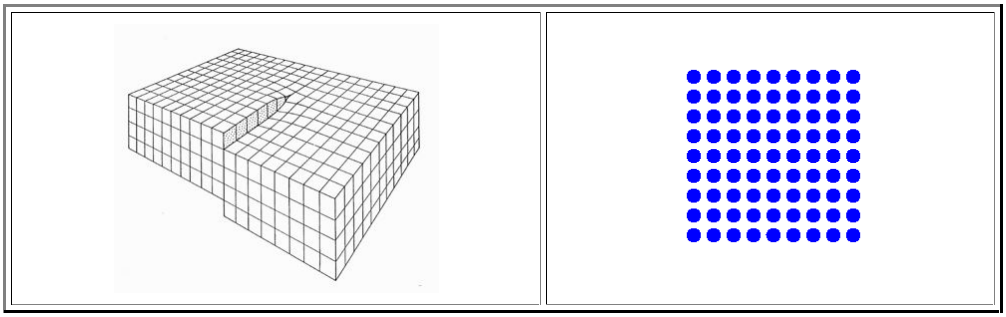
- This is easy to answer on a first glance. You can even take off your sharp glasses and just imagine a *perfect* **{100}** or **{111}** oriented surface of a **fcc** crystal, for example. There will be no steps. But look at the picture above. Steps in regular intervals are an unavoidable consequence of a substrate simply orientation that is off a bit from - in the picture - precise **{110}**.
- If d_{111} is the distance between **{111}** lattice planes, and d_{step} is the average distance between steps like in the picture above, the **misorientation** α is simply given by

$$\sin\alpha = \frac{d_{111}}{d_{step}}$$

- At this point, we get a first idea why semiconductor companies making **Si** chips order their wafers on occasion with a specification like "**2°** off **{100}**".

Not so easy perhaps, on a second glance, is to imagine that even surfaces with *no* misorientation from, for example, **{100}**, might still have a certain density of steps.

- This step density, to give you a hint, is identical to the density of screw dislocations terminating at that surface. Below is an picture of a screw dislocation ending at a surface; a Burgers circuit running around it shows that the Burgers vector is parallel indeed to the dislocations line direction.



If you now recall that **edge dislocations** that left the crystal, automatically produced a step in the surface, you finally get the picture:

- If your substrate contains defects - dislocations, grain boundaries, and so on, you might have steps even for a perfect low index surface orientation.

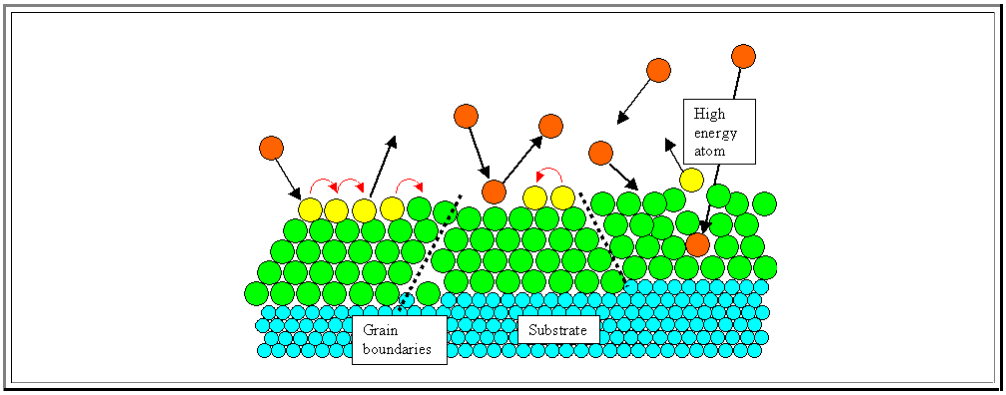
What happens in the beginning of a deposition process where some atoms (or molecules) are brought in contact with a substrate surface, thus depends on many parameters. All the processes that take place in parallel, and with some possible interaction, may produce an exceedingly complicated situation. Mind, its complicated - but not difficult.

During Growth

This module is not really the right place to discuss what happens *during* the growth of a layer, but this topic fits here very nicely with what has been described so far.

We simply ask ourselves: How does the general situation change after we - somehow - have already produced a thin layer of **B** and, keeping the process running, now just make it thicker?

The answer is. Not much changes. Look at the picture below:



- It's pretty much the same as the picture form above, For varieties sake one more process has been added: A relatively energetic **B** atoms hits the **B**-layer so hard that it penetrates into the layer, causing some damage, and kicking one (or more) already physisorped **B**-atom out again.
- It is also quite clear that you will produce grain boundaries and other defects in the **B**-layer because things - atom size, lattice constants, lattice types - just don't match.

- Of course, *sticking coefficients* and all other numbers you can attach to the atomic processes might be quite different in the "**B** hits **B**" case in comparison to the "**B** hits **A**" case.

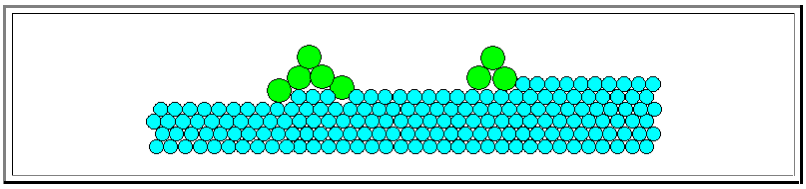
How Do We Know

- It's easy (and amazingly cheap) to look at what is going during thin film growth at atomic dimensions with your good glasses, the ones that allow you to see atoms. But how do we *know* that what we "see" is the *truth*?
  - Your choices are limited. If you are convinced that what you "see" is the truth, but you cannot corroborate your finding with independent evidence, you must either found a new religion or become a philosopher.
  - If you prefer to remain a scientist or engineer, you now must find some way to prove in a mathematical sense that what we discussed so far is the truth. Well - Good Luck to you!
- It's generally not so easy to *really* see on an atomic scale, and to make things worse, you want to see a dynamic process, where things change rapidly.
  - This is just not possible with the usual tool, the **High-Resolution Transmission Electron Microscope** (***HRTEM***) - it just can't give "surface dynamics".
  - The big boost to this topic came with the invention of the **Scanning Tunneling Microscope** (***STM***) in **1980** by **Gerd Binnig** and **Heinrich Rohrer** (Nobel prize **1986**) and the whole family of "Needle scrapers" following, in particular the **Atomic Force Microscope** (***AFM***). Before that, one could draw pictures like the ones here easily, but back in the Lab it was guess work.
- We will come back to these instruments, but already now you can appreciate how important the invention and development of these instruments was to the thin film communities (and others).
  - Finally, you may enjoy to look at some spectacular pictures produced by these instruments in the link to the company Omicron. A few selected examples are also found in this (future) advanced module.

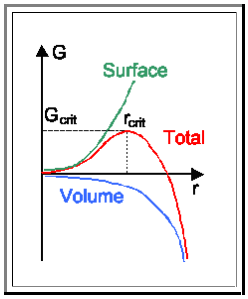
3.3.2 Nucleation and Growth Modes

Nuclation in General

- Let's suppose we have managed to get a few atoms "stuck", meaning *physisorbed* in polite terminology, on our substrate. At steps, or other defects, or whatever on our substrate **A**. We now ask two not-so-obvious questions:
  - *Question **1**:* Are our **nuclei** -that's what we call those small primary agglomerates or clusters - stable and willing to grow?
  - *Question **2**:* If they grow, what kind of general topology will the growing film assume?
- Those may not have been the questions foremost in you mind, so let's first see why those are not only valid questions, but important questions. Let's look at question **1** first.
  - This is what we consider our starting point. We have a few small clusters of **B**-atoms physisorbed to the **A**-substrate. Will they grow if there are more **B**-atoms around that could be added to the pile?



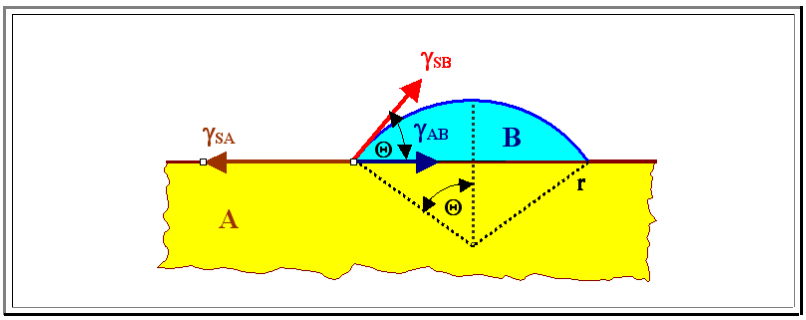
- The answer, as always, is straight forward and simple: As long as we look at thermal equilibrium (or thereabouts), they will grow *only* if they *free enthalpy* of the total system *decreases* if the nuclei get bigger.
  - That should remind us of a quite similar problem we considered earlier in our study course: Will a three-dimensional defect, e.g. a precipitate with radius ***r***, grow or not?
  - The problem was that the surface area and thus the surface energy always *increases* with r^2, while the *decrease* in the volume energy, which is the only reason why something should happen at all, goes with $-r^3$. When we say *energy*, we mean, of course, always *free enthalpy* ***G***.
  - The problem was that for small enough ***r*** the surface energy always "wins" and we found that growth is energetically only favored above a **critical size** expressed by r_{crit} and resulting from the interplay of volume and surface energy.



- The situation in our case here as shown in the picture is exactly the same, just a bit more complicated. If you increase in your mind the size of one of the clusters, you will have done the following with respect to energy:
  - You decreased the energy of the **A**-surface because that surface gets smaller if you increase the cluster size (in all directions, of course).
  - You increased the energy of the **A-B** interface because its area gets bigger.
  - You increased the energy of the **B** surface because its area gets bigger.
  - You decreased the energy of the **B**-nuclei because its volume increases (and we assume that this lowers the energy of a bunch of **B**-atoms relative to them being "single").
- If you think about this a bit, you will realize the essential curve for the free enthalpy will still look more or less like the red curve in the schematic drawing. If you want to look at the thinking already done, you activate the link.

The Wetting Angle and Its Significance

- While the balancing of all those energies (and we could easily add more for more complex cases) in order to find the arrangement with minimum free enthalpy seems to be rather tricky, it turns out that you don't have to do it if you don't want details. It is sufficient to consider just on "simple" quantity, the **wetting angle** Θ.
  - The name and the definition of the *wetting angle* comes, like many other names used in this context, from the early investigations of the (macroscopic) interaction of liquids and solids. You can't help noticing that a drop of water, if deposited on various substrates, forms ***spherical segments*** or culottes of various shapes; on greasy surfaces it is almost a sphere.
  - Solids that are deposited atom by atom on a substrate will behave quite similarly because you give them a choice of forming a shape for which surface and interface energies are minimized - and that's what the water drop does. If we assume that our material **B** deposited on substrate **A** actually assumes a shape that mimics to a large extent what a liquid would do, we end up with the following paradigmatical picture for the nucleation of thin films:



- The surface or interface energies involved can be understood as forces pulling in the way indicated at the edge of the cluster of **B** atoms. For the wetting angle Θ a simple relation follows:

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

$$\cos\Theta = \frac{\gamma_{SA} - \gamma_{AB}}{\gamma_{SB}}$$

- Now let's take it easy. What we obviously would like to have when we try to deposit a thin layer on a substrate is a wetting angle (also called **contact angle**) of $\Theta=0^o$ so that the "drop" spreads out completely, producing a thin film.
  - This we can get if $(\gamma_{SA} - \gamma_{AB})/\gamma_{SB}=1$ or $\gamma_{SA}=\gamma_{SB} + \gamma_{AB}$.
  - This is immediately clear. When we form the cluster as in the picture above, we reduce the total area of the substrate by some value $A(r)$ given by the radius r (in fact, we have $A(r)=\pi r^2\sin^2\Theta$). We thus take away the surface energy $\gamma_{SA}A(r)$ at the cost of having now some interface energy ($\gamma_{AB}A(r)$) plus some surface energy of the **B**-material. If we "pay" no more in energy then what we get, the wetting angle will be zero because wetting the whole substrate cost nothing.
  - In fact, if we can cover the whole substrate area at a gain in energy, the film will also wet the substrate completely. A wetting angle isn't even defined anymore because its cosine would have values **> 1**.
- The condition for getting smooth films, or the **Frank - van der Merve layer-by layer growth** mode, as we will call it from now on, is as follows:

$$\gamma_{SA} \geq \gamma_{SB} + \gamma_{AB}$$

- Now let's look at the other extreme; finite wetting angles (the maximum would be 180^o, meaning the cluster would be a little sphere just barely touching the substrate).
  - Obviously we simply need to have $\gamma_{SA} < \gamma_{SB} + \gamma_{AB}$, meaning that we have to make the interface area smaller than the substrate area.
  - If γ_{SA} is much smaller than the $\gamma_{SB} + \gamma_{AB}$, it just doesn't make energetic sense to cover the substrate; our thin film material **B** is better off just balling up.
- The condition for getting rough or patchy films, or the **Vollmer - Weber island growth mode**, as we will call it from now on, is as follows:

$$\gamma_{SA} < \gamma_{SB} + \gamma_{AB}$$

- That would be it if the world would be simple. Luckily, it's not. We have something called **Stranski-Krastanov layer-plus-island growth mode** which looks like this

$$\gamma_{SA} > \gamma_{SB} + \gamma_{AB}$$

plus strain

- Details of that growth mode aren't all that clear. We need more than just the simple surface and interface energies of the perfect bulk materials **A** and **B**; strain that develops during the deposition, for example.

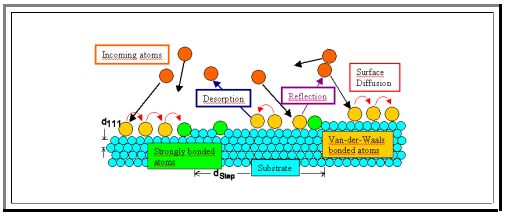
- Look at the interface energy γ_{AB} for a second. It will of course depend on the nature of **A** and **B**, but also on the exact structure of the interface. It might be epitaxial, for example and then possible strained so the lattices match. If that is so, there would be an increasing strain energy as the thickness increases.
- In other words, the three surface and interface energies γ considered are not absolute constants; they may depend on many things (surface reconstruction, for example), surface orientation, and so on.

Some Generalization

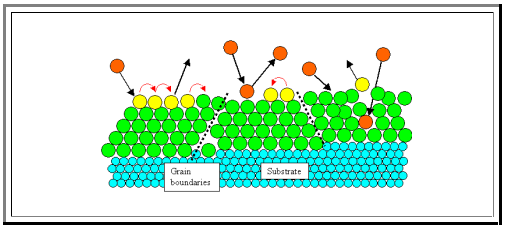
- In total, what we have learned is that it is just not good enough to take some nice and clean substrate **A** and throw some atoms **B** at it (in some vacuum machine) if you want to produce a nice and smooth thin film. It may or may not work - there are many parameters to consider coming solely from properties of **A** and B - and we haven't even considered the many ways of "throwing atoms" yet!
- For the grand final let's look at the first moments of film growth again, the nucleation phase, and then at a later stage, well after nucleation.
  - So we have a Frank - van der Merve layer-by-layer growth mode, let's say. Does that mean that the first thing we get is a complete layer, one atom thick, on our **300 mm Si** wafer? You must be kidding, That takes a lot of atoms, and even if they perfectly "wet" the surface of the **Si**, the best you can expect is that a lot of small patches, one atom thick, will nucleate and than grow laterally by adding more incoming atoms at their edges. Finally, those flat islands of atoms coalesce and form a closed layer.
  - Same story with the other growth modes. The Vollmer-Weber mode would lead to three-dimensional growth of lumps of atoms, that finally will coalesce into a closed, but bumpy layer.
  - Of course, the "coalescence" might no be so easy and provide all kinds of defects, and of course, it just may happen that a second layer already nucleated and started growing laterally on top of a big patch before full coverage had been achieved.
- Anyway, sooner or later the substrate **A** will be completely covered by **B** - by a nice smooth layer, or by a rough sorry looking patchwork.
  - How does it go on? We now deposit material **B** on material **B** and no longer on material **A**, after all.
  - It depends. The Stranski-Krastonov layer-plus-island growth mode already demonstrated that there might be more than meets the eye. Nevertheless, things get a bit easier. There are only two surface and interface energies to consider, for example. In any case, you must expect that what is going to happen during the phase of growing **B** on **B**, is always more or less influenced by what happened at the nucleation phase.
- Last question: Where are the numbers? We put a lot of importance in surface and interface energies, so how about a few examples?
  - Interestingly, neither the "Ohring" nor the "Smith" give (easily found) numbers. One simple reason for this is that there aren't any good numbers. Atomically clean surfaces do not exist under normal circumstances, making measurements difficult. (How would you measure a surface energy anyway?). Moreover, for crystals it also depends on **{hkl}**, the crystallographic orientation of the surface, and so on.

3.3.3 Summary to: 3.3 Nucleation and Growth

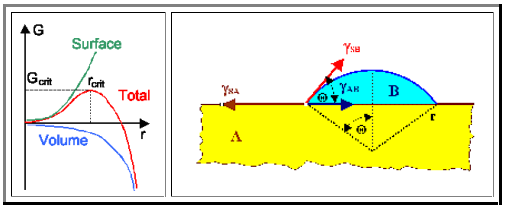
- Deposition of a thin layer must start with a "clean" substrate surface on which the first atomic / molecular layer of the film must nucleate.
- There are many possible interactions between the substrate and "first" incoming atoms.
  - As the interaction energy goes up we move from "some" absorption to physisorption (secondary bonds are formed) to chemisorption (full bonding)
  - The sticking coefficient is a measure of the likelihood to find an incoming atom in the thin film forming.
  - Immobilization by some bonding is more likely at defects (= more partners). The initial stage of nucleation is thus very defect sensitive.



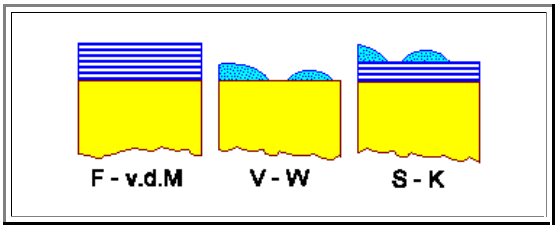
- Simple surface steps qualify as efficient "defects" for nucleation.
  - Small deviations from perfect orientation provide large step densities. Nucleation therefore can be very sensitive to the precise **{hkl}** of the surface
  - Intersections of (screw) dislocation lines with the surface also provide steps.
  - This may cause grain boundaries and other defects in the growing layer.
  - Scanning probe microscopy gives the experimental background



- There is always a nucleation barrier that has to be overcome for the first **B**-clusters" to form on **A**
  - the three involved interface energies, all expressed in the "wetting angle", plus possibly some strain are the decisive inputs for the resulting growth mode.



- **Frank - van der Merve**: Smooth layer-by-layer growth
- **Vollmer - Weber**: Island growth
- **Stranski - Krastanov**: Layer plus island growth



Exercise 3.3-1

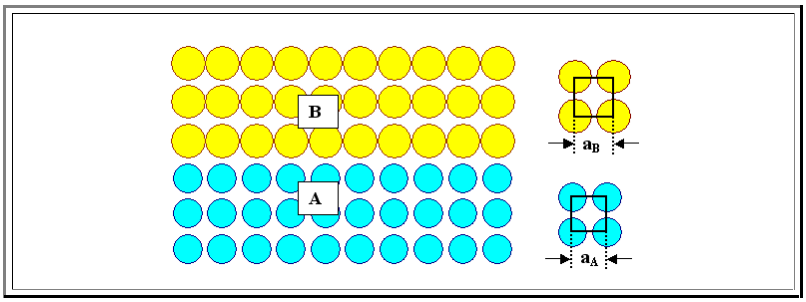
All Questions to 3.3

3.4 Thin Films: Structure, Interfaces and Some Properties

3.4.1 Single Crystalline Thin Films

Epitaxial Growth

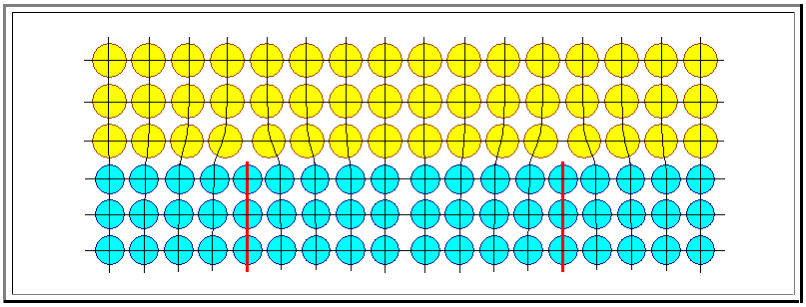
- In semiconductor technology we make all kinds of thin films. Often, we make thin films of some semiconducting material (cf. this picture) and then we want this thin film to be a perfect single crystal with no defects whatsoever (except what we intentionally put in, e.g. for doping).
  - This is a tough job description. How can you make a *single crystalline* thin film? Think about it. If you do it on a **Si** substrate wafer with **300 mm** diameter, you are actually demanding that any atom of your material **B** sitting somewhere on your substrate, has a precise geometrical relation to any other atom being many centimeters away, for example.
  - You may think as hard and as long as you like, but the only way to do this is to deposit your thin film on a *single crystalline substrate* to start with. If you are lucky, your **B**-atoms then just continue the lattice of the **A**-substrate and you have **epitaxial growth** or **epitaxy** for short. The term "epitaxy" comes from the Greek "*epi*" meaning "above", and "*taxis*" meaning "in ordered manner".
  - Let's look at a perfect epitaxial interface between **A** and **B** in a simple picture:



- That's what you want - but that is not what you will get. The reason is that we must assume that the two lattice constants ***a*** of **A** and **B** are *not* identical as shown on the right; we are lucky if the Bravais lattice types are.
  - In fact, it you look at our old "master picture", summarizing the basic properties of important semiconductors, you see that the lattice constants of those semiconductors are mostly quite different.
  - That leaves us with just two options:

    1. We may have a perfect epitaxial situation as shown above, but then we also have a lot of strain ϵ and stress σ in our layer **B**, because we must squeeze it to fit on the substrate. The strain energy stored in the stressed layer necessarily scales with $\epsilon \cdot \sigma$ and the thickness d_z of the layer. Stress and strain scale with the **misfit** $f = (a_A - a_B)/a_A$ of the situation, i.e. the relative difference of the lattice constants.
    2. We will not have a perfect epitaxial interface.

- It's as simple as that. The question now is how much strain energy can a thin film take before it is going to do a little stress relief by one means or another?
  - That depends on the means available fro stress relief. The most simple way of obtaining some stress relief is to suffer some **plastic deformation**, i.e. producing and moving some dislocations around.
  - This is exactly what will happen if the stress in a perfectly epitaxial thin film reaches some critical value: the film deforms plastically and introduces what we call **misfit dislocations**. A thin film may even do this if it cannot be plastically deformed as a bulk crystal like, e.g., **Si**. Misfit dislocations are easy to visualize:



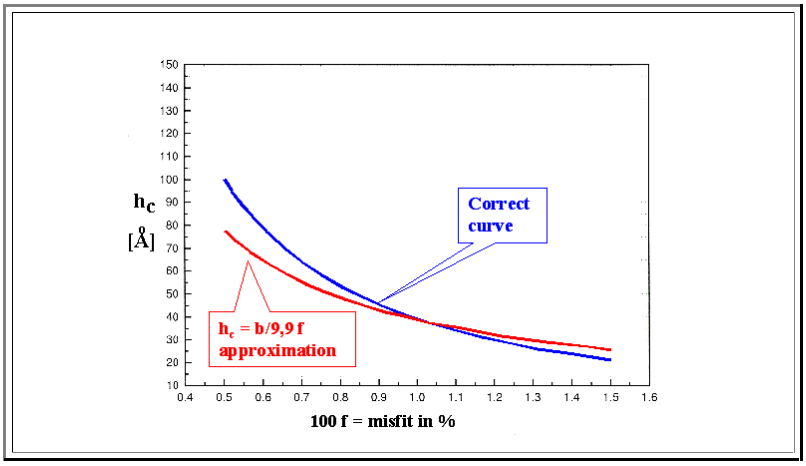
  - We have the same situation as above, but now we have misfit dislocations in the interface. They must by necessity be **edge dislocations** (as indicated by the red lines), or more generally, their Burgers vector must have a sizeable component in the interface.
  - Obviously you must have dislocations at regular distances d_{dis} (in at least two directions - you will get a network) given by $d_{dis} = a_A/f$.

- Now we have no more strain and strain energy in the thin film, but a lot of dislocations in the interface. We know or just believe that dislocations carry approximately the energy $E_{disl} = Gb^2$ per unit length (G = shear modulus). So what is better?
  - Same procedure every time. Look at the energy and compare:
    - *Perfect interface*: The total strain energy scales with $f \cdot d_z$.
    - *Interface with misfit dislocations*: The energy scales with the total dislocation length or with the dislocation density, and thus only with f and not (or just a little bit) with d_z.
- In other words: There is always some critical thickness d_{crit} of the thin layer so that for $d_z > d_{crit}$ the introduction of *misfit dislocations* in the interface is energetically favorable to having a strained layer.
  - Going through the energy comparison in detail (not quite as easy as it looks) produces a very important, very sad, and transcendent equation for the critical thickness; it is given below with a simple approximation added.

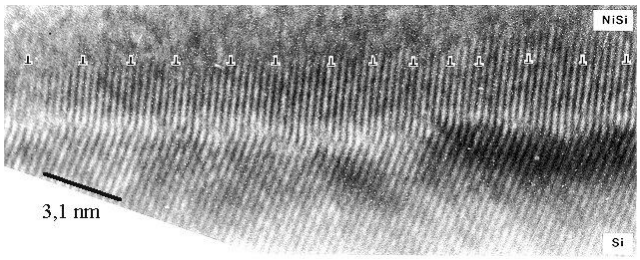
$$d_{crit} = \frac{b}{8\pi \cdot f \cdot (1 + \nu)} \cdot \ln \frac{e \cdot d_{crit}}{r_0}$$

$$d_{crit} \approx \frac{b}{9.9 \cdot f}$$

  - b is the magnitude of the Burgers vector, ν = Poisson's ratio, e = base of natural logarithms, r_0 the "size" of the dislocation core (about $1b$).
- Why is this a sad equation? Look at its graph. It shows the numerical solution to the equation from above and the approximate solution.



  - It also shows, sadly, that the critical thickness is rather small if there is any misfit at all. A misfit of just **1%** will lead to an interface full of dislocations if the thin layer thickness exceeds about **4 nm**!
  - It goes without saying, of course, that this is *seriously bad*. If you don't know why, look at what you have learned before. Defects like dislocations are never good news for electrons and holes in semiconductors. They reduce minority carrier diffusion lengths and life times, the efficiency of light generation, and the product life time of Laser diodes, for example.
- You just can't have misfit dislocations in your electronically active semiconductor - semiconductor interfaces. *None whatsoever*!
  - Now we understand why the lattice constant is the second important property besides the band gap and plotted in our master picture, and why there are so many combinations of semiconductors that we don't use.
- Being engineers, we now ask ourselves if there are some tricks to beat the critical thickness equation from above. The answer is: Yes - but you don't get very far at present. One of the neater ideas, going under the heading of "**compliant substrates**", can be accessed by the link.
- The picture below shows an extreme case of misfit dislocation for a misfit of about **15 %**. We have hexagonal **NiSi** with its **{0001}** plane in line with the **{111}** plane of **Si**. Since both lattice planes are hexagonal in this case, we can have an epitaxial relationship between tow different Bravais lattices.



© H. Foell

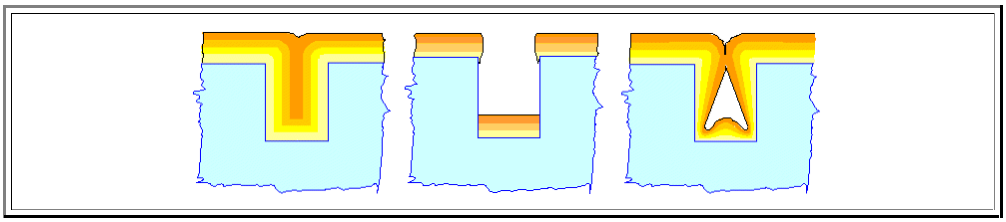
- As we must expect, we have an edge dislocation (seen as ending lattice fringe) about every **7** lattice constants. This is one of the first pictures of misfit dislocations at atomic resolution. The dislocations are so close that they could not be seen in "normal" **TEM**.

3.4.2 Thin Film Structure

- This will be short, because we will discuss this topic mostly when we come to deposition methods.
- Let's distinguish to separate points

  1. Geometry and morphology.
  2. Internal structure.

- The first topic has been mentioned before, but now we go beyond the points made there.
  - Let's consider to grow a thin film on a substrate that has a small hole in it - say half a **µm** in lateral size and **1 µm** deep. We deposit a thin film (atomic) layer by layer (symbolized by various shade of yellow). What we might get could have one of the topologies shown below or anything in between



  - The only difference between these quite different topologies is the deposition method; we will cover that in more detail in chapter 6.
- There is much more along this line, but we cover it when we get to it.
- As far as the internal structure goes, the situation is similar. Everything known from bulk materials goes:
  - *Poly-crystalline* thin films with grain sizes ranging from a few nanometers to **cm** (Are the **Zn**-covered steel lamp posts, letter boxes, etc. with huge grains products of the thin film industry?)
  - *Single crystalline* thin films, but full of defects like dislocations, precipitates, point defects.
  - *Nearly perfect* single crystalline thin films - what we often would like to have, but not always get.
- If we just look at polycrystalline thin films, we may have just regular grains, or all kinds of textures. Again, we deal with it when we run across it.
- Then we have some **thin film specialities**:
  - **Amorphous thin films**, like amorphous **Si** (***a-Si***) or many other materials. You just can't have amorphous *bulk* **Si** or most everything else that usually likes to form a crystal.
  - Mixtures of amorphous and crystalline phases; truly nanocrystalline structures (i.e. grain size around **10 nm**) - practically never found in bulk.
- A case in point is *Silicon* (what else?). We have:
  - **Amorphous Silicon**, used, e.g., in microelectronics. If it is heavily mixed (= "doped" with Hydrogen (**> 15 %**), we have the crucial thin film for **a-Si:H solar cells** or for the transistor matrix of **liquid crystal displays** (***LCD***). Another example is amorphous **SiO_2**, the work-horse of microelectronics.
  - *Amorphous-crystalline* mixes, like **a-Si:H** containing nanometer-sized embedded islands of crystalline silicon (***c-Si*** and then called **µc-Si:H**. This is the base of the so-called "microcrystalline **Si** thin-film solar cell", one of the hottest contender for the solar cell market of the future.

3.4.3 Special Properties of Thin Films

- The fact that there is a subchapter on thin film *properties* gives a hint that they might be different from the *bulk* properties of the same material.
  - The *bad* news is that there are a hell of a lot of special thin film properties - check the next subchapter where we discuss how to measure them - and there is no way that we can compare bulk and volume properties and discuss everything in detail.
  - The *good* news is that *some* thin film properties are often *far better* than the bulk properties. This may have trivial or tricky reasons.
- So let's only glance at some key properties and see if that helps to get a feeling for why thin films are "special" with respect to some properties.
- There is simple guideline of how to figure out if you could expect large property changes in a thin film: Ask yourself what *causes* the property in question.
  - Is it a property determined just by the *bonding* - for example, Young's modulus, the thermal expansion coefficient, or the dielectric constant - or does it have a *defect sensitive* part as, e.g., the carrier concentration in semiconductors or the minority carrier diffusion length?
  - Next, ask yourself, what *typical length scale* goes with the property. With "*bonding*" for example, goes the length scale "lattice constant"; with *minority carrier diffusion length* perhaps the average distance between point defects in the lattice.
  - Now you are done. If the thickness of your thin layer is far larger than the length scale in question, you cannot reasonably expect that its properties differ much from those of the bulk - and vice verse!
- This is not necessarily an easy ***recipe***, nor will it get everything right all the time. But the rule makes sense and provides a guideline.
- **Mechanical properties**
  - The **elastic moduli** shouldn't be all that different - they are coming from the atom-atom bonds which are the same in the bulk and in thin films. Only if the number of atoms at or close to the surface is comparable to the total number of atoms in your thin film, you may need to think twice about this. In other words: only if you consider *thin* to be in the order of atomic dimensions, your bonding situation is so severely disturbed that you might find large differences between bulk and thin films elastic moduli.
  - Parameters of plastic deformation like the ***critical yield strength*** (or hardness) can be far larger than bulk values. The reasons for this depend on many things (not least on the type of film), but if you look at what determines the critical yield strength in bulk crystals, you will find intrinsic length scales like the dislocation density (always ties up with some average distance between dislocations) or the grain size. In thi, film the grain size in one direction is at most the thin film thickness, and the dislocation density in areas with lateral extension some **10** times the film thickness is often zero - even for high dislocation densities, because the average distance between dislocations might be far larger than the film thickness.
  - Those are good news, because they mean that our thin films can take a lot of stress before they do something drastically.
  - There is a trivial, but perhaps unexpected property of thin films. If you deposit a perfectly brittle material like **Si** on a flexible substrate, you can roll up your substrate like a rollo - and your thin film will not break. It's simply a matter of the the radius of curvature being far larger than the film thickness; the link contains the equations.
- **Optical properties**
  - There is not much to say here. The **index of refraction** is tied to the bonding once more ("polarization mechanisms") and should not change much. If your bulk material is transparent at some wave length, the thin film will be even more so. Bulk materials that appear ***opaque*** because the absorption length of light is shorter than, say **5 µm**, may be fully transparent as a thin layer. Even some very thin metal layers (e.g. **Au**) become transparent to visible light.
- **Electrical properties**
- *Specific conductivity* σ:
  - We always have $\sigma = \Sigma_i(q_i \cdot n_i \cdot \mu_i)$ with q, n, μ = charge, carrier concentration, and mobility, respectively, of the carriers involved.
  - Going from bulk to a thin film may change the carrier concentration if the film is so thin that the system becomes a two-dimensional electron gas. What may change at larger thicknesses is the mobility μ. We expect something to happen as soon as the film thickness comes into the same order of magnitude as the mean free path of the carriers.
  - If you think about it, chances are good that the conductivity will *decrease*. That is not so good, but in real life the effect is usually not dramatic.
- *Electrical break down field strength* E_{BD}:

- Take a flat piece of quartz **1 mm** thick and put it between the two plates of a parallel-plate capacitor. Now crank up the voltage ***U***. At some (high) value of the voltage, the contraption will go up in smoke with a big bang because you have reached the critical break-down field strength $\boldsymbol{E_{BD} = U/1\ V/mm}$, which will be some **10.000 V/mm** in your experiment.
- Now do the same thing with a standard $\boldsymbol{SiO_2}$ layer from microelectronics, having a thickness of **5 nm**. You will find $\boldsymbol{E_{BD} \approx 10.000.000\ V/cm}$; a value far above the bulk number, allowing you to run your integrated circuit at unbelievably high voltages of up to **10 V**!
- Why do we have that large improvement? There are several possible reasons; but the issue is actually not all that clear, partially because the mechanisms of electrical break down in bulk materials are not so clear either

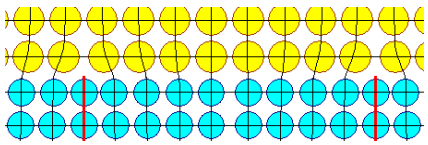
Critical current density $\boldsymbol{j_{crit}}$

- Take an **Al** or **Cu** wire with a cross-sectional area of $\boldsymbol{1\ mm^2}$ and run some current ***I*** through it. Crank up your current and watch what will happen. At some critical current density $\boldsymbol{j_{crit} = I/1\ mm^2}$ your wire will go up in smoke; before that it became a light bulb for a short time. You will find that $\boldsymbol{j_{crit}}$ will be around a few $\boldsymbol{1.000\ A/cm^2}$.
- Now do the same thing with a thin layer that you have structured into wires with a cross-section of about $\boldsymbol{1\ \mu m^2}$. You will find a critical current density of $\boldsymbol{> 10^5\ A/cm^2}$, again orders of magnitude larger than the bulk value, enabling you to run tremendous currents of up to **1 mA** through those interconnects in your integrated circuit.
- Again, why do we have that large improvement? In this case it is relatively clear. The volume to surface ratio of a thin film wire allows a much better transport of the heat generated in the wire to the large heat sink "substrate" and the environment.

In case you missed the point: You just learned that microelectronics is *only* possible because thin layers are so much better with respect to some important properties than the bulk materials!

3.4.4 Summary to: 3.4 Structure, Interface and Some Properties

- Epitaxial layers are crucial for semiconductor technology.
  - Misfit of lattice constants will produce strained layers upon epitaxial growth; strain relief happens by the formation of misfit dislocations.
  - Misfit dislocatipons must be avoided at all costs!
  - Below a usually rather small critical thickness d_{crit} of the the thin layer no misfit dislocations will occur.
  - Rule of thumb:
    0.5 % misfit $\Rightarrow$ $d_{crit} \approx$ **10 nm**



- The internal structure of thin films can be anything known from bulk materials plus some (important!) specialities.

a-Si: Micro electronics
a-Si:H: Solar cells, **LCD** displays
µc-Si:H: Solar cells

- Properties of thin films can be quite different from that of the bulk material
  - The reason can be differences in length scales.
  - Semiconductor technology relies to some extent on superior thin film properties

Much better in thin films

- **Electrical break-down field strength of dielectrics.**
- **Critical current densities in conductors.**

Exercise 3.4-1

All Questions to 3.4

3.5 Properties, Measurements and Characterization

3.5.1 More Properties

List of Properties to be Measured

- What are "**general properties**" and how do we measure them?
  - In a first approximation we take all the properties where a measurement produced just a single number. This always means that we measure some *average* property. What do we have - let's make a quick list:
- **Geometry and Topology**.
  - Under this heading we may list:
    - **Thickness and roughness**.
    - **Density and porosity.**
    - Bravais lattice type.
- **Mechanical properties**
  - Under this heading we may list:
    - **Elastic moduli**.
    - **Plastic deformation**. parameters like ***critical yield strength***, ultimate tensile strength, dislocation density, other defect densities.
    - **Fracture** parameters.
    - **Internal stress and strain**
- **Thermal properties**
  - What comes to mind is
    - **Thermal expansion coefficient**.
    - **Thermal conductivity**.
    - **Specific heat**.
    - **Thermoelectric coefficients**.
- **Chemical properties**
  - Things like
    - **Corrosion resistance** in various ambients.
    - Solid-state **reactivity** between substrate and top layers.
    - ???
- **Defect properties**
  - We have, outside of structural concerns, some simple numbers for, e.g.:
    - Formation and migration enthalpies of point defects.
    - Solubilities for impurities.
    - Stacking fault and surface energies.
    - Precipitate formation parameters.
- **Optical properties**
  - Not too much here on first sight. We have
    - **Index of refraction**.
    - **Absorption coefficient**.
    - **Non-linear optical properties**.
- **Electrical properties**
  - The list includes
    - **Conductivity**.
    - **Carrier mobility**.
    - **Hall coefficient**.
    - **Carrier type and concentration**.
    - **Work function**.
    - **Dielectric breakdown field strength**.
    - **Critical current density**.

- **Dielectric function.**

OK - you are right: This is getting ridiculous. Some of the properties listed you have never heard of before (which doesn't mean that they might not be crucial in some applications). Obviously, the list can be expanded to provide a short enumeration of everything dear to Material Scientists.

Well - yes. But the fact remains: All those properties (and more) are there and *need to be measured* if you either

- Really need to know them, and
- Have some reason to expect that they might be different from the (hopefully) known bulk values.

With that practical restriction in mind, we can cut out a lot of measurements contained in the list above. If we focus on the most essential general properties you almost always need to know *and* must measure, we are left with

- Film thickness.
- Density / porosity / uniformity.
- Conductivity.
- Special properties.

Let's discuss this list a bit in the next module and give some hints about methods of measurement (more will be found in the links provided)

3.5.2 Important Techniques

Thickness

- The film thickness is without doubt the first parameter you must measure and there are many methods for that. Neglecting exotic or expensive methods like, for example, **Rutherford back scattering** (***RBS***) and more indirect methods like "bevelling" plus optical microscopy, we are left with two basic cases:
  - The thin film is transparent in the optical region (including ***IR*** and ***UV***). Then we use *optical methods*.
  - The thin film is opaque (i.e. metal films). Then we have a problem.
- Let's deal with transparent films first. What we generally do is to use interference effects; which give clear signals if the phase difference between two interfering beams is just a fraction of the wavelength.
  - The most simple experiment, evaluating interference colors as shown in the link, already suffices to measure film thicknesses if you are not too demanding. Far more advanced methods exist, however, which we will treat very cursorily here.
  - **Reflectometry**: Measure the intensity of the reflected light as a function of the wavelength. You can even do this through a microscope, so you measure the thickness of a small spot. The intensity of the reflected light as a function of wavelength can have a complicated shape, depending on the film refractive index, its thickness, and the properties of the substrate (it might be transparent, too). The method still gives (even more complicated) spectra if several (more or less transparent) films are on top of each other. Except for the simplest cases, an evaluation of the spectra obtained needs powerful software. If such a software exists, matching the measured data against models of the layer system often allows to extract the information desired.
  - **Ellipsometry**. The probably most powerful technique - an **ellipsometer** is standard equipment in any thin film environment. A monochromatic polarized light beam is directed under a certain angle α on the specimen; the intensity and the change of the polarization cause by film(s) and substrate is measured. For simple systems (transparent film of suitable thickness on known flat substrate), it is comparatively easy to extract the film thickness *and* its index of refraction directly from the data. Even a film thickness as small as ≈ **1 nm** can be assessed. For more complicated systems - see above. If the wave length and / or the angle of incidence can be changed, too, the method becomes extremely powerful and extremely complicated.
- Now let's look at opaque films.
  - If you have an area on your sample where you have no film, just run a stylus across the sample and measure the step height as you hit the film. This is a major and mature technique that was (and still is) linked to the brand name "Dektak". Don't confuse a "Dektak" with an **AFM** or **STM**!
  - Measure the resistivity of the thin film - directly or indirectly. Assuming that its specific resistivity it is that of the bulk material, and knowing the lateral dimensions, you can calculate the thickness. There are many and partially quite sophisticated ways of measuring thin film resistivity. Even if you are not sure about the specific resistivity, you may simply calibrate your measure values for various thicknesses by comparison with an absolute standard. Of course, this works best it your thin film is on an insulating substrate.
  - Your absolute standard comes from looking at a cross section with an electron microscope that has sufficient resolution. Look at the **TEM** picture of the **SiO_2** layer in the link. You build-in ruler is the lattice constant; you can't get much more precise than that.

Density / Porosity / Uniformity

- The last topic - **uniformity** - is the simplest one in principle: just measure whatever at many points. It may not be so easy in practice, however, but we will not go into details because what you do depends very much on what you have and what you want.
- This leaves us with **density / porosity**. Both term are almost inverse synonyms: if the density is not what it should be, your film must be porous in a very general sense.
  - Since the index of refraction is simply a relatively straight forward function of porosity as long as the pores are far smaller than the wavelength of the light, ellipsometry gives you already some ideas about that.
  - There are many other methods. Often you take the property that is sensitive to porosity as the vehicle for characterization of porosity or density. For example, there are many ways of producing thin films of **SiO_2** that are very similar in many properties but may show quite different etch rates in **HF** (the almost only chemical dissolving **SiO_2**). Somehow the etching rate is tied to "porosity" (with "pores **< nm**) or the structural integrity, or to whatever you like to call it.

Conductivity

We will deal with this later.

Special Properties

Special properties are often the properties you are after when you make use of a thin film. We haven't listed any above, so let's do it here - and take note at the same time that the property and its measurement are mostly inextricably tied together. You can talk about the thickness of a thin film quite generally without having measured it, but you cannot talk about electrical breakdown field strength being special without having some number in mind

What do we have:

- **Electrical breakdown field strength**.
- **Critical current density**.
- Negative index of refraction.
- Electromigration resistance.
- Tunnel barriers.
- Diffusion barriers.
- Adhesion promoters.

OK. Enough of that - you can't possibly know what this is all about (but you can have a glimpse as soon as the links are in place). It is obvious that we will have to go deeper into properties of materials in general and of thin films in particular before we can discuss this topic in a sensible way.

Summary

The whole module just served to make you aware of a few basic and rather simple truth:

1. Thin films have just as many properties as bulk materials, but they might be far more difficult to measure.

2. Thin films may have additional properties not defined for bulk materials with specific measurement techniques

3. A lot of measurement and characterization techniques have been developed just for thin films

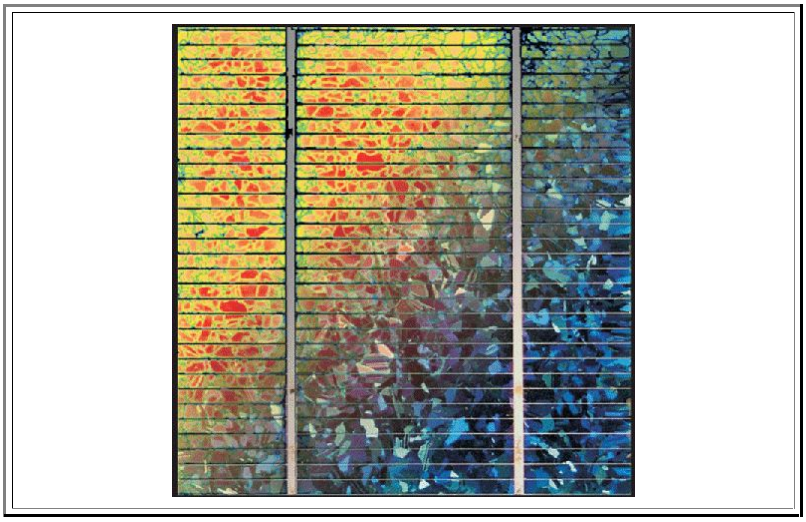
3.5.3 Local Properties

General

- Local properties, if seen as complementary to the *general* properties discussed before and where a single number was all you wanted, give pictures or maps of what is going on locally.
  - Of course, a number like "average grain size" may be seen as a local property, too - but what the hell, you know what is meant here.
  - In essence, here we are discussing methods that produce a kind of *picture* - directly like an optical microscope, or indirectly like a **STM**.
- Let's just enumerate the main techniques. Since they are covered in detail in other lecture courses, we will not go into any details here.

Optical Microscopy

- You can see a lot by just looking through an optical microscope *on* your thin film or *through* your thin film (if substrate and thin film are transparent) - provided that the things you see are ≥ **0.5 µm** or so.
  - You have many different imaging modes at your disposal today. From classical microscopy, to all kinds of contrast enhancers (polarizers, phase or interference contrast, scanning, ...).
  - You also may prepare your specimen in such a ways that things are better visible. Etching the surface with special etchants may delineate defects, for example (and may etch off your layer completely, so be careful).



  - The example here shows the microscope image (very low magnification) of a solar cell in the lower right hand corner blended electronically with a short-circuit current image, i.e. a property image of the **pn** junction formed because there is a thin **n**-type layer "on top" of the **p**-type substrate. The colors denote quantitatively the local photo-current (yellow - red: large currents; green - blue: low currents).

Electron Microscopy

- There are two kinds of electron microscopes:
  - **Scanning electron microscopes**, (***SEM)***. Using a **SEM** you look at the surface of your sample; in its analytical mode you may "see" to a depth of a fraction of a micron
  - **Transmission electron microscopes** (***TEM***). With a **TEM**, you look through a always thin sample. In one of the many imaging modes you can see and analyze what is inside your thin film at atomic resolution.

Scanning Probe Microscopy

- We can scan all kinds of probes across the surface of our thin film. if we desire very high lateral resolution we have
  - **Scanning tunneling microscopy**, (***STM)***. Using a **STM** you look at the electron density at the surface of your sample; this allows you to "see" single atoms
  - **Atomic force microscope** (***AFM***). With an **AFM**, you measure the force that surface atoms (or coarser structures) exert on a close by oscillating probe tip.
- There are many more scanning probe instruments, e.g. for measuring surface charge (....) look it up!

Summary

- The whole sub-chapter just served to make you aware of a few basic and rather simple truths:

1. Thin films have just as many properties as bulk materials, but they might be far more difficult to measure.

2. Thin films may have additional properties not defined for bulk materials with specific measurement techniques

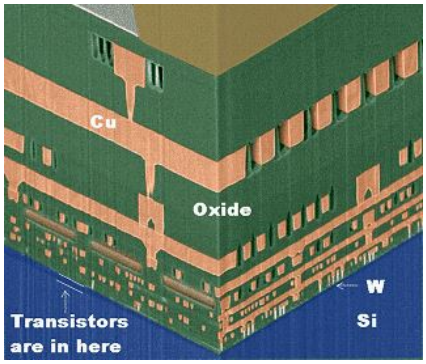
3. A lot of measurement and characterization techniques have been developed just for thin films

3.5.4 Summary to: 3.5 Properties, Measurements and Characterization

3.6 Summary

3.6.1 Summary to: 3. Thin Films

- Semiconductor technology is almost synonymous with thin film technology.
  - A thin film is always adhering to a substrate and (at least originally) continuous.
  - Thin films may still be found in the product or may have been "sacrificed" during the making of the product.
  - An **IC** is a study of thin films in and on the **Si** substrate.
  - The same is true for pretty much every semiconductor product.

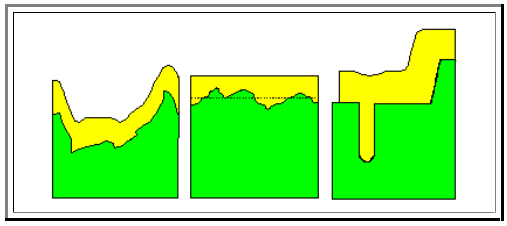


- Thin always means "thin" relative to some intrinsic (internal) length scale. Examples are:
  - Structural length scales.
  - Wavelength and Interaction length scales.
  - Transport parameter length scales.
  - Electrical scales.

• Dimensions $d_{x, y, z}$ • Grain size d_{grain} • Lattice constants a_0 • λ radiation (light, **IR**, **UV**) • Absorption depths • Mean free path lengths. • Diffusion length	• **SCR** width d_{SCR} • Debye length d_{Debye} • Critical thickness d_{crit} for electrical break down • Critical thickness d_{tu} for tunneling

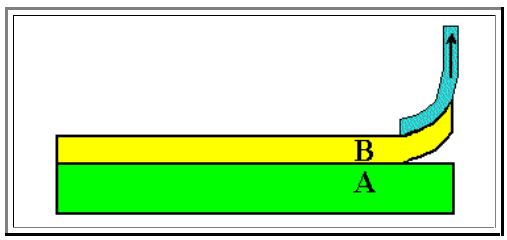
- There are many thin film applications outside of semiconductor technology:
  - Optical, electrical, chemical, mechanical, magnetic technologies use thin films.

- Thin films have other spatial properties besides their thickness.
  - Interface roughness and surface roughness ***R*** defined by their "root mean square":



$$R = \frac{1}{N} \sum_{i=1}^{N} |z_i|$$

- Thin films adhere to their substrate.
  - A direct measure of adhesion is the interfacial energy γ_{AB} between film **A** and substrate **B**.
  - The phase diagram provides some guideline. Complete miscibility=good adhesion, (eutectic)) decomposition=(?) low adhesion. Calculations of γ are difficult.
  - Full adhesion can only be obtained for films grown on a substrate. Adhesion energies can be measured.



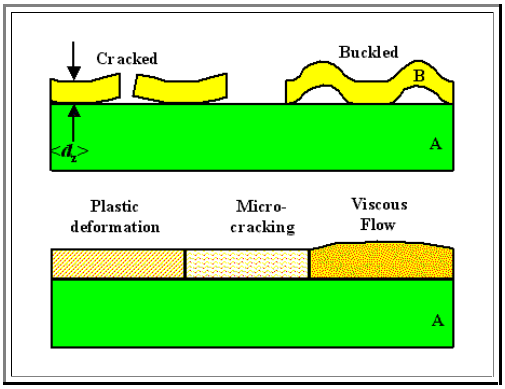
- Generally, there will be stress σ and strain ϵ in a thin film and its substrate.
  - A major source of strain is the difference of the thermal expansion coefficients α.

Stress and strain in thin films can be large and problematic!

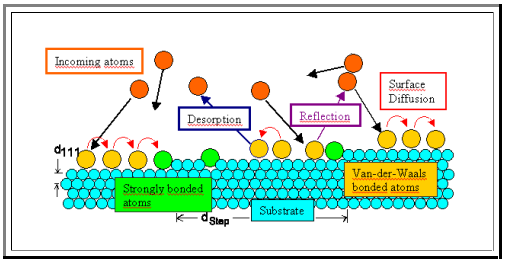
$$\epsilon_{TF} = \Delta T \cdot \Delta\alpha$$

$$\sigma_{TF} = Y \cdot \Delta T \cdot \Delta\alpha$$

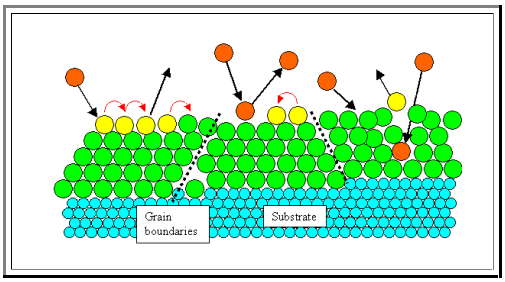
- Stress in thin film may relax by many mechanisms, and this might be good or bad:
  - Cracking or buckling.
  - plastic deformation.
  - Viscous flow.
  - Diffusion.
  - Bending of the whole system (Warpage).
- Warpage can be a serious problem in semiconductor technology.



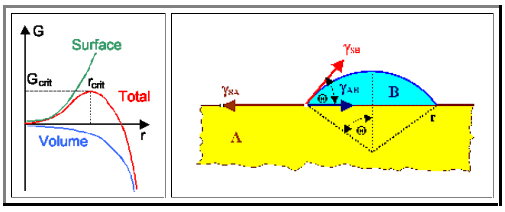
- Deposition of a thin layer must start with a "clean" substrate surface on which the first atomic / molecular layer of the film must nucleate.
- There are many possible interactions between the substrate and "first" incoming atoms.
  - As the interaction energy goes up we move from "some" absorption to physisorption (secondary bonds are formed) to chemisorption (full bonding)
  - The sticking coefficient is a measure of the likelihood to find an incoming atom in the thin film forming.
  - Immobilization by some bonding is more likely at defects (=more partners). The initial stage of nucleation is thus very defect sensitive.

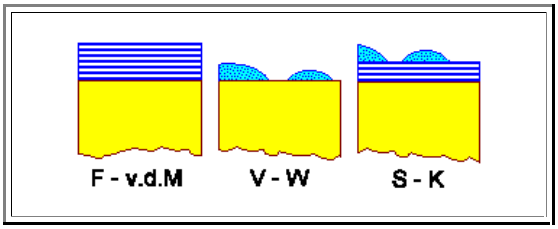


- Simple surface steps qualify as efficient "defects" for nucleation.
  - Small deviations from perfect orientation provide large step densities. Nucleation therefore can be very sensitive to the precise **{hkl}** of the surface
  - Intersections of (screw) dislocation lines with the surface also provide steps.
  - This may cause grain boundaries and other defects in the growing layer.
  - Scanning probe microscopy gives the experimental background

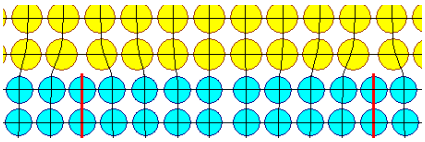


- There is always a nucleation barrier that has to be overcome for the first **B**-clusters" to form on **A**
  - the three involved interface energies, all expressed in the "wetting angle", plus possibly some strain are the decisive inputs for the resulting growth mode.
    - **Frank - van der Merve**: Smooth layer-by-layer growth
    - **Vollmer - Weber**: Island growth
    - **Stranski - Krastonov**: Layer plus island growth





- Epitaxial layers are crucial for semiconductor technology.
  - Misfit of lattice constants will produce strained layers upon epitaxial growth; strain relief happens by the formation of misfit dislocations.
  - Misfit dislocations must be avoided at all costs!
  - Below a usually rather small critical thickness d_{crit} of the the thin layer no misfit dislocations will occur.
  - Rule of thumb:
    0.5 % misfit $\Rightarrow$ $d_{crit} \approx$ **10 nm**



- The internal structure of thin films can be anything known from bulk materials plus some (important!) specialities.

a-Si: Micro electronics
a-Si:H: Solar cells, **LCD** displays
µc-Si:H: Solar cells

- Properties of thin films can be quite different from that of the bulk material.
  - The reason can be differences in length scales.
  - Semiconductor technology relies to some extent on superior thin film properties.

Much better in thin films

- **Electrical break-down field strength of dielectrics.**
- **Critical current densities in conductors.**

Exercise 3.6-1

All Questions to 3

4. Getting Started

4.1 Input to Si Processing in an Industrial Environment

4.1.1 The Factory

4.1.2 Producing Semiconductor-Grade Silicon

4.1.3 Silicon Crystal Growth and Wafer Production

4.1.4 Summary to: 4.1 Input to Si Processing in an Industrial Environment

4.2 Other Semiconductor Crystal Growth Technologies

4.2.1 Single Crystals Other Than Si

4.2.2 No Need For Single Crystals

4.2.3 Summary to: 4.2 Other Semiconductor Growth Technologies

4.3 Infrastructure

4.3.1 Cleanrooms and Defects

4.3.2 Packaging and Testing

4.3.3 Working in Chip Development and Production

4.3.4 Generation Sequences

4.3.5 Summary to: 4.3 Infrastructure

4.4 Summary

4.4.1 Summary to: 4. Getting Started

4. Getting Started

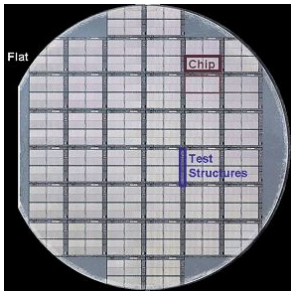
4.1 Input to Si Processing in an Industrial Environment

4.1.1 The Factory

A Chip Factory

- Let's have a quick look a factory where *semiconductor technology* is used to make a **product**. Behold the basic law of semiconductor processing in this context:

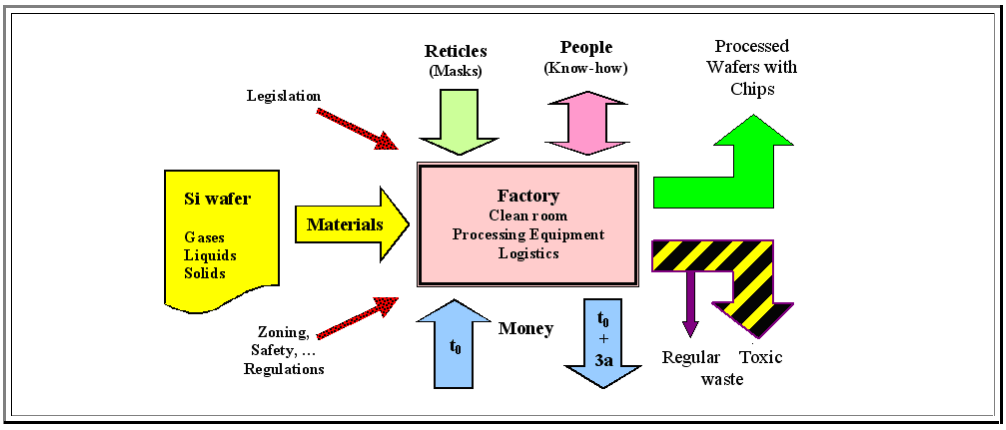
Semiconductor Technology is for making Products



Processed wafer (**16 Mbit DRAM**; **1990**)

- Let's pick the most advanced product, a **Si** microelectronic circuit - a **chip**, in other words.
  - There are several factories involved before you get a packaged chip that you can solder on a circuit board (actually you, personally, would have a hard time doing that, but some machines can). Here we look at the most expensive and complex part of the production chain: the huge **cleanroom** where **Si wafers** go in - and **Si** wafers come out, but now with completed chips on them.
  - The processed wafers as shown in the picture will be turned into packaged chips somewhere else - often somewhere really else, like **8.000 km** away.
  - We can look at the factory as a kind of a (**$3 \cdot 10^9$€**) *black box* and just look at inputs and outputs.

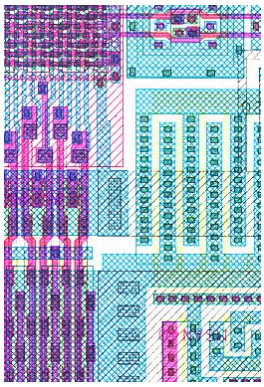
- Severely simplified, here is the basic semiconductor technology plant with its inputs and outputs:



- Let's look at the boring to annoying factors first:
  - You need a hell of a lot of **money** as *input* to get going and to keep going. Cost of building and equipping the factory (**some 10^9€**), of developing the product and technology needed, paying the **2.000 - 3.000** people working there (it's running **24 hours** a day, **7** days a week), paying for the input, and paying taxes etc. simply adds up. You can not reasonably expect the factory to actually sell anything and create money as an *output* before 3 years or so are over.
  - You need to know federal law, state law, local (zoning law), safety regulation, working condition regulation and so on, and you must strictly comply - otherwise you will be in deep trouble. Are you allowed to do you **24 hours** / **7** days shift? Overtime regulations? Safety issues?

- Obviously you need people - from lowly operators, via process engineers, team leaders, group leader to high-up management. You need people able to keep sophisticated equipment running, and you need people able to buy sophisticated equipment without getting over-charged and running up trouble with tax and custom authorities, and people able to sell your chips with a profit, if that is possible. The must bring in a lot of know-how, and they take out a lot of know-how by learning on the job.

- Now let's look at the more interesting input. Large groups of extremely knowledgeable and sophisticated people have defined what the product should be, what kind of performance it should have, and based on this, came up with the circuit diagrams and it's transformation into a layout, the precise structure in space that this or that layer is supposed to have on your wafer.
  - This is encoded into a set of "**reticles**" or masks - a kind of slide. one state-of-the art reticle may cost several **10.000 €** to make; you need about **20** for a product. So your box full of reticles is not exactly cheap, but the total costs of just making a set of reticles pales compared to the costs already incurred for coming up with the lay-out.
  - This is essentially the basic product input coming from you colleagues from electrical engineering and computer science, who have learned how to do this.



Small part of some lay-out

- Now let's look at the materials input side.
  - Foremost you need **Si** wafers; several **100** to several **1.000** per day; depending on the size of your factory and the complexity of your product. If your chip takes **500** processing steps before it is finished (normal number for normal chips) you cannot process as much per day compared to a more simple chip (e.g. power chips) needing only **300** process steps.
  - We skip the solid materials (the link provides some examples) and turn to liquids. You need large quantities of simple liquids like water, except that you need it ultrapure (the factory will have a rather large water processing plant in the basement), and extremely dangerous liquids like (ultrapure) hydrofluoric acid (**HF**) as well as other dangerous acids and bases.
  - Then you need a lot of special gases, in particular the most toxic inorganic gases known to mankind: Arsine (**AsH_3**) and Phosphine (**PH_3**). Other (ultrapure, of course) gases needed in high volume are merely explosive, e.g. Silane (**SiH_4**) or quite uncommon outside of semiconductor technology (e.g. **WF_3**). More about that in the link
- We will encounter most of these materials when we look at processes, so we will not go into any more details here.

- On the output side we produce wafers containing fully processed, but not yet measured and packaged chips. Some of the chips on a wafer may be fully functional, some may be faulty. The percentage of "good" chips is called the (wafer) **yield** η of your factors. It is the most important single number characterizing your factory, and it should be **100 %**, of course.
  - Haha! When you start to mass produce a new chip generation, you will be lucky if you make $\eta \approx$ **20 %**. From then on it's hard work to get it up to **80 %**, maybe **90 %** over the next few years before your chip is completely outdated anyway.
- In any case, you ship your finished wafers to a place where the get measured (a pretty complex task). The good ones (after slicing the wafer to separate them) then get packaged (you need another not too small factory for that - and plenty of know-how and specialized materials)
  - The they get measured again and, if still working, sold - typically for a few **€**; a few chosen ones (microprocessors) for a few **100 €**
  - Do you see the (economic) problem? By the time you sell your first chip of the new generation, you have already invested a few billion **€** in its development and production.

You need to sell one hell of a lot of chips (at a profit!) to recover your costs!

- In other words: *Semiconductor technology* always has a money aspect. It just doesn't make much sense to consider it only from a detached and purely scientific point of view.

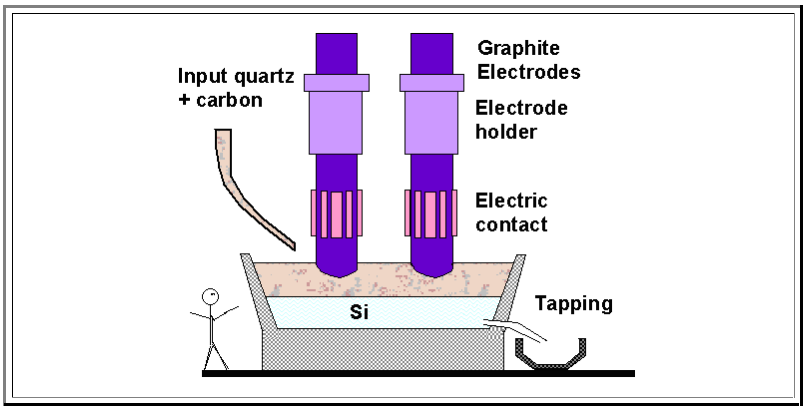
4.1.2 Producing Semiconductor-Grade Silicon

Introductory Remarks

- It is written somewhere that in the beginning God created heaven and the earth. It is not written from what.
  - We do not know for sure what the heaven is made of but we do know what the the earth is made of, at least as far as the upper crust is concerned. Interestingly enough, he (or she) created mostly **Silicon** and Oxygen with some dirt (in the form of the other **90** elements) thrown in for added value.
  - Indeed, the outer crust of this planet (lets say the first **100 km** or so) consists of all kinds of silicates - **Si** + **O** + something else - so there is no lack of **Si** as a raw material. **Si**, in fact, accounts for about **26 %** of the crust, while **O** weighs in at about **49 %**.
- However, it took a while to discover the element **Si**. **Berzellius** came up with some form of it in **1824** (probably amorphous), but it was **Deville** in **1854** who first obtained regular crystalline **Si**.
  - This is simply due to the very high chemical reactivity of **Si**. Pure **Si** (not protected by a thin layer of very stable SiO_2 as all **Si** crystals and wafers are) will react with *anything*, and that creates one of the problems in making it and keeping it clean.
  - *Liquid* **Si** indeed does react with all substances known to man - it is an universal solvent. This makes crystal growth from liquid **Si** somewhat tricky, because how do you contain your liquid **Si**? Fortunately, some materials - especially SiO_2 - dissolve only very slowly, so if you don't take too long in growing a crystal, they will do as a vessel for the liquid **Si**.
  - But there will always be some dissolved SiO_2 and therefore oxygen in your liquid **Si**, and that makes it hard to produce **Si** crystals with very low oxygen concentrations.
- What we *need*, of course, are **Si crystals** - in the form of **wafers** - with extreme degrees of perfection.
  - What we *have* are ***inexhaustible*** resources of **Silicondioxide**, SiO_2, fairly clean, if obtained from the right source. Since there is no other material with properties so precisely matched to the needs of the semiconductor industry, and therefore of the utmost importance for our modern society, the production process of **Si** wafers shall be covered in a cursory way.

Producing "Raw" Silicon

- Fortunately, the steel industry needs **Si**, too. And **Si** was already used as a crucial alloying component of steel before it started its career as the paradigmatic material of our times.
  - Most of the world production of **raw Si** still goes to the steel industry and only a small part is ***diverted*** for the semiconductor trade. This is why this stuff is commonly called "**metallurgical grade**" **Si** or ***MG-Si*** for short. The world production in **2006** was around **4** Mio tons per year.
  - How is **MG- Si** (meaning poly crystalline material with a purity of about **99%**) made? More or less like most of the other metals: Reduce the oxide of the material in a furnace by providing some reducing agent and sufficient energy to achieve the necessary high temperatures..
- Like for most metals, the reducing agent is *carbon* (in the form of coal or coke (**=** very clean coal)). The necessary energy is supplied electrically.
  - Essentially, you have a huge furnace (lined with **C** which will turn into very hard and inert **SiC** anyway) with three big graphite electrodes inside (carrying a few **10.000 A** of current) that is continuously filled with SiO_2 (**=** quartz sand) and carbon (**=**coal) in the right weight relation plus a few added secret ingredients to avoid producing **SiC**. This looks like this



  - The chemical reaction that you want to take place at about **2000 °C** is

$$SiO_2 + 2C \Rightarrow Si + 2CO$$

- But there are plenty of other reactions that may occur simultaneously, e.g. **Si + C** ⇒ **SiC**. This will not only reduce your yield of **Si**, but ***clog up*** your furnace because **SiC** is not liquid at the reaction temperature and extremely hard - your reactor ends up as a piece of junk if you make **SiC**.

Still, we do not have to worry about **MG-Si** - a little bit of what is made for the steel industry will suffice for all of **Si** electronics applications.

- What we do have to do is to *purify* the **MG-Si** - about 10^9 fold!

This is essentially done in three steps:

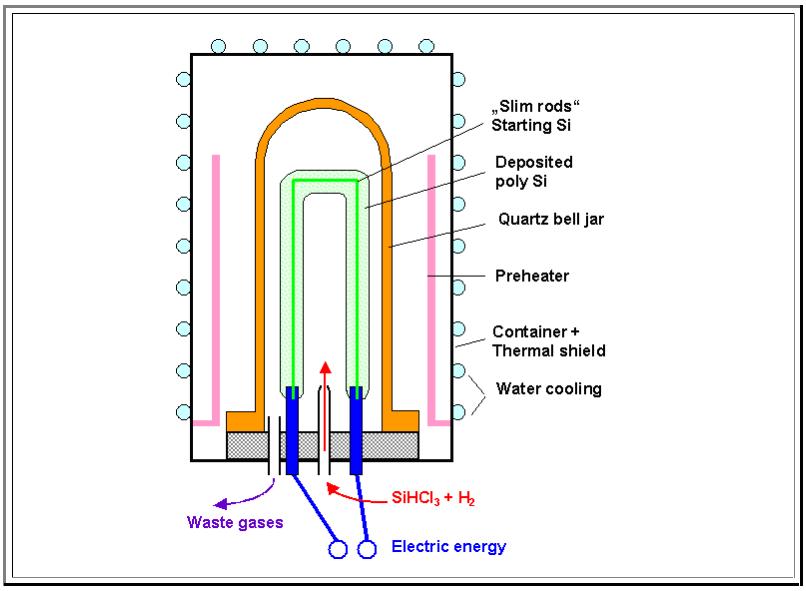
- *First*, **Si** is converted to $SiHCl_3$ in a "fluid bed" reactor via the reaction

$$Si + 3HCl \Rightarrow SiHCl_3 + H_2$$

- This reaction (helped by a catalyst) takes place at around **300 °C**. The resulting **Trichlorosilane** is already much purer than the raw **Si**; it is a liquid with a boiling point of **31.8 °C**.
- *Second*, the $SiHCl_3$ is distilled (like wodka), resulting in extremely pure Trichlorosilane.
- *Third*, high-purity **Si** is produced by the **Siemens process** or, to use its modern name, by a "**Chemical Vapor Deposition**" (***CVD***) process - a process which we will encounter more often in the following chapters.

Producing Doped Poly-Silicon

The doped **poly-Si** (not to be confused with the poly-**Si** layers on chips) used for the growth of single **Si** crystals is made in a principally simple way which we will discuss by looking at a poly-**Si CVD** reactor

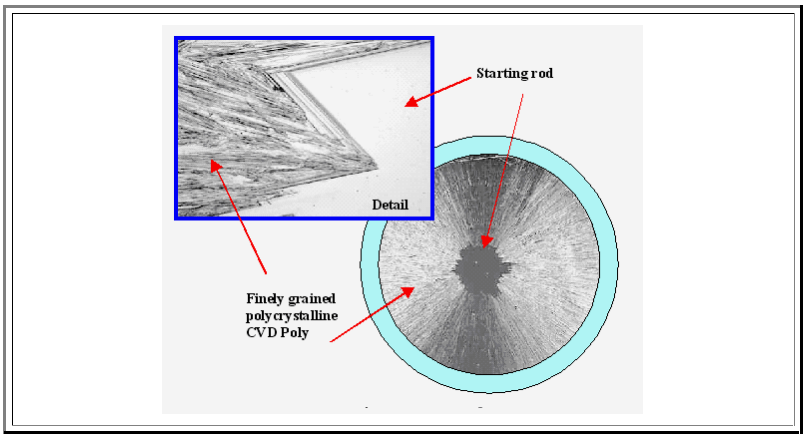


- In principle, we have a vessel which can be evacuated and that contains an "U" shaped arrangements of slim **Si** rods which can be heated from an outside heating source and, as soon as the temperature is high enough (roughly **1000 °C**) to provide sufficient conductivity, by passing an electrical current through it.
- After the vessel has been evacuated and the **Si** rods are at the reaction temperature, an optimized mix of $SiHCl_3$ (**Trichlorosilane**), H_2 and doping gases like AsH_3 or PH_3 are admitted into the reactor. In order to keep the pressure constant (at a typical value of some mbar), the reaction products (and unreacted gases) are pumped out at a suitable place.
- On *hot* surfaces - if everything is right this will only be the **Si** - a chemical reaction takes place, reducing the $SiHCl_3$ to **Si** and forming **HCl** (hydrochloric acid) as a new compound:

$$SiHCl_3 + H_2 \Rightarrow Si + 3\,HCl$$

- Similar reactions provide very small but precisely measured amounts of **As**, **P** or **B** that will be incorporated into the growing polysilicon
- The **Si** formed will adhere to the **Si** already present - the thin rods will grow as fresh **Si** is produced. The incorporation of the dopants will produce *doped polysilicon*.

- In principle this is a simple process, like all **CVD** processes - but not in reality. Consider the complications:
  - You have to keep the **Si** ultrapure - all materials (including the gases) must be specially selected.
  - The chemistry is extremely dangerous: **AsH_3** and **PH_3** are among the most poisonous substances known to mankind; **PH_3** was actually used as a toxic gas in world war **II** with disastrous effects. **H_2** and **$SiHCl_3$** are easily combustible if not outright explosive, and **HCl** (in gaseous form) is even more dangerous than the liquid acid and extremely corrosive. Handling these chemicals, including the safe and environmentally sound disposal, is neither easy nor cheap.
  - Precise control is not easy either. While the flux of **H_2** may be in the **100 liter/min** range, the dopant gases only require **ml/min**. All flow values must be precisely controlled and, moreover, the mix must be homogeneous at the **Si** where the reaction takes place.
  - The process is slow (about **1 kg/hr)** and therefore *expensive*. You want to make sure that your hyperpure (and therefore expensive) gases are completely consumed in the reaction and not wasted in the exhaust - but you also want high throughput and good homogeneity; essentially conflicting requirements. There is a large amount of optimization required!
  - And from somewhere you need the slim rods - already with the right doping.
- Still, it works and abut **10.000 tons** of poly-**Si** are produced at present (**2000**) with this technology, which was pioneered by **Siemens AG** in the sixties for the microelectronic industry. (in **2007** it is more like **21.000 to** plus another **30.000** tons for the solar industry).
  - **Electronic grade Si** is not cheap, however, and has no obvious potential to become very cheap either. The link provides todays specifications and some more information for the product. Here is an example for the poly-crystalline rods produced in the Siemens process:

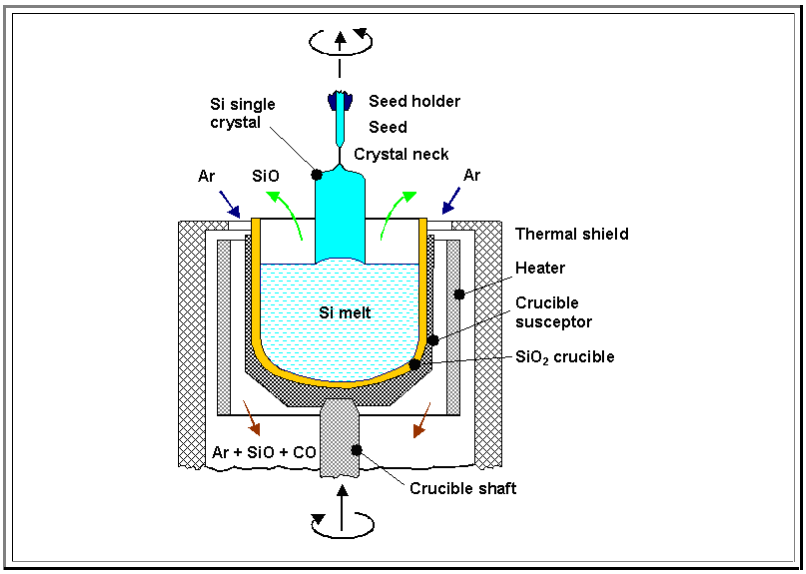


  - While this is not extremely important for the microelectronics industry (where the added value of the chip by far surpasses the costs of the **Si)**, it prevents other **Si** products, especially *cheap solar cells* (in connection with all the other expensive processes before and after the poly-**Si** process). Starting with the first oil crisis in **1976**, many projects in the USA and Europe tried to come up with a cheaper source of high purity poly-**Si**, so far without much success.
  - By now. i.e. in **2007**, demand for electronic grade **Si** si surging because of a booming solar cell industry. A short overview of the current Si crisis can be found in the link.

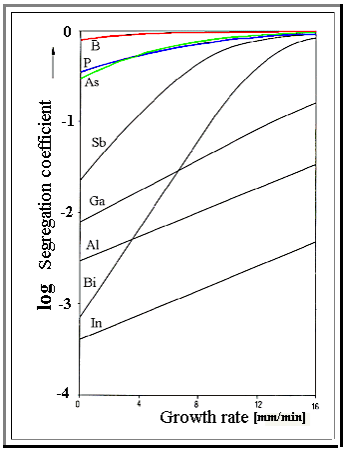
4.1.3 Silicon Crystal Growth and Wafer Production

Single Crystal Growth

- We now have hyperpure poly-**Si**, already doped to the desired level, and the next step must be to convert it to a *single crystal*. There are essentially two methods for crystal growth used in this case:
  - Crystals grown by the **Czochralski method**or ***crucible* grown** crystals (***CZ*** crystals).
  - **Float zone** or ***FZ*** crystals.
- The latter method produces crystals with the highest purity, but is not easily used at large diameters. **150 mm** crystals are already quite difficult to make and nobody so far has made a **300 mm** crystal this way. Float zone crystal growth, while the main method at the beginning of the **Si** age, is now only used for some specialties and therefore will not be discussed here; some details can be found in the link.
  - The Czochralski method, invented by the Polish scientist **J. Czochralski** in **1916**, is the method of choice for high volume production of **Si** single crystals of exceptional quality and shall be discussed briefly. Below is a schematic drawing of a crystal growth apparatus employing the Czochralski method. More details can be found in the link.



  - Essentially, a crystal is "pulled" out of a vessel containing liquid **Si** by dipping a **seed crystal** into the liquid which is subsequently slowly withdrawn at a surface temperature of the melt just above the melting point. There is a little trick in doing that, that is the key to dislocation -free crystal growth and thus to all of microelectronics; see the link above.
  - The *pulling rate* (usually a few mm/min) and the *temperature profile* determines the crystal diameter (the problem is to get rid of the heat of crystallization).
  - Everything else determines the quality and homogeneity - crystal growing is still as much an art as a science! Some interesting points are contained in the link.
- Here we only look at one major point, the **segregation coefficient k_{seg}** of impurity atoms.
  - The segregation coefficient in thermodynamic equilibrium gives the relation between the concentration of impurity atoms in the growing crystal and that of the melt. It is usually much lower than **1** because impurity atoms "prefer" to stay in the melt. This can be seen from the liquidus and solidus lines in the respective phase diagrams.
  - In other words, the *solubility* of impurity atoms in the melt is larger than in the solid.
  - "Equilibrium" refers to a growth speed of **0 mm/min** or, more practically, very low growth rates. For finite growth rates, k_{seg} becomes a function of the growth rate (called k_{seff}) and approximates **1** for high growth rates (whatever comes to the rapidly moving interface gets incorporated).
- This has a positive and a negative side to it:
  - On the positive side, the crystal will be *cleaner* than the liquid, crystal growing is simultaneously a purification method. Always provided that we discard the last part of the crystal where all the impurities are now concentrated. After all, what was in the melt must be in the solid after solidification - only the distribution may now be different.
  - This defines the negative side: The *distribution of impurities* - and that includes the doping elements and oxygen - *will change along the length of a crystal* - a homogeneous doping etc. is difficult to achieve.
  - That segregation can be a large effect with a sensitive dependence on the growth rate is shown below for the possible doping elements; the segregation coefficients of the unwanted impurities is given in a table.



Atom	Cu	Ag	Au	C	Ge	Sn	
k_{seg}	$4 \cdot 10^{-4}$	$1 \cdot 10^{-6}$	$2{,}5 \cdot 10^{-5}$	$6 \cdot 10^{-2}$	$3{,}3 \cdot 10^{-1}$	$1{,}6 \cdot 10^{-2}$	
Atom	**O**	**S**	**Mn**	**Fe**	**Co**	**Ni**	**Ta**
k_{seg}	1,25	$1 \cdot 10^{-5}$	$1 \cdot 10^{-5}$	$8 \cdot 10^{-6}$	$8 \cdot 10^{-6}$	$4 \cdot 10^{-4}$	$1 \cdot 10^{-7}$

- We recognize *one* reason why practically only **As**, **P**, and **B** is used for doping! Their segregation coefficient is close to **1** which assures half-way homogeneous distribution during crystal growth. Achieving homogeneous doping with **Bi**, on the other hand, would be exceedingly difficult or just impossible.

Present day single crystals of silicon are the most perfect objects on this side of Pluto - remember that perfection can be measured by using the second law of thermodynamics; this is not an empty statement! A very interesting and readable article dealing with the history and the development of **Si** crystal growth from W. **Zulehner** (Wacker Siltronic), who was working on this subject from the very beginning of commercial **Si** crystal growth until today, can be found in the link.

- What the finished crystal looks like can be seen in the link. What we cannot see is that there is no other crystal of a different material that even comes close in size and perfection.
- Our crystal does not contain dislocations - a unique feature that only could be matched by Germanium crystals at appreciable sizes (which nobody grows or needs)[1)]. It also does not contain many other lattice defects. With the exception of the doping atoms (and possible interstitial oxygen, which often is wanted in a concentration of about **30 ppm**), substitutional and interstitial impurities are well below a **ppb** if not **ppt** level (except for relatively harmless carbon at about **1 ppm**) - unmatched by most other "high purity" materials.
- Our crystal is homogeneous. The concentration of the doping atoms (and possibly interstitial oxygen) is radially and laterally rather constant, a feat not easily achieved.

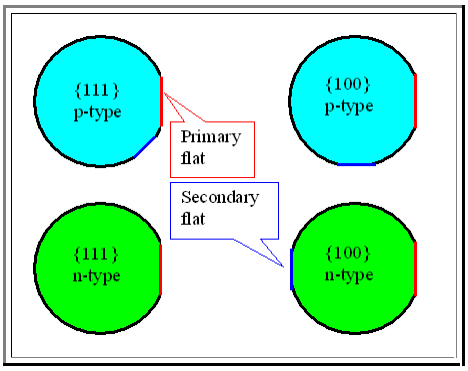
The crystal is now ready for cutting into wafers.

Wafer Technology

It may appear rather trivial now to cut the crystal into slices which, after some polishing, result in the **wafers** used as the starting material for chip production.

- However, it is *not* trivial. While a wafer does not look like much, its not easy to manufacture. Again, making wafers is a closely guarded secret and it is possibly even more difficult to see a wafer production than a single **Si** crystal production.
- First, wafers must all be made to exceedingly tight geometric specifications. Not only must the diameter and the thickness be precisely what they ought to be, but the flatness is constrained to about **1 µm**.
- This means that the polished surface deviates at most about **1 µm** from an ideally flat reference plane - for surface areas of more than **1000 cm^2** for a **300 mm** wafer!
- And this is not just true for one wafer, but for all **10.000** or so produced daily in *one* factory. The number of Si wafers sold in **2001** is about **100.000.000** or roughly **300.000** a day! Only tightly controlled processes with plenty of know-how and expensive equipment will assure these specifications. The following picture gives an impression of the first step of a many-step polishing procedure.

- In contrast to e.g. polished metals, polished **Si** wafers have a *perfect*surface - the crystal just ends followed by less than two **nm** of "**native oxide**" which forms rather quickly in air and protects the wafer from chemical attacks.
  - Polishing **Si** is not easy (but fairly well understood) and so is keeping the surface clean of particles. The final polishing and cleaning steps are done in a *cleanroom* where the wafers are packed for shipping.
  - Since chip structures are always aligned along crystallographic directions, it is important to indicate the crystallography of a wafer. This is done by grinding **flats** (or, for very large wafer - **200 mm** and beyond **notches**) at precisely defined positions.
  - The flats also encode the doping types - mix ups are very expensive! The convention for flats is as follows:



  - The main flat is always along a **<110>** direction. However, many companies have special agreements with wafer producers and have "customized" flats (most commonly no secondary flat on **{100} p**-type material).
  - More about flats in the link

- Typical wafer specifications may contain more than **30** topics, the most important ones are:
  - *Doping type*: **n** or **p**-type (**p**-type is by far the most common type) and *dopant* used (**P**, **As** or **B**). *Resistivity* (commonly between **100** Ω**cm** to **0,001** Ω**cm** with **(5 - 1)** Ω**cm** defining the bulk of the business. All numbers with error margins and homogeneity requirements
  - *Impurity concentrations* for metals and other "life time killers" (typically below $\mathbf{10^{12}\ cm^{-3}}$), together with the *life time* or diffusion length (which should be several **100 µm**).
  - *Oxygen* and *carbon* concentration (typically around $\mathbf{6 \cdot 10^{17}\ cm^{-3}}$ or $\mathbf{1 \cdot 10^{16}\ cm^{-3}}$, respectively. While the carbon concentration just has to be low, the oxygen concentration often is specified within narrow limits because the customer may use "**internal gettering**", a process where oxygen precipitates are formed intentionally in the bulk of the wafer with beneficial effects on the chips in the surface near regions.
  - *Micro defect densities* (after all, the point defects generated in thermal equilibrium during crystal growth must still be there in the form of small agglomerates). The specification here may simple be: ***BMD*** ("bulk micro defect") density $\mathbf{=0\ cm^{-3}}$. Which simply translates into: Below the detection limit of the best analytical tools.
  - *Geometry*, especially several parameters relating to flatness. Typical tolerances are always in the **1 µm** regime.
  - Surface *cleanliness*: No particles and no atomic or molecular impurities on the surface!
- This link provides a grapical overview of the complete production process - from sand to **Si** wafers - and includes a few steps not covered here.
  - Appreciate that the production of wafers - at last several thousands per day - with specifications that are always at the cutting edge of what is possible - is an extremely involved and difficult process.
  - At present (Jan. **2004**), there are only a handful of companies world wide that can do it. In fact, **4** companies control about **80%** of the market.

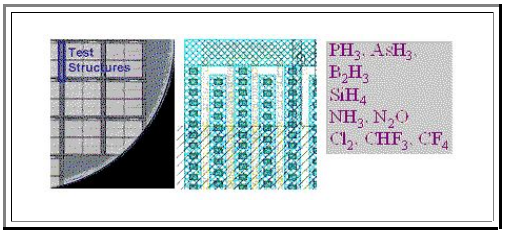
- This link leads to a recent (**1999**) article covering new developments in **Si CZ** crystal growth and wafer technology (from A.P. **Mozer**; Wacker Siltronic) and gives an impression of the richness of complex issues behind the production of the humble **Si** wafer.
- This link shows commercial wafer specifications.
  - To give an idea of the size iof the industry: In **2004** a grand total of about **4.000.000 m^2** of polished **Si** wafers was produced, equivalent to about **$1.25 \cdot 10^8$ 200 mm** wafers.

---

[1] No longer true in **2004**! Germanium wafers may (or may not) make a come-back; but they are certainly produced again.

4.1.4 Summary to: 4.1 Input to Si Processing in an Industrial Environment

- Semiconductor technology happens in factories. They need special materials, "reticles" (= structures), "know-how" and huge amoundt of money (= capital) as major inputs
  - It's always about money! Only mass production will recover large investments.
  - The materials side always contains semiconductor substrates ("wafers") and often very dangerous special "raw" materials.
  - A number tells it all: **500 - 1.000** wafers /day are processed in a large **Si** "wafer fab"



- Three big steps to **Si** wafers
  - **Si** single crystal growth is done by "Czochralski process" (**CZ**).
  - Dislocation-free crystals are possible but "bulk microdefects" and impurities cannot be totally avoided.
  - Nearly perfect **300 mm** wafers are standard.

Sand (SiO_2)	⇒	**Metallurgical Si**
Metallurgical Si	⇒	**clean (doped) poly-Si.**
Poly-Si	⇒	**Single crystal / wafer**

Exercise 4.1-1

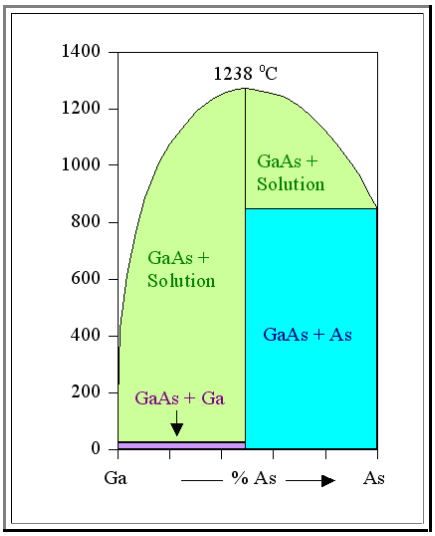
All Questions to 4.1

4.2 Other Semiconductor Crystal Growth Technologies

4.2.1 Single Crystals Other Than Si

Gallium Arsenide

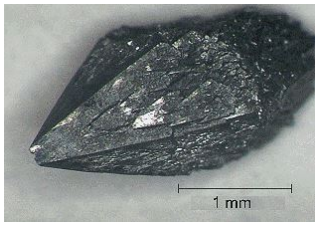
- We know that we need **III-V** semiconductors for optoelectronic products, so we have to make single crystals as input to our optoelectronics factory.
  - We may use those single crystal for the product or we may just use them as the substrate for the thin film that will contain the product. Whatever - they better be as perfect as possible because any defect present at the surface (e.g. a dislocation ending a the surface) will simply continue in any epitaxial layer we might produce.
  - Let's look at **GaAs** single crystal production as en example for other **III-V**'s like **InP** or **GaP**. However, not all **III-V**'s can be grown on a similar way; we will come to that.
- The Czochralski method for growing crystals should work for **GaAs**, too. So let's try it and see why it is far more difficult to grow compound single crystals than **Si** (or **Ge**) single crystals
  - The melting point of **GaAs**, however, is T_m**(GaAs)=1.238 °C**. This is a first indication that things might not be quite as simple as they look. Obviously, we now need a phase diagram of the **Ga** - **As** system. Here it is: We mix **Ga** and **As** in a (molar) ratio of **50:50** and melt it in a crucible. That should be easy, because both elements have low melting points (**Ga**: **30 °C**; **As**: **? °C** - it actually doesn't melt but evaporates at a sublimation or boiling point of **615 °C**).
  - So what happens if our mixture is not **50 : 50** but **49.999 : 50.001**? We will have little droplets of liquid in our growing crystal, that solidify later and probably cause defects at a concentration that is large - in comparison to what we call "low defect concentration".
  - To make things worse, whatever you mix together and melt, at the required **1.238 °C** the **As** in the melt will evaporate off quickly; the concentration thus changes as a function of time.



- The catchword now is "**liquid encapsulation technique**" (***LEC***).
  - The melt in the crucible will be enclosed in B_2O_3, a viscous liquid at the growth temperature. It encloses the entire melt, i.e. sits between crucible and melt and on the surface. Obviously, we have another case of getting surface energies right here. The seed crystal is now dipped through that layer into the melt - and so on.
  - There is no way that the crystal you get will come close in quality to **Si** crystals. The present state of the art is given in this relatively short article given at that Czochralski-Symposium 2003 in Poland
  - You will not understand some of the topics in the article, but you will get a feeling for the complexity involved in growing good **GaAs** single crystals.
- Of course, any other **III-V** compound crystals grown in a similar way runs into all the difficulties hinted at above plus some of its own. This is particularly true for **GaN**, which will be shortly discussed below.
- Less of course, there are some other ways of growing **GaAs** (and other) single crystals but we cam't go for that here.

Gallium Nitride

- In the (slightly changed) words of Dr. Yoke Khin **Yap**; Department of Physics, Michigan Technological University:
  - High-quality **GaN** single-crystals are still not possible because the melt growth of **GaN** single-crystal is prohibited due to the extremely high decomposition pressure (**~ 45000 atm**) involved at the melting point (**~2500 °C**). At present, **GaN** single-crystals can be grown by the solution method at **(1300 - 1600) °C** and N_2 gas pressure of **(10000 - 17000) atm**. Because of the need of high temperatures and gas pressures, the solution growth method is not suitable for mass-production of **GaN** single-crystals.
  - We have been growing high-quality **GaN** single-crystals by a **Na** flux method at **800 °C** and **<50 atm**. This technique is promising for mass production of **GaN** single-crystals at significantly low temperatures and pressures. A similar approach is also applicable for growing other nitride crystals like **w-AlN** and **h-BN**.



Single crystal of **GaN**

From Yap Research Lab, Michigan Tech

- So how come we have a flourishing **GaN** industry, making **UV** - blue **LED**'s and Lasers?
  - Because the input to a "commercial" **GaN** factory consists of substrates with a thin **GaN** layer on it. The substrate is either Al_2O_2 with a lattice misfit of **13.8%** and a thermal expansion coefficient mismatch of **25.5%**
  - Not good. So you look for a better substrate and find **SiC**. Not so hot either, look at our "master graph". Any you now need **SiC** single crystal (see below).
- Nevertheless, **GaN** made it into the ranks of serous semiconductors with billion **€** or so business attached. What we learn form this is that semiconductor technology, like politics, is the art of the possible (*"Politik ist die Kunst des Möglichen*; Otto von **Bismarck**)

Silicon Carbide

- We already know that **SiC** comes in many polytypes. That makes us suspect that growing a single crystal of just one polytype may not be so easy.
  - The fact that its melting point is **< 2.500 °C** and that **SiC** is extremely hard does not make the task to produce "perfect" single-crystalline wafers any easier.
  - On the other hand, inside the huge reactors designed for making absolutely imperfect poly-crystalline **SiC** of any polytype for products like grinding paper, one does find nice single crystals hanging on the wall on occasion, so single crystal growth should be possible.
- Here is how it is done.
- In most cases, large single crystals are grown from a melt or some solution (e.g. quartz, or sugar if you leave you coffee cup around too long), but this is not a feasible option for **SiC** single crystal growth since **SiC** does not have a liquid phase under normal conditions (i.e. without applying a large pressure). **SiC** is also extremely hard (close to diamond) and therefore has a high melting point (or is it the other way around?).
  - There is also, in principle, no crucible material that could contain molten **SiC** at is nominal melting point temperature of **< 2.500 °C**. Nevertheless, **SiC** was grown from a melt at **2200 °C** and **150 bar** in a recent study, but this is probably not a commercially viable process.
- We need a basically new method of crystal growth, and the main method used nowadays is **physical vapor transport** (***PVT***) also known as *seeded sublimation growth* or *modified Lely method*.
  - A piece of **SiC** is heated to **(1800-2600) °C** at low pressure. Due to the high sublimation rate, **SiC** vapor forms and deposits itself on a cooler single-crystalline seed crystal. Straightforward and basically simple, as shown in the schematic picture below.
- However, pondering the situation, some questions should come to mind:
  - What materials can you use for the crucible and everything else that gets hot? After all, not many materials can cope with temperatures above **2000 °C**! Well, you are basically stuck with graphite, and maybe a bit of **Ta** here and there. That means, of course, that you are forming **SiC** also on your crucible walls and everywhere else. If it flakes off, you will have a defect problem.
  - What kind of *growth rate* can you get? As you would expect: Not much! Growth rates depend on many parameters, but are in the range of **0.2 - 2 mm/hr**. That's about a factor of **50** slower than the growth rates for **Si** crystal pulling and that makes **SiC** crystal growing automatically expensive

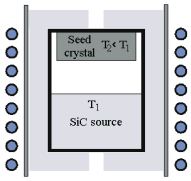
- What *polytype* will you get (hoping that it will not be a mixture)? What determines what you get? Can you control it and, if yes, how?
- Good questions! First, you might get mixtures as shown in the picture (*courtesy of the Erlangen group*). Otherwise, the following parameters are essential:

  - Polytype of the seed crystal (as you might have guessed).
  - "Face" of the seed crystal; i.e if the surface is a **C**- or a **Si** layer. If you start with a **4H-SiC** seed crystal, for example, you tend to get **4H-SiC** if you have a **C**-face, and **6H-SiC** if you have a **Si** face. Why? Nobody really knows.
  - Temperature difference and - gradient betwen **SiC** source and seed. Small values tend to favor **4H-SiC**, larger values **6H-SiC** growth.
  - Gas composition. Whatever gas you add will influence the polytype you obtain. **C**-rich gases, for example, promote **4H-SiC** growth
  - The pressure, oddly enough, seems not to have a large influence on polytypie.

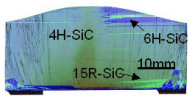
  Note that while the polytype **6H** is the easiest to grow, **4H** would be favored by the power electronics industry.

Last not least: What kind of crystal quality do you get? What is the dislocation density?

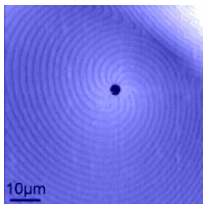
- The bad news is: the dislocation density is high. The good news is, you do not worry too much about that - you worry about something weird called "**micropipe** (and mixtures of polytypes, and all kinds of stacking faults or special boundary faults, and carbon inclusion, or **Si** inclusion, or big voids, ...).
- What are micropipes? Well, micropipes are hollow channels running through the lattice; the diameter of these pipes is **(0.1 - 5) µm**. One might consider them to be screw dislocations with a gigantic Burgers vector and a hollow core.
- Micropipes also will definitely kill any device that contains one of them. They thus must be avoided as much as possible! Micropipes are also somehow connected to the growth mechanism of the crystal. This is neatly illustrated in the picture on the right (taken with a scanning force microscope, courtesy of H. **Strunk**; Uni Erlangen) where typical growth spirals are visibly centered around a micropipe.



SiC crystal growth reactor (simplified!)



SiCsingle crystal - cross section



Micropipe with growth spirals

Let's look at the state of the art of what is around. To quote from the product sheet of the major **SiC** supplier Cree, Inc. (located somewhat ironically in Silicon Drive 4600 in Durham, North-Carolina, **USA**):

- At present, wafer diameters are **50.8 mm** or **76.2 mm**; doping (usually with **N** for **n**-type and **Al** for **p**-type) at high levels produces resistivities in the **0.0x mΩcm** region. Or there is no doping for semi-insulating stuff. **4H**- and **6H-SiC** polytypes are sold

The **2003** state of the art (mostly in the laboratories and not necessarily on the market) is summarized in the following table:

Diameter		**100 mm** "Four-inch"	For **Si**, **100 mm** was the standard back in the late **70**ties/early **80**ties).
Defects	**Micropipes**	**< 1 cm^{-2}**for **3"** **< 30 cm^{-2}**for **100 mm**	Increasing wafer size usually dramatically increases micropipe density
	Dislocations	**$3 \cdot 10^3 cm^{-2}$** achieved	Factor **10** reduction

Of course, in the many laboratories (university and industrial) devoted to **SiC**, some data might be even better.

The Rest

- Just kidding. All we can do here is to look at a list of what is left over
- **Germanium**.
  - Obviously, **Ge** wafer technology can follow the **Si** lead in principle. It should be even somewhat easier to grow **Ge** single crystals because the melting point of **Ge** is lower (**938 ^{o}C** vs. **1410 ^{o}C**, resp.)
  - However, while there is a lot of **SiO_2** around to start from, where do you find **Ge**-minerals? Indeed, the price of germane,**GeH_4**, needed for starting, is orders of magnitudes higher that that of silane, **SiH_4**.
  - All things considered, you can get pretty good dislocation free **Ge** wafers - up to one or two **300 mm** diameters - but for a price. The total quality, however, is not as good as the mass-produced **300 mm** Si wafer.
- **II-VI Semiconductors**
  - There are many kinds, and you may have to find a specific way of growing single crystals for each kind.
  - The state of the art in terms of what you can buy are small and rather imperfect single crystals of **CdSe**, **ZnO** and .....
- **Everything else** mentioned in chapter 2.
  - Forget it .No single crystals to speak of - yet.
- **Special Wafers**
  - Some **Si** microelectronic factories have switched from regular **Si** wafers as input to "***SOI***" wafers; "**Silicon on Insulator**"; e.g. AMD in Dresden
  - Read the article in the link if you want to find out what **SOI** is all about

4.2.2 No Need For Single Crystals

- There is an important product line of modern semiconductor technology could not exist at all if you would always need single crystal substrates:
  - Large area displays in the form of "**liquid crystal displays**" (**LCD**) and - quite new - "**OLED**" displays based on organic semiconductors.
  - **Solar cells** based on large-area thin film technology and employing as the main functional semiconductor:
    - **H**-rich, doped, amorphous **Si** (**a-Si:H**).
    - **H**-rich, doped, micro-crystalline **Si** (**µc-Si:H**).
    - "**CIS**" and "**CIGS**" based solar cells.
    - **CdTe** based solar cells.
- Then we have one major product line where single crystals compete with other structural forms of the semiconductor:
  - Solar cells made from single crystal wafers and solar cells made from multi-crystalline wafers
- In the laboratory is much more along these lines. There is a unifying element however, in all of this:

Large areas are needed; preferably or necessarily in one piece

- This is self-evident for solar cells, let's have a quick look at displays:
  - Any display processes light in pixels. There are two basic ways of doing this:
    1. There is always plenty of light in the "back" of the display; processing consists of allowing only the proper amount of light (and the proper color) to pass through the pixel at the proper time.
    2. The pixel actively generates the right amount of light at the right time.
  - **LCD's** and all beamers belong in the first category; **OLED** displays in the second. Of course, the second category has the option of being more energy efficient (since you don't waste most of the light you generate) but the first category is presently (**2008**) more advanced
- In both case, however, an individual pixel needs to be "told" what to do.
  - At the crossing of a matrix of (**n × m**) conductors each pixel can be individually addressed by its coordinate (**n,m**). For good display, you need to have at least one transistor at the cross-point to allow efficient addressing.
  - You thus need a large substrate that allows to make at least simple transistors with **100 %** yield (one "dead" transistor = one dead pixel!). Single crystals are mostly not large enough, would be prohibitively expensive and very troublesome because they are not transparent to light.
  - Thin layers of **a-Si:H** or sometimes microcrystalline **Si** on glass, while not good enough for, e.g., the kind of transistor needed for microprocessors or memory chips, are good enough for **LCD** displays - since about **1990**. Before that time, there simply were no (affordable) flat panel displays!
- Remember the credo: Products only sell if they are cheaper and / or better.
  - If you can make a transistor matrix on a large size substrate, you can make a flat panel display, and that is certainly *better* than the good old picture tube (***CRT***).
  - The transistor matrix is then one of the **enabling technologies** you need for the **LCD** display (the other one are the liquid crystals)

4.2.3 Summary to: 4.2 Other Semiconductor Growth Technologies

- Growing single crystals of compound semiconductors is far more difficult than for elemental semiconductors
  - Precise stoichiometry is important
  - Vapor pressures if the constituents at the melting point might be very different
  - New kinds of defects might be encountered
  - Polytypie might be encountered
- Major techniques are
  - Encapsulated **CZ**
  - Sublimation growth

GaAs:
150 mm wafers, encapsulation technique, disl. density (**10^3 - 10^6**) **cm^{-2}**

GaP, **InP**
as **GaAs** but smaller and more expensive

SiC:
100 mm wafers, sublimation technique, several polytypes available, "pipe" defects

Exercise 4.2-1

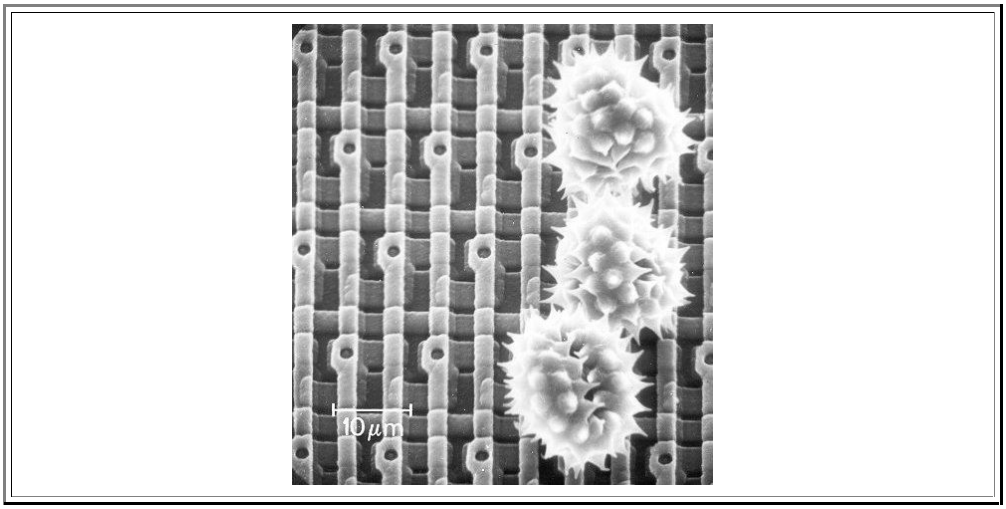
All Questions to 4.2

4.3 Infrastructure

4.3.1 Cleanrooms and Defects

Particles

- Normal air is full of "dirt" usually called **particles**. The fact that *we* cannot see them (except the bigger ones in a bright beam of light) does not mean that the air is clean. What happens when a particle (e.g. pollen, scrapings from whatever, unknown things) falls on a chip is shown in the picture below.
  - Anything that can "fall" on a chip is called a *particle*; independent of its size and of what it is. Particles smaller than some **10 µm** usually do not "feel" gravity anymore (other forces dominate); so they do not "fall" on a chip. However, they may be attracted electrostatically and that makes it quite difficult to remove them.
  - Often anything that disturbs the structure of a chip by lying *on* the layers of the integrated circuit is called a **"defect"**. Defects may not only be particles, but all kinds of other mishaps too, e.g. small holes in some coating.
  - However, we will *not* use that terminology here, but restrict the name "*defect*" to **crystal lattice defects** *in* the **Si**, i.e. *in*, not *on*, the integrated circuit.
  - A pretty old chip (a **256k** memory as sold around **1985**) was chosen for the following illustration because its structures are clearly visible. It has a few pollen grains (from "Gänseblümchen") on its surface (which essentially shows the wiring matrix of a memory array). What would have happened if a pollen grain would have fallen on the chip while it was made needs no long discussion: The chip would be dead!



- At feature sizes **< 0,2 µm**, everything that falls on a chip with sizes **> 0,1 µm** or so will be deadly. All those defects- the **particles** - must be avoided at all costs. There are three major sources of particles:
- **1.** The *air* in general. Even "clean" mountain air contains very roughly **10^6** particles **> 1 µm** per *cubic foot* (approximately **30** liters). We need a "**cleanroom**" serving two functions:
  - It provides absolutely clean air (usually through filters in the ceiling), and
  - It immediately removes particles generated somewhere in the cleanroom by pumping large amounts of clean air from the ceiling through the (perforated) floor.
  - Avoiding and removing particles while processing **Si** wafers has grown into a science and industry of its own. The link provides some information about cleanrooms and cleanroom technology.
- **2.**The *humans* working in the cleanroom.
  - Wiping your (freshly washed) hair just once, will produce some **10 000** particles. If you smoke, you will exhale thousands of particles (size about **0,2 µm**) with every breath you take. A ***TI*** add once said: Work for us and we will turn you into a non-smoker.
  - The solution is to pack you into **cleanroom garments**; what this looks like can be seen in the link. It is not as uncomfortable as it looks; but it is not pure fun either. Graphic examples of humans as a source of particles can be found in the link.
- **3.** The *machines* (called "**equipment**") that do something to the chip, may also produce particles.
  - As a rule, *whenever something slides on something else* (and this covers most mechanical movements), particles are produced. Layers deposited on chips are also deposited on the inside of the equipment; they might flake off. There is no easy fix, but two rules:

- Use special engineering and construction to avoid or at least minimize all possible particle sources, and
- Keep your equipment clean - frequent "special" cleaning is required!

But even with state-of-the art cleanrooms, completely covered humans, and optimized equipment, particles can not be avoided - look at the picture gallery to get an idea of what we are up to. The most frequent process in chip manufacture therefore is "cleaning" the wafers.

- Essentially, the wafers are immersed in special chemicals (usually acids or caustics or in combination with various special agents), agitated, heated, rinsed, spin-dried, ... , its not unlike a washing machine cycle.
- This cleaning process in all kinds of modifications is used not only for removing particles, but also for removing unwanted atoms or layers of atoms which may be on the surface of the wafers. This brings us to the next point:

Contamination and Crystal Lattice Defects

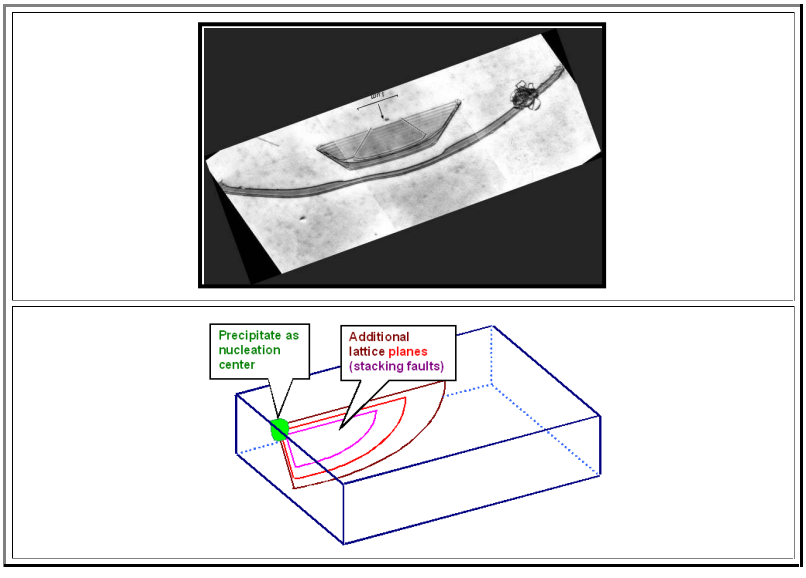
The **Si** single crystals used for making integrated circuits are the (thermodynamically) most perfect objects in existence - at least on this side of Pluto. They are in particular completely free of dislocations and coarser defects as, e.g., grain boundaries or precipitates of impurities, and have impurity concentrations typically in the **ppt** (parts per trillion) or **ppqt** (parts per quadrillion) range - many orders of magnitude below of what is normally considered "high-purity".

Defects (always now in the meaning of "*crystal lattice defects*") will without fail kill the device or change properties!

- *Dislocations* or *precipitates* in the electronically active region of the device; e.g. in or across **pn**-junctions, simply "kill" it - the junction will be more or less short-circuited.
- *Point defects* in solid solution (e.g. **Cu**, **Au**, **Fe**, **Cr**, ...most metals) in the **Si** crystal reduce the *minority carrier lifetime* and thus influence device characteristics directly - usually it will be degraded. Alkali atoms like **Na** and **K** in the gate oxides kill **MOS** transistors, because they move under the applied electrical field and thus change the charge distribution and therefore the transistor characteristics.
- But point defects do more: If they precipitate (and they all have a tendency to do this because their solubility at low temperatures is low) close to a critical device part (e.g. the interface of **Si** and $\mathbf{SiO_2}$ in the **MOS** transistor channel), they simply kill that transistor. Possibly worse: Even very small precipitates of impurities may act as the nuclei for large defects, e.g. dislocations or stacking faults, that without help at the nucleation stage would not have formed.

This simply means that we have to keep the **Si**-crystal free of so-called **process-induced defects** during processing, something not easily achieved. Cleaning helps in this case, too.

- Below a picture of what a process-induced defect may look like. It was taken by a *transmission electron microscope* (***TEM***) and shows the projection of a systems of stacking faults (i.e. additional lattice planes bounded by dislocations) extending from the surface of the wafer into the interior. The schematic picture outlines the three-dimensional geometry

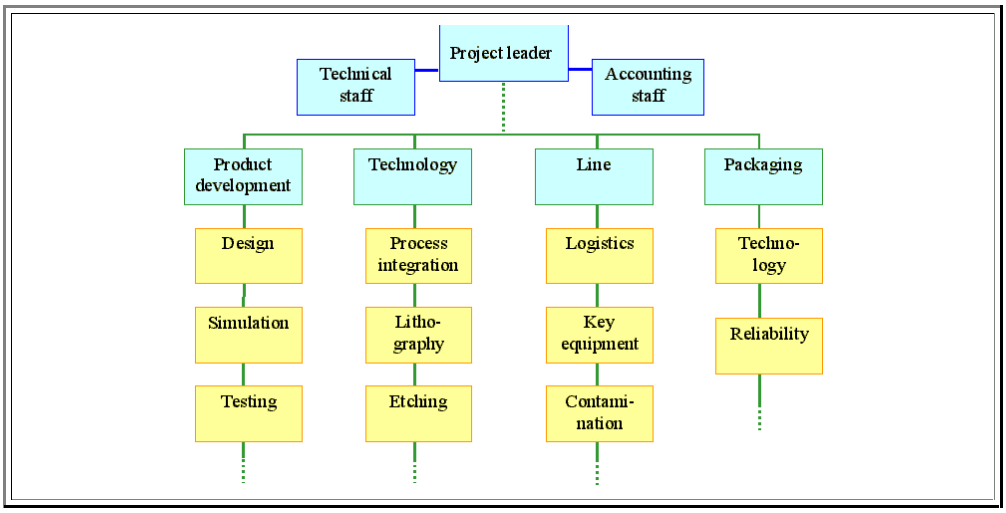


- The central ***precipitate*** that nucleated the stacking fault system is visible as a black dot. The many surplus **Si** atoms needed to form the excessive lattice planes were generated during an oxidation process.
  - Oxidation liberates **Si** interstitials which, since in supersaturation, tend to agglomerate as stacking faults.
  - However, without "help", the nucleation barrier for forming an extended defects can not be overcome, the interstitials then diffuse into the bulk of the crystal were they eventually become immobile and are harmless.
- Defects like the one above are known as "**oxidation induced stacking faults**" or ***OSF***. They form in large densities if even trace amounts of several metals are present which may form precipitates. In order to provide enough metal atoms, it is sufficient to hold the wafer just once with a metal tweezer and subject it to a high temperature process afterwards.
- There are many more ways to generate lattice defects, but there are two golden rules to avoid them:
  - *1. Keep the crystal clean!*
    Even **ppt** of **Fe**, **Ni**, **Cr** (i.e. stainless steel) **Cu**, **Au**, **Pt** or other notorious metals will, via a process that may develop through many heat cycles, eventually produce large defects and kill the device.
  - *2. Keep temperature gradients low!*
    Otherwise mechanical stress is introduced which, if exceeding the yield strength of **Si** (which decreases considerably if impurity precipitates are present), will cause plastic deformation and thus the introduction of large amounts of **dislocations**, which kill your device.
- Via the link a gallery of **process-induced defects** can be accessed together with short comments to their nature and how they were generated.
- There is a simple lecture that can be learned from this: *Electronic Materials* in the context of microelectronics comprise not only the semiconductors, but
  - Anything that can be found in the finished product - the casing (plastics, polymers. metal leads, ..), the materials on and in the **Si**, the **Si** or **GaAs** or... .
  - Anything directly used in making the chip - materials that are "sacrificial", e.g. layers deposited for a particular purpose after which they are removed again, the wet chemicals used for cleaning and etching, the gases, etc.
  - Anything used for handling the chip - the mechanisms that hold the **Si** in the apparatus or transport it, tweezers, etc.
  - Anything in contact with these media - tubing for getting gases and liquids moved, the insides of the processing equipment in contact with the liquid or gaseous media, e.g. furnace tubes.
  - Anything in possible contact with these parts - and so on! It never seems to end - make one mistake (the wrong kind of writing implement that people use in the cleanroom to make notes (not on (dusty) paper, of course, but on clean plastic sheets) - and you may end up with non-functioning chips.
- The link provides a particularly graphic example of how far this has to go!

4.3.2 Packaging and Testing

4.3.3 Working in Chip Development and Production

- Most material scientists and engineers in the **Si** semiconductor industry will be involved in chip development and production.
  - They will be part of a *large* team that also includes colleagues from electrical engineering (design, testing), computer engineering (on-chip software, functionality, testing routines) physicists and chemists and, not to forget, "money" people.
- Three major tasks can be distinguished:
  - **1.** Development of the next chip generation up to the point where the factory takes over.
  - **2.** Improving yield and throughput in the factory for the respective technology (making money!)
  - **3.** Introducing new products based on the existing technology.
- However, this three fields started to grow together in the late eighties:
  - Development of new technologies takes place in a factory because pure research and developments "**lines**" - a cleanroom with the complete infrastructure to process (and characterize) chips - are far too expensive and must produce some sellable product at least "at the side". More important, without a "base load" produced at a constant output and high quality, it is never clear if everything works at the required level of perfection!
  - Improving the yield (and cutting down the costs) is easily the most demanding job in the field. It is hard work, requires lots of experience and intimate knowledge of the chip and its processes. The experts that developed the chip therefore often are involved in this task, too.
  - There are not only new products based on the new technology that just vary the design (e.g. different memory types), but constant additions to the technology as well. Most important the "shrink" designs (making the chips smaller) that rely on input from the ongoing development of the next generation and specific processes (e.g. another metallization layer) that need development on their own.
- A large degree of interaction therefore is absolutely necessary, demanding flexibility on the part of the engineers involved.
- Lets look briefly on the structure and evolution of a big chip project; The development of the **16 Mbit DRAM** at the end of the eighties. The project structure may look like this:

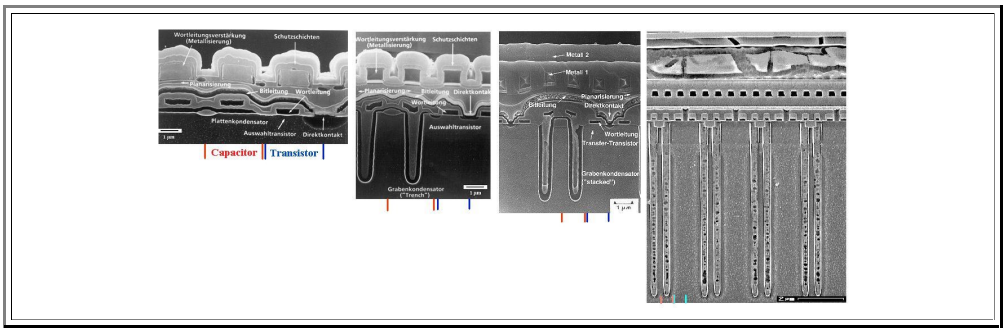


  - The number of experts working in a project like this may be **100 - 200**; they rely on an infrastructure (e.g. clean room personnel) that counts in the thousands (but these people only spend part of their time for the project).
- While there are many tasks that just need to be done on a very high level of sophistication, some tasks involve topics never done before: New technologies (e.g. trench- or stacked capacitor process modules, metallization with chemical-mechanical polishing (***CMP***, one of the key processes of the nineties), new materials (e.g. silicides in the eighties or **Cu** in the nineties), new processes (always lithography, or, e.g., plasma etching in the eighties, or electrodeposition in the nineties).
  - The problem is that nobody knows if these new ingredients will work at all (in a mass production environment) and if they will run at acceptable costs. The only way of finding out is to try it - with very high risks involved.
  - It is here were you - a top graduate in materials science of a major university - will work after a brief training period of **1 - 2** years.
- One big moment in the life of the development team is the so-called "**First Silicon**".
  - This means the first chips ever to come out of the line. Will they work - at least a little bit? Or do we have to wait for the next batch, which will be many weeks behind and possibly suffer from the same problems that prevents success with the first one?

- Waiting for first **Si** can be just as nerve racking as waiting for the answer to your job applications, your research proposal or the result of presidential elections (this was written on Nov. **17th** in **2000**, when **10** days after the election nobody knows if Bush or Gore will be the next president of the **USA**).
- In the link, the results of first Silicon for the**16 Mbit DRAM** at Siemens are shown, together with how it went on from there.

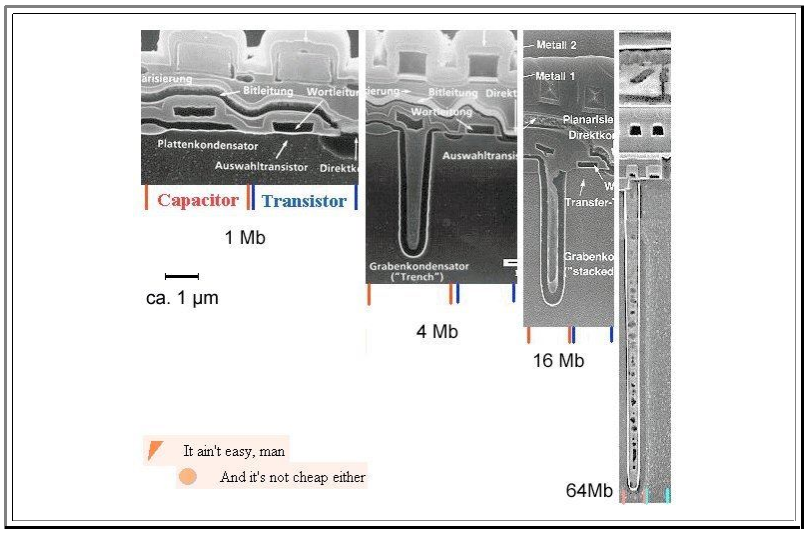
4.3.4 Generation Sequences

- It is quite instructive, if difficult to arrange, to look at several **generations** of **DRAMs** in direct comparison.
  - The picture below shows cross sections though the transistor - capacitor region necessary to store **1** bit - from the **1 Mbit DRAM** to the **64 Mbit DRAM** (all of **Siemens** design)
  - The pictures have been scaled to about the same magnification; the assembly is necessarily quite large. It starts with the **1 Mbit DRAM** on the left, followed by the **4 Mbit**, **16 Mbit** and **64 Mbit** memory.



- Decrease in **feature size** and some new key technologies are easily perceived. Most prominent are:
  - Planar capacitor ("Plattenkondensator") for the **1 Mbit DRAM**; **LOCOS** isolation and **1** level of metal ("Wortleitungsverstärkung") parallel to the **poly-Si** "Bitleitung" (bitline) and at right angles to the **poly-Si/Mo-silicide** "Wortleitung" (wordline).
  - Trench capacitor for the **4 Mbit DRAM**, "FOBIC" contact, and **TiN** diffusion barrier.
  - Two metal levels for the **16 Mbit DRAM**, *poly-ONO-poly* in trench; improved planarization between bitline - metal **1**, and metal **1** - metal **2**.
  - Box isolation instead of **LOCOS** for the **64 Mbit DRAM**, very deep trenches, **W**-plugs, and especially complete planarization with **chemical mechanical polishing** (***CMP***), the key process of supreme importance for the **64 Mbit** generation and beyond.
- If some of the technical expressions ***eluded*** you - don`t worry, be happy! We will get to them quickly enough.
- Parallel to a reduction in feature size is always an increase in **chip size**; this is illustrated the link.
  - You may ask yourself: Why do we not just make the chip bigger - instead of **200 4 Mbit DRAMs** on a wafer we just as well produce **50 16 Mbit Drams**?
  - Well. Le's say you have a very high yield of **75 %** in your **4 Mbit** production. This gives you **150** good chips out of your **200** - but it would give you a yield close to zero if you now make **16 Mbit DRAMs** with that technology.
  - What's more: Even if you solve the yield problem: Your **16 Mbit** chips would be exactly **4** times more expensive then your **4 Mbit** chip - after all your costs have not changed and you now produce only a quarter of what you had before. Your customer would have no reason to buy these chips, because they are not only not cheaper per bit, but also not faster or less energy consuming.
  - Progress performance only can come from reducing the feature size.
- The cost per bit problem you also can address to some extent by using larger wafers, making more chips per process run.
  - This has been done and is being done: Wafer sizes increased from **< 2** inch in the beginning of the seventies to **300 mm** now (**2002**) - we also went metric on the way.

4.3.5 Summary to: 4.3 Infrastructure



Exercise 4.3-1

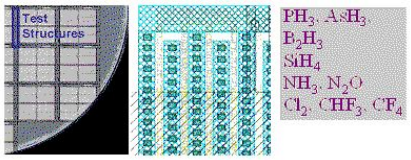
All Questions to 4.3

- At the core of lithography are the **steppers** - optical machines for around **5 Mio €** a piece
- Resist technology, too, is a highly developed part of lithography
- For some big problems simple solutions have been found. Example: reticles with pellicles

4.4 Summary

4.4.1 Summary to: 4. Getting Started

- Semiconductor technology happens in factories. They need special materials, "reticles" (= structures), "know-how" and huge amoundt of money (= capital) as major inputs
  - It's always about money! Only mass production will recover large investments.
  - The materials side always contains semiconductor substrates ("wafers") and often very dangerous special "raw" materials.
  - A number tells it all: **500 - 1.000** wafers /day are processed in a large **Si** "wafer fab"



- Three big steps to **Si** wafers
  - **Si** single crystal growth is done by "Czochralski process" (**CZ**).
  - Dislocation-free crystals are possible but "bulk microdefects" and impurities cannot be totally avoided.
  - Nearly perfect **300 mm** wafers are standard.

Sand (SiO_2)	⇒	**Metallurgical Si**
Metallurgical Si	⇒	**clean (doped) poly-Si.**
Poly-Si	⇒	**Single crystal / wafer**

- Growing single crystals of compound semiconductors is far more difficult than for elemental semiconductors
  - Precise stoichiometry is important
  - Vapor pressures if the constituents at the melting point might be very different
  - New kinds of defects might be encountered
  - Polytypie might be encountered

GaAs:
150 mm wafers, encapsulation technique, disl. density (10^3 - 10^6) cm^{-2}

GaP, **InP**
as **GaAs** but smaller and more expensive

SiC:
100 mm wafers, sublimation technique, several polytypes available, "pipe" defects

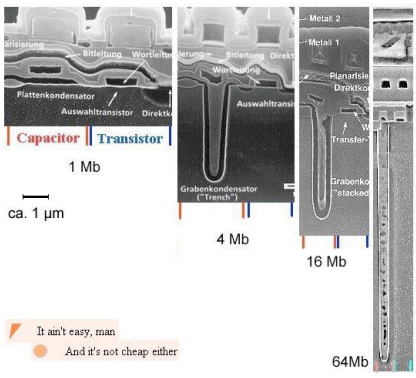
- Major techniques are
  - Encapsulated **CZ**
  - Sublimation growth

Exercise 4.4-1

All Questions to 4.

Questionnaire

Multiple Choice questions to 4.



5. Integrated Circuits - Process Integration

5.1 Basic Considerations for Process Integration

5.1.1 What is Integration?

5.1.2 Basic Concepts of Integrating Bipolar Transistors

5.1.3 Basic Concepts of Connecting Transistors

5.1.4 Integrated MOS Transistors

5.1.5 Integrated CMOS Technology

5.1.6 Summary to: 5.1 Basic Considerations for Process Integration

5.2 Process Integration

5.2.1 Chips on Wafers

5.2.2 Packaging and Testing of Silicon Chips

5.2.3 Summary to: Chips on Wafers

5.3 Money, Markets, and Management

5.3.1 Moore's Law and what it Means

5.4 Summary

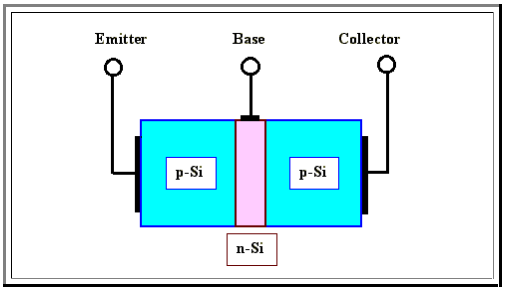
5.4.1 Summary to 5: Integrated Circuits - Process Integration

5. Integrated Circuits - Process Integration

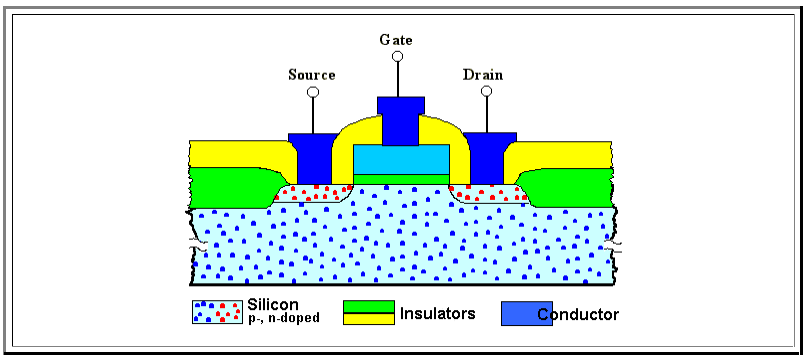
5.1 Basic Considerations for Process Integration

5.1.1 What is Integration?

- The key element of electric engineering, computer engineering, or pretty much everything else that is remotely "technical" in the last thirty years of the **2nd** millennium, is the **integrated transistor** in a **Silicon crystal** - everything else comes in second - at best.
  - *Integrated* means that there is more than one transistor on the same piece of **Si** crystal and thus in the same package. And "*more than one*" means at the present stage of technology some 10^7 *transistors per* cm^2 of Silicon.
  - *Silicon crystal* means that we use huge, extremely perfect single crystals of **Si** to do the job. Why **Si** and not. for example**Ge**, **GaAs** or **SiC**? Because if you look at the sum total of the most important properties you are asking for (crystal size and perfection, bandgap, extremely good and process compatible dielectric, ...) **Si** and its oxide, SiO_2 are so vastly superior to any possible contender that there is simply no other semiconductor that could be used for complex integrated circuitry.
- The lowly **integrated circuit** (***IC***), mostly selling for a few Dollars, is the most marvelous achievement of Materials Science in the second half of the **20th** century. Few people have an idea of the tremendous amount of science and engineering that was (and still is) needed to produce a state-of-the art **chip**, the little piece of **Si** crystal with some other materials in precise arrangements, that already starts to rival the complexity of the brains of lower animals and might at some day in the not so distant future even rival ours.
- If we want to make a *circuit* out of many transistors (some of which we use as resistors) and maybe some capacitors, we need *three basic ingredients* - no matter if we do this in an integrated fashion or by soldering the components together - and on occasion some "spices", some special additions:
- **1. Ingredient:** *Transistors.* "Big" and "small" ones (with respect to the current they can switch), for low or high voltage, fast or not so fast - the whole lot. We have two basic types to chose from:
  - **Bipolar transistors** (the hopefully familiar **pnp**- or **npn**-structures) *are usually drawn as follows*



  - Note right here that no real transistor looks even remotely like this structure! It's only and purely a schematic drawing to show essentials and and has nothing whatsoever to do with a real transistor.
  - The name *bipolar* comes from the fact that *two* kinds of carriers, the negatively charged electrons and the positively charged holes, are necessary for its function.
  - ***MOS*** **Transistors** or **unipolar Transistors**, more or less only exist in integrated form and usually are drawn as follows:



- **2. Ingredient:** *Insulation.* Always needed between transistors and the other electrically active parts of the circuit.
  - In contrast to circuits soldered together where you simply use air for insulation, it does not come "for free" in **ICs** but has to be made in an increasingly complex way.
- **3. Ingredient:** *Interconnections* between the various transistors or other electronic elements - the wires in the discrete circuit.

- The way you connect the transistors will determine the function of the device. With a large bunch of transistors you can make everything - a microprocessor, a memory, anything - only the interconnections must change!

Then we may have *special elements*

- These might be capacitors, resistors or diodes on your chip. Technically, those elements are more or less subgroups of transistors (i.e. if you can make a transistor, you can also make these (simpler) elements), so we will not consider them by themselves.

If you remember your introduction into Materials Science , you should be familiar with the basic physics of the two transistor types; *otherwise read it up right now!!!* Here are the links to the German Hyperscript

- **Bipolare Transistoren**
- **MOS Transistoren**

If you prefer the English (and slightly advanced) version, use these links

- **Basic bipolar Transistors**
- **Basic MOS Transistors**

The list of necessary ingredients given above automatically implies that we have to use several different materials. At the very minimum we need a *semiconductor* (which is practically always Silicon; only **GaAs** has a tiny share of the **IC** market, too), an *insulator* and a *conductor*. As we will see, we need many more materials than just those three basic types, because one kind of material cannot meet all the requirements emerging from advanced **Si** technology.

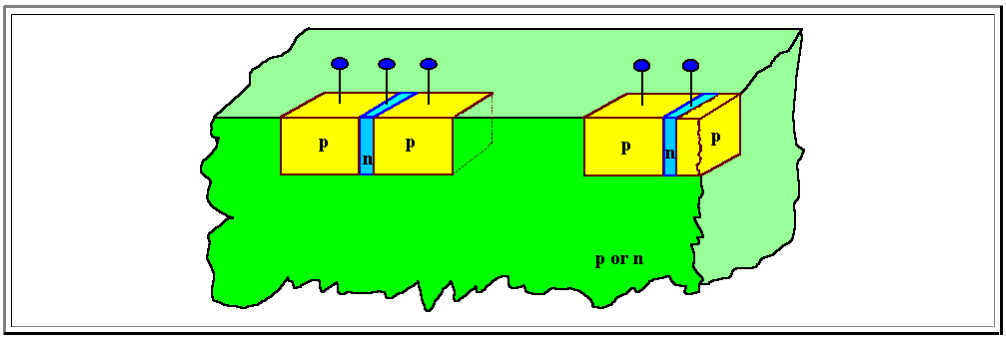
- Since this lecture course is about *electronic materials*, it may appear that all we need now is a kind of list of suitable materials for making integrated circuits. But that would be far too short sighted. In **IC** technology, *materials and processes* must be seen as a unit - one cannot exist without the other.
- We therefore have to look at both, materials with their specific properties and their integration into a **process flow**.

Todays integrated circuits contain mostly **MOS** transistors, but we will start with considering the integration of bipolar transistors first. That is not only because historically bipolar transistors were the first ones to be integrated, but because the basic concepts are easier to understand.

5.1.2 Basic Concepts of Integrating Bipolar Transistors

How *Not* to Make an Integrated Bipolar Transistor

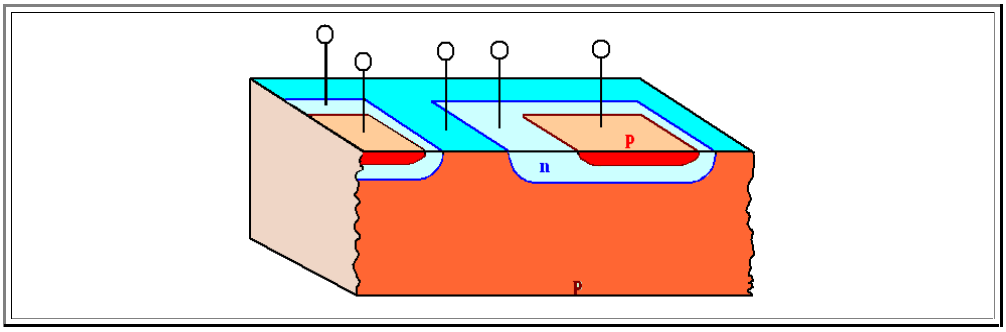
Obviously, embedding the three slices of Si that form a bipolar transistor into a **Si** crystal will not do you any good - we just look at it here to see just how ludicrous this idea would be:



What is the problem with this approach? Many points

- The transistors would not be insulated. The **Si** substrate with a certain kind of doping (either **n**- or **p**-type) would simple short-circuit all transistor parts with the same kind of doping.
- There is not enough place to "put a wire down", i.e. attach the **leads**. After all, the base width should be very small, far less than **1 µm** if possible. How do you attach a wire to that?
- How would you put the sequence of **npn** or **pnp** in a piece of **Si** crystal? After all, you have to get the right amount of , e.g. **B**- and **P**-atoms at the right places.

So we have to work with the really small dimensions in ***z***-direction, into the **Si**. How about the following approach?



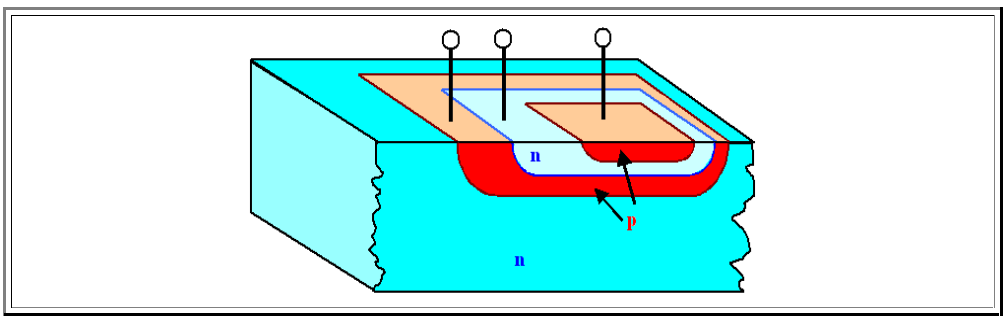
This is much better, but still not too convincing. The *pro* arguments are:

- Enough space for *leads*, because the lateral dimensions can be as large as you want them to be.
- It is relatively easy to produce the doping: Start with **p**-type **Si**, diffuse some **P** into the **Si** where you want the Base to be. As soon as you overcompensate the **B**, you will get **n**-type behavior. For making the emitter, diffuse lots of **B** into the crystal and you will convert it back to **p**-type.
- The base width can be very small (we see about this later).

But there is a major shortcoming:

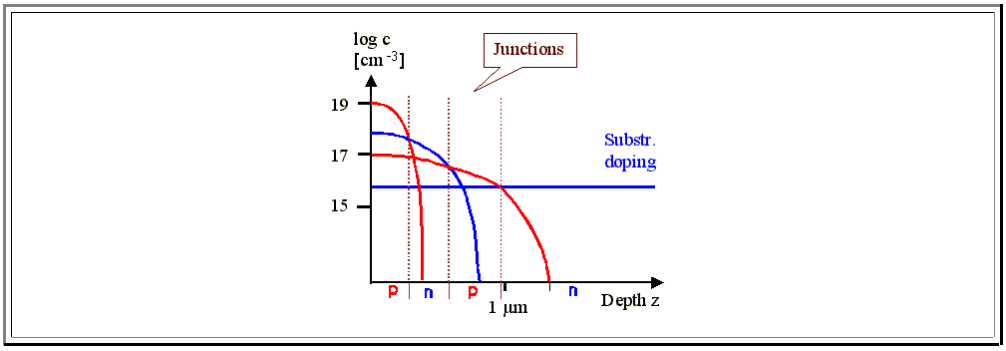
- Still *no insulation* between the collectors - in fact the **Si** crystal is the collector of all transistors and that is not going to be very useful.

Easy you say, lets add another layer of doped **Si**:



This would be fine in terms of insulation, because now there is always a **pn**-junction between two terminals of different transistors which is always blocked in one direction for all possible polarities.

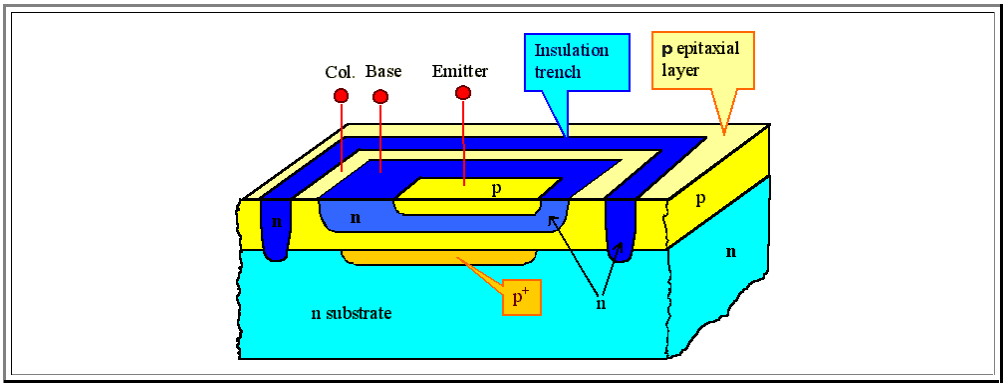
- However, you now have to change an **n**-doped substrate to **p**-doping by over-compensating with, e.g. **B**, then back to **n**-type again, and once more back to **p**-type. Lets see, how that would look in a **diffusion profile** diagram:



- The **lg** of the concentration of some doping element as shown in the illustration above is roughly what you must have - except that the depth scale in modern **ICs** would be somewhat smaller.
  - It is obvious that it will be rather difficult to produce junctions with precisely determined depths. Control of the base width will not be easy.
  - In addition, it will not be easy to achieve the required doping by over-compensating the doping already present three times. As you can see from the diagram, your only way in the resistivity is *down*. If the substrate, e.g., has a doping of **10** Ω**cm**, the collector can only have a lower resistivity because the doping concentration must be larger than that of the substrate, so lets have **5** Ω**cm**. That brings the base to perhaps **1** Ω**cm** and the emitter to **0,1** Ω**cm**. These are reasonable values, but your freedom in designing transistors is severely limited
  - And don't forget: It is the relation between the doping level of the emitter and the base that determines the amplification factor γ
- There must be *smarter* way to produce integrated bipolar transistors. There is, of course, but this little exercise served to make clear that integration is far from obvious and far from being easy. It needs *new ideas*, *new processes*, and *new materials* - and that has not changed from the first generation of integrated circuits with a few **100** transistors to the present state of the art with some **100** million transistors on one chip.
  - And don't be deceived by the *low costs* of integrated circuits: Behind each new generation stands a huge effort of the "best and the brightest" - large scale integration is still the most ambitious technical undertaking of mankind today.

How to Make an Integrated Bipolar Transistor

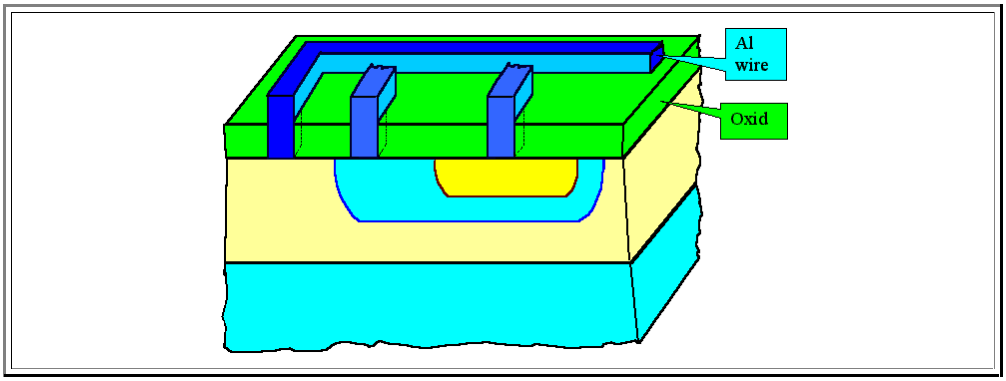
- So how is it done? *By inventing special processes*, first of all: **Epitaxy**, i.e. the deposition of thin layers of some material on a substrate of (usually, but not necessarily) the same kind, so that the lattice continues undisturbed.
  - Lets look at a cross-section and see what *epitaxy* does and why it makes the production of **IC**s easier.



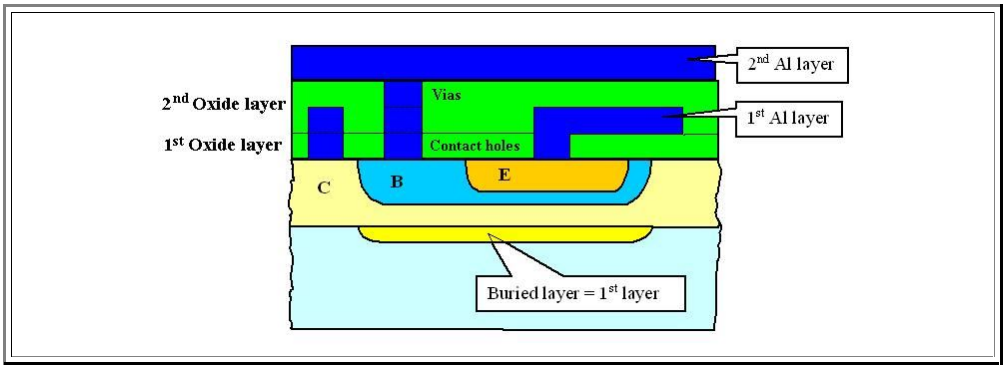
- We start with an **n**-doped wafer (of course you can start with a **p**-doped wafer, too; than everything is reversed) and diffuse the $\mathbf{p^+}$ layer into it. We will see what this is good for right away.
  - On top of this wafer we put an *epitaxial layer of **p**-doped Silicon*, an *epi-layer* as it is called for short. Epitaxial means that the crystal is just continued without change in orientation. The epitaxial layer will always be the collector of the transistor.
  - Next, we diffuse a closed ring of **n**-material around the area which defines the transistor deeply into the **Si**. It will insulate the transistors towards its neighbours because no matter what voltage polarity is used between the collectors of neighbouring transistors, one of the two **pn**-junctions is always in reverse; only a very small leakage current will flow.
  - Then we diffuse the **n**-base- and **p**-emitter region in the epi-layer.
- Looks complicated because it is complicated. But there are many advantages to this approach:
  - We only have *two "critical" diffusions*, where the precise doping concentration matters.
  - The transistor is in the epitaxial layer which, especially in the stone age of integration technology (about from **1970 - 1980**) had a much better quality in terms of crystal defects, level and homogeneity of doping, minority carrier lifetime τ, ...) than the **Si** substrate.
  - We get one level of wiring *for almost free*, the $\mathbf{p^+}$ layer below the transistor which can extend to somewhere else, contacting the collector of another transistor!
- This leads us to the next big problem in integration: The "wiring", or how do we connect transistors in the right way?

5.1.3 Basic Concepts of Connecting Transistors

- How do we connect a few million transistors - i.e. run signal wires from transistor **x** to transistor **y** (**x** and **y** being arbitrary integers between **1** and about **50 000 000**?) and connect *all* transistors to some voltage and current supply - and all that *without wires crossing*?
  - For state-of-the-art **ICs** this is one of the bigger challenges. Obviously you must have wiring on several planes because you cannot avoid that connections must cross each other.
  - The first level is simple enough - in principle! Lets see this in a schematic drawing.



- So all you do is to cover everything with an insulator. For that you are going to use **SiO_2**, which is not only one of the best insulators there is, but is easily produced and fully compatible with **Si**.
  - On top of this oxide you now run your "wires" from here to there, and wherever you want to make a contact to a transistor, you make a **contact hole** in the **SiO_2** layer.
  - Every transistor needs three contact holes - and as you can see in the drawing, you rather quickly run into the problem of crossing connections.
- What we need is a **multi-level metallization**, and how to do this is one of the bigger challenges in integration technology.
  - Fortunately, we already have a second level in the **Si** - it is the "**buried layer**" that we put down before adding the epitaxial layer. It can be structured to connect the collectors of all transistors where this makes sense. And since the collectors are often simply connected to the power supply, this makes sense for most of the transistors.
  - But this is not good enough. We still need more metallization layers on top. So we repeat the "putting oxide down, making contact holes, ..etc". procedure and produce an second metallization layer:



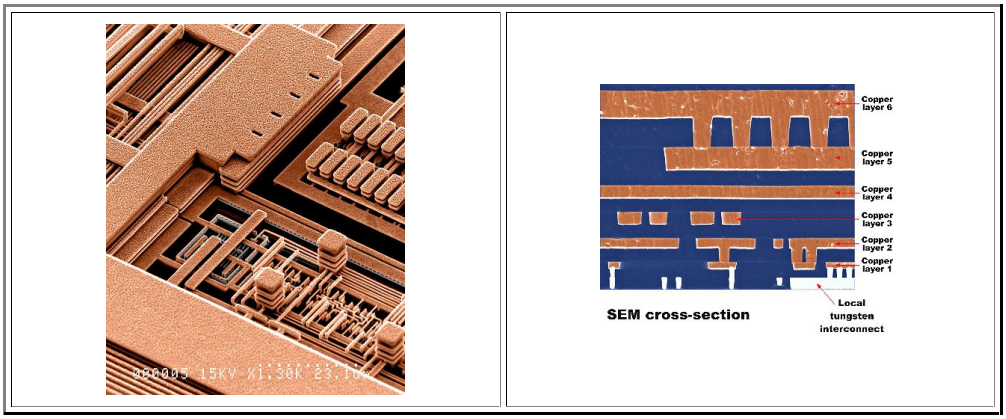
- If you get the idea that this is becoming a trifle complicated, you get the right idea. And you haven't seen anything yet!
  - State-of-the-art **ICs** may contain **7** or more connection (or metallization) layers. For tricky reasons explained later, besides Aluminium (**Al**), Tungsten (**W**) is employed, too, and lately **Al** is being replaced by Copper (**Cu**).
  - Between the metal layers we obviously need an "**intermetal dielectric**". We could (and do) use **SiO_2**; but for modern chips we would rather use something better. In particular, a material with a smaller dielectric constant (**SiO_2** has a value of about **3.7**). Polymers would be fine, in particular polyimides, a polymer class that can "take the heat", i.e. survives at relatively high temperatures. Why we do not have polyimides in use just now is an interesting story that can serve as a prime example of what it means to introduce a new material into an existing product.
- Why are we doing this - replacing trusty old **Al** by tricky new **Cu** - at considerable costs running in the billion **$** range?
  - Because the total resistance ***R*** of an **Al** line is determined by the specific resistivity **ρ = 2,7 $\mu\Omega$cm** of **Al** and the geometry of the line. Since the dimensions are always as small as you can make it, you are stuck with **ρ**.

- Between neighbouring lines, you have a parasitic capacitance ***C***, which again is determined by the geometry and the dielectric constant ϵ of the insulator between the lines. Together, a **time constant** $R \cdot C$ results, which is directly proportional to $\rho \cdot \epsilon$. This time constant of the wiring - found to be in the **ps** region - gives an absolute upper limit for signal propagation. If you don't see the probem right away, turn to this basic module.
- In other words: Signal delay in **Al** metallization layers insulated by **SiO_2** restricts the operating frequency of an **IC** to about **1 GHz** or so.

This was no problem before **1998** or so, because the transistors were far slower anyway. But it is a problem *now* (**2000 +**)!

- Obviously, we must use materials with lower ρ and ϵ values. Choices are limited, however - **Cu (ρ = 1,7 μΩcm)** is one option that has been chosen; the last word about a suitable replacement for **SiO_2** (having ϵ **= 3,7**) is not yet in.

Here are famous pictures of an advanced **IBM** chip with **7** metallization layers, completely done in **W** and **Cu**. In the picture on the left, the dielectric between the metals has been etched off, so only the metal layers remain.

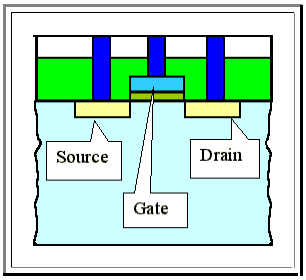


The transistors are not visible at this magnification - they are too small. You would find them right below the small "local tungsten interconnects" in the cross sectional view.

Before we go into how one actually does the processes mentioned (putting down layers, making little contact holes, ...), we have to look at how you make *MOS transistors* as opposed to *bipolar* transistors. We will do that in the next sub-chapter.

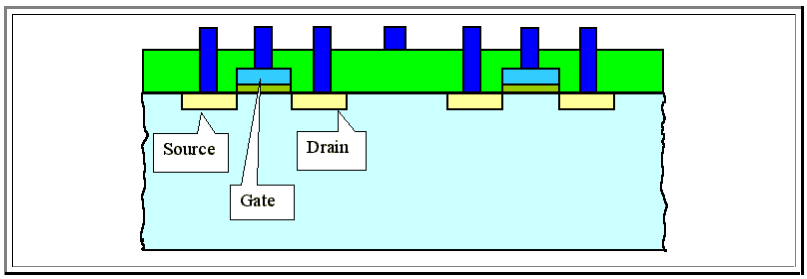
5.1.4 Integrated MOS Transistors

MOS transistors are quite different from bipolar transistors - not only in their basic function, but also in the way they are integrated into a **Si** substrate. Lets first look at the basic structure

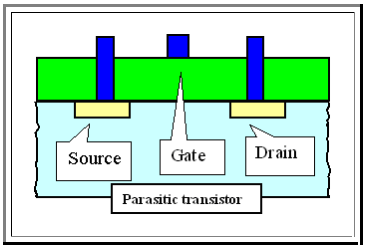


- We have a **source** and **drain** region in the **Si** (doped differently with respect to the substrate) with some connections to the outside world symbolically shown with the blue rectangles. Between source and drain is a thin **gate dielectric** - often called **gate oxide** - on top of which we have the **gate electrode** made from some conducting material that is also connected to the outside world.
- To give you some idea of the real numbers: The thickness of the gate dielectric is below **10 nm**, the lateral dimensioon of the source, gate and drain region is well below **1 µm**.
- You know, of course, what a **MOS** transistor is and how it works - at least in principle. If not: Use the link **Basic MOS transistor**.

If we integrate **MOS** transistors now, it first appears (wrongly!) that we can put them into the same **Si** substrate as shown below:



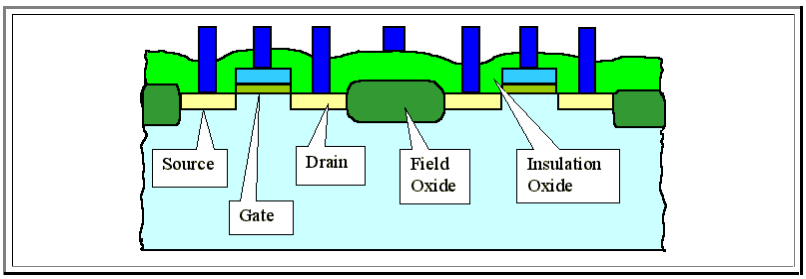
There seems to be no problem. The transistors are insulated from each other because one of the **pn**-junctions between them will always be blocking. *However:* We must also consider "**parasitic transistors**" not intentionally included in our design!



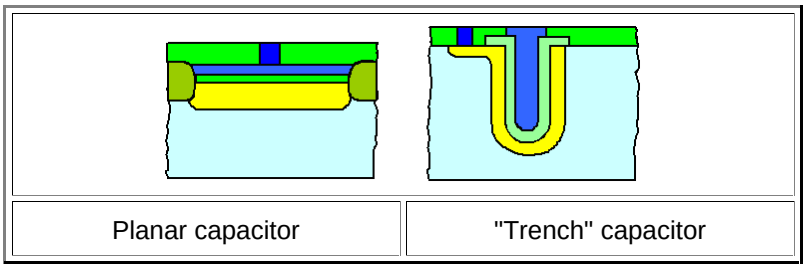
- If in the space between transistors a wire is crossing on top of the insulating layer as shown in the illustration, it will, on occasion be at high potential. The drain of the left transistor together with the source of the right transistor will now form a *parasitic transistor* with the insulating layer as the gate dielectric, and the overhead wire as the gate electrode.
- Everything being small, the threshold voltage may be reached and we have a current path where there should be none.
- This is *not* an academic problem, but a typical effect in *integrated* circuit technology, which is not found in *discrete* circuits: Besides the element you want to make, you may produce all kinds of unwanted elements, too: parasitic transistors, capacitors, diodes, and even thyristors.

The solution is to make the threshold voltage larger than any voltage that may occur in the system. The way to do this is to increase the *local thickness* of the insulating dielectric.

- This gives us the structure in the next illustration



- How we produce the additional insulator called **field oxide** between the transistors will concern us later; here it serves to illustrate two points:
  - Insulation is not just tricky for bipolar transistors, it is a complicated business with **MOS** technology, too
  - There is now some "topology" - *the interfaces and surfaces are no longer flat*. Looks trivial, but constitutes one of the major problems of large-scale integration!
- Note that the gate - substrate part of a **MOS** transistor is, in principle, a **capacitor**. So we can now make capacitors, too.
  - However, if we need a *large* capacitance - say some **50 fF** (femto Farad) - we need a *large* area (several **μm^2**) because we cannot make the dielectric arbitrarily thin - we would encounter tunneling effects, early breakdown, or other problems. So we have to have at least a thickness of around **5 nm** of **SiO_2**. If the capacitor area than gets too large, the escape road is the *third dimension*: You fold the capacitor! Either into the substrate, or up into the layers on top of the **Si**.
  - The "simple" way of folding integrated capacitors into the substrate is shown in the right hand side of the next illustration



Planar capacitor	"Trench" capacitor

- The planar capacitor (on the left) and the "**trench**" capacitor (on the right) have a doped region in the **Si** for the second electrode, which must have its own connection - in the drawing it is only shown for the trench capacitor. We learn two things from that:
  - **1.** Large scale integration has long since become three-dimensional - it is no longer a "planar technology" as it was called for some time. This is not only true for truly three-dimensional elements like the trench capacitor, but also because the processes tend to make the interfaces rough as we have seen already in the case of the field oxide.
  - **2.** The names for certain features generally accepted in the field, are on occasion simply **wrong**! The capacitor shown above is not folded into a *trench* (which is something deep and long in one lateral direction, and small in the other direction), but into a *hole* (deep and small in both lateral directions). Still, everybody calls it a **trench capacitor**.
- The key processes for **ICs** more complex than, say, a **64 Mbit** memory, are indeed the processes that make the surface of a chip halfway flat again after some process has been carried out.
- Again, there is a special message in this subchapter: Integrating **MOS** transistors, although supposedly simpler than bipolar transistors (you don't need all those **pn**-junctions), is far from being simple or obvious. It is again ***intricately*** linked to specific combinations of materials and processes and needs lots of ingenuity, too.
  - But we are still not done in trying to just get a very coarse overview of what integration means. If you take an arbitrary chip of a recent electronic product, changes are that you are looking at a ***CMOS*** chip, a chip made with the "*Complementary Metal Oxide Semiconductor*" technology.
  - So lets see what that implies.

5.1.5 Integrated CMOS Technology

Power Consumption Problem

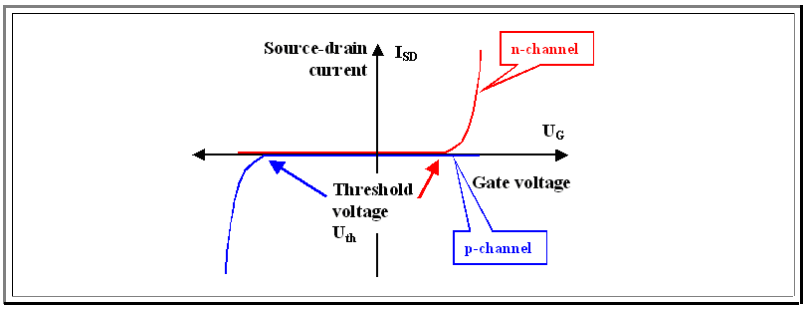
- The first integrated circuits hitting the markets in the seventies had a few **100** transistors integrated in bipolar technology. **MOS** circuits came several years later, even though their principle was known and they would have been easier to make.
  - However, there were insurmountable problems with the *stability* of the transistor, i.e. their threshold voltage. It changed during operation, and this was due to problems with the gate dielectric (it contained minute amounts of alkali elements which are some of many "**IC** killers", as we learned the hard way in the meantime).
  - But **MOS** technology eventually made it, mainly because bipolar circuits need a lot of power for operation. Even for all transistors being "off", the sum of the leakage current in bipolar transistors can be too large for many application.
- **MOS** is principally better in that respect, because you could, in principle, live with only switching voltages; current per se is not needed for the operation. **MOS** circuits do have lower power consumption; but they are also slower than their bipolar colleagues. Still, as integration density increased by an average **60%** per year, power consumption again became a problem.
  - If you look at the data sheet for some state of the art **IC**, you will encounter power dissipations values of up to **1 - 2 Watts** (before **2000**)! Now (**2004**) its about **10** times more. If this doesn't look like a lot, think again!
  - A chip has an area of roughly **1 cm^2**. A power dissipation of **1 Watt/cm^2** is a typical value for the hot plates of an electrical range! The only difference is that we usually do not want to produce french fries with a chip, but keep it cool, i.e. below about **80 ^{o}C**.
- So power consumption is a big issue in chip design. And present day chips would not exist if the ***CMOS*** technique would not have been implemented around the late eighties. Let's look at some figures for some more famous chips:
  - Early Intel microprocessors had the following power rating:

Type	Architecture	Year	No. transistors	Type	Power
4004	4bit	1971	2300	PMOS	
8086	16bit	1978	29000	NMOS	1,5W/8MHz
80C86	16bit	1980	?50000?	**CMOS**	250mW/30Mhz(?)
80386	16bit	1985	275000	**CMOS**	
Pentium 4		2004		**CMOS**	80 W/3 GHz

  - **CMOS** seems to carry the day - so what is **CMOS** technology?

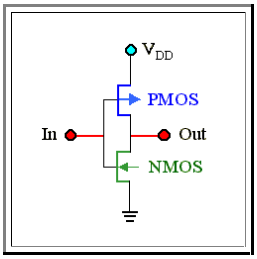
CMOS - the Solution

- Lets first see what "**NMOS**" and ***PMOS***" means. The first letter simply refers to the kind of carrier that carries current flow between source an drain as soon as the threshold voltage is surpassed:
  - *PMOS* stands for transistors where *positively* charged carriers flow, i.e. *holes*. This implies that source and gate must be **p**-doped areas in an **n**-doped substrate because current flow begins as soon as inversion sets in, i.e. the **n**-type **Si** between source and drain is inverted to **Si** with holes as the majority carriers
  - *NMOS* then stands for transistors where negatively charged carriers flow. i.e. electrons. We have **n**-doped source and drain regions in a **p**-doped substrate.
- The characteristics, i.e. the source-drain-current vs. the gate voltage, are roughly symmetrical with respect to the sign of the voltage:



- The red curve may stand for a **NMOS** or **n**-channel transistor, the blue one then would be the symmetrical **PMOS** or **p**-channel transistors. The threshold voltages are not fully symmetric if the same gate electrode is used because it depends on the difference of the Fermi energies of the gate electrode materials and the doped **Si**, which is different in the two cases.
- Anyway, for a given gate voltage which is larger than either threshold voltage applied to the transistor, one transistor would be surely "on", the other one "off".

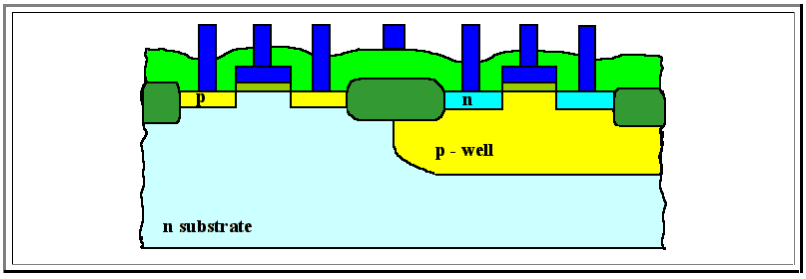
So if you always have a **NMOS** and a **PMOS** transistor in series, there will *never* be any static current flow; we have a small dynamic current component only while switching takes place.



- Can you make the necessary logical circuits this way?
- Yes you can - at least to a large extent. The illustration shows an **inverter** - and with inverters you can create almost anything!
- Depending on the right polarities, the blue **PMOS** transistor will be closed if there is a gate voltage - the output then is zero. For gate voltage zero, the green **NMOS** transistor will be closed, the **PMOS** transistor is open - the output will be $\mathbf{V_{DD}}$ (the universal abbreviation for the *supply voltage*).

So now we have to make *two* kinds of transistors- **NMOS** and **PMOS** - which needs substrates with different kind of doping - in *one* integrated circuit. But such substrates do not exist; a Silicon wafer, being cut out of an homogeneous crystal, has always *one* doping kind and level.

- How do we produce differently doped areas in an uniform substrate? We remember what we did in the bipolar case and "simply" add another diffusion that converts part of the substrate into the different doping kind. We will have to diffuse the right amount of the compensating atom rather deep into the wafer, the resulting structure is called a **p**- or **n**-**well**, depending on what kind of doping you get.
- If we have a **p**-type substrate, we will have to make a **n**-well. The **n**-well then will contain the **PMOS** transistors, the original substrate the **NMOS** transistors. The whole thing looks something like this:



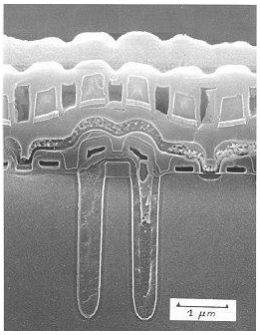
By now, even the "simple" **MOS** technology starts to look complicated. But it will get even more complicated as soon as you try to put a metallization on top. The gate structure already produced some "roughness", and this roughness will increase as you pile other layers on top.

- Let's look at some specific metallization problems (they are also occurring in bipolar technology, but since you start with a more even surface, it is somewhat easier to make connections).
- A cross-section through an early ***16 Mbit DRAM*** (**DRAM=***Dynamic Random Access Memory*; the work horse memory in your computer) from around **1991** shown below illustrates the problem: The surface becomes exceedingly wavy. (For enlarged views and some explanation of what you see, click on the image or the link)

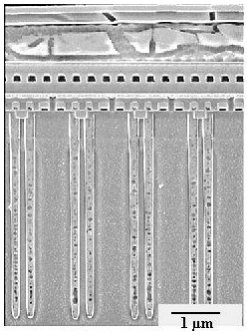
Adding more metallization layers becomes nearly impossible. Some examples of the difficulties encountered are:

- **1.** With wavy interfaces, the thickness between two layers varies considerably, and, since making connection between layers need so-called "**via**" holes, the depths of those *vias* must vary, too. This is not easily done! And if you make all vias the same (maximum) depth, you will etch deeply into the lower layer at places where the interlayer distances happens to be small.
- **2.** It is very difficult to deposit a layer of anything *with constant thickness* on a wavy surface.
- **3.** It is exceedingly difficult to fill in the space between **Al** lines with some dielectric without generating even more waviness. The problem then gets worse with an increasing number of metallization layers.

The **64 Mbit DRAM**, in contrast, is very flat. A big break-through in wafer processing around **1990** called "**Chemical mechanical Polishing**" or ***CMP*** allowed to planarize wavy surfaces.



Cross section 16 Mbit DRAM (Siemens)



Cross section 64 Mbit DRAM (Siemens)

It's time for an exercise:

Exercise 5.1-1
Integrating transistors

State of the Art

Lets get some idea about the state of the art in (**CMOS**) chip making in the beginning of the year **2000**. Above you can look at cross-sectional pictures of a **16 Mbit** and a **64 Mbit** memory; the cheap chip and the present work horse in memory chips. The following data which come from my own experience are not extremely precise but give a good impression of what you can buy for a few Dollars.

Property	**Number**
Feature size	**0,2 µm**
No. metallization levels	**4 - 7**
No. components	**> 6 · 10^8 (Memory)**
Power	**several W/cm^2**
Speed	**600 MHz**
Lifetime	**> 10 a**
Price	**$2 (memory) up to $ 300 (microprocessor)**
Complexity	**> 500 Process steps**
Cost (development and **1** factory)	**ca. $ 6 · 10^9**

How will it go on? Who knows - but there is always the official semiconductor roadmap from the **Semiconductor Industry Associaton** (***SIA***)

- That's it. Those are holy numbers which must not be doubted. Since they are from **1993**, the predictive power can be checked.

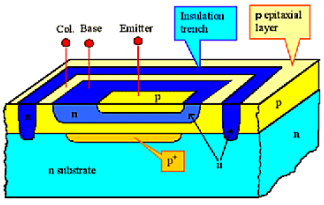
Semiconductor Industry Association Roadmap (1993)							
		1992	1995	1998	2001	2004	2007
Feature size (µm)		0.5	0.35	0.25	0.18	0.12	0.1
Bits/Chip	DRAM	16M	64M	256M	1G	4G	16G
	SRAM	4M	16M	64M	256M	1G	4G
Chip size (mm^2)	Logic / microprocessor	250	400	600	800	1000	1250
	DRAM	132	200	320	500	700	1000
Performance (MHz)	on chip	120	200	350	500	700	1000
	off chip	60	100	175	250	350	500
Maximum power (W/chip)	high performance	10	15	30	40	40-120	40-200
	portable	3	4	4	4	4	4
Power supply voltage (V)	desktop	5	3.3	2.2	2.2	1.5	1.5
	portable	3.3	2.2	2.2	1.5	1.5	1.5
No. of interconnect levels - logic		3	4-5	5	5-6	6	6-7
Number of I/Os		500	750	1500	2000	3500	5000
Wafer processing cost ($/cm^2)		$4.00	$3.90	$3.80	$3.70	$3.60	$3.50
Wafer diameter (mm)		200	200	200-400	200-400	200-400	200-400
Defect density (defects/cm^2)		0.1	0.05	0.03	0.01	0.004	0.002

5.1.6 Summary to: 5.1 Basic Considerations for Process Integration

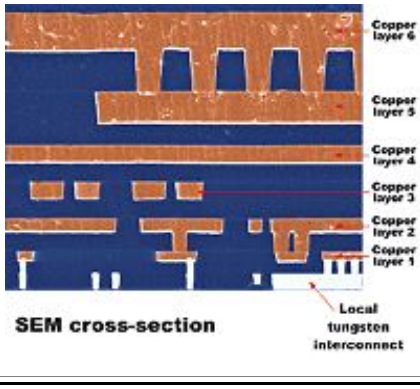
- Integration means:
  - **1.** Produce a large number (up to **1.000.000.000**) of transistors (bipolar or **MOS**) and other electronic elements on a **cm^2** of **Si**
  - **2.** Keep thoses elements electrically insulated from each other.
  - **3.** Connect those elements in a meaningful way to produce a system / product.

It ain't easy!

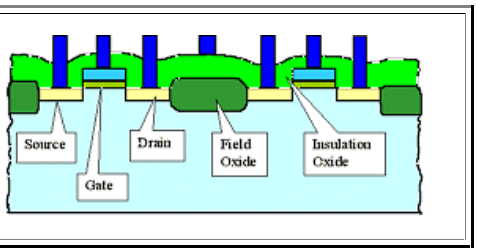
- An integrated bipolar transistor does not resemble the textbook picture at all, but looks far more complicated ⇒.
  - This is due to the insulation requirements, the process requirements, and the need to interconnect as efficiently as possible.
  - The epitaxial layer cuts down on the number of critical diffusions, makes insulation easier, and allows a "buried contact" structure.



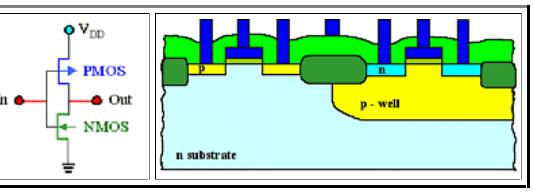
- Connecting transistor / elements is complicated; it has to be done on several levels
  - Materials used are **Al** ("old"), **Cu** ("new"), **W**, (highly doped) poly-**Si** as well as various silicides.
  - Essential properties are the conductivity σ of the conductor, the dielectric constant ϵ_r of the intermetal dielectric, and the resulting time constant $\tau = \sigma \cdot \epsilon_r$ that defines the maximum signal transmision frequency through the conducting line.



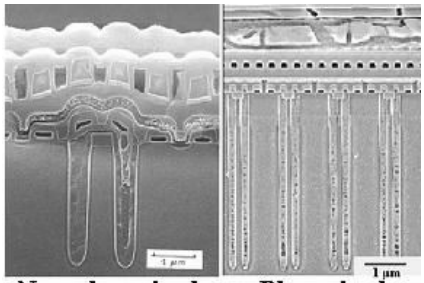
- Integrating **MOS** transistors requires special measures for insulation (e.g. a field oxide) and for gate oxide production
  - Since a **MOS** transistor contains intrinsically a capacitor (the gate "stack"), the technology can be used to produce capacitors, too.



- **CMOS** allows to reduce power consumption dramatically.
  - The process, however, is more complex: Wells with different doping type need to be made.



- Using the third dimension (depth / height) might become necessary for integrating "large" structures into a small projected are (example: trench capacitor in **DRAMs** ⇒).
  - Unwanted "topology", however, makes integration more difficult.
  - Planarized technologies are a must since about 1995! ⇒



- It ain't neither easy nor cheap!

Property	Number
Feature size	**0,2 µm**
No. metallization levels	**4 - 7**
No. components	**> 6 · 10^8 (Memory)**
Complexity	**> 500 Process steps**
Cost (development and **1** factory)	**ca. $ 6 · 10^9**

Questionaire

Multiple Choice questions to 5.1

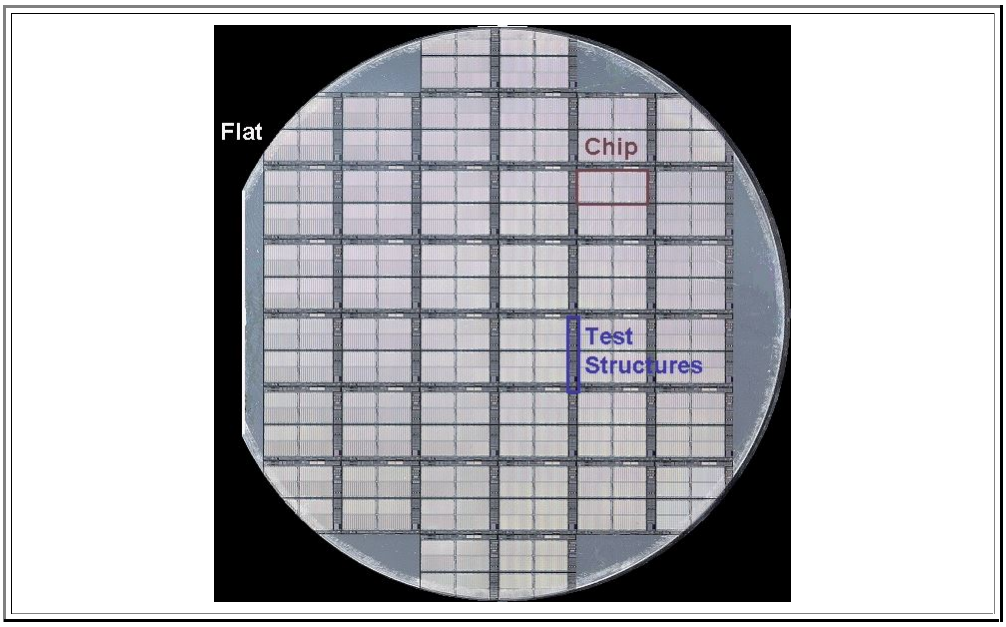
Exercise 5.1-6

All Questions to 5.1

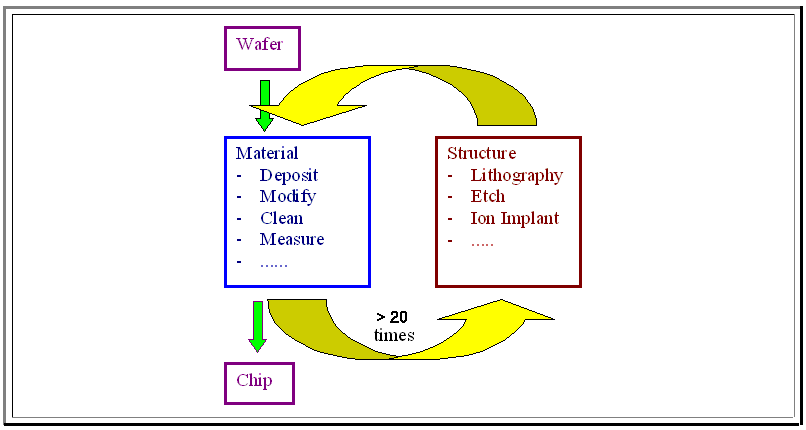
5.2 Process Integration

5.2.1 Chips on Wafers

- We now have a crude idea of *what* we want to make. The question now is *how* we are going to do it.
  - We start with a suitable piece of a **Si** crystal, a **Si wafer**. A wafer is a thin (about **650 µm)** round piece of rather perfect **Si** single crystal with a typical diameter (in the year **2000**) of **200 mm**. Nowadays (**2007**) you would build you factory for **300 mm**.
  - On this wafer we place our **chips**, square or rectangular areas that contain the complete integrated circuit with dimensions of roughly **1 cm^2**.
  - The picture below, which we have seen before, shows a **150 mm** wafer with (rather large **1st** generation) **16 Mbit DRAM** chips and gives an idea about the whole structure.



- The chips will be cut with a diamond saw from the wafer and mounted in their casings.
  - Between the chips - in the area that will be destroyed by cutting - are test structures that allow to measure certain technology parameters.
  - ´The (major) **flat**" of the wafer is aligned along a **<110>** direction and allows to produce the structures on the wafer in perfect alignment with crystallographic directions. It also served to indicate the crystallography and doping type of the wafer; consult the link for details
  - Don't forget: **Si** is *brittle like glass*. Handling a wafer is like handling a thin glass plate - if you are not careful, it breaks.
- How to get the chips on the wafer? In order to produce a **CMOS** structure as shown before, we essentially have to go back and forth between two two basic process modules:
- *Material module*
  - Deposit some material on the surface of the wafer (e.g. **SiO_2**), or
  - Modify material already there (e.g. by introducing the desired doping), or
  - Clean the material present, or
  - Measure something relative to the material (e.g. its thickness), or
  - - well, there are a few more points like this but which are not important at this stage.
- *Structuring module*
  - Transfer the desired structure for the relevant material into some light sensitive layer called a photo-resist or simply **resist** (which is a very special *electronic material*!) by **lithography**, i.e. by projecting a slide (called a **mask** or more generally **reticle**) of the structure onto the light sensitive layer, followed by developing the resist akin to a conventional photographic process, and then:
  - Transfer the structure from the resist to the material by **structure etching** or other techniques.
- Repeat the cycle *more than **20** times* - and you have a wafer with fully processed chips.
  - This is shown schematically in the drawing:



- For the most primitive transistor imaginable, a minimum of **5** lithographic steps are required. Each process module consists of many individual **process steps** and it is the art of **process integration** to find the optimal combination and sequence of process steps to achieve the desired result in the most economic way.

It needs a lot of process steps - most of them difficult and complex - to make a chip.

- Even the most simple **5** mask process requires about **100** process steps.
- A **16 Mbit DRAM** needs about **19** masks and **400** process steps.

To give an idea what this contains, here is a list of the ingredients for a **16 Mbit DRAM** at the time of its introduction to the market (with time it tends to become somewhat simpler):

- **57** layers are deposited (such as **SiO_2** (**14** times), **Si_3N_4**, **Al**, ...).
- **73** etching steps are necessary (**54** with "plasma etching", **19** with wet chemistry).
- **19** lithography steps are required (including deposition of the resist, exposure, and development).
- **12** high temperature processes (including several oxidations) are needed.
- **37** dedicated cleaning steps are built in; wet chemistry occurs **150** times altogether.
- **158** measurements take place to assure that everything happened as designed.

A more detailed rendering can be found in the link.

Two questions come to mind:

- *How long does it take to do all this?* The answer is: *weeks* if everything always works and you never have to wait, and *months* considering that there is no such thing as an uninterrupted process flow all the time.
- *How large is the success rate?* Well, lets do a back-of-the-envelope calculation and assume that each process has a success rate of ***x* %**. The overall **yield *Y*** of working devices is then
  Y*=(*x*/100)^N %** with ***N=number of process steps. With ***N*=450** or **200** we have

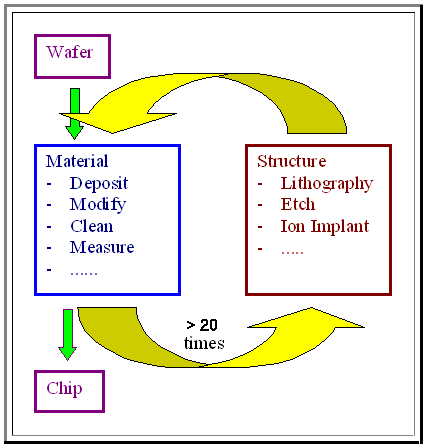
x	***Y* for *N* =450**	***Y* for *N*=200**
95%	$9{,}45 \cdot 10^{-9}$ %	$3{,}51 \cdot 10^{-3}$ %
99%	1,09 %	13,4 %
99,9%	63,7 %	81,9 %

- **N=200** might be more realistic, because many steps (especially controls) do not influence the yield very much.

But whichever way we look at these numbers, there is an *unavoidable conclusion*: Total perfection at each process step is absolutely necessary!

5.2.2 Packaging and Testing of Silicon Chips

5.2.3 Summary to: Chips on Wafers



- Typical wafer size for new factories (**2007**) : **300 mm** diameter, **775 µm** thickness, flatness in lower **µm** region
  - Chip size a few **cm^2**, much smaller if possible
  - Yield ***Y*** = most important parameter in chip production = % of chips on a wafer that function (= can be sold).
  - ***Y* = 29 %** is a good value for starting production
- Chip making = running about **20** times (roughly!!) through "materials" - "structuring" loop.
  - About **400 - 600** individual processing steps (= in / out of special "machine") before chip is finished on wafer
  - More than **30** processing steps for packaging (after separation of chips by cutting)
  - Simple estimate: **99.9%** perfection for each processing step means ***Y* < 70 %.**

Questionaire

Multiple Choice Questions to 5.2

5.3 Money, Markets, and Management

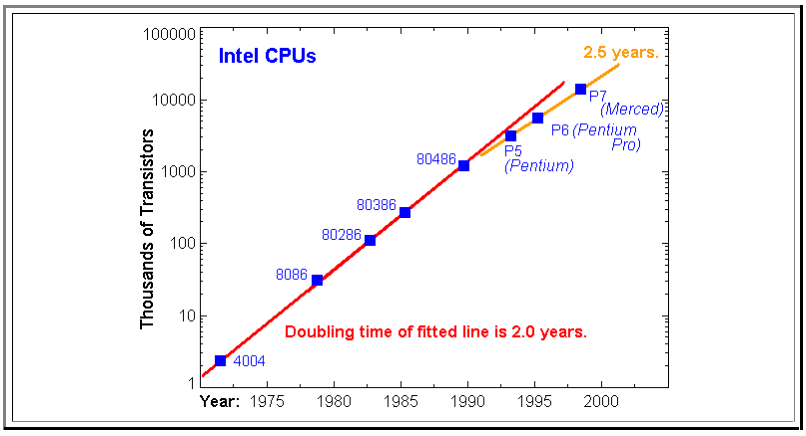
5.3.1 Moore's Law and what it Means

- This module more or less continues chapter "4.3 Infrastructure" - it is essentially about money!
  - Semiconductor technology thrives because it makes money by offering products that become exponentially cheaper and / or better. That is the *only* criterium for a new product that will make money; it is the **central credo**:

Cheaper and / or better!

- Let's look at a simple example: You have already spent something like **2 × 10^9 €** when you produce your very frist **1 Bbit DRAM** memory that you can actually sell to a customer. It has four times more bits that the **256 Mb DRAM** presently on the market. It is definitely *better* in the eyes of your engineers; it's just so much cooler, smaller, faster; uses the latest materilas and so on.
  - However, for you customers it's a memory. If the price is not substantially smaller than **4** times that of the tried-and-proven **256 Mb DRAM**, they are not going to buy it. All they care for is the *price per bit*!

- You may be the foremost *materials* expert in the world, but if you try to leave your mark on chip development without regard to some boundary conditions of a more *economical* nature, you will not achieve much. And if you are the *manager* (which you should be with the kind of education you get here), you better be aware of the following points that are special to research, development and manufacture of (memory) chips.
  - There is no other product with quite such brutal requirements, even considering that *all* technical product development must follow similar (but usually much more relaxed) rules.
- ***1.** A new generation with four-fold capacity will appear on the market every three years*.
  - That is an expression of "**Moore's law**". It is, of course, not a "law" but an extrapolation from observation and bound to break down in the not so distant future (with possible disastrous consequences to the economy of the developed countries).
  - The original observation made in **1965** by Gordon **Moore**, co-founder of **Intel**, was that the number of transistors per square inch on integrated circuits had doubled every year since the integrated circuit was invented. Moore predicted that this trend would continue for the foreseeable future. In subsequent years, the pace slowed down a bit, but data density has doubled approximately every **18** months, and this is the current definition of Moore's Law, which Moore himself has blessed. Most experts, including Moore himself, expect Moore's Law to hold for at least another two decades.
  - Here is a graphic representation for microprocessors:

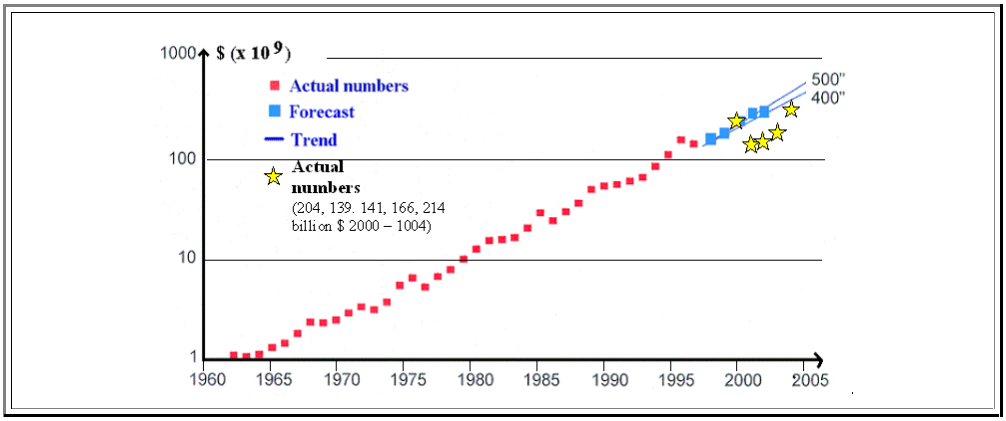


  - Not bad - but of course Moore's law must break down sometime (at the very latest when the feature size is akin to the size of an atom (when would that be assuming that the feature size of the **P7** is **0,2 µm**?). This is illustrated in a separate module.
  - Still, as long as it is true, it means that you either have your new chip generation ready for production at a well-known time in the future, or you are going to loose **large amounts of money**. There are some immediate and unavoidable consequences:
  - You must spent *large amounts of money* to develop the chip and to built the new factory **2 - 3** years before the chip is to appear on the market, i.e. at a time were you do not know if chip development will be finished on time. And *large* means several billion **$**.

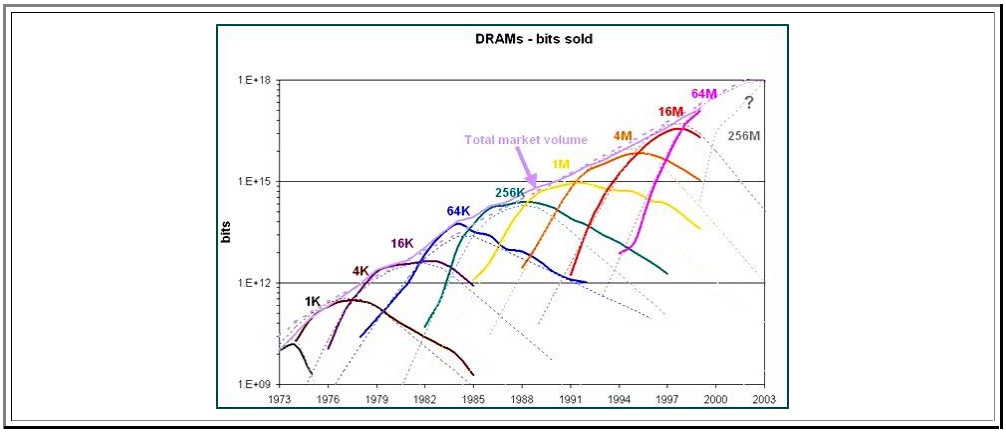
- The time allowed for developing the new chip generation is a constant: You can't start early, because everything you need (better lithography, new materials,) does not exist then. But since chip complexity is ever increasing, you must do more work in the same time. The unavoidable conclusion is more people and *shift work*, even in research and development.
- It follows that you need ever increasing amounts of money for research and development of a new chip generation (there is a kind of Moore's law for the costs of a new generation, too). Look at it in another way in a separate module.

2. The market for chips grows exponentially

- That is an expression of the insatiable demand for chips - as long as they provide more power for less money! That this statement was true is shown below. Note *that the scale is logarithmic*!



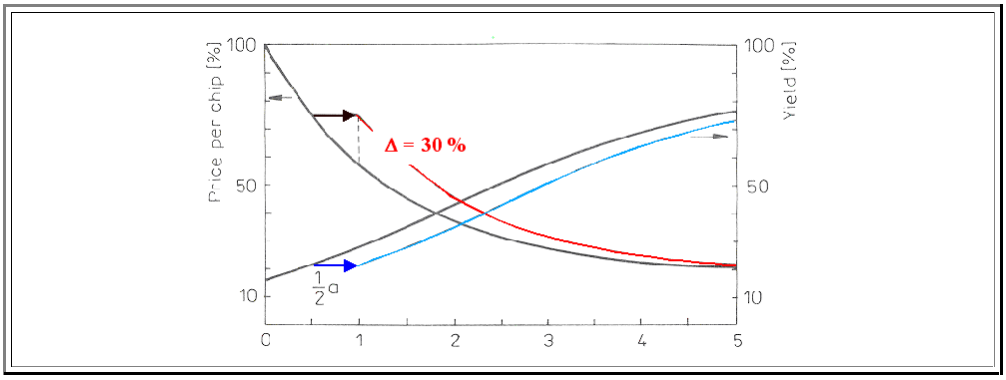
- Shown is the total amount of money grossed in the semiconductor market from **1960 - 2000** (Source: Siemens / Infineon).
  - The essentially straight line indicates exponential growth - including the foreseeable future. Extrapolations, however, are still difficult. The two extrapolated (in **2000)** blue lines indicating a difference of **100.000.000.000 $** of market volume in **2005** are rather close together. The error margins in the forecast thus correspond to the existence or non-existence of about **10** large semiconductor companies (or roughly **150.000** jobs).
  - We also see the worst downturn ever in **2001**. Sales dropped from **204** billion $ in **2000** to **139** billion $ in **2001**, causing major problems throughout the industry.
- More specific, we can see the exponential growth of the market by looking at sales of **DRAMs**. Shown is the total number of memory *bits* sold. Note that just a fourfold increase every three years would keep the number of *chips* sold about constant because of the fourfold increase of storage capacity in one chip about every three years.



- The unavoidable consequence is: Your production capacity must grow exponentially, too, if you just want to keep your share of the market. You must pour an *exponentially growing amount of money* into investments for building factories and hiring people, while the returns on these investments are delayed for at least **2 - 3** years. In other words: the difference of what you must spent and what you earn increases, very roughly, exponentially. This is not a healthy prospect for very long, and you must make *large amounts of money* every now and then (e.g. by being the first one on the market with a new chip or by having a quasi monopoly).
- You must make (and sell at a profit) an *exponentially increasing number of chips* (since the prize per chip is roughly constant) to recover your ever increasing costs. Since chip sizes increase and prices must stay halfway constant, you must use larger wafers to obtain more chips per processing. This puts a lot of pressure on developing larger **Si** crystals and equipment to handle them.
- You must produce as many chips as you can during the product life time (typically **5** years). Continuous shift work in the factory (**7** days a week, **24** hours a day) are absolutely mandatory!

*3. Chip prices for memories decay exponentially from the time of their market introduction (roughly **$60**) by two orders of magnitude within about **5** years (i.e. at the end you get **$1**).*

- The prize development up to the **16 Mbit DRAM** can be seen in an illustration via the link. Microprocessors may behave very differently (as long as Intel has a quasi monopoly). The rapid decay in prizes is an expression of fierce competition and mostly caused by:
- **1.** The "**learning curve**", i.e. the increase of the percentage of good chips on a wafer (the **yield**) from roughly **15%** at the beginning of the production to roughly **90%** at the end of the product life time (because you keep working like crazy to improve the yield).
- **2.** Using a "**shrink strategy**" . This means you use the results of the development efforts for the next two generations to make your present chip smaller. Smaller chips mean more chips per wafer and therefore cheaper chips (the costs of making chips are mostly the costs of processing wafers).
- An immediate consequence is that if you fall behind the mainstream for **6** months or so - you are dead! This can be easily seen from a simple graph:



- The descending black curve shows the expected trend in prizes (it is the average exponential decay from the illustration). The ascending curve is the "learning curve" needed just to stay even - the cost of producing one chip then comes down exactly as the expected prize.
- Now assume you fall behind **6** month - your learning curve, i.e. your yield of functioning chips does not go up. You then move on the modified blue learning curve. The prize you would have to ask for your chip is the modified red prize curve - it is **30 %** above the expected world market prize in the beginning (where prizes are still moderately high).
- Since nobody will pay more for your chips, you are now **30 %** behind the competition (much more than the usual profit margins) - you are going to loose *large amounts of money*!
- In other word: You *must* meet the learning curve goals! But that is easier said then done. Look at real yield curves to appreciate the point

To sum it up: If you like a quiet life with just a little bit of occasional adrenaline, you are better off trying to make a living *playing poker in Las Vegas*.

- Developing and producing new chip generation in our times is a quite involved gamble with billions at stake! There are many totally incalculable ingredients and you must make a lot of million **$** decisions by feeling and not by knowledge!
- So try at least to know what can be known. And don't forget: *It's fun*!

To give a few more data, here is a table with many numbers:

Type	**4 kb**	**16 kb**	**64 kb**	**256 kb**	**1 Mb**	**4 Mb**	**16 Mb**	**64 Mb**	**256 Mb**	**1 Gb**	**4 Gb**
Begin of production	1974	1976	1979	1982	1985	1988	1991	1994	1997	2001	2004
Equivalent of type written pages	0,23	1	4	16	64	250	1000	4000	16000	64000	250000
	Growth per year about **+ 60 %**										
Prize for 1 Mbit memory (DM)	150000.-	50000.-	10000.-	800.-	240.-	60.-	10.-	1.-	0.25.-	0.11.-	0.05.-
	Growth about **– 40%** per year										
Chip size (mm^2)	24	16	25	45	54	91	140	190	250	400	?
Structure size (μm^2)	6	4	2	1.5	1.2	0.8	0.6	0.4	0.3	0.2	0.15
Number of process steps	70	80	8	120	280	400	450	500	600	?	?

Size of "killer" particles (>µm^2)	1.5	1.3	0.8	0.6	0.4	0.2	0.15	0.1	0.07	0.05	0.03
Total development costs (M$)	(90)	(140)	200	450	650	1000	2000	3500	5000	7000	?

5.4 Summary

5.4.1 Summary to 5: Integrated Circuits - Process Integration

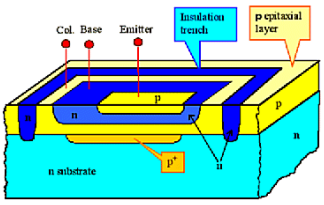
Integration means:

- **1.** Produce a large number (up to **1.000.000.000**) of transistors (bipolar or **MOS**) and other electronic elements on a **cm^2** of **Si**
- **2.** Keep thoses elements electrically insulated from each other.
- **3.** Connect those elements in a meaningful way to produce a system / product.

It ain't easy!

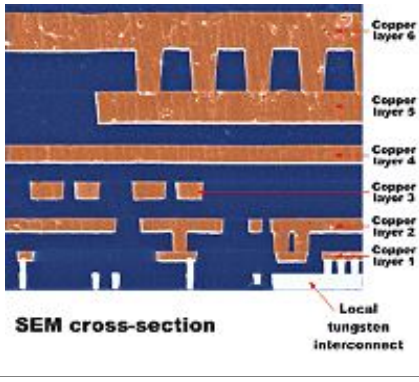
An integrated bipolar transistor does not resemble the textbook picture at all, but looks far more complicated ⇒.

- This is due to the insulation requirements, the process requirements, and the need to interconnect as efficiently as possible.
- The epitaxial layer cuts down on the number of critical diffusions, makes insulation easier, and allows a "buried contact" structure.



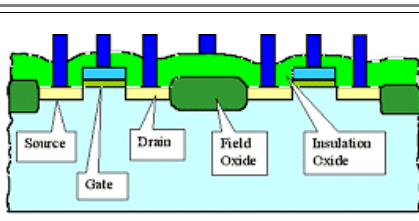
Connecting transistor / elements is complicated; it has to be done on several levels

- Materials used are **Al** ("old"), **Cu** ("new"), **W**, (highly doped) poly-**Si** as well as various silicides.
- Essential properties are the conductivity σ of the conductor, the dielectric constant ϵ_r of the intermetal dielectric, and the resulting time constant $\tau = \sigma \cdot \epsilon_r$ that defines the maximum signal transmision frequency through the conducting line.



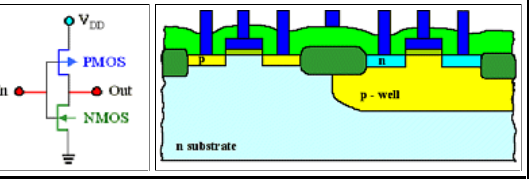
Integrating **MOS** transistors requires special measures for insulation (e.g. a field oxide) and for gate oxide production

- Since a **MOS** transistor contains intrinsically a capacitor (the gate "stack"), the technology can be used to produce capacitors, too.

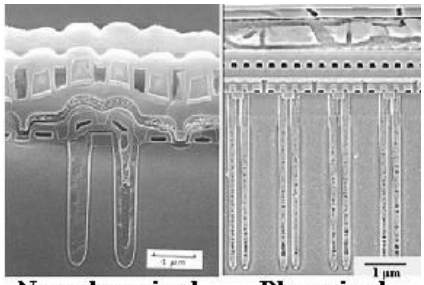


CMOS allows to reduce power consumption dramatically.

- The process, however, is more complex: Wells with different doping type need to be made.



- Using the third dimension (depth / height) might become necessary for integrating "large" structures into a small projected are (example: trench capacitor in **DRAMs** ⇒).
  - Unwanted "topology", however, makes integration more difficult.
  - Planarized technologies are a must since about 1995! ⇒



- It ain't neither easy nor cheap!

Property	Number
Feature size	**0,2 µm**
No. metallization levels	**4 - 7**
No. components	**> 6 · 10^8 (Memory)**
Complexity	**> 500 Process steps**
Cost (development and **1** factory)	**ca. $ 6 · 10^9**

Questionaire

Multiple Choice questions to 5.

Exercise 5.4-1

All Questions to 5.

6. Materials and Processes for Silicon Technology

6.1 Si Oxides and LOCOS Process

6.1.1 Si Dioxide

6.1.2 LOCOS Process

6.1.3 Summary to: 6.1 Si Oxide and LOCOS Process

6.2 Chemical Vapor Deposition

6.2.1 Silicon Epitaxy

6.2.2 Oxide CVD

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

6.2.4 Summary to: 6.2 Chemical Vapor Deposition

6.3. Physical Processes for Layer Deposition

6.3.1 Sputter Deposition and Contact Hole Filling

6.3.2 Ion Implantation

6.3.3 Miscellaneous Techniques and Comparison

6.3.4 Summary to: 6.3 Physical Processes for Layer Deposition

6.4 Etching Techniques

6.4.1 General Remarks

6.5.2 Chemical Etching

6.4.3 Plasma Etching

6.4.4 Summary to: 6.4 Etching Techniques

6.5 Lithography

6.5.1 Basic Lithography Techniques

6.5.2 Resist and Steppers

6.5.3 Summary to: 6.5 Lithograpy

6.6 Summary:

6.6.1 Materials and Processes for Silicon Technology

6. Materials and Processes for Silicon Technology

6.1 Si Oxides and LOCOS Process

6.1.1 Si Dioxide

The Importance of Silicon Dioxide

- Silicon would not be the "miracle material" without its oxide, ***SiO_2***, also known as ordinary **quartz** in its bulk form, or as **rock crystal**if you find it in single crystal form (and relatively pure) somewhere out there in the mountains.
  - Not only the properties of **Si** - especially the achievable crystalline perfection in combination with the bandgap and easy doping - but also the properties of **SiO_2** are pretty much as one would have wished them to be for making integrated circuits and other devices.
  - From the beginning of integrated circuit technology - say **1970** - to the end of the millennium, **SiO_2** was a key material that was used for many different purposes. Only now (around **2004**), industry has started to replace it for some applications with other, highly "specialized" dielectrics.
- What is so special about **SiO_2**?
  - First of all, it comes in many forms: There are several *allotropes* (meaning different crystal types) of **SiO_2** crystals; the most common (stable at room temperature and ambient pressure) is "low quartz" or α-quartz, the quartz crystals found everywhere. **But SiO_2** is also rather stable and easy to make in amorphous form. It is amorphous **SiO_2**- homogeneous and isotropic - that is used in integrated circuit technology. The link provides the phase diagram of **SiO_2** and lists some of its allotropes.
  - **SiO_2** has excellent *dielectric properties*. Its dielectric constant ϵ_r is about **3.7 - 3.9** (depending somewhat on its structure). This *was* a value *large* enough to allow decent capacitances if **SiO_2** is used as capacitor dielectric, but *small* enough so that the time constant $R \cdot C$ (which describes the time delay in wires having a resistance ***R*** and being insulated by **SiO_2**, and thus endowed with a parasitic capacitance ***C*** that scales with ϵ_r) does not limit the maximum frequency of the devices. It is here that successors for **SiO_2** with larger or smaller ϵ_r values are needed to make the most advanced devices (which will hit the market around **2002**).
  - It is among the *best insulators* known and has one of the highest break-through field strengths of all materials investigated so far (it can be as high as **15 MV/cm** for very thin layers; compare that with the values given for normal "bulk" dielectrics).
  - The *electrical properties of the **Si - SiO_2** interface are excellent*. This means that the interface has a very low density of energy states (akin to surface states) in the bandgap and thus does neither provide recombination centers nor introduce fixed charges.
  - **SiO_2** is relatively *easy to make* with several quite different methods, thus allowing a large degree of process leeway.
  - It is also relatively *easy to structure*, i.e. unwanted **SiO_2** can be removed selectively to **Si** (and some other materials) without many problems.
  - It is very *stable and chemically inert*. It essentially protects the **Si** or the whole integrated circuit from rapid deterioration in a chemically hostile environment (provided simply by humid air which will attack most "unprotected" materials).
- What are the uses of **SiO_2**? Above, some of them were already mentioned, here we just list them a bit more systematically. If you do not quite understand some of the uses - do not worry, we will come back to it.
  - **Gate oxides**: As we have seen before, we need a thin dielectric material to insulate the gate from the channel area. We want the channel to open at low threshold voltages and this requires large dielectric constants and especially no charges in the dielectric or at the two interfaces. Of course, we always need high break through field strength, too. No dielectric so far can match the properties of **SiO_2** in total.
  - *Dielectrics in integrated capacitors*. Capacitors with high capacitance values at small dimensions are needed for so-called **dynamic random access memories** (***DRAM***), one of the most important integrated circuits (in terms of volume production). You want something like **30 fF** (femtofarad) on an area of **0.25 μm^2**. The same issues as above are crucial, except that a large dielectric constant is even more important. While **SiO_2** was the material of choice for many **DRAM** generations (from the **16 kbit**

DRAM to the **1 Mbit DRAM**), starting with the **4 Mbit** generation in about **1990**, it was replaced by a triple layer of **SiO_2 - Si_3N_4 - SiO_2**, universally known as "***ONO***" (short for oxide - nitride - oxide); a compromise that takes not only advantage of the relatively large dielectric constant of silicon nitride (around **7.5**) while still keeping the superior quality of the **Si - SiO_2** interface, but has a few added benefits - at added costs, of course.

- *Insulation*: Some insulating material is needed between the transistor in the **Si** as well as between the many layers of wiring on the chip; cf. the many pictures in chapter five, starting with the one accessible via the link. **SiO_2** was (and still is) the material of choice. However, here we would like to have a material with a *small* dielectric constant, ideally **1**, minimizing the parasitic capacitance between wiring and **SiO_2** may have to be replaced with a different kind of dielectric around **2003**.
- *Stress relieve* layer: **SiO_2** becomes "*viscous*" at high temperatures - it is a glass, after all. While it is a small effect, it is large enough to absorb the stress that would develop between unwielding materials, e.g. **Si_3N_4** on **Si**, if it is used as a "**buffer oxide**", i.e. as a thin intermediary layer.
- **Masking**: Areas on the **Si** which are not be exposed to *dopant diffusion* or *ion implantation* must be protected by something impenetrable and that also can be removed easily after "use". It's **SiO_2**, of course, in many cases,
- **"Screen oxides"** provide one example of so-called **sacrificial layers** which have no direct function and are disposed off after use. A screen oxide is a thin layer of **SiO_2** which stops the low energy ***debris*** that comes along with the high-energy ion beam - consisting, e.g., of metal ions that some stray ions from the main beam banged off the walls of the machine. All these (highly detrimental) metal and carbon ions get stuck in the screen oxide (which will be removed after the implantation) and never enter the **Si**. In addition, the screen oxide scatters the main ion beam a little and thus prevents "*channeling*", i.e. deep penetration of the ions if the beam happens to be aligned with a major crystallographic direction.
- **Passivation**: After the chip is finished, it has to be protected from the environment and all bare surfaces need to be electrically passivates - its done with **SiO_2** (or a mixture of oxide and nitride).

- **Gate oxide for Transistors**
- **Dielectric in Capacitors**
- **Insulation**
- **Stress relieve layer**
- **Masking layer**
- **Screen oxide during Implantation**
- **Passivation**

Enough reasons for looking at the oxide generation process a little more closely? If you think not - well there are more uses, just consult the list of processes for a **16 Mbit DRAM**: You need **SiO_2** about **20** times!

How is **SiO_2** made - in thin layers, of course? There are essentially *three* quite distinct processes with many variations:

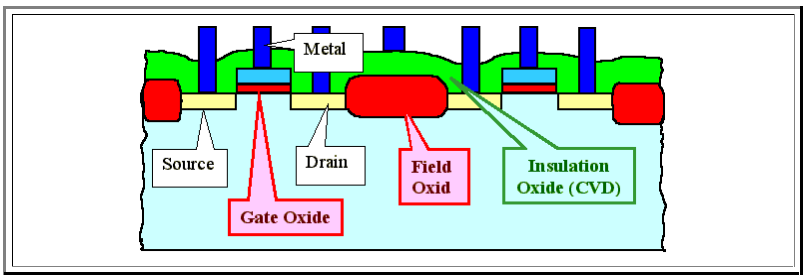
- **Thermal oxidation**. This means that a *solid state reaction* (**$Si + O_2 \Rightarrow SiO_2$**) is used: Just expose **Si** to **O_2** at sufficiently high temperatures and an oxide will grow to a thickness determined by the temperature and the oxidation time.
- **CVD oxide deposition**. In complete analogy to the production of poly-**Si** by a **CVD** (= chemical vapor depositions) process, we can also produce **SiO_2** by taking the right gases and deposition conditions.
- **Spin-on glass** (***SOG***). Here a kind of polymeric suspension of **SiO_2** dissolved in a suitable solvent is dropped on a rapidly spinning wafer. The centrifugal forces spread a thin viscous layer of the stuff on the wafer surface which upon heating solidifies into (not extremely good) **SiO_2**. Its not unlike the stuff that was called "***water* glass**" or "liquid glass" and that your grandma used to conserve eggs in it.

There is one more method that is, however, rarely used - and never for mass production: **Anodic oxidation**.

- Anodic oxidation uses a current impressed on a **Si** - electrolyte junction that leads to an oxidation reaction. While easy to understand in principle, it is not very well understood in practice and an object of growing basic research interest.
- In the module **2.5.1** "Porous semiconductors more information to this point can be found.

Thermal Oxidation

- In this paragraph we will restrict ourselves to **thermal oxidation**. It was (and to a large extent still is) one of the key processes for making integrated circuits. While it may be used for "secondary" purposes like protecting the bare **Si** surface during some critical process (remember the "screen oxide" from above?), its major use is in three areas:
  - **Gate oxide** (often known as "***GOX***")
  - **Capacitor dielectric**, either as "simple" oxide or as the "bread" in an "**ONO**" (**=** oxide-nitride-oxide sandwich)
  - **Field oxide** (***FOX***) - the lateral insulation between transistors.
- We can use a picture from chapter 5.1.4 to illustrate **GOX** and **FOX**; the capacitor dielectric can also be found in this chapter.



  - We must realize, however, that those drawing are *never to scale*. The gate oxide is only around **10 nm** thick (actually, it "just" (**2007**) petered out at **1.2 nm** accoding to Intel and is now replaced by a thicked HfO_2), whereas the field oxide (and the insulating oxide) is in the order of **500 nm**. What it looks like at atomic resolution in an electron microscope is shown in this link.
- There are essentially *two* ways to do a thermal oxidation.
- "**Dry oxidation**", using the reaction
.

$$2\,Si + O_2 \Rightarrow 2\,SiO_2 \quad (800\,^oC - 1100\,^oC)$$

  - This is the standard reaction for thin oxides. Oxide growth is rather slow and easily controllable.
  - To give an example: Growing **700 nm** oxide at **1000 °C** would take about **60 hr** - far too long for efficient processing. But **7nm** take only about **15 min** - and that is now too short for precise control; you would want to lower the temperature.
- "**Wet oxidation**", using the reaction

$$Si + 2\,H_2O \Rightarrow SiO_2 + 2\,H_2 \quad (800\,^oC - 1100\,^oC)$$

  - The growth kinetics are about **10x** faster than for dry oxidations; this is the process used for the thick *field oxides*.
  - Growing **700 nm** oxide at **1000 °C** now takes about **1.5 hr** - still pretty long but tolerable. Thin oxides are never made this way.
- In both cases the oxygen (in the form of **O**, O_2, OH^-, whatever,...) has to *diffuse through the oxide already formed* to reach the **Si - SiO_2** interface where the actual reaction takes place.
  - This becomes more difficult for thick oxides, the reaction after *some time of oxide formation* is always **diffusion limited**. The thickness d_{ox} of the growing oxide in this case follow a general "*square root*" law, i.e. it is proportional to the diffusion length $L=(Dt)^{1/2}$ (***D***=diffusion coefficient of the oxygen carrying species in SiO_2; ***t*****=** time).
  - We thus have a general relation of the form .

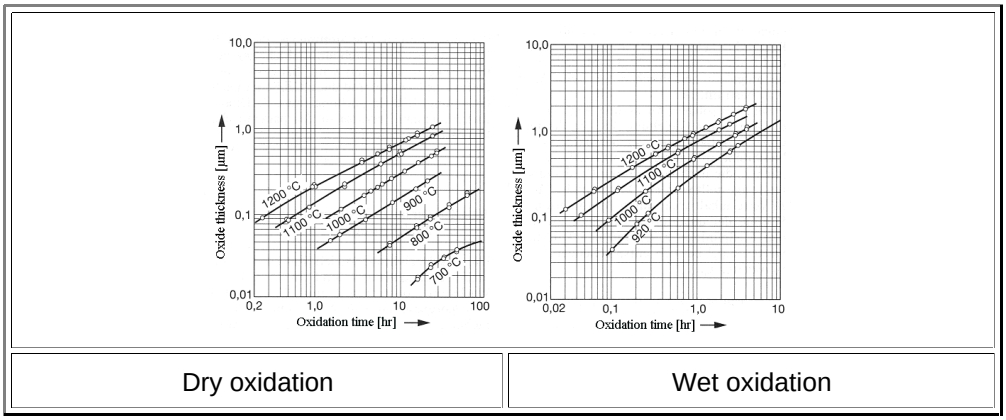
$$d_{thick\text{-}ox} = const. \cdot (D \cdot t)^{1/2}$$

  - For *short times or thin oxide thicknesses* (about **< 30 nm**), a *linear* law is found

$$d_{thin\text{-}ox} = const. \cdot t$$

  - In this case the limiting factor is the rate at which **O** can be incorporated into the **Si - SiO_2** interface.

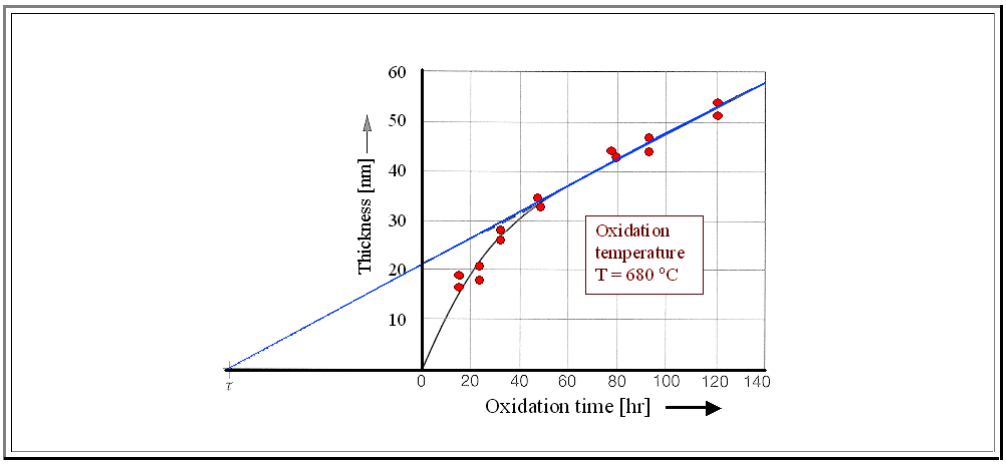
This kind of behavior - linear growth switching to square root growth - can be modelled quite nicely by a not too complicated theory known as the **Deal-Grove model**. Some results of experiments and modeling are shown below.



Dry oxidation | Wet oxidation

- The left diagram shows dry, the right hand one wet oxidation. The solid curves were calculated with the Deal-Grove model after parameter adjustment, the circles indicate experimental results. *The theory seems to be pretty good !*.

However, for *very thin oxides* - and those are the ones needed or **GOX** or capacitors - things are even more complicated as shown below.

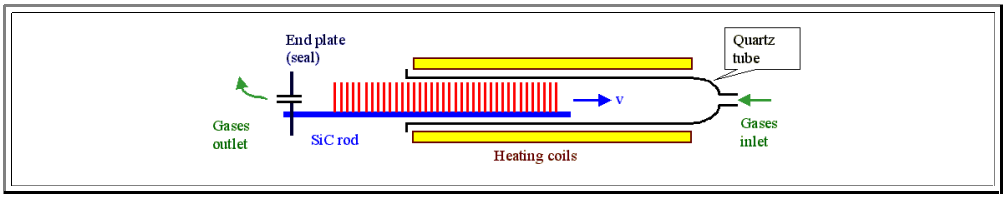


- The red points are data points from an experiment (at an unusually low temperature); the blue curve is the best Deal-Grove fit under the (not justified) assumption that at ***t*=0** an oxide with a thickness of about **20 nm** was already present.
- The Deal-Grove model is clearly inadequate for the technologically important case of very thin oxides and experimentally an *exponential law* is found for the dependence of the oxide thickness on time for very short times
- Moreover, the detailed kinetics of oxide growth are influenced by many other factors, e.g. the *crystallographic orientation* of the **Si** surface, the *mechanical stress* in the oxide (which in turn depends on many process variables), the substrate *doping*, and the *initial condition* of the **Si** surface.

And this is only the thickness! If we consider the *properties* of the oxide, e.g. the amount of fixed charge or interface charge, its etching rate, or - most difficult to ***assess*** - how long it will last when the device is used, things become most complicated. An oxide with a nominal thickness d_{ox} can be produced in many ways: dry or wet oxidation, high temperatures and short oxidation times or the other way around - its properties, however, can be very different.

- Learning about microelectronic processes involves very little Math; and "theory" is needed only at an elementary to medium level! But this does *not* make the issue trivial - quite the contrary. If you would have a theory - however complicated - that predicts all oxide properties as a function of all variables, process development would be easy. But presently, even involved theories are mostly far too simple to come even close to what is needed. On an advanced level of real process development, it is the interplay of a solid foundation in materials science, lots of experience, usage of mathematical models as far as they exist, and finally some luck or "feeling" for the job, that will carry the day.
- So do not consider microelectronic processes "simple" because you do not find lots of differential equations here. There are few enterprises more challenging for a materials scientist then to develop key processes for the next chip generation!

- How is a thermal oxidation done in real life? Always inside an oxidation furnace in a **batch process**; i.e. many wafers (usually **100**) are processed at the same time.
  - Oxidation furnaces are complicated affairs, the sketch below does not do justice to the intricacies involved (nowadays they are usually no longer horizontal as shown below, but vertical). For some pictures of real furnaces use the link.



  - First of all, temperature, gas flow etc., needs not only to be very constant but precisely adjustable to your needs. Generally, you do not load the furnace at the process temperature but at some lower temperature to avoid thermal shock of the wafers (inducing temperature gradients, leading to mechanical stress gradients, leading to plastic deformation, leading to the generation of dislocations, leading to wafers you have to throw away). After loading, you "ramp up" the temperature with a precisely defined rate, e.g. **15 ^{o}C/min**, to the process temperature selected.
  - During loading and ramping up, you may not want to start the oxidation, so you run **N_2** or **Ar** through the furnace. After the process temperature has been reached, you switch to **O_2** for dry oxidation or the right mixture of **H_2** and **O_2** which will immediately burn to **H_2O** if you want to do a wet oxidation.
  - After the oxidation time is over, you ramp down the temperature and move the wafers out of the furnace.
- Moving wafers in and out of the furnace, incidentally, is not easy.
  - First you have to transfer the wafers to the rigid rod system (usually **SiC** clad with high purity quartz) that moves in and out, and than you have to make sure that the whole contraption - easily **2 m** long - moves in and out at a predetermined speed **v** without ever touching anything - because that would produce particles.
  - There is quite a weight on the rods and they have to be absolutely unbendable even at the highest process temperature around **1150 ^{o}C**.
  - Of course, the rods, the quartz furnace tube and anything else getting hot or coming into contact with the **Si**, must be ultrapure - hot **Si** would just lap up even smallest traces of deadly fast-diffusing metals.
- And now to the *difficult* part: After the process "works", you now must make sure that it works exactly the same way (with tolerances of **< 5%**) on all **100** wafers in a run in, from run to run, and independently of which one of the **10** or so furnaces is used.
  - So take note: Assuring *stable process specifications* for a *production environment* may be a more demanding and difficult job than to develop the process in the first place.
- You see: At the final count, a "simple" process like thermal oxidation consists of a process recipe that easily calls for more than **20** precisely specified parameters, and changing anyone just a little bit may change the properties of your oxide in major ways.

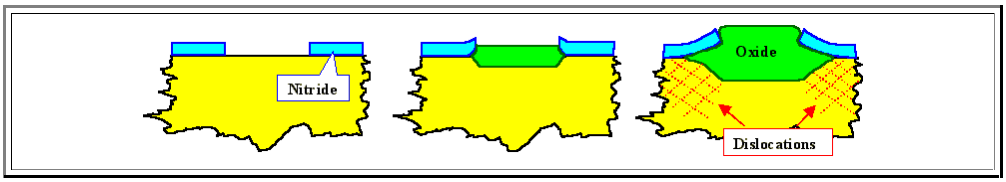
All these points were emphasized to demonstrate that even seemingly simple processes in the production of integrated circuits are rather complex.

  - The processes to be discussed in what follows are no less complex, rather more so. But we will not go into the finer points at great depth anymore.
- There are two more techniques to obtain **SiO_2** which are so important that we have to consider them in independent modules:
  - *Local oxidation* - this will be contained in the following module, and
  - ***CVD** deposition* of oxide - this will be part of the **CVD** module.

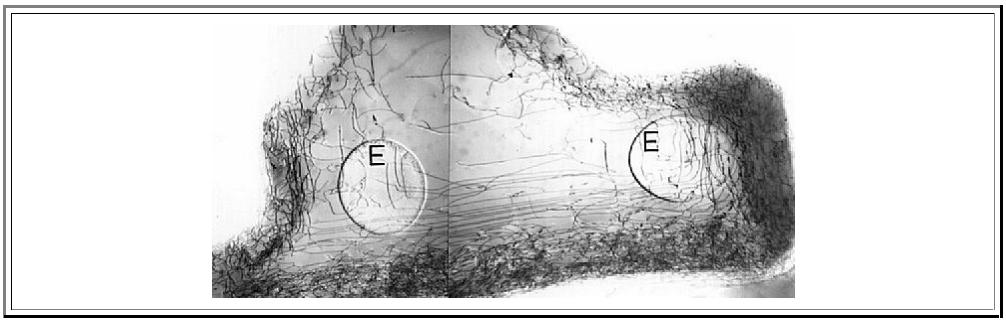
6.1.2 LOCOS Process

Basic Concept of Local Oxidation

- The abbreviation "***LOCOS***" stands for "**Local Oxidation of Silicon**" and was almost a synonym for **MOS** devices, or more precisely, for the insulation between single transistors. **LOCOS** makes the isolation between **MOS** transistors considerably easier then between bipolar transistors, cf. the drawings discussed before:
  - For bipolar transistors, you have to separate the collectors. This involves an epitaxial layer and some deep diffusion around every transistor.
  - For MOS transistors, no isolation would be needed weren't it for the possible parasitic transistors. And this problem can be solved by making the "gate oxide" of the parasitic transistors - which then is called **field oxide** - sufficiently thick.
- The thick field oxide has been made by the **LOCOS** process from the beginning of **MOS** technology until presently, when **LOCOS** was supplanted by the "box isolation technique", also known as "***STI***" for "**Shallow trench isolation**".
  - Since the **LOCOS** technique is still used, and gives a good example of how processes are first conceived, are optimized with every generation, become very complex, and are finally supplanted with something different, we will treat it here in some detail
- As the name implies, the goal is to oxidize **Si** only *locally*, wherever a field oxide is needed. This is necessary for the following reason:
  - *Local* (thermal) oxide penetrates into the **Si** (oxidation is using up **Si**!), so the **Si - SiO_2** interface is *lower* than the source - drain regions to be made later. This could not be achieved with oxidizing all of the **Si** and then etching off unwanted oxide.
  - For device performance reasons, this is highly beneficial, if not absolutely necessary.
- For a *local* oxidation, the areas of the **Si** that are not to be oxidized must be protected by some *material* that does not allow oxygen diffusion at the typical oxidation temperatures of **(1000 - 1100) ^{0}C**. We are talking electronic materials again!
  - The *only material* that is "easily" usable is **Silicon nitride**, ***Si_3N_4***. It can be deposited and structured without too much problems and it is compatible with **Si**.
  - However, **Si_3N_4** introduces a major new problem of its own, which can only be solved by making the process more complicated by involving yet another materials. This gives a ***succinct*** example of the statement made before: That materials and processes have to be seen as a unit.
  - Lets see what would happen with just a **Si_3N_4** layer protecting parts of the **Si** from thermal oxidation.



  - Oxygen diffusion through the oxide already formed would also oxidize the Si under the **Si_3N_4**; i.e. there would be some amount of lateral oxidation. Since a given volume of **Si** expands by almost a factor of **2** upon oxidation (in other words: Oxidizing **1cm^3** of **Si** produces almost **2 cm^3** of **SiO_2**), the nitride mask is pressed upwards at the edges as illustrated.
  - With increasing oxidation time and oxide thickness, pressure under the nitride mask increases, and at some point the *critical yield strength* of **Si** at the oxidation temperature is exceed. *Plastic deformation* will start and dislocations are generated and move into the **Si**. Below the edges of the local oxide is now a high density of dislocations which kill the device and render the **Si** useless - throw it out.
  - This is not "theory", but eminently practical as shown in the ***TEM*** picture from the early days of integrated circuit technology:



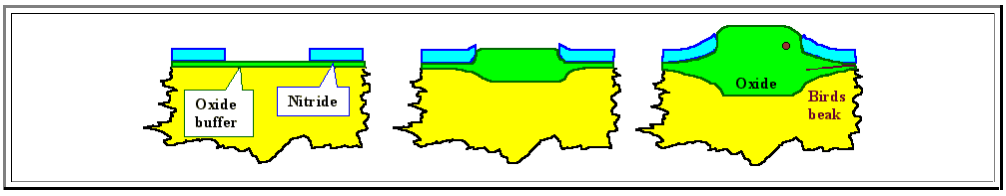
- We are looking through a piece of **Si**. The dark lines are the projections of single dislocations, the "dislocations tangles" corresponds to oxide edges; "**E**" shows contact areas (emitters) to the **Si**. Another picture can be found in the link.
- Actually, it doesn't even need the oxidation to produce dislocations. Si_3N_4 layers are always under large stresses at room temperature and would exert great shear stresses on the **Si**; something that can not be tolerated as soon as the nitride films are more than a few **nm** thick.

We arrive at a simple rule: You *cannot* use Si_3N_4 directly on **Si** - never ever. What are we to do now, to save the concept of local oxidation?

Buffer Oxide

We need something *between* the Si_3N_4 mask and the **Si**; a thin layer of a material that is compatible with the other two and that can *relieve the stress* building up during oxidation. Something like the oil in you motor, a kind of ***grease***.

- This "grease" material is SiO_2, as you might have guessed - it was already mentioned before under its proper name of "**buffer oxide**". The hard Si_3N_4 (which is a ceramic that is very hard not yielding at a "low" temperature of just about **1000 °C**), is now pressing down on something "soft", and the stress felt by the **Si** will not reach the yield stress - if everything is done right.
- The situation now looks like this



- No more dislocations, but a comparatively large lateral oxidation instead, leading to a configuration known as "**birds beak**" for the obvious reason shown in the picture to the right (the inserts just are there to help you see the bird).

So we got rid of one problem, but now we have another one: The lateral extension of the field oxide via the birds beak is comparable to its thickness and *limits the minimum feature size*.

- While this was not a serious problems in the early days of **IC** technology, it could not be tolerated anymore around the middle of the eighties.
- One way out was the use of a poly-**Si** layer as a sacrificial layer. It was situated on top of the buffer oxide below the nitride mask and was structured with the mask. It provided some sacrificial **Si** for the "birds beak" and the total dimension of the field oxide could be reduced somewhat.
- This process is shown in comparison with the standard process in the link.

But even this was not good enough anymore for feature sizes around and below **1 µm**. The **LOCOS** process eventually became a very complicated process complex in its own right; for the Siemens **16 Mbit DRAM** it consisted of more than **12** process steps including:

- **2** oxidations, **2** poly-**Si** deposition, **1** lithography, **4** etchings and **2** cleaning steps.
- It was one of the decisive "secrets" for success, and we can learn a simple truth from this:

Before new materials and processes are introduced, the existing materials and processes are driven to extremes! And that is not only true for the **LOCOS** process, but for all other processes.

- Still, with feature sizes shrinking ever more, **LOCOS** reached the end of its useful life-span in the nineties and had to be replaced by "Box isolations", a simple concept in theory, but hellishly difficult in reality.
- The idea is clear: Etch a hole (with vertical sidewalls) in the **Si** wherever you want an oxide, and simple "fill" it with oxide next. More about this process can be found in the link above.

6.1.3 Summary to: 6.1 Si Oxide and LOCOS Process

- Silicondioxide (SiO_2) has been the "ideal" dielectric with many uses in chip manufacture
  - Only recently (**2007**) is it replaced by "low **k"** and "high **k**" dielectrics, i.e. dielectrics with a dielectric constant either lower or larger than that of SiO_2
  - "Low **k**" dielectrics (polymers, porous SiO_2, ..; the ideal material has not yet been found) are used for intermetal insulation; low **k** is important here to keep the **RC** time constants small
  - "High **k**" dielectrics (the present front runner is HfO_2) will replace the gate oxides. They can be somewhat thicker than SiO_2 without sacrificing capacity, while strongly reducing tunneling currents.

- **Gate oxide for Transistors**
- **Dielectric in Capacitors**
- **Insulation**
- **Stress relieve layer**
- **Masking layer**
- **Screen oxide during Implantation**
- **Passivation**

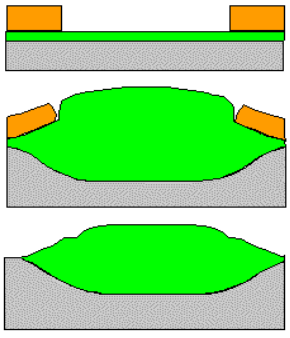
- SiO_2 can be made in several ways:
  - Dry oxidation is relatively slow but gives best oxide qualities as defined by:
    - Uniformity
    - thickness control
    - Break down field strangt
    - Interface quality
    - Reliability

    Typical use: Highest quality gate oxid.
  - Wet oxidation is about 10 times faster; it is used whenever relatively thick oxides are needed.
    Typical use: Field oxide.
  - The other methods are needed whenever there is no **Si** available for oxidation (e.g. intermetal dielectrics).

- Dry thermal oxidation:
  $2\,Si + O_2 \Rightarrow 2\,SiO_2$
- Wet thermal oxidation:
  $Si + 2\,H_2O \Rightarrow SiO_2 + 2\,H_2$
- "Chemical Vapor Deposition" (next sub-chapter)
- "Spin-on techniques (next sub-chapter)
- "Anodic oxidation (presently not used in technology)

- As long as the process is diffusion controlled (i.e. the time it takes oxygen to diffuse through the already formed oxide determines rates, the thickness increases protorional to $t^{1/2}$
  - For thin oxides the growth rate is reaction controlled and the thickness - time dependence becomes complicated.

- Growing oxide only locally ("**LOCOS**") was a key process for field oxides.
  - Without a "buffer" oxide below the masking nitride, large mechanical strain develops, producing plastic deformation and thus dislocations around the oxide edges.
  - These "Oxide edge dislocations" kill the transistor.
  - Buffer oxides solve the problem, but create new problems: A "birds beak" develops, increasing lateral dimensions beyond the mask dimension.
- **LOCOS** is a good example for a universal feature of **Si** technology: Solutions to old problems create new problems. Solutions to the new problems... and so on. It follows:
  - Process complexity increases all the time.
  - New materials are needed all the time.



Questionaire

Multiple Choice questions to all of 6.1

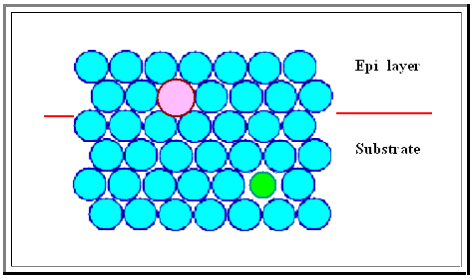
6.2 Chemical Vapor Deposition

6.2.1 Silicon Epitaxy

We have encountered the need for epitaxial layers before, and we also have seen a **Si CVD** process for making poly-crystalline material good enough for growing crystals. All we have to do now is to put both things together.

- We can use essentially the same **CVD** process as before, but instead of thin rods of poly-**Si** which we want to grow in diameter as fast as possible, we now want to make a thin, but absolutely perfect **Si** layer on top of a wafer.

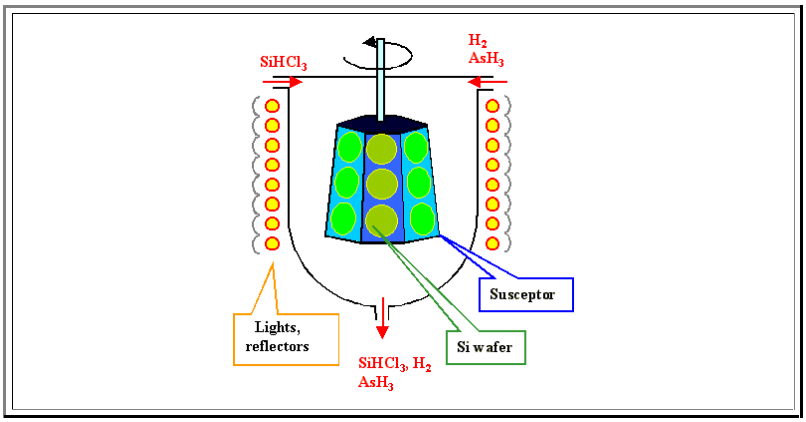
We now must have tremendous process control. We require:



- A *precise continuation* of the substrate lattice. There should be no way whatsoever to identify the interface after the epitaxial layer has been deposited. This means that *no lattice defects whatsoever* should be generated.
- *Doping* of the epitaxial layer with high precision (e.g. **5 Ωcm ± 5%**), and the doping is usually very different from that of the substrate. The picture on the right symbolizes that by the two differently colored doping atoms.
- *Precise thickness control*, e.g. ***d* = 1.2 µm ± 10%** over the entire wafer, from wafer to wafer and from day to day. Now there is a challenge: If you met the first point and thus can't tell where the interface is - how do you measure the thickness? (The answer: Only electronically, e.g. by finding the position of the **pn**-junction produced).
- *Cleanliness*: No contaminants diffusing into the substrate and the epitaxial layer are allowed.

This looks tough and it is indeed fairly difficult to make good epitaxial layers. It is also quite expensive and is therefore avoided whenever possible (e.g. in mainstream **CMOS** technology). It is, however, a must in bipolar and some other technologies and also a good example for a very demanding process with technical solutions that are far from obvious.

Lets look at a typical epitaxial reactor from around **1990** (newer ones tend to be single wafer systems). It can process several wafers simultaneously and meets the above conditions. Here is a muchly simplified drawing:



The chemical reaction that produces the **Si** is fairly simple:

$$SiCl_4 + 2\,H_2 \Rightarrow Si + 4\,HCl \quad (1000\,^{o}C - 1200\,^{o}C)$$

- The dopant gases just decompose or react in similar ways. However, instead of $SiCl_4$ you may want to use SiH_xCl_{4-x} .
- The essential point is that the process needs high temperatures and the **Si** wafer will be at high temperature! In an Epi reactor as shown above, the **Si** wafer surfaces (and whatever shows of the susceptor) are the only hot surfaces of the system!

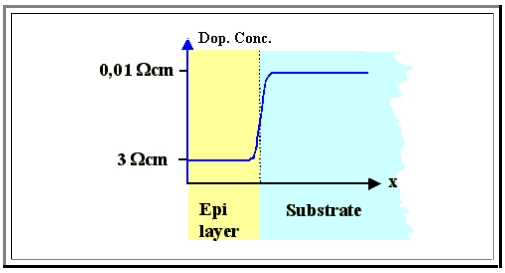
How is the process actually run? You need to meet some tight criteria for the layer specifications, as outlined above, and that transfers to tight criteria for process control.

- **1.** *Perfectly clean* **Si** surface before you start. This is not possible by just putting clean **Si** wafers inside the Epi-reactor (they always would be covered with SiO_2), but requires an in-situ cleaning step. This is done by first admitting only H_2 and Cl_2 into the chamber, at a very high temperature of about **1150 °C**. **Si** is etched by the gas mixture - every trace of SiO_2 and especially foreign atoms at the surface will be removed.
- **2**. Temperature gradients of at most **(1 - 2) °C**. This is (better: was) achieved by **heating with light** as shown in the drawing. The high intensity light bulbs (actually rods) consume about **150 kW** electrical power (which necessitates a **30 kW** motor running the fan for air-cooling the machinery).
- **3.** Extremely tightly controlled gas flows within a range of about **200 l/min** H_2, **5 l/min** $SiCl_4$ (or **Si** HCl_3), and fractions of **ml/min** of the doping gases.

Not to forget: Epi-reactors are potentially very dangerous machines with a lot of "dirty" output that needs to be cleaned. All things taken together make Epi-reactors very expensive - you should be prepared to spend several million **$** if you want to enter this technology.

- **Si** epitaxy thus is a process that is avoided if possible - it costs roughly **$5** per wafer, which is quite a lot. So when do we use epitaxy?

Epitaxy is definitely needed if a *doping profile* is required where the *resistivity in surface near regions is larger than in the bulk*. In other words, a profile like this:



- By diffusion, you can always lower the resistivity and even change the doping type, but *increasing the resistivity by diffusion* is not realistically possible.
- Consider a substrate doping of 10^{16} **cm**^{3}. Whatever resistivity it has (around **5 - 10** Ω**cm**), if you diffuse **2 · 10**^{16} **cm**^{3} of a dopant into the substrate, you *lowered* the resistivity of the doped layer by a factor of **2**.
- To *increase* the resistivity you have to compensate half of the substrate doping by diffusing a dopant for the reverse doping type with a concentration of **5 · 10**^{15} **cm**^{3}. Not only does that call for much better precision in controlling diffusion, but you will only get that value at a particular distance from the surface because you always have a diffusion profile. So all you can do by diffusion is to increase the resistivity somewhat near the surface regions; but you cannot make a sizeable layer this way.

You also may use epitaxial layers if you simply need a *degree of freedom in doping* that is not achievable otherwise.

- While **DRAMs** were made without epitaxy up to the **16 Mbit** generation (often to the amazement of everybody, because in the beginning of the development work epitaxy seemed to be definitely necessary), epitaxial **Si** layers are now included from the **64 Mbit DRAM** upwards.

6.2.2 Oxide CVD

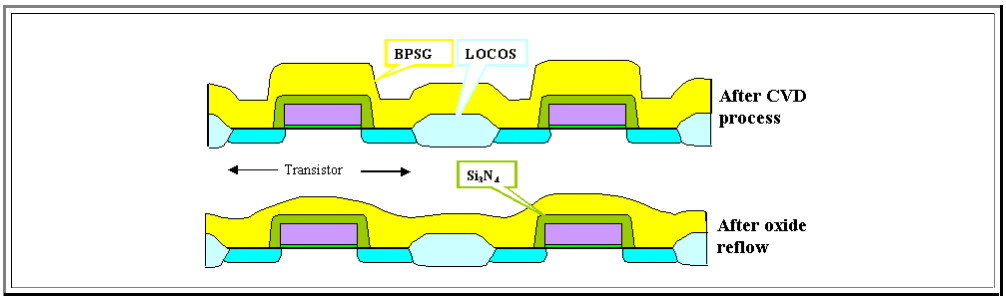
- Whenever we need **SiO_2** layers, but can not oxidize **Si**, we turn to **oxide CVD** and deposit the oxide on top of the substrate - whatever it will be
  - Again, we have to find a suitable chemical reaction between gases that only occurs at high temperature and produces **SiO_2**. There are several possibilities, one is

$$SiH_2Cl_2 + 2NO_2 \Rightarrow SiO_2 + 2HCl + 2N_2 \quad (900\ ^oC)$$

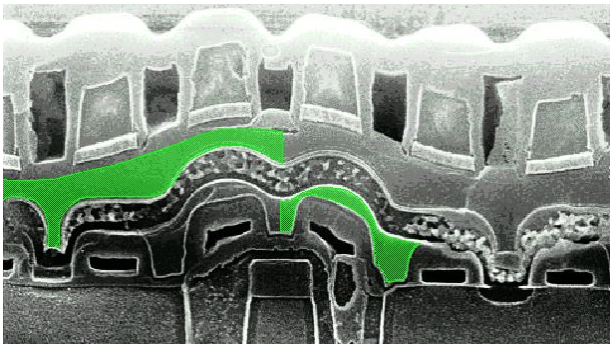
  - While this reaction was used until about **1985**, a better reaction is offered by the "***TEOS***" process:

$$Si(C_2H_5O)_4 \Rightarrow SiO_2 + 2H_2O + C_2H_4 \quad (720\ ^oC)$$

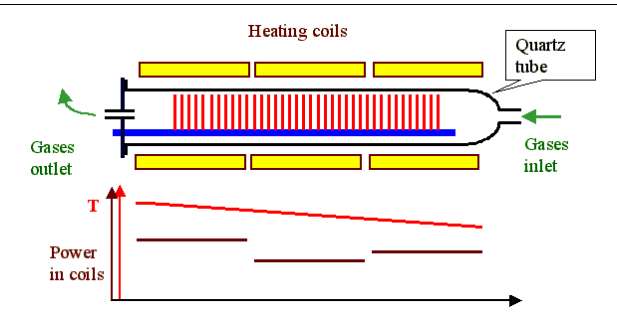
  - **$Si(C_2H_5O)_4$** has the chemical name **Tetraethylorthosilicate**; abbreviated **TEOS**. It consists of a **Si** atom with the four organic molecules bonded in the four tetrahedral directions. The biggest advantage of this process is that it can be run at lower temperatures, but it is also less dangerous (no **HCl**), and it produces high quality oxides.
- Low temperature processes are important after the transistors and everything else in the **Si** has been made. Every time the temperature must be raised for one of the processes needed for metallization, the dopant atoms will move by diffusion and the doping profiles change.
  - Controlling the "**temperature budget**" is becoming ever more important as junction depths are getting smaller and smaller.
- **CVD** techniques allow to tailor some properties of the layers deposited by modifying their chemistry. Often, an oxide that "flows" at medium temperature, i.e. evens out the topography somewhat, is needed. Why is shown below.



  - After the transistor has been made, there is a veritable mountain range. Here it is even worse than before, because the whole "gate stack" has been encapsulated in **Si_3N_4** - for reasons we will not discuss here. (Can you figure it out? The process is called "**FOBIC**", short for "Fully Overlapping Bitline Contact").
  - It is important for the next processes to flatten the terrain as much as possible. While this is now done by one of the major key process complexes introduced around **1995** (in production) called "***CMP***" for "**Chemical-mechanical polishing**", before this time the key was to make a "**flow glass**" by doping the **SiO_2** with **P** and/or **B**. Conventional glass, of course is nothing like **SiO_2** containing ions like **Na** (which is a no-no in chip making), but **P** and **B** are also turning quartz into glass.
  - The major difference between *glass* and *quartz* is that glass becomes a kind of viscous liquid above the *glass temperature* which depends on the kind and concentration of ions incorporated.
- So all you have to do during the **SiO_2** deposition, is to allow some incorporation of **B** and/or **P** by adding appropriate gases.
  - As before phosphine (**PH_3**) is used for **P**, and "**TMB**" (=**$B(OCH_3)_3$=** trimethylborate) for **B**. Concentrations of both elements may be in the **%** range (**4% P** and **2% B** are about typical), the resulting glass is called "***BPSG***" (=Bor-Phosphorous Silicate Glass). It "flows" around **850 ^oC**, i.e the viscosity of **BPSG** is then low enough to allow the surface tension to reduce the surface areas by evening out peaks and valleys.
  - How much it "flows" can be further influenced by the atmosphere during the annealing: **O_2** or even better, **H_2O** like in wet oxidation, enhances the viscosity and helps to keep the thermal budget down
  - The **BPSG** process was a key process to ***VLSI*** (=**Very Large Scale Integration**), this can be seen in any cross section of a real device. Lets look at the cross section of the **16 Mbit DRAM** again that was shown before:



- Two layers of **BPSG** are partially indicated in green
  - The lower layer has been etched back to some extent; it only fills some deep crevices in some places.
  - Both layers smoothed the topography considerably; but there can never be complete planarization with **BPSG** glasses, of course.
- How do we carry out an oxide **CVD** process? Of course, we could use a "*tool*" like an epi-reactor, but that would be an overkill (especially in terms of money).
- For "simple" oxide **CVD**, we simply use a furnace as in the thermal oxidation process and admit the process gases instead of oxygen. However, there are certain (costly) adjustments to make:
  - **CVD** processes mostly need to be run at *low pressure* (then often abbreviated **LPCVD**) - some **mbar** will be fine - to ensure that the layers grow smoothly and that the gas molecules are able to penetrate into every nook and cranny (the mean free path length must be large). The furnace tube thus must be vacuum tight and big pumps are needed to keep the pressure low.
  - We want to *minimize the waste* (dangerous gases not used up) which at the same time maximizes the conversion of the (costly) gases to SiO_2. But this means that at the end of the tube the partial pressure of the process gases is lower than at the beginning (most of it has been used up by then). To ensure the same layer thickness for the last wafer than for the first one, requires a higher temperature at the end of the furnace tube because that leads to a higher reaction rate countering the lower gas concentration.
  - The first wafers to be exposed to the gas flow are "air-cooled" by the process gas to some extent. We therefore need to raise the temperature a bit at the front end of the furnace.
- Essentially, we must be able to run a *defined temperature gradient* along the **CVD** furnace tube! This calls for at least three sets of heating coils which must be independently controlled.
  - The whole thing looks like this



- Again, we see that there are many "buttons" to adjust for a "simple" **CVD** oxide deposition.
  - Base pressure and temperature, flow rates of the gases, temperature profile of the furnace with the necessary power profile (which changes if a gas flow is changed), ramping up and ramping down the temperature, etc., - all must be right to ensure constant thickness of the deposited layer for every wafer with minimum waste of gases.
  - Changing any parameter may not only change the oxide thickness, but also its properties (most important, maybe, its etch rate in some etching process).
  - Developing a "new" oxide **CVD** process thus is a lengthy undertaking, demanding much time and ingenuity. But since this is true for every process in microelectronics, we will from now on no longer emphasize this point.
- **CVD** furnaces have a major disadvantage: Everything inside (including the quartz tube) is hot and will become covered with oxide (including the wafer back sides). This is not quite so bad, because the quartz tube will simply grow in thickness. Nevertheless, in regular intervals everything has to be **cleaned** - in special equipment inside the cleanroom! Very annoying, troublesome and costly!
- A "conventional" **CVD** furnace is, however, not the only way to make **CVD** oxides. Several dedicated machines have been developed just for **BPSG** or other variants of oxides.

- One kind, adding also something new to the process, merits a short paragraph: ***PECVD*** or "**Plasma Enhanced CVD**"

Plasma Enhanced CVD

- As the thermal budget gets more and more constrained while more and more layers need to be added for multi-layer metallization, we want to come down with the temperature for the oxide (or other) **CVD** processes.
  - One way for doing this is to supply the *necessary energy for the chemical reaction* not by heating everything evenly, but just the gas. The way to do this is to pump electrical energy into the gas by exposing it to a suitable electrical field at high frequencies. This could induce "dielectric losses", but more important is the *direct energy transfer* by collisions as soon as the *plasma* stage is reached.
  - In a gas plasma, the atoms are ionized and both free electrons and ions are accelerated in the electrical field, and thus gain energy which equilibrates by collisions. However, while the average kinetic energy and thus the temperature of the heavy ions is hardly affected, it is quite different for the electrons: Their temperature as a measure of their kinetic energy may attain **20.000 K**.
  - (If you have problems with the concept of *two* distinctly different temperatures for *one* material - you're doing fine. Temperature is an *equilibrium* property, which we do not have in the kind of plasma produced here. Still, in an approximation, one can consider the electrons and the ions being in equilibrium with the other electrons and ions, respectively, but not among the different species, and assign a temperature to each subgroup separately.)
  - The chemical reactions thus may occur at low nominal temperatures of just a few **100 ^{o}C**.
- There are many kinds of **PECVD** reactors, with **HF** frequencies from **50 kHz to >10 MHz** and electrical power of several **100 W** (not to be sneered at in the **MHz** range!).
  - Since after the first **Al** deposition, the temperature has to be kept below about **400 ^{o}C**, (otherwise a **Si - Al** eutectic will form), **PECVD** oxide is the material of choice from now on, rivaled to some extent by spin-on glass.
  - However, its properties are not quite as good as those of regular **CVD** oxide (which in turn is inferior to thermal oxide).

Footnote:	There are certain names used for the "hardware" needed to make chips that are not immediately obvious to the uninitiated:
	Simple people - e.g. you and me or operators - may talk of "*machinery*" or "*machines*" - which is what those big boxes really are.
	More sophisticated folks - process engineers or scientists - talk about their "*equipment*"
	Really sophisticated people - managers and **CEOs** - will contemplate the "*tools*" needed to make chips.

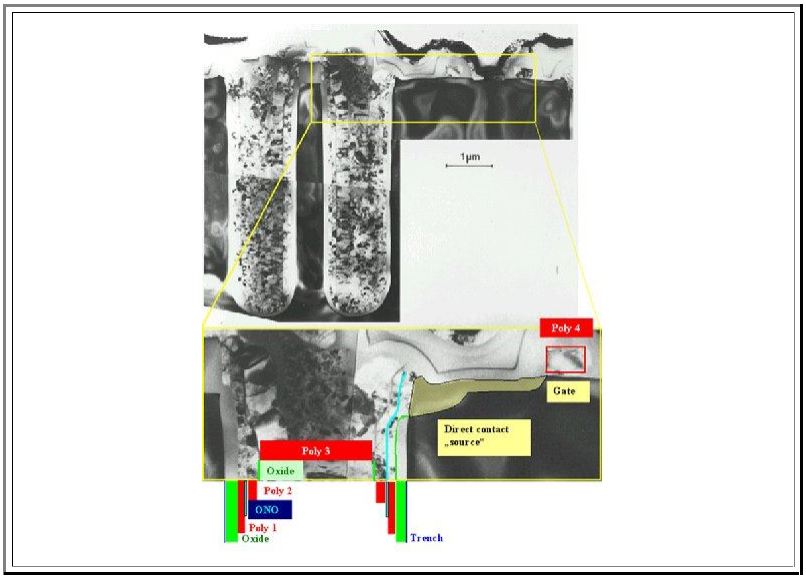
6.2.3 CVD for Poly-Silicon, Silicon Nitride and Miscellaneous Materials

Poly Silicon CVD

- If we were to use an epitaxial reactor for wafers covered with oxide, a layer of **Si** would still be deposited on the hot surface - but now it would have no "guidance" for its orientation, and poly-crystalline **Si** layers (often just called "**poly**" or "polysilicon") would result.
  - **Poly-Si** is one of the key materials in microelectronics, and we know already how to make it: Use a **CVD** reactor and run a process similar to epitaxy.
  - If doping is required (it often is), admit the proper amounts of dopant gases.
- However, we also want to do it cheap, and since it we want a polycrystalline layer, we don't have to pull all the strings to avoid crystal lattice defects like for epitaxial **Si** layers.
  - We use a more simple **CVD** reactor of the furnace type shown for oxide **CVD**, and we employ smaller temperatures (and low pressure, e.g. **60 Pa** since we only need thin layers and can afford lower deposition rates). This allows to use SiH_4 instead of$SiCl_4$; our process may look like this:

$$SiH_4 \xrightarrow[630^oC]{60\ Pa} Si + 2\ H_2$$

  - Much cheaper! The only (ha ha) problem now is: Cleaning the furnace. Now you have poly-**Si** all over the place; a little bit nastier than SiO_2, but this is something you can live with.
- What is poly-**Si** used for and why it is a key material?
  - Lets look at a **TEM** (=transmission electron microscope) picture of a memory cell (transistor and capacitor) of a **16 Mbit DRAM**. For a larger size picture and additional pictures click here.



- All the speckled looking stuff is poly-**Si**. If you want to know exactly what you are looking at, turn to the drawing of this cross section. We may distinguish **4** layers of poly **Si**:
  - "*Poly **1***" coats the inside of the trench (after its surface has been oxidized for insulation) needed for the capacitor. It is thus one of the "plates" of the capacitor. In the **4 Mbit DRAM** the substrate **Si** was used for this function, but the space charge layer extending into the **Si** if the capacitor is charged became too large for the **16 Mbit DRAM**.
  - The "*Poly **2***" layer is the other "plate" of the capacitor. The **ONO** dielectric in between is so thin that it is practically invisible. You need a ***HRTEM*** - a high resolution transmission electron microscope - to really see it.
  - Now we have a capacitor folded into the trench, but the trench still needs to be filled. Poly-**Si** is the material of choice. In order to insulate it from the poly capacitor plate, we oxidize it to some extent before the "*poly **3***" plug is applied.
- One plate of the capacitor needs to be connected to the source region of the transistor. This is "simply" done by removing the insulating oxide from the inside of the trench in the right place (as indicated).
  - Then we have a *fourth poly layer*, forming the gates of the transistors.
  - And don't forget: there were two sacrificial poly-**Si** layers for the LOCOS process!
- That makes **6** poly-**Si** deposition (that we know of). Why do we like poly-**Si** so much?

- Easy! It is perfectly compatible with single crystalline **Si**. Imagine using something else but poly-**Si** for the plug that fills the trench. If the thermal expansion coefficient of "something else" is not quite close to **Si**, we will have a problem upon cooling down from the deposition temperature.
- No problem with poly. Moreover, we can oxidize it, etch it, dope it, etc. (almost) like single crystalline **Si**. It only has *one* major drawback: Its conductivity is not nearly as good as we would want it to be. That is the reason why you often find the poly-gates (automatically forming one level of wiring) "re-enforced" with a **silicide** layer on top.
- A silicide is a metal silicon compound, e.g. **Mo_2Si**, **PtSi**, or **Ti_2Si**, with an almost metallic conductivity that stays relatively inert at high temperatures (in contrast to pure metals which react with **Si** to form a silicide). The resulting double layer is in the somewhat careless slang of microelectronics often called a "*polycide*" (its precise grammatical meaning would be the killing of the poly - as in fratricide or infanticide).
- Why don't we use a silicide right away, but only in conjunction with poly-**Si**? Because you would loose the all-important high quality interface of (poly)-**Si** and **SiO_2**!

Si_3N_4 Deposition

- We have seen several uses for silicon nitride layers - we had **LOCOS**, **FOBIC** (and there are more), so we need a process to deposit **Si_3N_4** .
  - Why don't we just "nitride" the **Si**, analogous to oxidations, by heating the **Si** in a **N_2** environment? Actually we do - on occasion. But **Si_3N_4** is so impenetrable to almost everything - including nitrogen - that the reaction stops after a few **nm**. There is simply no way to grow a "thick" nitride layer thermally.
  - Also, don't forget: **Si_3N_4** is always producing tremendous stress, and you don't want to have it directly on the **Si** without a buffer oxide in between. In other words: We need a **CVD** process for nitride.
- Well, it becomes boring now:
  - Take your **CVD** furnace from before, and use a suitable reaction, e.g.
    .

$$3\ SiH_2Cl_2 + 4NH_3 \Rightarrow Si_3N_4 + 2HCl + 1{,}5\ H_2 \quad (\approx 700\,^{o}C))$$

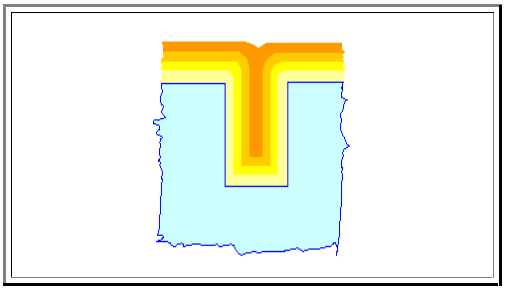
  - Nothing to it - except the cleaning bit. And the mix of hot ammonia (**NH_3**) and **HCl** occurring simultaneously if you don't watch out. And the waste disposal. And the problem that the layers, being under internal stresses, might crack upon cooling down. And, - well, you get it!

Tungsten CVD

- For reasons that we will explain later, it became necessary at the end of the eighties, to deposit a metal layer by **CVD** methods. Everybody would have loved to do this with **Al** - but there is no good **CVD** process for **Al**; nor for most other metals. The candidate of choice - mostly by default - is **tungsten** (chemical symbol ***W*** for "Wolfram").
  - Ironically, **W-CVD** comes straight form nuclear power technology. High purity **Uranium** (chemical symbol ***U***) is made by a **CVD** process not unlike the **Si** Siemens process using **UF_6** as the gas that decomposes at high temperature.
  - W is chemically very similar to **U**, so we use **WF_6** for **W-CVD**.
- A **CVD** furnace, however, is not good enough anymore. **W-CVD** needed its own equipment, painfully (and expensively) developed a decade ago.
  - We will not go into details, however. **CVD** methods, although quite universally summarily described here, are all rather specialized and the furnace type reactor referred to here, is more an exception than the rule.

Advantages and Limits of CVD Processes

- **CVD** processes are ideally suited for depositing thin layers of materials on some substrate. In contrast to some other deposition processes which we will encounter later, **CVD** layers always follow the contours of the substrate: They are conformal to the substrate as shown below.



- Of course, conformal deposition depends on many parameters. Particularly important is which process dominates the reaction:
  - *Transport controlled process* (in the gas phase). This means that the rate at which gas molecules arrive at the surface controls how fast things happen. This implies that molecules react immediately wherever they happen to reach the hot surface. This condition is always favored if the *pressure is low enough*.
  - *Reaction controlled kinetics*. Here a molecule may hit and leave the surface many times before it finally reacts. This reaction is dominating at high pressures.
- Controlling the partial pressure of the reactants therefore is a main process variable which can be used to adjust layer properties.
  - It is therefore common to distinguish between ***APCVD*** (**=** atmospheric pressure **CVD**) and ***LPCVD*** (**=** low pressure **CVD**).
  - **LPCVD**, very generally speaking, produces "better" layers. The deposition rates, however, are naturally much lower than with **APCVD**.
- **CVD** deposition techniques, though quite universal and absolutely essential, have certain *disadvantages*, too. The two most important ones (and the only ones we will address here) are
  - They are not possible for some materials; there simply is no suitable chemical reaction.
  - They are generally not suitable for *mixtures* of materials.
- To give just one example: The metallization layers for many years were (and mostly still are) made from **Al** - with precise additions of **Cu** and **Si** in the **0,3% - 1%** range
  - There is no suitable **Al**-compound that decomposes easily at (relatively low) temperatures. This is not to say that there is none, but all **Al**-organic chemicals known are too dangerous to use, to expensive, or for other reasons never made it to production (people tried, though).
  - And even if there would be some **Al CVD** process, there is simply no way at all to incorporate **Si** and **Cu** in the exact quantities needed into an **Al CVD** layer (at least nobody has demonstrated it so far).
- Many other materials, most notably perhaps the silicides, suffer from similar problems with respect to **CVD**. We thus need alternative layer deposition techniques; this will be the subject of the next subchapter.

Footnote:	The name "*Poly Silicon*" is used for at least three qualitatively very different kinds of materials:
	1. The *"raw material" for crystal growth*, coming from the "Siemens" **CVD** process. It comes - after breaking up the rods - in large chunks suitable for filling the crucible of a crystal grower.
	2. Large ingot of cast **Si** and the *thin sheets made from them*; exclusively used for solar cells. Since the grains are very large in this case (in the cm range), this material is often referred to as "**multi crystalline Si**".
	3. The *thin layers of poly **Si*** addressed in this sub-chapter, used for micro electronics and micro mechanical technologies. Grain sizes then are µm or less.
	In addition, the term poly **Si** might be used (but rarely is) for the *dirty stuff coming out of the **Si** smelters*, since this **MG-Si** is certainly poly-crystalline

6.2.4 Summary to: 6.2 Chemical Vapor Deposition

- Chemical Vapor Deposition (CVD) is simple in principle
  - Find to gases that react ot the desired material at elevated temperatures
  - Put your wafer(s) into some machine, evacuate, heat to the desired temperature (preferably only the wafers) and admit the gases (and remove undesired reaction products).
  - There are many quute different technical ways (all of them expensive) to realize a **CVD** apparatus

- Major **CVD** process are
- Deposition of epitaxial **Si** layers - obviously always on (atomically clean) **Si** substrates. By admitteing some gases carrying doping atoms (e.g. **AsH_3, AsH_3**) the layer can be doped in-situ.
- Depositon of poly crystalline **Si** layers.
  - Chemically similar to epitaxial layers, in reality quite different because the CVD reactos can be simpler.
  - Poly-**Si** is needed for many uses: Gate electrode, interconnect, filling of holes, sacrifical layer.
  - Its great advantage is its full compatibility with **Si** and **SiO_2**; its great disadvantage is its mediocre conductivity (for heavy doping).
- Deposition of **Si_3N_4**
  - Very important. Always prone to produce mechanical stress (**Si_3N_4** is an unyielding ceramic!).
- **W** (and Silicides, and ...)
  - Not "good" processes, but sometimes unavoidable!

Epitaxial Si layer

$$SiCl_4 + 2\,H_2 \Rightarrow Si + 4\,HCl \quad (1000\,^oC - 1200\,^oC)$$

Polycrystalline Si layer

$$SiH_4 \xrightarrow[630^oC]{60\ Pa} Si + 2\,H_2$$

SiO_2 layer ("TEOS process)

$$Si(C_2H_5O)_4 \Rightarrow SiO_2 + 2H_2O + C_2H_4 \quad (720\,^oC)$$

Si_3N_4 layer

$$3\,SiH_2Cl_2 + 4NH_3 \Rightarrow Si_3N_4 + 2HCl + 1{,}5\,H_2 \quad (\approx 700\,^oC))$$

W layer

$$WF_6 + H_2 \xrightarrow[470^oC]{10^4\ Pa} W + \text{(nasty) gases}$$

Questionaire

Multiple Choice questions to all of 6.2

6.3. Physical Processes for Layer Deposition

- The technologies discussed in this subchapter essentially cover *physical processes* for layer deposition as opposed to the more *chemical* methods introduced in the preceding subchapters. In other words, we deal with some more techniques for the material module in the process cycle for **ICs**.
  - While still intimately tied to the electronic materials to be processed, these technologies are a bit more of a side issue in the context of electronic materials, and will therefore be covered in *far less detail* then the preceding, more material oriented deposition techniques.
  - On occasion, however, a particular tough problem in **IC** processing is elucidated in the context of the particular deposition method associated with it. *So don't skip these modules completely!*
- Essentially, what you should know are the basic technologies employed for layer deposition, and some of the major problems, advantages and disadvantages encountered with these techniques in the context of chip manufacture.

6.3.1 Sputter Deposition and Contact Hole Filling

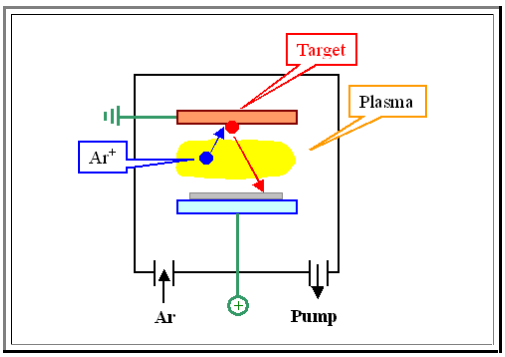
General Remarks

- It should be clear by now that the *deposition of thin layers* is the key to all microelectronic structures (not to mention the emerging micro electronic and mechanical systems (**MEMS**), or **nano technology**).
  - *Chemical* vapor deposition, while very prominent and useful, has severe limitations and the more alternative methods exist, the better.
  - *Physical* methods for controlled deposition of thin layers do exist too; and in the remainder of his subchapter we will discuss the major ones.
- What are *physical* methods as opposed to *chemical* methods? While there is no ironclad distinction, we may simply use the following rules
  - If the material of the layer is produced by a *chemical reaction* between some primary substances in-situ, we have a chemical deposition process. This does not just include the **CVD** processes covered before, but also e.g. galvanic layer deposition.
  - If the material forming the layer is sort of transferred from some substrate or source to the material to be coated, we have a *physical process*. The most important physical processes for layer deposition which shall be treated here are

    - **Sputtering techniques**
    - **Ion implantation**
    - **Spin on coating**

Basic Sputter Process

- "**Sputtering**" or "**sputter deposition**" is a conceptually simple technique:
  - A "**target**" made of the material to be deposited is bombarded by energetic ions which will dislodge atomes of the target, i.e., "***sputter*** them off".
  - The dislodged atoms will have substantial kinetic energies, and some will fly to the substrate to be coated and stick there.
- In practice, this is a lot easier than it appears. The basic set-up is shown below:



- The ions necessary for the bombardment of the target are simply extracted from an **Ar plasma** burning between the target and the substrate.

- Both target and substrate are planar plates arranged as shown above. They also serve as the cathode and anode for the gas discharge that produces an **Ar** plasma, i.e. ionized **Ar** and free electrons, quite similar to what is going on in a fluorescent light tube.
- Since the target electrode is always the cathode, i.e. negatively charged, it will attract the $\mathbf{Ar^+}$ ions and thus is bombarded by a (hopefully) constant flux of relatively energetic **Ar** ions.
- This ion bombardment will liberate atoms from the target which issue forth from it in all directions.

Reality is much more complex, of course. There are many ways of generating the plasma and tricks to increase the deposition rate. Time is money, after all, in semiconductor processing.

- Some short comments about sputtering technologies might be provided in an "advanceed" module in the future.

Some target atoms will make it to the substrate to be coated, others will miss it, and some will become ionized and return to the target. The important points for the atoms that make it to the substrate (if everything is working right) are:

- **1.** The target atoms hit the substrate with an energy large enough so they "*get stuck*", but not so large as to liberate substrate atoms. Sputtered layers therefore usually stick well to the substrate (in contrast to other techniques, most notably evaporation).
- **2.** *All* atoms of the target will become deposited, in pretty much the same composition as in the target. It is thus possible, e.g., to deposit a silicide slightly off the stoichiometric composition (advantageous for all kinds of reason). In other words, if you need to deposit e.g. $\mathbf{TaSi_{2-x}}$ with $\mathbf{x \approx 0.01 - 0.1}$, sputtering is the way to do it because it is comparatively easy to change the target composition.
- **3.** The target atoms hit the substrate coming from *all directions*. In a good approximation, the flux of atoms leaving the target at an angle Φ relative to the normal on the target is proportional to **cos** Φ. This has profound implications for the coverage of topographic structures.
- **4.** Homogeneous coverage of the substrate is relatively easy to achieve- just make the substrate holder and the target big enough. The process is also relatively easily scaled to larger size substrates - simply make everything bigger.

Of course, there are problems, too.

- Sputtered layers usually have a very bad crystallinity - very small grains full of defects or even amorphous layers result. Usually some kind of annealing of the layers is necessary to restore acceptable crystal quality.
- Sputtering works well for metals or other somewhat conducting materials. It is not easy or simply impossible for insulators. Sputtering $\mathbf{SiO_2}$ layers, e.g., has been tried often, but never made it to production.
- While the **cos** Φ relation for the directions of the sputtered atoms is great for over-all homogeneity of the layers, it will prevent the filling of holes with large **aspect ratios** (aspect ratio **=** depth/width of a hole). Since contact holes and vias in modern **ICs** always have large aspect ratios, a serious problem with sputtering A**l(Si Cu)** for contacts came up in the nineties of the last century. This is elaborated in more detail below.

More or less by default, sputtering is the layer deposition process of choice for **Al**, the prime material for metallization.

- How else would you do it? Think about it. We already ruled out **CVD** methods. What is left?
- The deposition of a metallization layer on a substrate with "heavy" topography - look at some of the drawings and pictures to understand this problem - is one of the big challenges **IC** technology and a particularly useful topic to illustrate the differences between the various deposition technologies; it will be given some special attention below.

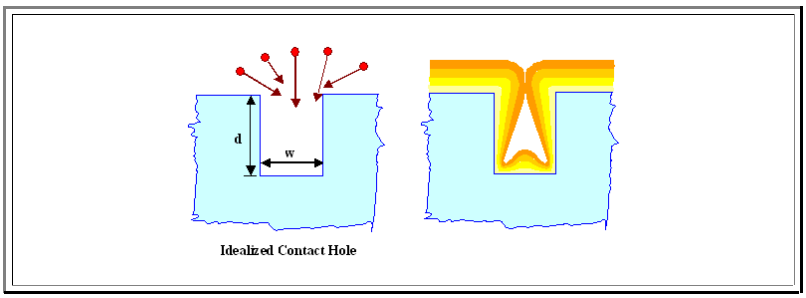
The Contact Hole Problem

The metallization of chips for some **30** years was done with Aluminum as we know by now - cf. all the drawings in chapter 5 and the link.

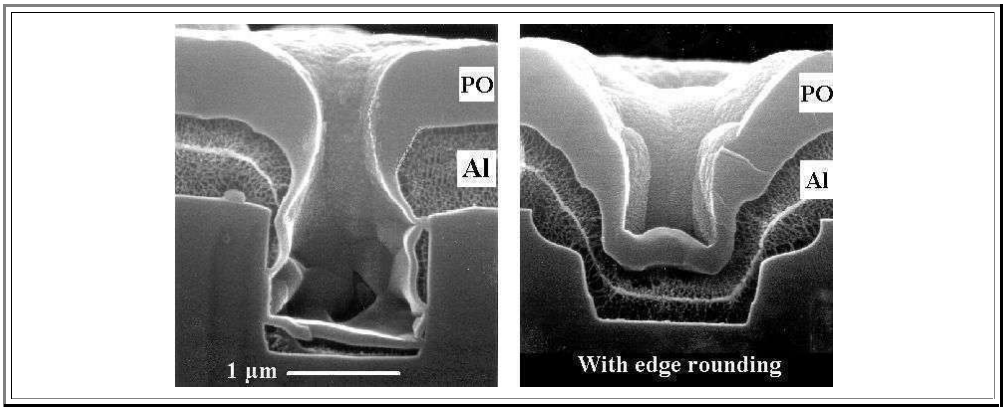
- **Al**, while far from being optimal, had the best over-all properties, including less than about **0,5 % Si** and often a little bit (roughly **1 %**) of **Cu**, **V** or **Ti**. These elements are added in order to avoid deadly "spikes", to decrease the contact resistance by avoiding epitaxial Si precipitates and to make the metallization more resistant to electromigration.
- While you do not have to know what that means (you might, however, look it up via the links), you should be aware that there are even more requirements for a metallization material than listed before.

Sputtering is the only process that can deposit an **Al** layer with a precisely determined addition of some other elements on a large **Si** substrate. There is an unavoidable problem however, that becomes more severe as features get smaller, related to the so-called "edge coverage" of deposition processes.

- Many **Al** atoms hitting the substrate under an oblique angle, will not be able to reach the bottom of a contact hole which thus will have less Al deposited as the substrate surface.
- To make it even worse, the layer at the edge of the contact hole tends to be especially thick, reducing the opening of the hole disproportionately and thus allowing even less **Al** atoms to bottom of the hole. What happens is illustrated below.



- Fort aspect ratios ***A=w/d*** smaller than approximately **1**, the layer at the edge of the contact hole will become unacceptably thin or will even be non-existing - the contact to the underlying structure is not established.
  - The real thing together with one way to overcome the problem is shown below (the example is from the **4Mbit DRAM** generation, around **1988**).



  - ***PO*** denotes "**Plasma oxide**", referring to a PECVD deposition technique for an oxide. This layer is needed for preparation purposes (the upper **Al** edge would otherwise not be clearly visible)
- Clearly, the **Al** layer is interrupted in the contact hole on the left. **Al** sputter deposition cannot be used anymore without "tricks".
  - One possible trick is shown on the right: The edges of the contact hole were "rounded". "**Edge rounding**", while not easy to do and consuming some valuable "real estate" on the chip, saved the day for the at or just below the **1µm** design rules.
  - But eventually, the end of sputtering for the **1st** metallization layer was unavoidable - despite valiant efforts of sputter equipment companies and **IC** manufacturers to come up with some modified and better processes - and a totally new technology was necessary
- This was **Tungsten CVD** just for filling the contact hole with a metal.
  - In some added process modules, the wafer was first covered with tungsten until the contact holes were filled (cf. the drawing in the **CVD** module). After that, the tungsten on the substrate was polished off so that only filled contact holes remained.
  - After that, **Al** could be deposited as before.
- However, depositing **W** directly on **Si** produced some new problems; related to interdiffusion of **Si** and **W**.
  - The solution was to have an intermediary **diffusion barrier** layer (which was, for different reasons, already employed in some cases with a traditional **Al** metallization).
  - Often, this diffusion barrier layer consisted of a thin **$TiSi_2$/Ti /TiN** layer sequence. The **$TiSi_2$** formed directly as soon as **Ti** was deposited (by sputtering, which was still good enough for a very thin coating), the **Titanium Nitride** was formed by a **reactive sputtering process**.
  - Reactive sputtering in this case simply means that some **N_2** was admitted into the sputter chamber which reacts immediately with freshly formed (and extremely reactive) **Ti** to **TiN**.
- A typical contact to lets say **p**-type **Si** now consisted of a **p-Si/p^+-Si/$TiSi_2$/Ti/TiN/W/Al** stack, which opened a new can of worms with regard to contact reliability. Just imagine the many possibilities of forming all kinds of compounds by interdiffusion of whatever over the years.
  - But here we stop. Simply because meanwhile (i.e. **2001**), contacts are even more complicated, employing **Cu** (deposited galvanically after a thin **Cu** layer necessary for electrical contact has been sputter deposited), various barrier layers, possibly still **W**, and whatnot.
  - So: Do look at a modern chip with some awe *and remember*: We are talking electronic materials here!

Questionaire

Multiple Choice questions to 6.4.1

6.3.2 Ion Implantation

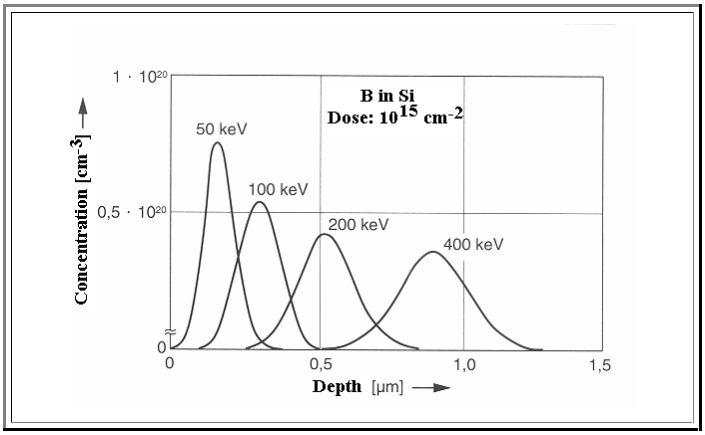
Ion Implantation Basics

What is **ion implantation**, often abbreviated ***I²*** ? The name tells it succinctly: *Ions* of some material - almost always the dopants **As**, **B**, **P** - are *implanted*, i.e. shot into the substrate.

Ion implantation may be counted among layer deposition processes because you definitely produce a layer of something different from the substrate even so you do not deposit something in the strict meaning of the term.

How is it done? Obviously you need an ion beam, characterized by three basic parameters:

- **1.** The *kind of the ions*. Almost everything from the periodic table could be implanted, but in practice you will find that only **Sb** (as dopant) and occasionally **Ge** and **O** are being used besides the common dopants **As**, **B**, and **P**.
- **2**. The *energy of the ions* in **eV**. This is directly given by the accelerating voltage employed and is somewhere in the range of **(2 - 200) kV**, always allowing for extremes in both directions for special applications. The energy of the ion together with its mass determine how far it will be shot into a **Si** substrate. The following graph gives an idea of the distribution of **B** atoms after implantation with various energies. The curves for **As** or **P** would be similar, but with peaks at smaller depth owing to the larger mass of these atoms.



- There are several interesting points to this graph: Obviously everything happens at dimensions ≤ **1 µm**, and a **dose** ***D*** of 10^{15} **cm**$^{-2}$ gives a pretty high doping density in the peak value. Moreover, by changing the implantation energy, all kinds of concentration profiles could be produced (within limits). That is completely impossible by just diffusing the dopant into the **Si** from the outside.
- **3.** The *flux* (number of ions per **cm²** and second), i.e. the current (in **µA** or **mA**) carried by the ion beam. In order to obtain some rough idea about the current range, we assume a certain beam cross section ***A*** and a constant current density ***j*** in this cross section. A total current ***I*** then corresponds to a current density $j = I/A$ and the implanted dose is

$$D = \frac{j \cdot t}{e} = \frac{I \cdot t}{e \cdot A}$$

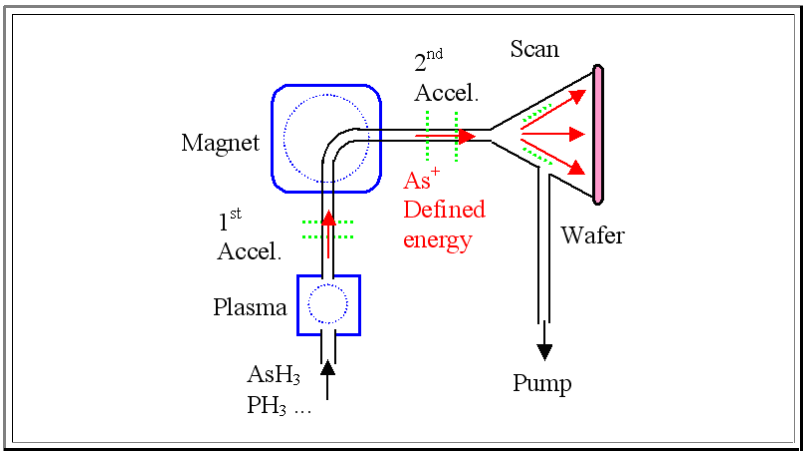
- ***t*** is the implantation time, **e** the elementary charge **= 1,6 · 10^{-19}C**.

For an implantation time of **1 s** for an area $A = 1$ **cm²** and a dose $D = 10^{15}$ **cm**$^{-2}$, we obtain for the beam current

$$I = \frac{D \cdot e \cdot A}{t} = 10^{15} \cdot 1{,}6 \cdot 10^{-19}\ \mathrm{C \cdot s^{-1}} = 1{,}6^{-4}\ \mathrm{A}$$

- Total implantation time for *one* **200 mm** wafer than would be about **300 s**, far too long. In other words, we need implanters capable to deliver beam currents of **10 mA** and more for high doses, and just a few precisely controlled **µA** for low doses.
- Think a minute to consider what that means: **10 mA** in **1 cm^{-2}** at **200 kV** gives a deposited power of **2 kW** on **1 cm^{-2}**. That is three orders of magnitude larger than what your electric range at home has to offer - how do you keep the **Si** cool during implantation? If you do nothing, it will melt practically instantaneously.

- Since the beam diameter is usually much smaller than the **Si** wafer, the ion beam must be scanned over the wafer surface by some means. For simplicities sake we will dismiss the scanning procedure, even so it is quite difficult and expensive to achieve with the required homogeneity. The same is true for all the other components - ion beam formation, accelaration, beamshaping, and so on.
  - If we take everything together, we start to see why ion implanters are very large, very complicated, and very expensive (several million **$)** machines. Their technology, while ingenious and fascinating, shall not concern us here, however, besides giving an extremely simplifed schematic diagram.



- Why do we employ costly and complex ion implantation?
  - Because there simply is no other way to dope selected areas of a **Si** substrate with a precisely determined amount of some dopant atoms and a controlled concentration profile.
  - In our simple drawing of a **CMOS** structure we already have three doped regions (in reality there are much more). Just ask yourself: How do you get the dopants to their places?
- With ion implantation it is "easy":
  - Mask with some layer (usually **SiO_2** or photo resist) that absorbs the ions, and shoot whatever you need into the open areas.
- The only alternative (used in the stone age of **IC** semiconductor technology) is to use **diffusion** from some outside source.
  - Again, mask every area where you do not want the dopants with some layer that is impenetrable for the atoms supposed to diffuse, and expose the substrate to some gas containing the desired atoms at high temperature. After some time, some atoms will have diffused into the **Si** - you have your doping.
  - But there are so many problems that direct diffusion is not used anymore for complex **ICs**: Accuracy is not very good, profiles are limited, the necessary high temperatures change the profiles already established before, and so on. Simply forget it. Ion implantation is what you do.
- But like everything (and everybody), implantation has it limits.
  - For example: How do you dope around the trench shown before in the context of integrated capacitors? Obviously, you can't just shoot ions into the side wall. Or can you? Think about how you would do it and then turn to the advanced module.

Defects and Annealing

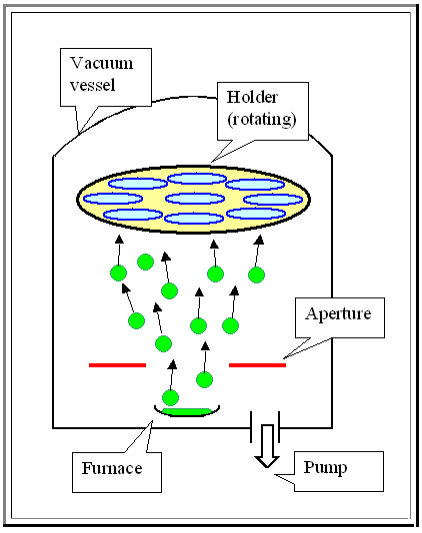
- After implanting the ions of your choice with the proper dose and depth distribution, you are not yet done.
  - Implantation is a violent process. The high energy ion transfers its energy bit by bit to lattice atoms and thus produces a large number of defects, e.g. vacancies and interstitials. Often the lattice is simply killed and the implanted layer is **amorphous**. This is shown in an illustration module.
  - You must restore order again. Not only are **Si** crystal lattice defects generally not so good for your device, but only dopant atoms, which have become neatly incorporated as substitutional impurities, will be electrically active.
- Implantation, in short, must always be followed by an annealing process which hopefully will restore a perfect crystal lattice and "activate" the implanted atoms.
  - How long you have to anneal at what temperature is a function of what and how you implanted. It is a necessary evil, because during the annealing the dopants will always diffuse and your neat implanted profiles are changing.
  - Much research has been directed to optimal annealing procedures. It might even be advantageous to anneal for very short times (about **1 s**) at very high temperatures, say **(1100 - 1200) °C**. Obviously this cannot be done in a regular furnace like the one illustrated for oxidation, and a whole new industry has developed around "**rapid thermal processing**" (***RTP***) equipment.

- Some of the more interesting issues around ion implantation and annealing can be found in future advanced module.

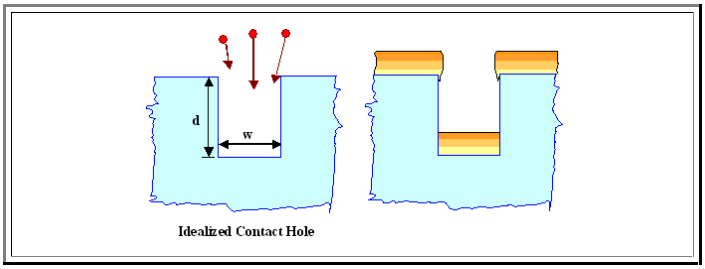
6.3.3 Miscellaneous Techniques and Comparison

Evaporation

- By now you may have wondered why the time-honored and widely used technique of evaporation has not been mentioned in context with **Si** technology.
  - The answer is simple: *It is practically not used*. This is in contrast to other technologies, notably optics, where evaporation techniques played a major role.
  - In consequence, this paragraph shall be kept extremely short. It mainly serves to teach you that there are more deposition techniques than meets the eye (while looking at a chip).
- What is the evaporation technique? If your eye glasses or your windshield ever fogged, you have seen it: Vapor condenses on a cold substrate.



  - What works with water vapor also works with all other vapors, especially metal vapor.
  - All you have to do is to create the vapor of your choice, always inside a vacuum vessel kept at a good vacuum. The (usually) metal atoms will leave the crucible or "boat" with an kinetic energy of a few **eV** and sooner or later will condense on the (cooled) substrate (and everywhere else if you don't take special precautions).
  - Your substrate holder tends to be big, so you can accommodate several wafers at once (Opening up and loading vacuum vessels takes expensive time!)
- The technique is relatively simple (even taking into account that the heating nowadays is routinely done with high power electron beams hitting the material to be evaporated), but has major problems with respect to **IC** production:
  - The atoms are coming from a "point source", i.e. their incidence on the substrate is nearly perpendicular. Our typical contact hole filling problem thus looks like this:

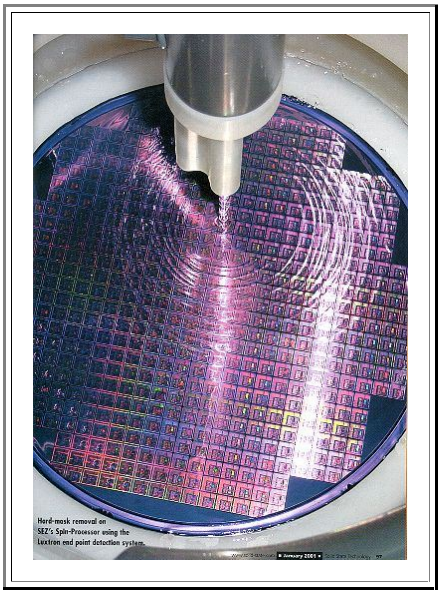


  - In other words: Forget it!
  - It is also clear that it is very difficult to outright impossible to produce layers with arbitrary composition, e.g. **Al** with **0,3% Si** and **0,5% Cu**. You would need three independently operated furnaces to produce the right mix.
- All things considered, sputtering is usually better and evaporation is rarely used nowadays for microelectronics.

Spin-on Techniques

- **Spin-on techniques**, a special variant of so-called *sol-gel techniques*, start with a liquid (and usually rather viscous) source material, that is "painted" on the substrate and subsequently solidified to the material you want

the substrate and subsequently solidified to the material you want.



- The "*painting*" is not done with a brush (although this would be possible), but by spinning the wafer with some specified **rpm** value (typically **5000 rpm**) and dripping some of the liquid on the center of the wafer. Centrifugal forces will distribute the liquid evenly on the wafer and a thin layer (typically around **0,5 µm**) is formed.
- Solidification may occur as with regular paint: the solvent simply evaporates with time. This process might be accelerated by some heating. Alternatively, some chemical reaction might be induced in air, again helped along by some "baking" as it is called.

As a result you obtain a thin layer that is rather smooth - nooks and crannies of the substrate are now planarized to some extent. The film thickness can be precisely controlled by the angular velocity of the spin process (as a function of the temperature dependent viscosity of the liquid).

- Spin-on coating is the technique of choice for producing the light-sensitive **photo resist** necessary for lithography. The liquid resist rather resembles some viscous paint, and the process works very well. It is illustrated on the right.
- Most other materials do not have suitable liquid precursors, the spin-on technique thus can not be used.

A noteworthy exception, however, is **spin-on glass**, a form of SiO_2 mentioned before.

- The liquid consists basically of Silicon-tetra-acetate ($Si(CH_2COOH)_4$) (and some secret additions) dissolved in a solvent. It will solidify to an electronically not-so-good SiO_2 layer around **200 °C**.
- Using spin-on glass is about the only way to fill the interstices between the **Al** lines with a dielectric at low temperatures. The technique thus has been developed to an art, but is rather problematic. The layers tend to crack (due to shrinkage during solidifications), do not adhere very well, and may interact detrimentally with subsequent layers.

A noteworthy example of a material that can be "spun on", but nevertheless did not make it so far are **Polyimides**, i.e. polymers that can "take" relatively high temperatures

- They look like they could be great materials for the intermetal dielectric - low ϵ_r, easy deposition, some planarizing intrinsic to spin-on, etc. They are great materials - but still not in use. If you want to find out why, and how new materials are developed in the real world out there, use this link.

Other Methods

Deposition techniques for thin layers is a rapidly evolving field; new methods are introduced all the time. In the following a couple of other techniques are listed with a few remarks

Molecular Beam Epitaxy (***MBE***). Not unlike evaporation, except that only a few atoms (or molecules) are released from a tricky source (an "effusion cell") at a time.

- **MBE** needs ultra-high vacuum conditions - i.e. it is very expensive and not used in **Si-IC** manufacture. **MBE** can be used to deposit single layers of atoms or molecules, and it is relatively easy to produce multi layer structures in the **1 nm** region. An example of a **Si-Ge** multilayer structure is shown in the link
- **MBE** is the method of choice for producing complicated epitaxial layer systems with different materials as needed, e.g., in advanced optoelectronics or for superconducting devices. An example of what you can produce with **MBE** is shown in the link

Laser Ablation. Not unlike sputtering, except that the atoms of the target are released by hitting it with an intense Laser beam instead of **Ar** ions extracted from a Plasma.

- Used for "sputtering" ceramics or other non conducting materials which cannot be easily sputtered in the conventional way.

Bonding techniques. If you bring two ultraflat **Si** wafers into intimate contact without any particles in between, they will just stick together. With a bit of annealing, they fuse completely and become bonded.

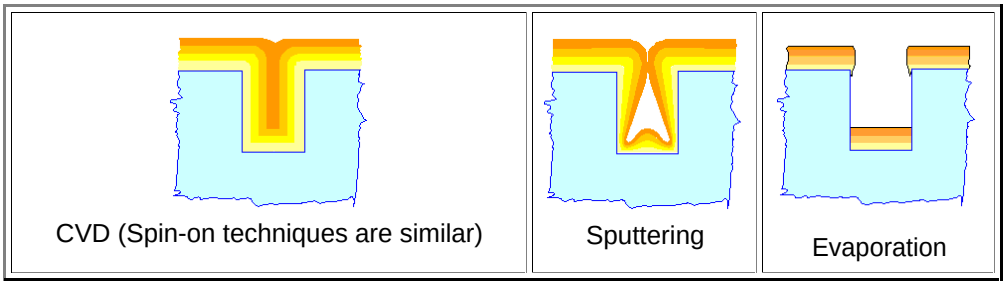
- Glass blowers have done it in a crude way all the time. And of course, in air you do not bond **Si** to **Si**, but SiO_2 to SiO_2. One way to use this for applications is to produce a defined SiO_2 layer first, bond the oxidized wafer to a **Si** wafer, then polish off almost all of the **Si** except for a layer about **1 µm** thick
- Now you have a regular wafer coated with a thin oxide and a perfect single crystalline **Si** layer - a so-called "**silicon on insulator**" (***SOI***) structure. The **Si** industry in principle would love **SOI** wafers - all you have to do to become rich immediately, is to make the process cheap. But that will not be easy. You may want to check why **SOI** is a hot topic, and how a major company is using wafer bonding plus some more neat tricks, including mystifying electrochemistry, to make **SOI** affordable.

- Bonding techniques are rather new; it remains to be seen if they will conquer a niche in the layer deposition market.

- **Galvanic techniques**, i.e. electrochemical deposition of mostly metals. Galvanizing materials is an old technique (think of chromium plated metal, anodized aluminium, etc.) normally used for relatively thick layers.
  - It is a "dirty" process, hard to control, and still counted among the black arts in materials science. No self-respecting **Si** process engineer would even dream of using galvanic techniques - except that with the advent of **Cu** metallization he was not given a choice.
  - Using **Cu** instead of **Al** for chip metallization was unavoidable for chips hitting the market around **1998** and later - the resistivity of the **Al** was too high.
  - As it turned out, established techniques are no good for **Cu** deposition - galvanic deposition is the method of choice. **Cu** metallization calls for techniques completely different from Al metallization - the catchword is "damascene technology". The link takes you there - you may also enjoy this module from the *"Defects" Hyperscript* because it contains some other interesting stuff in the context of (old) materials science.
- And not to forget: Galvanic techniques are also used in the packaging of chips

Comparison of Various Layer Deposition processes

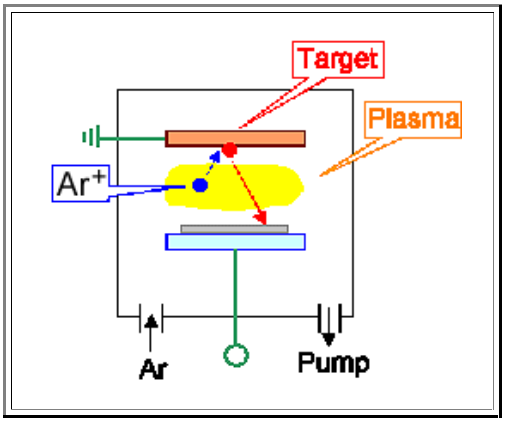
- *First*, lets look at **edge coverage**, i.e. the dependence on layer thickness on the topography of the substrate. This is best compared by looking at the ability to fill a small contact hole with the layer to be deposited.
  - We have the following schematic behavior of the major methods as shown before.



- *Second*, lets look at what you can deposit.
  - **CVD** methods are limited to materials with suitable gaseous precursors. While it is not impossible to deposit mixtures of materials (as done, e.g. with doped poly **Si** or flow glass), it will not generally work for arbitrary compositions.
  - Sputter methods in practice are limited to conducting materials - metals, semiconductors, and the like. Arbitrary mixtures can be deposited; all you have to do is make a suitable target. The target does not even have to be homogeneous; you may simply assemble it by arranging pie-shaped wedges of the necessary materials in the required composition into a "cake" target.
  - Evaporation needs materials that can be melted and vaporized. Many compounds would decompose, and some materials simply do not melt (try it with **C**, e.g.). If you start with a mixture, you get some kind of distillation - you are only going to deposit the material with the highest vapor pressure. Mixtures thus are difficult and can only be produced by co-evaporation from different sources.

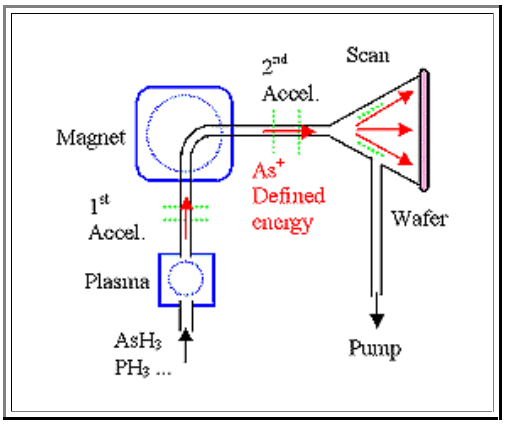
6.3.4 Summary to: 6.3 Physical Processes for Layer Deposition

Sputter deposition



- Plasma technique ⇒ Vacuum + high voltage (and possible high frequency): complicated and expensive
- Layers amorphous to highly defective ⇒ needs usually annealing after deposition.
- Very versatile because of easy control of layer composition by target composition
- Decent depositioen rates possible. Particularly suited to conductors.
- Coverage is *not* conformal!

Ion implantation

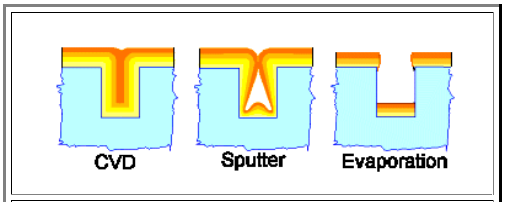


- Depth (**< ca. 1 µm)** and dose precisely controllable.
- Very compley and expensive
- Method od choice for making doped layers.
- Introduces defects or destroys crystallinity ⇒ annealing at high ***T*** (**> 800 °C**) is a must

There are many more techniques for producing thin layers

- **Evaporation**. Relatively simple but limited as to materials and edgencoverage
- **Molecular beam epitaxy**. (**MBE**) Standard for **III-V's**
- **Spin-on techniques** ("Sol- Gel"). Used for making photo resist layers; occasionally for others
- **Galvanics**. Kind of crude but necessary for **Cu** interconnects in modern **IC's**

Edge coverage may be the decisive property!



Comparing edge coverage

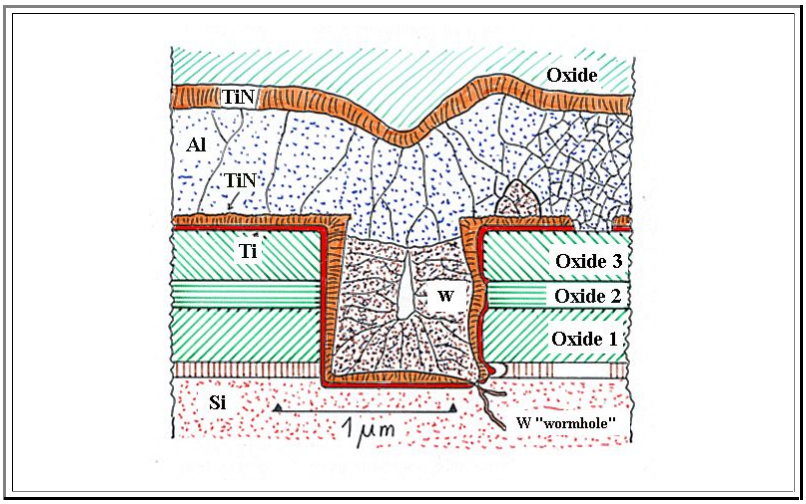
Questionaire

Multiple Choice questions to all of 6.3

6.4 Etching Techniques

6.4.1 General Remarks

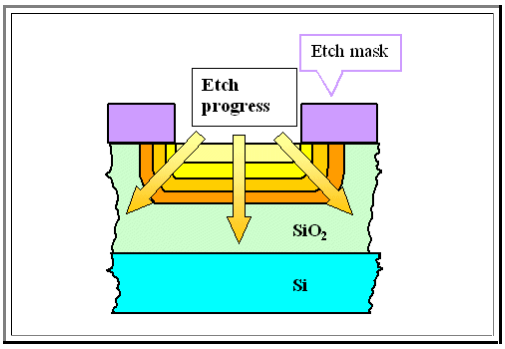
- After we have produced all kinds of layer, we must now proceed to the structure module of our basic process cycle. First we discuss **etching techniques**.
  - Lets see what it means to produce a structure by etching. Lets make, e.g., a contact hole in a somewhat advanced process (and do some of the follow-up processes for clarity).
  - What the stucture contains may look like this:



- Obviously, before you deposit the **Ti/TiN** diffusion barrier layer (and then the **W**, and so on), you must etch a hole through **3** oxide layers and an **Si_3N_4** - and here we don't care why we have all those layers. (The right hand side of the picture shows a few things that can go wrong in the contact making process, but that shall not concern us at present).
- There are some obvious requirements for the etching of this contact hole that also come up for most other etching processes.
  - **1.** You only want to etch *straight down* - not in the lateral direction. In other words, you want strongly **anisotropic etching** that only affects the bottom of the contact hole to be formed, but not the sidewalls (which are, after all, of the same material).
  - **2.** You want to *stop* as soon as you reach the **Si** substrate. Ideally, whatever you do for etching will not affect **Si** (or whatever material you want not to be affected). In other words, you want a large **selectivity** (**=** ratio of etch rates).
  - **3.** You also want reasonable *etching rates* (time is money), the ability to etch through several *different* layers in *one* process (as above), no *damage* of any kind (including rough surfaces) to the layer where you stop, and sometimes *extreme* geometries (e.g. when you etch a trench for a capacitor: **0,8 µm** in diameter and **8 µm** deep) - and you want perfect *homogeneity and reproducibility* all the time (e.g. all the about **200.000.000.000** trenches on *one* **300 mm** wafer containg **256 Mbit DRAMs** must be identical to the ones on the other ***500 - 1000*** wafers you etch today, *and* to the *thousands* you etched before, or are going to etch in the future).
- Lets face it: *This is tough*! There is no single technique that meets all the requirements for all situations.
  - Structure etching thus is almost always a search for the best compromise, and new etching techniques are introduced all the time.
  - Here we can only scratch at the surface and look at the two basic technologies in use: *Chemical or wet etching* and *plasma or dry etching*.

6.5.2 Chemical Etching

- Chemical etching is simple: Find some (liquid) chemical that dissolves the layer to be etched, but that does not react with everything else.
  - Sometimes this works, sometimes it doesn`t. Hydrofluoric acid (**HF**), for example will dissolve **SiO_2**, but not **Si** - so there is an etching solution for etching **SiO_2** with extreme selectivity to **Si**.
  - The other way around does not work: Whatever dissolves **Si**, will always dissolve **SiO_2**, too. At best you may come up with an etchant that shows somewhat different etching rates, i.e. some (poor) selectivity.
- Anyway, the thing to remember is: Chemical etchants, if existing, can provide extremely good selectivity and thus meet our *second* request from above.
  - How about the *first* request, anisotropy? Well, as you guessed: It is rotten, practically non-existent. A chemical etchant always dissolves the material it is in contact with, the forming of a contact hole would look like this:



- There is a simple and painful consequence: As soon as your feature size is about **2 µm** or smaller, *forget chemical structure etching*.
  - Really? How about making the opening in the mask smaller, accounting for the increase in lateral dimensions?
  - You could do that - it would work. But it would be foolish: If you *can* make the opening smaller, you also want your features smaller. In real life, you put up a tremendous effort to make the contact hole opening as small as you can, and you sure like hell don't want to increase it by the structure etching!
- Does that mean that there is no chemical etching in **Si** microelectronics? Far from it. There just is no chemical *structure* etching any more. But there are plenty of opportunities to use chemical etches (cf. the statistics to the **16 Mbit DRAM** process). Lets list a few:
  - *Etching off whole layers*. Be it some sacrificial layer after it fulfilled its purpose, the photo resist, or simply all the **CVD** layers or thermal oxides which are automatically deposited on the wafer backside, too - they all must come off eventually and this is best done by wet chemistry.
  - Etching *coarse structures*, e.g. the opening in some protective layers to the large **Al** pads which are necessary for attaching a wire to the outside world.
  - Etching off unwanted *native oxide* on all **Si** or poly-**Si** layers that were exposed to air for more than about **10 min**.
  - All *cleaning steps* may be considered to be an extreme form of chemical etching. Etching off about **1,8 nm** of native oxide might be considered cleaning, and a cleaning step where nothing is changed at the surface, simply has no effect.
- While these are not the exciting process modules, experienced process engineers know that this is where trouble lurks. Many factories have suffered large losses because the yield was down - due to some problem with wet chemistry.
- A totally new field, just making it into production for some special applications, is **electrochemical etching**. A few amazing (and not yet well understood) things can be done that way; the link provides some samples.

6.4.3 Plasma Etching

Plasma etching, also known as **dry etching** (in contrast to *wet* etching) is the universal tool for structure etching since about **1985**. In contrast to all other techniques around chip manufacture, which existed in some form or other *before* the advent of microelectronics, plasma etching was practically unknown before **1980** and outside the microelectronic community.

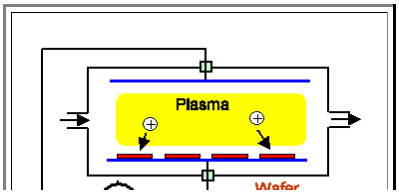
- What is Plasma etching? In the most simple way of looking at it, you just replace a liquid etchant by a plasma. The basic set-up is not unlike sputtering, where you not only deposit a layer, but *etch the target* at the same time.
- So what you have to do is to somehow produce a plasma of the right kind between some electrode and the wafer to be etched. If all parameters are right, your wafer might get etched the way you want it to happen.

If we naively compare chemical etching and plasma etching for the same materials to be etched - lets take **SiO_2** - we note major differences:

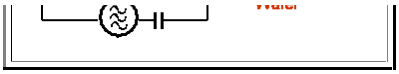
Chemical etching of **SiO_2**	*Plasma etching* of **SiO_2**
Etchant: **HF + H_2O** (for etching **SiO_2**.	*Gases:* **CF_4 + H_2** (or almost any other gas containing **F**).
Species in solution:: **F^-**, **HF^-**, **$H^+SiO_4^{2-}$**, **SiF_4**, **O_2** - whatever chemical reactions and dissociation produces.	*Species in plasma and on wafer:* **CF_x^+ ($x \leq 3$)**, and all kinds of unstable species not existent in wet chemistry. Carbon based polymers, produced in the plasma which may be deposited on parts of the wafer.
Basic processes: **SiO_2** dissolves	Etching of **SiO_2**, formation of polymers, deposition of polymers (and other stuff) and etching of the deposited stuff, occurs simultaneously
Driving force for reactions: Only "chemistry", i.e. reaction enthalpies or chemical potentials of the possible reactions; essentially equilibrium thermodynamics	*Driving force for reactions:* "Chemistry", i.e. reaction enthalpies or chemical potentials of the possible reactions, including the ones never observed for wet chemistry, near equilibrium, *and* non-equilibrium physical processes", i.e. mechanical ablation of atoms by ions with high energies.
Energy for kinetics: Thermal energy only, i.e. in the **1 eV** range	*Energy for kinetics:* Thermal energy, but also kinetic energy of ions obtained in an electrical field. High energies (several **eV** to hundreds of **eV**) are possible.
Anisotropy: None; except some possible **{hkl}** dependence of the etch rate in crystals.	*Anisotropy:* Two major mechanisms **1.** Ions may have a preferred direction of incidence on the wafer. **2.** Sidewalls may become protected through preferred deposition of e.g. polymers Completely isotropic etching is also possible
Selectivity: Often extremely good	*Selectivity:* Good for the chemical component, rather bad for the physical component of the etching mechanism. Total effect is open to optimization.

If that looks complicated, if not utterly confusing - that's because it is (and you thought just chemistry by itself is bad enough).

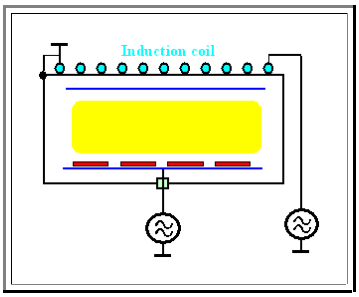
- Plasma etching still has a strong black art component, even so a lot of sound knowledge has been accumulated during the last **20** years.
- It exists in countless variants, even for just *one* material.



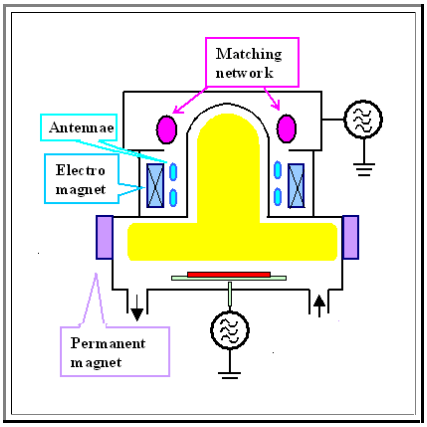
- The many degrees of freedom (all kind of gases, pressure and gas flux, plasma production, energy spread of the ions, ...), or more prosaically, the many buttons that you can turn, make process development tedious on the one hand, but allow optimization on the other hand.



- The two perhaps most essential parameters are: **1.** the relative strength of chemical to physical etching, and **2.** the deposition of polymers or other layers on the wafer, preferably on the sidewalls for protection against lateral etching.
  - The physical part provides the absolutely necessary anisotropy, but lacks selectivity
  - The chemical part provides selectivity.



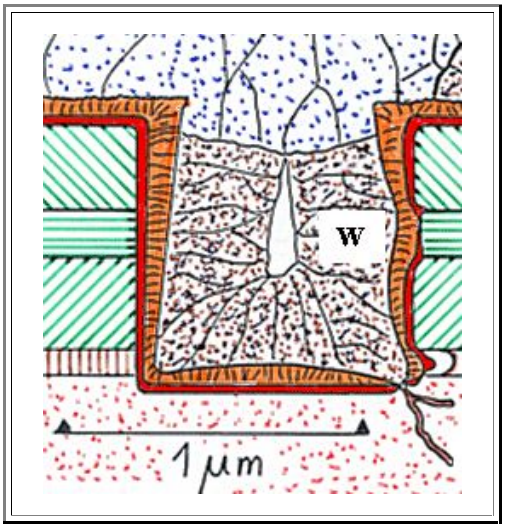
- Polymer deposition, while tricky, is often the key to optimized processes. In our example of **SiO_2** etching, a general finding is:
  - **Si** and **SiO_2** is etched in this process, but with different etch rates that can be optimized
  - The (chemical) etching reaction is always triggered by an energetic ion hitting the substrate (this provides for good anisotropy).
  - The tendency to polymer formation scales with the ratio of **F/H** in the plasma. The etching rate increases with increasing **F** concentration; the polymerization rate with increasing **H** concentration.
  - Best selectivity is obtained in the border region between etching and polymer formation. This will lead to polymer formation (and then protecting the surface) with **Si**, while **SiO_2** is still etched. The weaker tendency to polymer formation while etching **SiO_2** is due to the oxygen being liberated during **SiO_2** etching which oxidizes carbon to **CO_2** and thus partially removes the necessary atoms for polymerization



- Enough about plasma etching. You get the idea.
  - The thee pictures show extremely schematically different set-ups of plasma etchers. Essentially the big deal is how the plasma is made. For sufficiently large etching rates, you want a dense homogenous plasma with control of the ion energy.
  - A taste treat of what it really implies and some explanations to the pictures can be found in the advanced module in the link

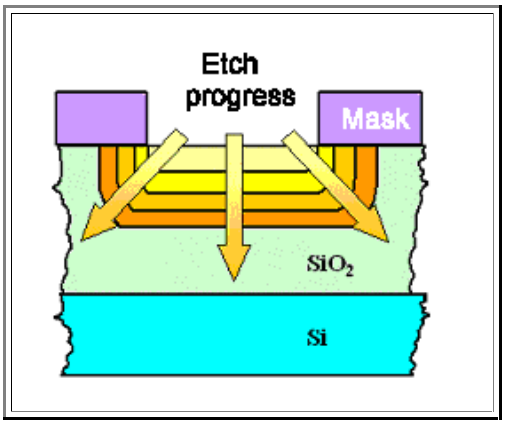
6.4.4 Summary to: 6.4 Etching Techniques

Structuring means selective removal of material (through a mask) by etching. There are three main conditions for etching:



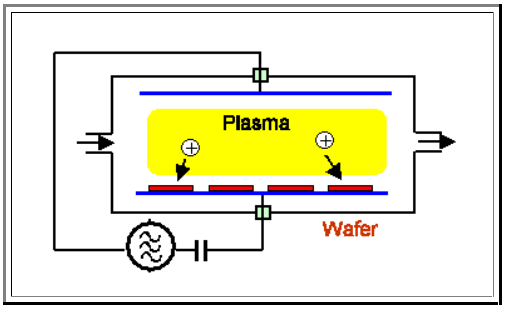
- **1.** Must attack material to be etched ⇒ *etching rate*.
- **2.** Must *not* attack everything else ⇒ *selectivity*.
- **3.** Must conserve structure of mask (good on left side of picture, not so good on right side).

Chemical etching:



- Can be near perfect for points **1.** and **2.**. Example: **HF** attacks only **SiO_2** but not **Si** and most other materials.
- Fails miserably on point **3.**
- Underetching is unavoidable. Can't be used for lateral structure sizes **< ≈ 2 µm**

Plasma etching ("Dry" etching)



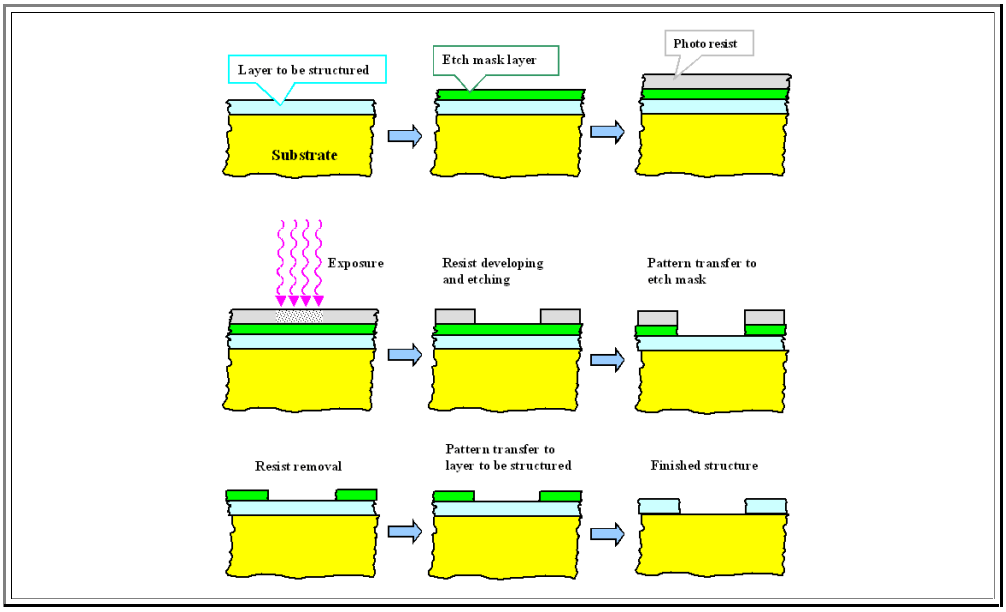
- In a plasma quite unusual reactions can take place - including reactions never seen in normal chemistry. Many materials can be etched in a suitable plasma
- Etching might preserve the lateral mask dimensions - for reasons not always entirely clear
- There is tremendous potential in plasma etching because of the tremendously large parameter space - and tremendous problems and costs for the same reasons
- Allmost all "small" structures in semiconductor technology are obtained by plasma etching

6.5 Lithography

6.5.1 Basic Lithography Techniques

Process flow of Lithography (and Pattern Transfer)

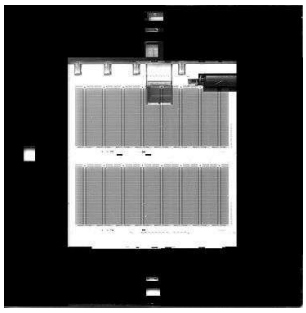
- Lets start by considering the basic processes for the complete structuring module.
  - Shown is a more complex process flow with a special etch mask layer (usually **SiO_2**).
  - Often, however, you just use the *photo resist as masking layer for etching*, omitting deposition, structuring, and removal of the green layer. A photo resist mask generally is good enough for ion implantation (provided you keep the wafer cool) and many plasma etching processes.



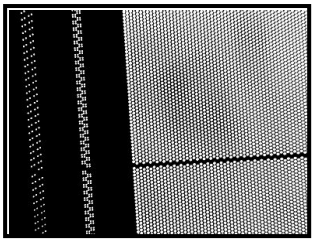
- As far as lithography is concerned, it is evident that we need the following key ingredients:
  - A **photo resist** [1)], i.e. some light sensitive material, not unlike the coating on photographic film.
  - A **mask** (better known as **reticle** [2)]) that contains the structure you want to transfer - not unlike a slide.
  - A **lithography unit** that allows to project the pattern on the mask to the resist on the wafer. Pattern **No. x** must be *perfectly aligned* to pattern **No. x - 1**, of course. Since about **1990** one (or just a few) chips are exposed at one time, and than the wafer is moved and the next chip is exposed. This step-by-step exposure is done in machines universally known as **steppers**.
  - Means to **develop** and structure the resist. This is usually done in such a way that the *exposed* areas can be removed by some etching process (using **positive resist**). For some special purpose, you may also use **negative resists**, i.e. you remove the *unexposed* areas.
- In principle, it is like projecting a slide on some photosensitive paper with some special development.
  - However, we have some very special requirements. And those requirements make the whole process very complex!
  - And with *very complex* I mean *really* complex, super-mega-complex - even in your wildest dreams you won't even get close to imagining what it needs to do the lithography part of a modern chip with structures size around **0,13 µm**.
- But relax. We are not going to delve very deep into the intricacies of lithography, even though there are some advanced material issues involved, but only give it a cursory glance.

Reticles

- For any layer that needs to be structured, you need a ***reticle***. Since the projection on the chip usually reduces everything on the reticle fivefold, the reticle size can be about **5** times the chip size
  - A reticle then is a glass plate with the desired structure etched into a **Cr** layer. Below, a direct scan of an old reticle is shown, together with a microscope through-light image of some part.
  - "Obviously", the regular lattice of small opening in the non-transparent **Cr** layer is the array for the trenches in a memory chip. The smallest structures on this reticle are about **5 µm**.

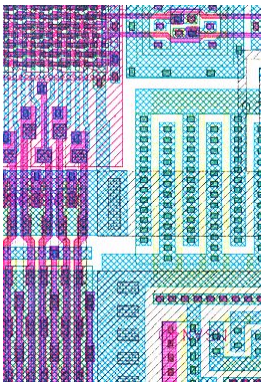


Typical reticle, about original size



Enlargement (**x 100**)

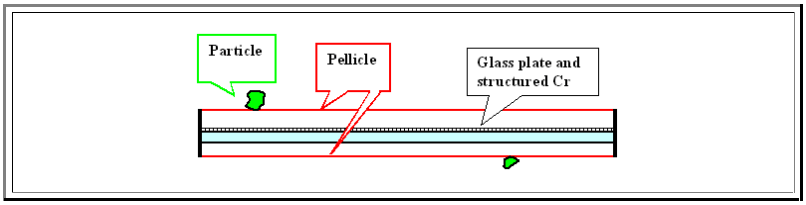
- Before we look at the requirements of reticles and their manufacture, lets pause for a moment and consider how the structure on the reticle comes into being.



- First, lets look at these structure, or the **lay-out** of the chip. Shown on the left is a tiny portion of a **4 Mbit DRAM**.
- Every color expresses *one* structured layer (and not all layers of the chip are shown).
- A print-out of the complete layout at this scale would easily cover a soccer field.
- The thing to note is: it is *not* good enough to transfer the structure on the reticle to the chip with a resolution *somewhat better* than the smallest structures on the chip, it is also necessary to superimpose the various levels with an **alignment accuracy** *much better* than the smallest structure on the chip!
- And remember: We have about **20** structuring cycles and thus reticles for one chip.

- The *lay-out* contains the *function* of the chip. It established where you have transistors and capacitors, how they are connected, how much current they can carry, and so on.
  - This is determined and done by the product people - electrical engineers, computer scientists - no materials scientists are involved.
  - The *technology*, the making of the chip, determines the *performance* - speed, power consumption, and so on. This is where material scientists come into their own, together with semiconductor physicists and specialized electrical engineers (who e.g., can simulate the behavior of an actual transistor and thus can tell the process engineers parameters like optimal doping levels etc.).
  - In other words, the reticles are the primary input of the product engineers to chip manufacturing. But they only may contain structures that can actually be made. This is expressed in **design rules** which come from the production line and must be strictly adhered to. Only if *all* engineers involved have some understanding of *all* issues relevant to chip production, will you be able to come up with aggressive and thus competitive design rules!

- What are the requirements that reticles have to meet (besides that their structures must not contain mistakes from the layout. e.g. a forgotten connection or whatever).
  - Simple: They must be *absolutely* free of defects *and* must remain so while used in production! Any defect on the reticle will become transferred to *every* chip and more likely than not will simply kill it.
  - In other words: Not a single particle is ever allowed on a reticle!
- This sounds like an impossible request. Consider that a given reticle during its useful production life will be put into a stepper and taken out again a few thousand times, and that every mechanical movement tends to generate particles.
  - Lithography is full of "impossible" demands like this. Sometimes there is a simple solution, sometimes there isn't. In this case there is:
  - First. make sure that the freshly produced reticle is defect free (you must actually check it pixel by pixel *and* repair unavoidable production defects).
  - Then encase it in **pellicles** [3] (**=** fully transparent thin films) with a distance of some mm between reticle and pellicle as shown below.



  - One of the bigger problems with steppers - their very small (about **1 µm**) *depth of focus* - now turns to our advantage: Unavoidable particles fall on the pellicles and will only be imaged as harmless faint blurs.
- How do we make reticles?
  - By writing them pixel by pixel with a finely focussed electron beam into a suitable sensitive layer , i.e. by direct writing **electron-beam lithography**.
  - Next, this layer is developed and the structure transferred to the **Cr** layer.
  - Checking for defects, repairing these defects (using the electron beam to burn off unwanted **Cr**, or to deposit some in a kind of e-beam triggered **CVD** process where it is missing), and encasing the reticle in pellicles, finishes the process.
- Given the very large pixel size of a reticle (roughly $\mathbf{10^{10}}$), *this takes time* - several hours just for the electron beam writing!
  - This explains immediately why we don't use electron beam writing for directly creating structures on the chip: You have at most a few seconds to "do" one chip in the factory, and e-beam writing just can`t deliver this kind of throughput.
  - It also gives you a vague idea why reticles don't come cheap. You have to pay some **5000 $ - 10 000 $** for *making* one reticle (making the lay-out is not included!). And you need a set of about 20 reticles for one chip. And you need lots of reticle sets during the development phase, because you constantly want to improve the design. You simply need large amounts of money.

---

[1] Something as a protective coating that resists or prevents a particular action (Webster, second meaning)

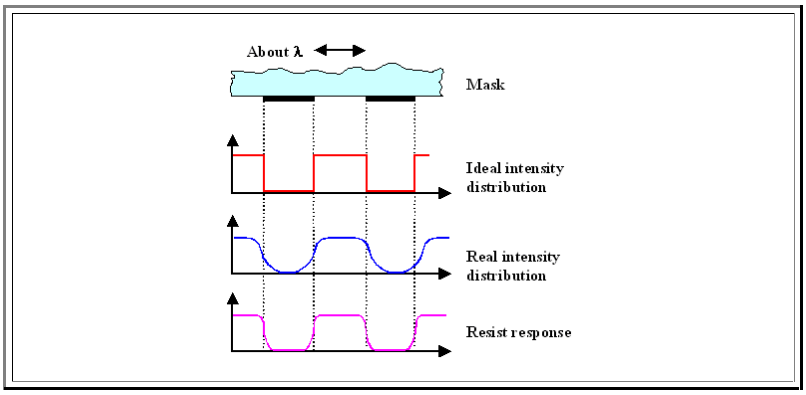
[2] A system of lines, dots, cross hairs, or wires in the focus of the eyepiece of an optical instrument (Webster)

[3] A thin skin or film, especially for optical uses

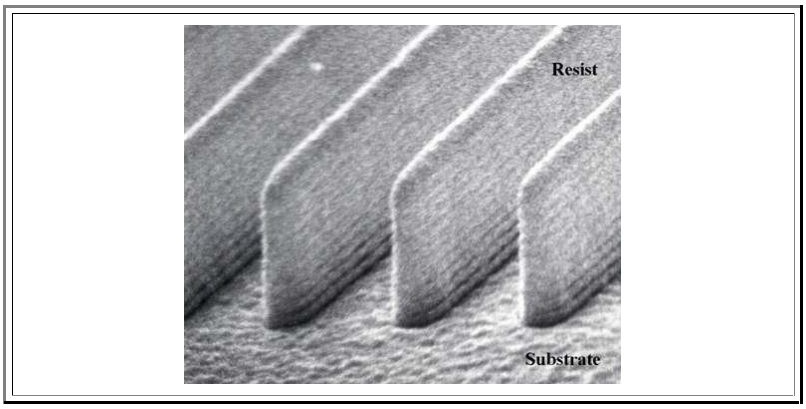
6.5.2 Resist and Steppers

Photo Resists

- Lets just look at a list of requirements for resists. We need to have:
  - *High sensitivity* to the wavelength used for imaging, but not for all optical wave lengths (you neither want to work in the dark, nor expose the resist during optical alignment of the reticle which might be done with light at some other wave length). Not easy to achieve for the short wave lengths employed today.
  - *High contrast*, i.e. little response (= "blackening") to intensities below some level, and strong response to large intensities. This is needed to sharpen edges since diffraction effects do not allow sharp intensity variations at dimensions around the wavelength of the light as illustrated below.



  - *Compatibility* with general semiconductor requirements (easy to deposit, to structure, to etch off; no elements involved with the potential to contaminate **Si** as e.g. heavy metals or alkali metals (this includes the developer), no particle production, and so on).
  - *Homogeneous "blackening" with depth* - this means little absorption. Simply imagine that the resist is strongly absorbing, which would mean only its top part becomes exposed. Removal of the "blackened" and developed resist than would not even open a complete hole to the layer below.
  - No *reflection of light*, especially at the interface resist - substrate. Otherwise we encounter all kinds of interference effects between the light going down and the one coming up (known as "**Newton fringes**"). Given the highly monochromatic and coherent nature of the light used for lithography, it is fairly easy to even produce *standing* light waves in the resist layer as shown below. While the ripple structure clearly visible in the resist is not so detrimental in this example, very bad things can happen if the substrate below the resist is not perfectly flat.



  - This would call for a *strongly absorbing* resist - in direct contradiction to the requirement stated above. Alternatively, an **anti-reflection coating** (***ARC***) might be used between substrate and resist, adding process complexity and cost.
  - Suitability of the resist as *direct mask* for ion-implantation or for plasma etching.
  - Easy *stripping* of the resist, even after it was turned into a tough polymer or carbonized by high-energy ion bombardment. Try to remove the polymer that formed in your oven from some harmless organic stuff like plum cake after it was carbonized by some mild heat treatment without damaging the substrate, and you know what this means.
- Enough requirements to occupy large numbers of highly qualified people in resist development!
  - Simply accept that resist technology will account for the last **0,2 µm** or so in minimum structure size. And if you do not have the state-of-the-art in resist technology, you will be a year or two behind the competition - which means you are loosing *large amounts of money*!

Stepper

- A stepper is a kind of fancy slide projector. It projects the "picture" on the reticle onto the resist-coated wafer. But in contrast to a normal slide projector, it does not *enlarge* the picture, but *demagnifies* it - exactly fivefold in most cases. Simple in principle, ***however:***
- **1.** We need the *ultimate in optical resolution*!
  - As everybody knows, the **resolution limit** of optical instruments is equal to about the wave-length λ. More precisely and quantitatively we have

$$d_{min} \approx \frac{\lambda}{2NA}$$

  - With $\mathbf{d_{min}}$ **= minimal distinguishable feature size**; i.e the distance between two **Al** lines, and ***NA*** = **numerical aperture** of the optical system (the ***NA*** for a single lens is roughly the quotient of focal length / diameter; i.e. a crude measure of the size of the lens).
  - Blue light has a wave length of about **0.4 µm**, and the numerical apertures ***NA*** of very good lenses are principally **< 1**; a value of **0.6** ~~**is**~~ was about the best you can do (consider that all distortions and aberrations troubling optical lenses become more severe with increasing ***NA***). This would give us a minimum feature size of

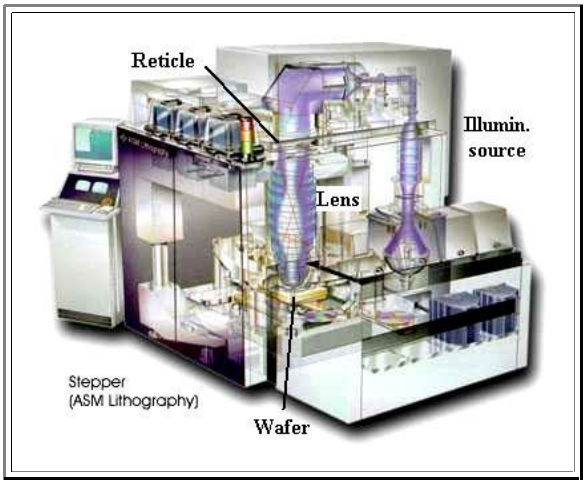
$$d_{min} \approx \frac{0.4}{1.2} = 0.33\ \mu m$$

  - Since nowadays you can buy chips with minimum features of **0.18 µm** or even **0.13 µm**; we obviously must do better than to use just the visible part of the spectrum.
- **2.** Resolution is not everything, we need some *depth of focus*, too. Our substrate is not perfectly flat; there is some topography (not to mention that the **Si** wafer is also not perfectly flat).
  - As anyone familiar with a camera knows, your depth of focus Δf *decreases*, if you *increase* the aperture diameter, i.e. if you *increase* the ***NA*** of the lens. In formulas we have

$$\Delta f \approx \frac{\lambda}{(NA)^2} = \frac{0.4}{0.6^2} = 1.11\ \mu m$$

  - Tough! What you gain in resolution with larger numerical apertures, you loose (quadratically) in focus depth. And if you decrease the wavelength to gain resolution, you loose focus depth, too!
- **3.** We need to *align* one exposure exactly on top of the preceding one. In other words, we need a wafer stage that can move the wafer around with a precision of lets say **1/5 of** $\mathbf{d_{min}}$ - corresponding to **0.18/5 µm = 0.036 µm =** ***36 nm***.
  - And somehow you have to control the stage movement; i.e. you must measure where you are with respect to some alignment marks on the chip *with the same kind of precision*. We need some **alignment module** in the stepper.
  - Alignment is done optically, too, as an integral (and supremely important) part of stepper technology. We will, however, not delve into details.
- **4.** We need to do it *fast*, *reliable* and *reproducible* - **10 000** and more exposures a day in one stepper.
  - Time is money! You can't afford more than a few seconds exposure time per chip.
  - And you also can not afford that the machine breaks down frequently, or needs frequent alignments. Therefore you will put your stepper into separate temperature and humidity controlled enclosures, because the constancy oft these parameters in the cleanroom (Δ**T** ≈ **1** $^{\mathbf{o}}$**C**) is not good enough. You also would need to keep the atmospheric pressure constant, but ingenious engineers provided mechanism in the lens which compensates for pressure variations of a few **mbar**; still, when you order a stepper, you specify your altitude and average atmospheric pressure).

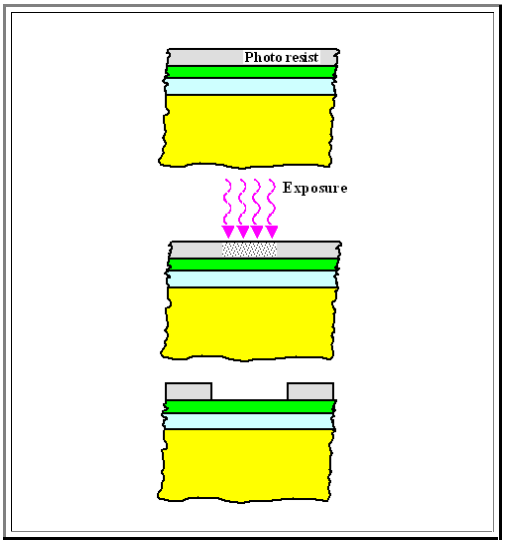
- How do we built a stepper? By combining elements from the very edge of technology in a big machine that costs around **10.000.000.000 $** and that can only be produced by a few specialized companies.
  - The picture below gives an impression. The basic imaging lens of the stepper is a huge assembly of many lenses; about **1 m** in length and **300 kg** in weight.



  - We need intense monochromatic light with a short wave length. If you use colored light, there is no way to overcome the chromatic aberrations inherent in all lenses and your resolution will suffer.
  - The wave lengths employed started with the so-called **g**-line (**436 nm**) of **Hg**, fairly intense in a **Hg** high pressure arc lamp and in the deep blue of the spectrum. It was good down to about **0.4 nm** as shown in the example above.
  - Next (around **1990**) came the **365 nm i**-line in the near ultra violet (***UV***). This took us down to about **0.3 µm**.
  - Next came *a problem*. There simply is no "light bulb" that emits enough intensity at wavelengths considerably smaller than **365 nm**. The (very expensive) solution were so-called *excimer Lasers*, first at **248 nm** (called deep **UV** lithography), and eventually (sort of around right now (**2001**)), at **194 nm** and **157 nm**.
  - Next comes *the end*. At least of "conventional" stepper technology employing lenses: There simply is no material with a sizeable index of refraction at wavelengths considerably below **157 nm** that can be turned into a high-quality lens. After 2000, lots of people worried about using single crystals of **CaF_2** for making lenses for the **157 nm** stepper generation.
  - Now, in **2007**, **CaF_2** is out again - too many problems, it seems.
- What do you do then? First you raise obscenely *large amounts of money*, and than you work on alternatives, most notably
  - **Electron-beam lithography**. We encountered it before; the only problem is to make it much, much faster. As it appears today (Aug. **2001**), this is not possible.
  - **Ion beam lithography**. Whatever it is, nobody now would bet much money on it.
  - **X-ray lithography**. Large-scale efforts to use **X**-rays for lithography were already started in the eighties of the **20** century (involving huge electron synchrotons as a kind of light bulb for intense **X**-rays), but it appears that it is pretty dead by now.
  - **Extreme UV lithograpy** at a wave length around **10 nm**. This is actually soft **X**-ray technology, but the word "**X**-ray lithography" is loaded with negative emotions by now and thus avoided. Since we have no lenses, we use mirrors. Sounds simple - but have you ever heard of mirrors for **X**-rays? Wonder why not? This is what the US and the major US companies favor at present.
- Well, lets stop here. Some more advanced information can be found in the link.
  - *But note*: There are quite involved materials issues encountered in lithography in general, and in making advanced steppers in particular. **CaF_2** is an electronic material! And the success - or failure - of the global enterprise to push minimum feature size of chips beyond the **100 nm** level, will most likely influence *your* professional life in a profound matter.
  - This is so because the eventual break down of Moores law will influence in a major way *everything* that is even remotely tied to technology. And what will happen is quite simply a question if we (including you) succeed in moving lithography across the **100 nm** barrier.

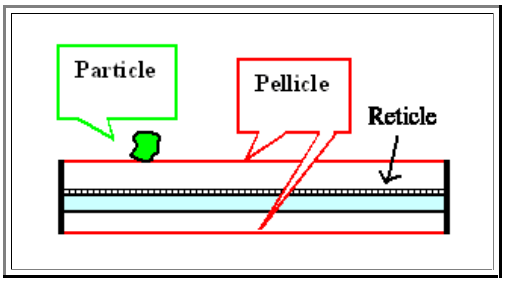
6.5.3 Summary to: 6.5 Lithograpy

Lithography comprises three elementary steps:



- Cover the layer to be structured with a light-sensitive layer called (photo) **resist**
- Expose the resist to light only where you want the structure by a "slide projector" called **stepper** (always demagnify the "slide" called **reticle**.
- Develop the exposed resist in such a way that unexposed parts are etched off.
- The structure has now be transferred into the resist; the process is rather similar to regular old-fashioned analog photography.

The problem is that we want to make structures with lateral sizes in the **30 nm** region, far smaller than the wavelength of light. This necessitates extreme measures in all components involved



- At the core of lithography are the **steppers** - optical machines for around **5 Mio €** a piece
- Resist technology, too, is a highly developed part of lithography
- For some big problems simple solutions have been found. Example: reticles with pellicles

6.6 Summary:

6.6.1 Materials and Processes for Silicon Technology

- Silicondioxide (**SiO_2**) has been the "ideal" dielectric with many uses in chip manufacture
  - Only recently (**2007**) is it replaced by "low **k"** and "high **k**" dielectrics, i.e. dielectrics with a dielectric constant either lower or larger than that of **SiO_2**
  - "Low **k**" dielectrics (polymers, porous **SiO_2**, ..; the ideal material has not yet been found) are used for intermetal insulation; low **k** is important here to keep the **RC** time constants small
  - "High **k**" dielectrics (the present front runner is **HfO_2**) will replace the gate oxides. They can be somewhat thicker than **SiO_2** without sacrificing capacity, while strongly reducing tunneling currents.

- **Gate oxide for Transistors**
- **Dielectric in Capacitors**
- **Insulation**
- **Stress relieve layer**
- **Masking layer**
- **Screen oxide during Implantation**
- **Passivation**

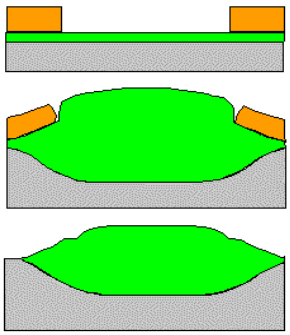
- **SiO_2** can be made in several ways:
  - Dry oxidation is relatively slow but gives best oxide qualities as defined by:
    - Uniformity
    - thickness control
    - Break down field strangt
    - Interface quality
    - Reliability

    Typical use: Highest quality gate oxid.
  - Wet oxidation is about 10 times faster; it is used whenever relatively thick oxides are needed.
    Typical use: Field oxide.
  - The other methods are needed whenever there is no **Si** available for oxidation (e.g. intermetal dielectrics).

- Dry thermal oxidation:
  $2\,Si + O_2 \Rightarrow 2\,SiO_2$
- Wet thermal oxidation:
  $Si + 2\,H_2O \Rightarrow SiO_2 + 2\,H_2$
- "Chemical Vapor Deposition" (next sub-chapter)
- "Spin-on techniques (next sub-chapter)
- "Anodic oxidation (presently not used in technology)

- As long as the process is diffusion controlled (i.e. the time it takes oxygen to diffuse through the already formed oxide determines rates, the thickness increases protorional to $t^{1/2}$
  - For thin oxides the growth rate is reaction controlled and the thickness - time dependence becomes complicated.

- Growing oxide only locally ("**LOCOS**") was a key process for field oxides.
  - Without a "buffer" oxide below the masking nitride, large mechanical strain develops, producing plastic deformation and thus dislocations around the oxide edges.
  - These "Oxide edge dislocations" kill the transistor.
  - Buffer oxides solve the problem, but create new problems: A "birds beak" develops, increasing lateral dimensions beyond the mask dimension.
- **LOCOS** is a good example for a universal feature of **Si** technology: Solutions to old problems create new problems. Solutions to the new problems... and so on. It follows:
  - Process complexity increases all the time.
  - New materials are needed all the time.



- **Chemical Vapor Deposition** (**CVD**) is simple in principle
  - Find to gases that react ot the desired material at elevated temperatures
  - Put your wafer(s) into some machine, evacuate, heat to the desired temperature (preferably only the wafers) and admit the gases (and remove undesired reaction products).
  - There are many quute different technical ways (all of them expensive) to realize a **CVD** apparatus

- Major **CVD** process are
- Deposition of epitaxial **Si** layers - obviously always on (atomically clean) **Si** substrates. By admitteing some gases carrying doping atoms (e.g. **AsH_3, AsH_3**) the layer can be doped in-situ.
- Depositon of poly crystalline **Si** layers.
  - Chemically similar to epitaxial layers, in reality quite different because the CVD reactos can be simpler.
  - Poly-**Si** is needed for many uses: Gate electrode, interconnect, filling of holes, sacrifical layer.
  - Its great advantage is its ull compatibility with **Si** and **SiO_2**; its great disadvantage is its mediocre conductivity (for heavy doping).
- Deposition of **Si_3N_4**
  - Very important. Always prone to produce mechanical stress (**Si_3N_4** is an unyielding ceramic!).
- **W** (and Silicides, and ...)
  - Not "good" processes, but sometimes unavoidable!

Epitaxial Si layer

$$SiCl_4 + 2\,H_2 \Rightarrow Si + 4\,HCl \quad (1000\,^oC - 1200\,^oC)$$

Polycrystalline Si layer

$$SiH_4 \xrightarrow[630^oC]{60\ Pa} Si + 2\,H_2$$

SiO_2 layer ("TEOS process)

$$Si(C_2H_5O)_4 \Rightarrow SiO_2 + 2H_2O + C_2H_4 \quad (720\,^oC)$$

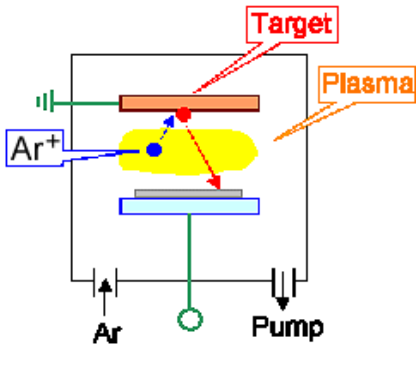
Si_3N_4 layer

$$3\,SiH_2Cl_2 + 4NH_3 \Rightarrow Si_3N_4 + 2HCl + 1{,}5\,H_2 \quad (\approx 700\,^oC))$$

W layer

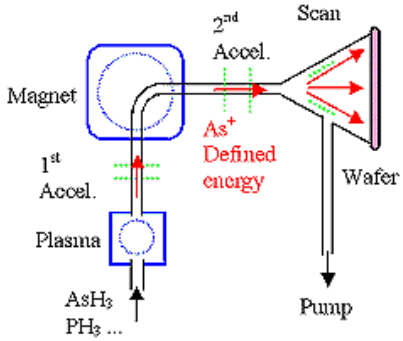
$$WF_6 + H_2 \xrightarrow[470^oC]{10^4\ Pa} W + \text{(nasty) gases}$$

- **Sputter deposition**
  - Plasma technique ⇒ Vacuum + high voltage (and possible high frequency): complicated and expensive
  - Layers amorphous to highly defective ⇒ needs usually annealing after deposition.
  - Very versatile because of easy control of layer composition by target composition
  - Decent depositioen rates possible. Particularly suited to conductors.
  - Coverage is *not* conformal!



Ion implantation

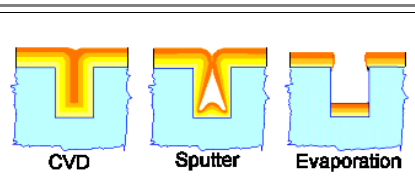
- Depth (**< ca. 1 µm)** and dose precisely controllable.
- Very compley and expensive
- Method od choice for making doped layers.
- Introduces defects or destroys crystallinity ⇒ annealing at high ***T*** (**> 800 °C**) is a must



There are many more techniques for producing thin layers

- **Evaporation**. Relatively simple but limited as to materials and edgencoverage
- **Molecular beam epitaxy**. (**MBE**) Standard for **III-V's**
- **Spin-on techniques** ("Sol- Gel"). Used for making photo resist layers; occasionally for others
- **Galvanics**. Kind of crude but necessary for **Cu** interconnects in modern **IC's**

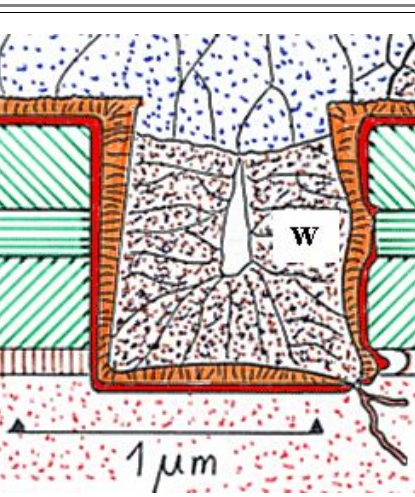
Edge coverage may be the decisive property!



Comparing edge coverage

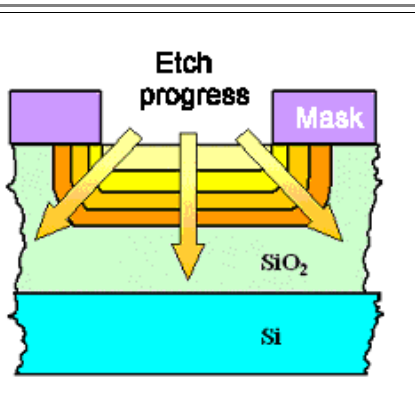
Structuring means selective removal of material (through a mask) by etching. There are three main conditions for etching:

- **1.** Must attack material to be etched ⇒ *etching rate*.
- **2.** Must *not* attack everything else ⇒ *selectivity*.
- **3.** Must conserve structure of mask (good on left side of picture, not so good on right side).

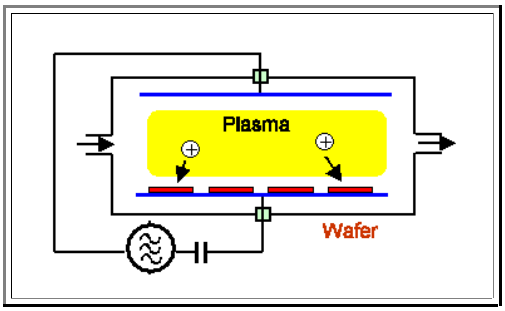


Chemical etching:

- Can be near perfect for points **1.** and **2.**. Example: **HF** attacks only **SiO_2** but not **Si** and most other materials.
- Fails miserably on point **3.**
- Underetching is unavoidable. Can't be used for lateral structure sizes **<** ≈ **2 µm**

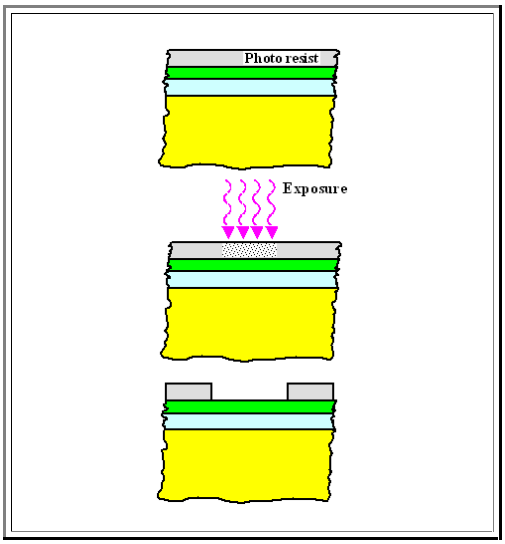


Plasma etching ("Dry" etching)



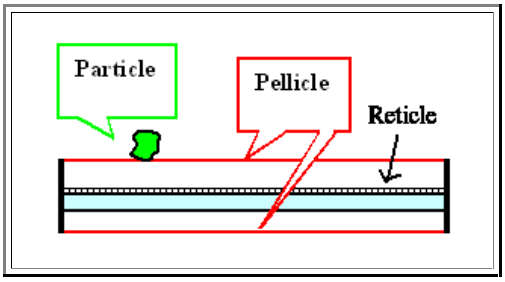
- In a plasma quite unusual reactions can take place - including reactions never seen in normal chemistry. Many materials can be etched in a suitable plasma
- Etching might preserve the lateral mask dimensions - for reasons not always entirely clear
- There is tremendous potential in plasma etching because of the tremendously large parameter space - and tremendous problems and costs for the same reasons
- Allmost all "small" structures in semiconductor technology are obtained by plasma etching

Lithography comprises three elementary steps:



- Cover the layer to be structured with a light-sensitive layer called (photo) **resist**
- Expose the resist to light only where you want the structure by a "slide projector" called **stepper** (always demagnify the "slide" called **reticle**.
- Develop the exposed resist in such a way that unexposed parts are etched off.
- The structure has now be transferred into the resist; the process is rather similar to regular old-fashioned analog photography.

The problem is that we want to make structures with lateral sizes in the **30 nm** region, far smaller than the wavelength of light. This necessitates extreme measures in all components involved



- At the core of lithography are the **steppers** - optical machines for around **5 Mio €** a piece
- Resist technology, too, is a highly developed part of lithography
- For some big problems simple solutions have been found. Example: reticles with pellicles

Questionaire

Multiple Choice questions to all of 6.

7. Si MEMS

7.1 Products and Developments

7.1.1 What is MEMS?

7.1.2 A Closer Look at a Gyro

7.1.3 MEMS Sensors and Actuators

7.1.4 Summary to: 7.1 MEMS - Products and Developments

7.2 Processes and Specialities

7.2.1 Processes and Materials - General Consideration

7.2.2 Important MEMS Processes

7.2.3 Example for Process Integration

7.2.4 Summary to: 7.2 Processes and Specialities

7.3 Summary

7.3.1 Summary to 7: MEMS

7. Si MEMS

7.1 Products and Developments

7.1.1 What is MEMS?

MEMS Basics

- ***MEMS*** is the abbreviation for **Micro Electro Mechanical Systems**. It is not a good abbreviation because it omits, for example, all those microoptical, micromaterial, microchemical, or microfluidic systems that are also part of **MEMS** technology. The old-fashioned name "Microsystems" (or, to stay in touch with the spirit of the times, "Nanosystems"), would have said it all much better and simpler.
  - The word "system" is not unambiguous either. According to Wikipedia, a *"system" (from Latin "systema", in turn from Greek "συστημα"=systema) is a set of interacting or interdependent entities, real or abstract, forming an integrated whole*. OK - so an integrated circuit is a system too, because it is a set of interacting transistors, forming an integrated whole, like a memory chip? It is a matter of taste, but a the feeling is that a memory *chip* is not yet a system. If you add input and output and the interface for something that uses the memory, you might call it a memory *system*.
  - Another way of looking at it would be to equate "system" with "little machine", an entity that can do something or interacts with something else on its own.
  - A more practical point of view is simply to say that a **MEMS** *chip* is something produced on a **Si** substrate that has more functions than just microelectronics. It may *contain* some microelectronics, but it must also have some non-electronic component, be it mechanical, optical, or whatever.
  - A somewhat oversimplified but helpful point of view is that **MEMS** products invariably contain **sensors**, **actuators** or both; and that they are small (in the range of a few **µm** to a few **100 µm** across).
- **MEMS** is an "**emerging technology**", meaning that while I'm writing this, new developments undreamed off a few years ago take place, and in a few years form now this text will be outdated on many counts. Nevertheless, **MEMS** "devices" gross about **$ 5,5 · 10^9** per year in **2007**, while **MEMS** "systems" will cross the **$ 100 · 10^9** per year barrier shortly after **2010** or so (more to the MEMS market in this link). The difference between device and system is shown below.

MEMS based Gyrometer="*system*".

MEMS "chips", *devices*, or components

  - So be careful when numbers concerning money come up. A "system" like a car contains "systems" like gyros that contain **MEMS**-Systems and so on.
  - For all emerging technologies there is also a lot of ***hyperbole*** concerning applications and money around, and that makes it not so easy to assess what is really going on. So let's focus on the *essentials* here.
- Let's have a quick look at **MEMS products**. There are three kinds.

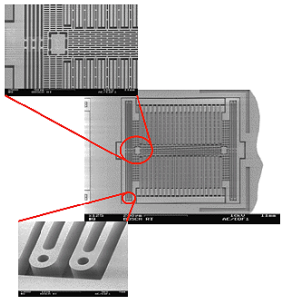
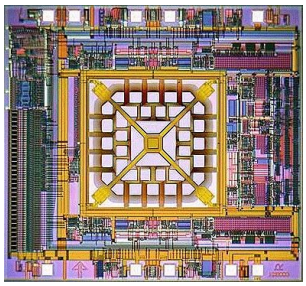
  - Products you most likely own (possibly without being aware of that) or experience in the world around you (possibly without being aware of that), and that you and I could buy right now.
  - Products not yet available to everyone or not ***ubiquitous***, but already marketed or just about to hit the market place.
  - Intended or projected products that may or may not make it to the market.
  - The list given below is neither complete nor does it mirror the present day (**2008**) market. It just serves to give a flavor of what is, or might be, around.

Example of MEMS Products

- What kind of **MEMS** products are on the market and possibly in your possession?
- Very prominent in this respect are **acceleration sensors**.

- If you have a car equipped with **air bags**, you need acceleration sensors and they will be based on **MEMS**.
- Knowing that, and being in Germany, you naturally type "air bag acceleration sensor bosch" into Google Pictures - and get almost nothing. Only if you follow text pages and turn to detailed product information, you may find something about **MEMS** and acceleration sensors from **Robert Bosch GmbH**; an example is shown in the link.

Here are two pictures with short descriptions from the producers of acceleration sensors (also called "**g-sensors**").

 **Bosch g-sensor** Scale ? (several **mm**)	 **MEMSIC g-sensor** Scale ? (about **1 cm**)
The common sensing principle of the accelerometers is capacitive. An acceleration in the lateral direction deflects the proof mass that is suspended by folded springs in the x-sensing element. One set of electrodes is attached to the proof mass and moves with acceleration. These movable electrodes form capacitors with two sets of fixed electrodes opposing them with a small air gap in between. The use of such a differential capacitive arrangement with two capacitors reduces the nonlinearity of the transfer function of the device. Overrange stops are implemented for shock protection that avoids the direct contact of the fingers at large accelerations. The mechanical sensitivity (in ***fF/g***) can be adjusted by the thickness and/or the length of the springs. The differential capacitance signal is evaluated by an ***ASIC*** which is electrically connected to the sensor by chip-to-chip wire bonds. A change of $\mathbf{C^1}$ and $\mathbf{C_2}$ is detected and transformed into a corresponding analog voltage by a capacitance/voltage converter. (Image courtesy of Bosch) .	MEMS-Based Accelerometer MEMSIC's dual-axis thermal accelerator is a MEMS-based semiconductor device that works conceptually like the air bubble in a construction level. The square in the middle of the chip is a resistor that heats up a gas bubble. The next larger squares contain **thermal couples** that sense the location of the heated bubble as the device is tilted or accelerated. (Image courtesy of MEMSIC, Inc.) .

- The first thing we learn from the pictures is that obviously two quite different principles are used for measuring acceleration. The second thing we learn is that both accelerometers are quite complicated little things; the **MEMSIC** one evidently coming with some integrated microelectronics around the actual sensor in the center of the device.

Accelerations sensors are not restricted to cars and airbags, of course. Your car might have some electronic *system* that keeps it from ***skidding*** and ***rolling*** when you move on ice and snow or negotiate a curve with too much speed. This "***ESP***" system (or whatever cute abbreviation your car manufacturer use for that) needs **MEMS** to tell it that something is not as it should be, and this includes information about acceleration.

- Pretty much everything that flies (air crafts, rockets, missiles, unmanned aircraft (drones), balloons, war heads, ...) need accelerometers; in particular for calculating where you are or how fast you are.
- More and more "***gadgets***" contain accelerometers, like Nintendo consoles for measuring movement and tilt to complement its pointer functionality. Sports watches for runners may contain accelerometers to help determine the speed and distance for the runner wearing the unit. A small number of modern notebooks and cell phones ("iPhone") feature accelerometers to automatically align the screen depending on the direction the device is held.

Besides sensors for *linear* or translational acceleration, cars with an **EPS** system (and anything that flies) will also need sensors for angular movements, or ***gyroscopes***.

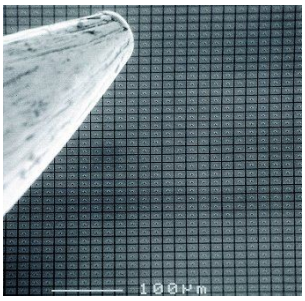
- The market for linear accelerometer and gyroscope chips is around **$ 650 · 10^6**, with single chips going for roughly **$ 20**. The market is expected to grow massively, driven by cost reductions and the many uses coming from ever more sophisticated gadgets as explained above.
- We will look at gyroscopes in more detail in the next module, so nothing more about accelerometers here

The next **MEMS** device - or better ***OMEMS*** device (an abbreviation that is sometimes used), is the "**Digital Micromirror Device**" from Texas Instrument (***TI***) called "**Digital Light Processing**" (***DLP***). A **DLP** chip is the "heart" of the **beamers**, a *product* virtually non-existing **10** years ago and by now (**2007**) quite ubiquitous.

- The **DLP** chip (or the principles behind it) was invented in **1987** by **Larry Hornbeck** and made it to the market around **1996**. Ten years later, in **2006**, according to **TI** a grand total of $\mathbf{10 \cdot 10^6}$ chips have been sold. Not a lot, if you compare it to memory chips or microprocessors, but still an unparalleled success for a rather outlandish device back in the final days of the second millennium. Here it is:

Courtesy of Texas Instruments

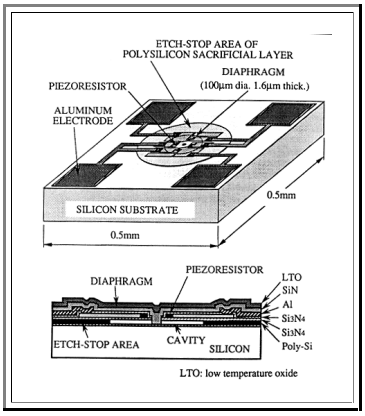
The **DLP** chip - looks like a mirror.


Courtesy of Texas Instruments

DLP chip close up with the tip of a needle

- All you see by looking a the chip directly is a mirror. Only if you look closely (with a microscope), you realize that this mirror consists of something like **1 280 · 720=921 600** separate small (about **15 µm · 15 µm**) mirrors that can move independently.
- Exactly how it works (with more picture) is explained in a separate module. Note that in this example **MEMS** is another **enabling technology**; there is simply no other way to do what a **DLP** chip can do.

The last major and rather common **MEMS** product is the **pressure sensor**. Not something you use conscientiously, but something that is quite important in many existing products and machinery and probably inside several things you do use. In fact, measuring the pressure of gases and liquids is about as common as measuring the temperature - there is a tremendous market for this out there!



- Moreover, there would be an even larger market for **MEMS** pressure sensors if they could be used in in places where they would be very useful, but face major problems. Inside your car tires, for example, or even inside *you* - if you suffer from high or low blood pressure. The problem, of course, about to be solved right now, is how to supply power and how to interface the sensor with the outside world (at negligible cost, of course).
- The picture shows one kind of **MEMS** pressure sensor (probably from Toyota). We recognize a few by now familiar materials like **Si**, $\mathbf{Si_3N_4}$, **Al** and **poly-Si**, but also new parts like "piezoresistors".
- The working principle is easy to guess - words like "***diaphragm***", "cavity" and and "piezoresistor", if taken together, should make it quite clear.

If you give the matter a bit more thought, you will start to realize that there is a tremendous diversity of requirements for pressure measurements, just given as alternatives:

1. Gas - liquids.
2. High pressures - low pressures.
3. At room temperature or at extremely high or low temperatures.
4. In a ***benign*** or corrosive environment.
5. With high or low demands on accuracy.

- Just consider the **4th** point. A pressure sensor is always exposed to the medium the pressure of which it is supposed to measure. How about ***liquid manure***? If you want to run a large biogas facility automatically, you must measure the pressure inside your (more or less) liquid ingredients.
- How about measuring the pressure inside a car engine at working temperatures? Well, if it is done at all, it will not be easy. If you use a **MEMS** device consisting of **Si**, $\mathbf{Si_3N_4}$, $\mathbf{SiO_2}$, and so on, it might be able to take the heat, but you cannot integrate electronics on this chip because *transistors* can't. This teaches us that **MEMS** may not be the answer to *all* your system needs but must compete with other solutions to a given problem.

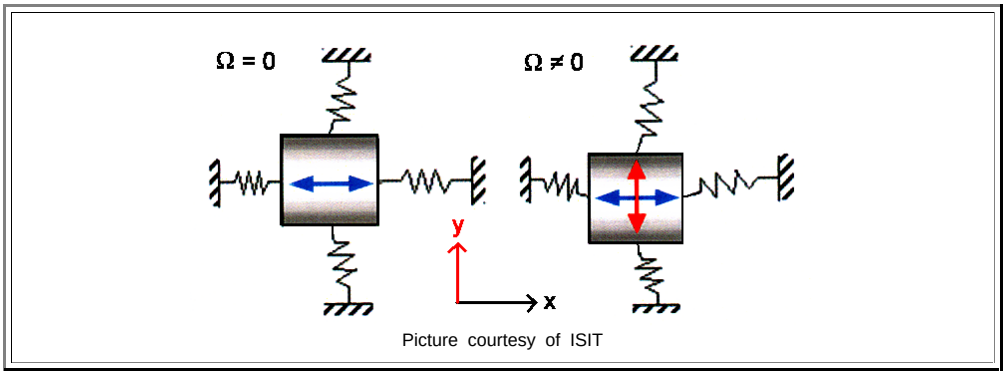
- Finally. let's just list some more products, possibly not yet on the market or not yet of monetary importance. It is a save bet, however, that during *your* professional life as an engineer, you will see most of those products around - and a lot more, not even imagined by now. We will just give some catch words with a few explanatory lines. More to all of that you will learn in a separate lecture course and in the links given
- ***RF MEMS*** or "radio frequency microelectromechanical systems".
  - "***RF***" here includes frequencies in the higher **GHZ** region (in German it would be ***HF***=high frequency). If you consider that just making a switch that opens and closes a **GHZ** circuit is a non-trivial enterprise because your parasitic capacities more or less short-circuit everything, you get the idea why **MEMS** could be useful.

    Class Exercise: *How large is the resistance of a **1 pF** capacitor at **10 GHZ**?*
  - What **RF MEMS** can do is to produce high quality switches, varactors (variable capacitors or reactances), high ***Q*** capacitors and inductors, resonators, filters and phase shifters, couplers and power dividers; offering the advantages of improved isolation, lower power dissipation, and reduced cost, size, and weight. **RF MEMS** is seen as being on the verge of revolutionizing wireless communication, but presently is mostly investigated at universities.
- **Optical MEMS** (besides the **DLP** chip).
  - While the **DLP** chip is digital, i.e. the mirrors have only two positions, analog mirrors with precisely adjustable positions and two axes would allow to project any picture via scanning a Laser beam. Other intended applications are displays, ***IR*** imagers, spectrometers, bar code readers, maskless lithography, adaptive optics, head-up displays, and so on.
- **BioMEMS** and **Microfluidics**
  - Put a drop of blood (or any other body liquid) on the appropriate **BioMEMS** chip, and it will give you either a full analysis ("lab-on-chip") or answer a particular question (do I have Aids?). Grow and connect neurons on a **bioMEMS** chip, or use it to re-connect nerves. Implant an artificial retina into the eyeball of the blind, or the proper device into the ear to allow hearing for the deaf. Micropumps under the skin soon might deliver continuously the proper amount of a drug according to some sensor (a huge improvement of the quality of life for people with diabetes!).
  - There is no end to keen visions here, and all of the above is being vigorously pursued at present
- Then there are **specialties**.
  - Reading and writing heads for memory discs. Relatively simple devices but accountiung for a large part of the MEMS market.
  - "Silicon microphones" are, if you like, very special pressure sensors. Consider that small and cheap microphones are needed in amazingly large quantities - there is one in every cell-phone.
  - Nozzles for ink-jet printers. Many very small nozzles in parallel for faster and more precise printing are being investigated and might enable "printing on demand" on a large scale.
  - Control of package shipping, where **MEMS** shock sensors would rest inside packages to monitor time and any type of damage that may occur while the package is in transit.
  - **MEMS** Oscillators, replacing the good old "quartz crystals", flow sensors, finger print sensors. And so on ....
- Like with any emerging technology, it is rather hard to figure out what is fact (in the sense that a product exists or will come into the market for sure) and what is fiction (in the sense that it would be nice to have, but is too expensive or that it can't really be made).
  - Time will tell and *you* will see.
- What can be predicted with certainty, however, is that feature sizes will shrink. Things will tend to become smaller for the same reasons as in microelectronics: More chips per ware=more performance per Dollar.
  - That will necessitate some rather outlandish materials science or rather quantum physics. A taste of what's around the corner can be found in the link.

7.1.2 A Closer Look at a Gyro

Working Principle

- We have already seen that there is more than one way to make a **MEMS** gyro. Here we look at a concept that is used quite a bit and pursued at the ***ISIT***, the **Fraunhofer Institute for Si Technology** in Itzehoe, a partner of the Inst. of Materials Science and Technology of the ***CAU***.
  - The basic concept is to use the **Coriolis force**, always present in rotating systems, to induce a response in a moveable mass held by springs.
  - This is done by having a mass suspended on two orthogonal set of springs as schematically shown below. The mass is driven by some mechanism to oscillate in ***x***-direction with constant amplitude. The frequency of the driving force is chosen in such a way that the system is close to its **resonance** frequency to ensure large amplitudes and therefore signals. As long as the the car or whatever moves with constant speed (in the ***x***-***y***-plane to keep things easy), that is all that happens: The sensor mass oscillates in ***x***-direction only.



Picture courtesy of ISIT

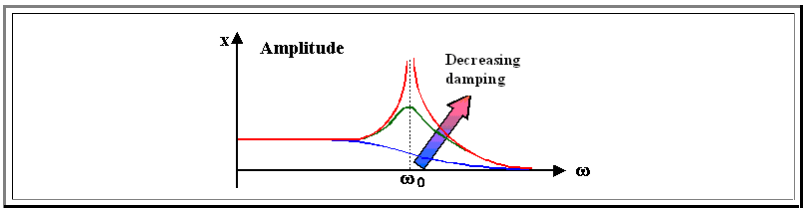
- If we now imagine some rotation with angular velocity Ω around the ***z***-axis perpendicular to the ***x***-***y***-plane (our car is driving into a curve) a **Coriolis** force F_{Cor} develops, given by

$$F_{Cor} = 2 \cdot m \cdot (\underline{v} \times \underline{\Omega})$$

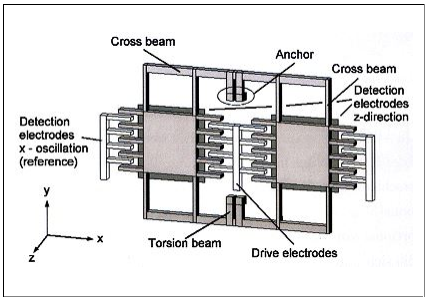
  - This Coriolis force drives our mass into the ***y***-direction and thus will produce an oscillation in ***y***-direction as shown above. The oscillation amplitude in ***y***-direction then will "somehow" be coupled to the angular velocity Ω. Again, we try to work under resonance conditions because that gives higher sensitivity.
  - The resonance frequencies of the two oscillation modes (one in ***x***-direction, the other one in ***y***-direction) do not have to identical. In fact, we will take care to make them somewhat different.
- So all we have to do now is to "somehow" measure the amplitude of the ***y***-direction oscillation and figure out how it relates to Ω Also, not to forget, we have to figure out how todrive the oscillation in ***x***-direction with constant amplitude "somehow". And after all that figuring is done, we have to make and package the system "somehow" - and sell it for **10 €** or so.
  - For the first "somehow" we need a lot of physics and Math. At the most simple level we need to delve into the "**driven damped linear oscillator**" theory, contained in the following differential equation expressing Newtons first law:

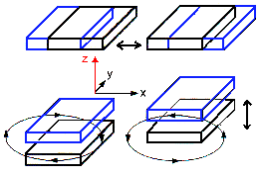
$$m \cdot \frac{d^2x}{dt^2} + k_F \cdot m \cdot \frac{dx}{dt} + k_S \cdot x = F_0 \cdot \cos(\omega_D \cdot t)$$

  - We have ***m***=mass of the oscillating body, k_F=friction or damping constant, k_S=spring constant, F_0=amplitude of the driving force, ω_D=frequency of the driving force.
  - This is a standard problem in more elementary mechanics and this link contains all you actually should know about this. Solving the equation gives the momentary location $x(\omega, t)$, the amplitude $X_0(\omega)$, and the velocity $\underline{v}=dx(\omega, t)/dt$ needed in the Coriolis force equation.
  - Looking just at the amplitude of a driven damped linear oscillator, we have the following, hopefully familiar picture:



- Depending on the degree of damping or friction, we will have very large to relatively small amplitudes at the resonance frequency ω_0
- Obviously we need to know ***m***, k_S, and k_F to go on. So let's look at an actual gyro in a very simple and by now (**2007**) outdated version:





The two oscillation modes.
The ellipses shows the combined effect

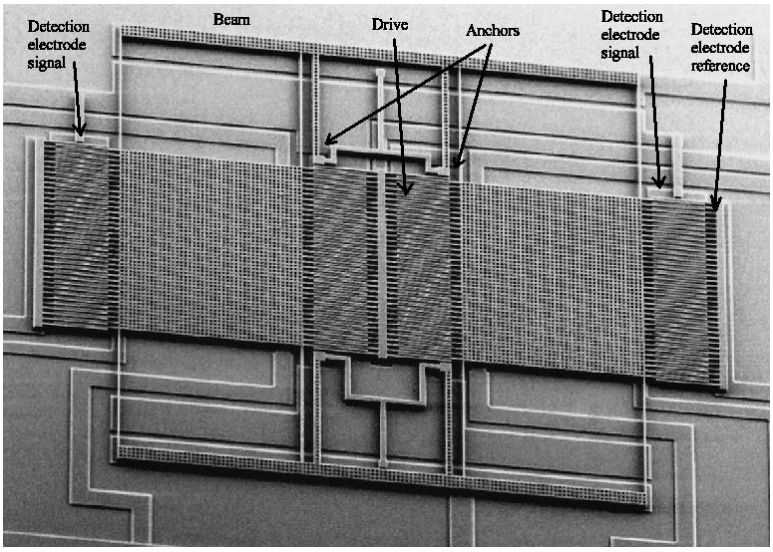
Pictures courtesy of ISIT

First thing to realize is that the oscillating mass is not the orderly *mass point* of basic mechanics but some extended piece of probably **Si** with a complex shape. The second thing to realize is that there are no neat little springs. The "beams" that define parts of the oscillating mass are also part of the "spring".

- Being sophisticated Materials Science students, we know that the points raised above are no major problems. There are ways to calculate (numerically if necessary) how that contraption oscillates. The "spring" results from the mechanical properties of **Si**; ***Young's modulus*** will come up for sure. That takes care of ***m*** and k_S "somehow".
- Good enough. But what about the k_F, the friction or damping constant? Now we are in deep water - or can you make an educated guess on that topic? No, you cannot - neither can I. Suffice it to say that the main friction or energy dissipation mechanism, is "air damping" i. e. the kind of damping you experience if you ***fawn*** a ***fan*** around.
- That offers an opportunity and a problem: Adjusting the pressure inside the packaged chip allows to pick just the right damping for optimal functionality (that we "somehow" determined). That is the opportunities and that is what we actually really do. The problem is: How can you guarantee that this pressure will stay at its necessary value for **20** years or so? This is a first aspect of one of the tougher problems in **MEMS**: How to ensure long-time **reliability**.

Closer to Reality

Now let's take a look at the real thing in a relatively simple, but by todays R&D standards outdated, version:

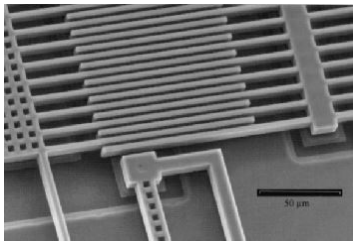


Picture courtesy of ISIT

What you see is

1. A formidable problem in modelling the oscillatory behavior of that contraption with the aim of coming up with an optimal design.
2. An equally formidable challenge in making what you designed.

And we haven't even discussed how the driving and the signal detection works. But that is obvious from just looking at the picture above or the detail of the drive given below:



Picture courtesy of ISIT

- Driving the oscillation is *obviously* done electrostatically. The fixed part of the comb is supplied with some alternating voltage, changing the charge from plus to minus at the driving frequency, while the movable part is kept at a fixed charge=potential. This will lead to alternating attractive and repelling forces - the oscillator is driven accordingly
- Detection is done capacitively. The interdigitated finger structure at the outside for the reference signal detection is just a capacitor with an (alternating) capacity that depends on how deeply the fingers penetrate in the ***x***-direction,
- The two "large" plate under the outside fingers also form a capacitor with an (alternating) capacity that depends on how far the fingers are away in the ***z***-direction.
- In total we get easy-to-process ***AC*** signals with frequencies around the two resonance frequencies of the major oscillation modes. One signal we use for a feed-back loop that keeps the oscillation amplitude in the ***x***-direction constant (or at whatever value we like), the other one contains the information we are after: the angular velocity Ω.

Reality

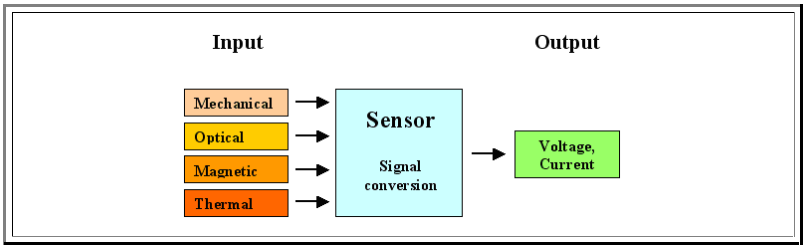
What does a real gyro, one we can buy now (**2007**) look like, and what are the salient points in designing and making it?

- Forget it. Let's just say that there is a lot more to designing and making a competitive gyro than we can discuss in the limited space here. There are people out there (called Materials Scientists and Engineers) who know how to do this. One day, if you keep at it, you might join their ranks.
- Just *one* example for what that could mean: The pictures here are mostly form the **2007 PhD** thesis of Dr. O. **Schwarzelbach** (**ISIT**) who got his degree with an involved analysis (plus first designs) of a gyro relying on *non-linear* oscillatory behavior.

7.1.3 MEMS Sensors and Actuators

Sensors General

- Many if not all MEMS devices could be described as being either a **sensor** or an **actuator**
  - Accelerometers and gyros are *sensors* because the convert the non-electrical input "acceleration" or "angular velocity" into electrical signals - and that's what sensors do.
  - The DLP Chip is an *actuator* because it converts electrical signals to mechanical displacements of mirrors.
- It is thus a good idea to take a general look at some of the principles of sensors and actuators incorporated in MEMS devices. Let's look at sensors first.
  - In general, we have the following simple situation:



  - We have all kinds of **input** signals - mostly but not always from the four categories shown - and want an electrical signal as output.
  - The sensor is supposed to have two major properties.

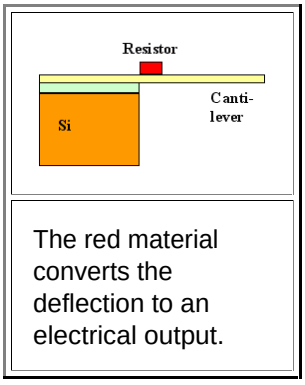
    1. Maximum response to whatever is to be detected - in other words: large **sensitivity**
    2. No or very small response to all other inputs - in other words: very small **cross-sensitivity** or a high **selectivity**.

  - What we need to have inside the sensor is some kind of detector coupled to something that produces an electrical signal.
- Let's look at some examples to make this less abstract. We have a mechanical input - pressure, acceleration, angular velocity, vibration, ..., whatever.
  - Inside the sensor something will respond by moving - a membrane bows according to pressure, a cantilever bends upon acceleration, a vibrating gyro mass starts to wobble when encountering angular velocity, and so on.
- Converting this **movement** to electrical signals can now be done in a number of ways.
- The **stress** or **strain** in the moving part of the sensor is measured. That can be done, for example, by using the following "effects".

  - **Piezoresistive effect**. *All* materials change their resistivity if their dimensions change; with *some* materials (including **Si**) the effect is far larger than what would be expected from geometry alone. This is called piezo*resistive* effect and the reasons for this effect cannot be simply explained; you have to delve deeply into band theory for this.

    Class Exercise: ***Calculate*** $\Delta R/R$ *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*

  - The change of the resistance if a piece of poly-**Si** sitting on a cantilever as shown may be **10 - 50** times larger than what on would it expect from the geometry change alone. Measuring this change allows to determine ϵ and thus the deflection of the beam.
  - **Piezoelectric effect**. Some materials, alway insulators, if "squeezed" become electrically polarized, i.e. develop a potential difference between the surfaces perpendicular to the stress direction; details can be found in the link.
  - Typical piezoelectric materials are quartz (crystalline SiO_2) and, most important "***PZT***"; short for a mixture of $PbTiO_3$ and $ZrTiO_3$; some details can be found in the link.
  - Piezoelectric materials behave like a charged capacitor with the charge depending on the strain ϵ and thus allow to determine the deflection

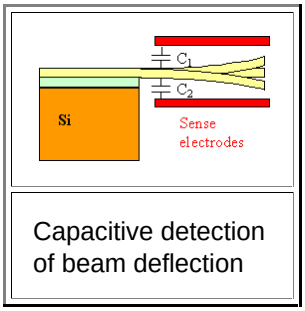


The red material converts the deflection to an electrical output.

- The word "piezo" implies that there is some connection between the two effects, but that is not so. Materials showing large piezo-resistive effects are not piezo-electric and vice verse

The **deflection** in the moving part of the sensor is measured. That can be done, for example, in the following ways:

- **Capacitive** sensing. The moving part is close to some fixed-position electrode (on one side or on both sides). Both parts form a capacitor with a capacity ***C*** that depends on the precise geometry and thus changes whenever the cantilever or the membrane moves.
- The movement can be "up and down" as shown in the figure or "in and out". The gyro dealt with before contains two types of capacitive sensors as can be easily seen: comb structures and "large" electrodes on the substrate.
- **Magnetic** and inductive sensing is possible if one uses a ferromagnetic material - e.g. some **Ni** on the cantilever. This is not a method easily implemented in **Si** MEMS, however.
- **Optical** methods might be used (as in any ***AFM***); but once more not easily in **Si** MEMS technology
- **Thermal transfer**, in contrast, might be used with **Si** MEMS - but not so much for detecting deflections of beams and membranes



Capacitive detection of beam deflection

In essence, we are left either with capacitive methods for the detection of the deflection of some moveable part, or we use "piezo" effects.

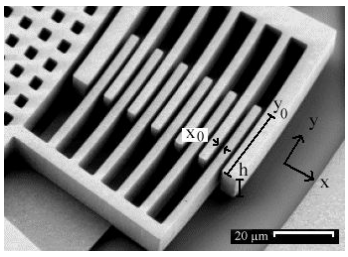
We will not discuss the other three input case here. It is clear that we have many ways in principle to realize the sensing of optical, magnetic, thermal or other inputs, but that the realization and optimization of a suitable MEMS device will take a lot of **R&D** for every individual case.

- It is equally clear that we need special materials for this, and that progress depends to some extent on the discovery of new effects and materials. One example for this is the discovery that **Si** nanowires show a "giant" piezoresistive effect (about **40** times large than bulk **Si**[1)]) and thus might help to increase the sensitivity of MEMS devices in years to come.

Actuators General

Take the diagram from above and read it backwards - now you have an *actuator*. Some electrical input produces an action.

- Of course, the "action" might be the production of magnetic field, light, or some heat, but usually it is a *mechanical movement* we want. This can be induced by producing a force that pulls or pushes at something directly, or by just producing some pressure that acts on all surfaces the same.



Let's see what we have for that.

- Very prominent are **electrostatic actuators**. Just use your capacitor structure, apply a voltage, and there will be forces. If we look at the structure at right, the question is, of course, what force?
- Well, there is some capacity C_{comb} between a finger of the comb and a fixed plate. If we apply a potential difference ***U*** to the capacitor, some electrostatic energy E_C is stored in the capacitor given by

$$E_C = \tfrac{1}{2}\, C \cdot U^2$$

- This is the key equation for "capacitive forces" and you should try to derive it (in case of doubt look at the exercise below).

The force F_{Comb} pulling or pushing on one element of the comb in some direction then is simply given by the proper (negative) derivative:

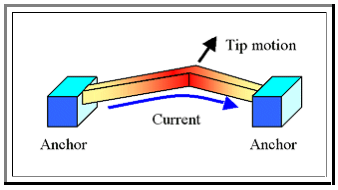
$$F_{Comb} = -\frac{U^2}{2} \cdot \frac{dC}{dy} = \frac{\epsilon \cdot h \cdot U^2}{x_0}$$

This calls for an exercise, of course.

Exercise 7.1-2

Capacitors and Forces

Besides using capacitors for driving some (small!) movement, **thermal actuators** are in use.

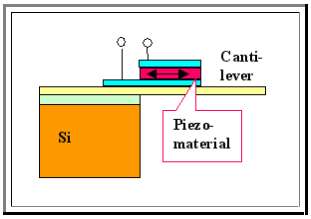


- The principle, as shown in the figure, is simple and self-explaining. Thermal expansion of a double-anchored beam will produce deflection as shown.
- Advantages are very high forces and relatively large deflections (depending on the geometry, of course).
- Disadvantages are relatively high energy consumption and relatively sluggish response times.

With double layers ("bimetal principle") movement with just one anchor point can be obtained, too.

- Thermal actuation is used for a variety of applications; most prominent, perhaps, are **inkjet** systems for printers

Relatively large mechanical forces can also be produced be **piezoelectric materials**. For this, we use the piezoelectric sensing mentioned above in reverse



- The same materials as for sensing can be used, i.e. crystalline quartz or **PZT**, but there are many other piezoelectric materials.
- There is a problem, however. Only materials with no inversion symmetry can be piezoelectric, and that condition excludes all "simple" crystals. Materials with a Perowskite structure (often of the type $\mathbf{ABO_3}$) like $\mathbf{LiNiO_3}$ are piezoelectric, but also others like **ZnO** or **AlN**.
- This should give you a hint why piezoelectric drivers are difficult to implement. We have materials that are hard to deposit. While a very thin layer that can be deposited by sputtering in reasonable times might be good enough for sensing, the amount of force built up in a piezoelectric material is tied to its volume, and we may need thicker layers, not easily handled by present day (**2007**) technology.
- However, progress has been made and piezoelectric drives in MEMS may have a bright future.

All of the above can only give a taste treat of MEMS sensors and actuators. How important the topic is, become clear if you consider that all the computing power of microelectronics and so on comes to nought if there is no input and no output.

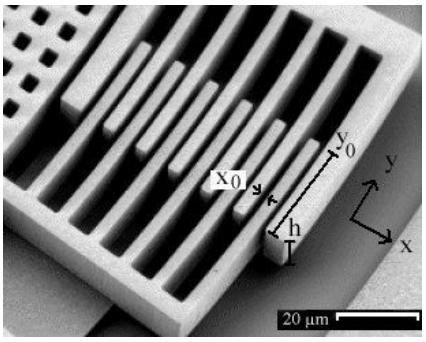
- By necessity, input and output means sensors and actuators,. Not necessarily only mechanical output, and not necessarily MEMS devices. A flat panel display or a **DLP**-based beamer is an output system; the first one without MEMS technology, the second one without it. Your keypad is an input device. Is there MEMS inside? We leave it at that.

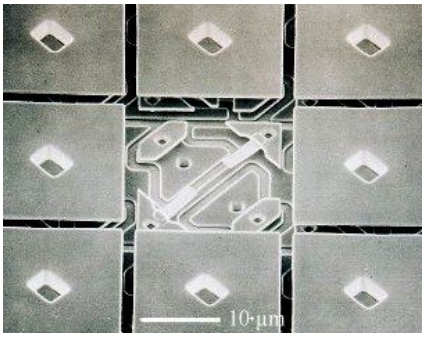
---

[1] R. He, P.Yang Nature Nanotechnology 1(2006) 42-46, Oct. 2006)

7.1.4 Summary to: 7.1 MEMS - Products and Developments

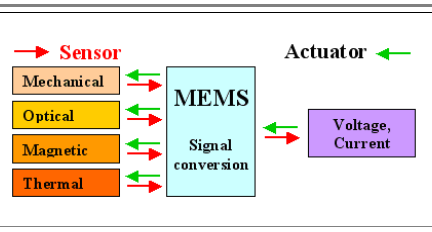
- **MEMS** are "Micro Electro Mechanical Systems" including also micro optics, micro fluidics and generally meaning micro systems.
  - **MEMS** uses **Si** substrates and technologies because "it is there and cheap" for the non-electronic part *and* because electronic components can be integrated on the same chip.
  - Examples of high-volume **MEMS** products are
    - (Pressure) sensors.
    - Accelerometers.
    - Gyros
    - "Beamer" chips (**DLP**)
  - More products are to come; **MEMS** is an *emerging* and often an *enabling* technology





- **Gyros** are particular complex **MEMS** sensor products with a huge range of applications.
  - There must be a physical principle behind the sensor design; different approaches can be used.
  - One approach uses the **Coriolis force** causing detectable additional vibrations in an oscillator with two degrees of freedom if some rotation is experienced.

- Many **MEMS** devices are either sensors or actuators.
  - Looking only at mechanical **MEMS**, there is a need to couple mechanical movements to electrical signals and vice verse.
  - Ways to do this include.
    - Capacitive coupling
    - Piezoelectric and piezoresisitive coupling.
    - Thermal coupling (expansion, resistivity changes).
    - Magnetic coupling.
    - Optical coupling.
  - There is no "ideal" coupling; all methods suffer from certain problems.



Exercise 7.1-1

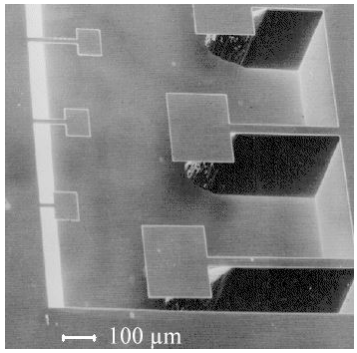
All Questions to 7.1

7.2 Processes and Specialities

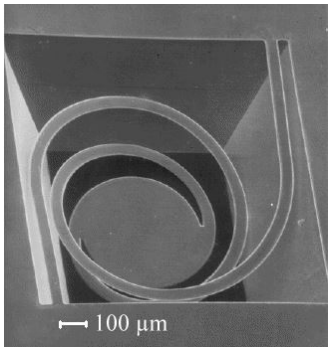
7.2.1 Processes and Materials - General Consideration

General Processing

- How do we make a MEMS device? By using all processes and tricks we know from microelectronics *plus* a growing number of special processes and materials developed and optimized for one of the many MEMS products.
- There are a few general processes and materials specific for most of MEMS technology as can be appreciated by just looking at the pictures of MEMS structures already shown. In particular we often have free-standing structures, only supported at a few points or not at all
  - Let's look at two examples from early MEMS research to clarify this point:

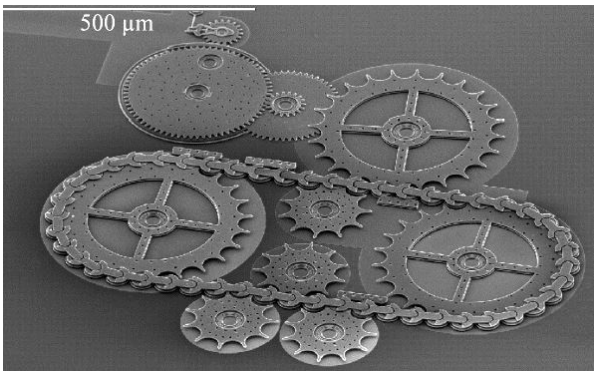


Courtesy of ISIT



Courtesy of ISIT

  - 1. We need to be able to make deep **cavities** into the Si substrate.
    2. We need to be able to make "**cantilevers**" as the paradigm of a free-standing something.
    3. We need to be able to make **membranes**.
  - The pictures from very early (**1982**) MEMS **R&D** illustrate the first two points quite nicely; the third point follows as a kind of ***corollary*** if you consider devices like pressure sensors.
  - The left picture also shows one of the many problems we may encounter: There is still some **Si** under the three large cantilevers - whatever etching procedure produced the cavity below the cantilevers was not yet quite done below the larger ones.
- If we contemplate these structures (and maybe this one) again, and consider what the finished product must look like, we will also be forced to conclude the following:
  - We must avoid **stress** in the various layers at all cost.
  - We must be able to keep things absolutely **planar**- even on top of "topology".
  - We must find special ways to hermetically seal some devices (like the gyro) without impeding the movement of the oscillating parts; i.e. we must have special **packaging technologies**.
  - Why is it essential to avoid stress? Find out yourself!
    Class Exercise: *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?*
- All things considered, MEMS is really a system technology. While microelectronics can go a long way with standardized technologies, procedures and software tools, for MEMS many processes are component and application specific, with strong interdependencies between design, electronic interfaces, system integration, packaging and testing.
  - This means that even a very large box of tools and processes - far more than you would need to produce any sophisticated **IC** - might not yet be good enough for that particular MEMS device you are after and that means that we cannot even get close to a comprehensive survey here.
- Let's finish this paragraph by considering what we do *not* want to make: Scaled down versions of mechanical gadgets, a sort of ultra precision mechanics as shown in this picture:



Courtesy of SANDIA Labs.

- Pictures like that are very popular and shown a lot in "public science" magazines. Everybody can immediately see what that is and marvel at its tiny dimensions, while a picture like the one in the link would be received with a shrug.

But scaled down versions of **gear wheels**, and so on, are rather useless. They don't work for very long, if at all, because in the micro- or nano-world, things often behave quite differently from what would expect.

Gravitational forces, for example, become unimportant and other forces take over (why do small dust particles not fall to the ground?).

- **Friction** and **surface tension** may become extremely important because the surface to volume ratio increases if things get smaller, and forces transmitted via the surface (including friction) may dominate over forces scaling with the volume (like gravity or ***inertia***).

The key word now is "**stiction**", a newly coined word, meaning that parts of a MEMS device that should be able to move *stick* to other parts, effectively locking all movement.

- The word also alludes to "friction" because whenever microscopic entities slide across each other, friction, as always, does occurs - but in ways quite different from what we experience in the macroscopic world. Friction may then lead to stiction and "dead" MEMS devices.
- Avoiding stiction is a major concern in designing and making of MEMS structures; more about it in the link. Generally, one should avoid sliding movements - and that is the reason why designing device functionality by having a piece **A** sliding on a piece **B**, as in the gear wheel picture above, is not a good idea.
- The problem is that there is no simple way of obtaining efficient **lubrication** in those devices. You should now take a minute to think about the overwhelming importance of **lubrication** in everyday mechanical systems like cars, and what our civilization would look like if stiction would be a tough-to-fight macroscopic phenomena, too.

Special Materials

In terms of materials we use more or less everything microelectronics has to offer, and then some. We have already encountered some MEMS specific materials directly or indirectly. Let's see:

- Thermal couples came up in the context of a gyro. We need two different conductors (metals) to make one.
- The DLP chip works with mirrors. We need a highly reflective material (= metal) for that.
- Piezoresistors came up a few times, the first time in a pressure sensor.
- Special packaging needs have appeared a few times, necessitating special materials.
- And we can be sure that there is a lot we have not yet covered.

- We know from microelectronics that the introduction of a new material into a process sequence takes a lot of **R&D** and therefore time and money. While in microelectronics we always have "Moore's law" as a powerful economic driver, MEMS materials are usually very application specific and progress may be slower due to economical restraints.

But lest's be clear about one thing:

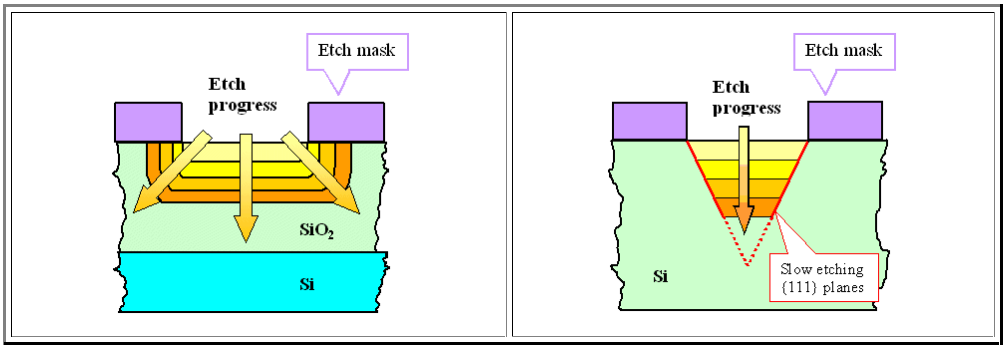
> Progress in MEMS technology comes from specific **Materials** and **Processes**

- The reason for that is simple: Progress in general **Si** processing comes from microelectronics - and is taken over by MEMS as soon as it become affordable. MEMS by itself could never "pay" for, let's say, the next generation lithography technology.
- For Materials Scientists and Engineers this are good news - you are going to be needed!

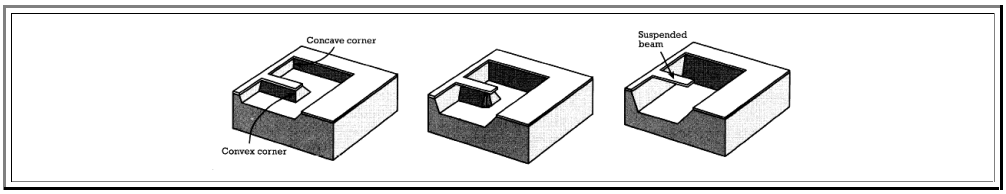
7.2.2 Important MEMS Processes

Cavities

- As we have seen, we need to make relatively large cavities. For doing that we have to remove relatively large volumes of **Si**, and this we can do by either wet chemical etching or by plasma etching - as in microelectronics.
  - However, plasma etching in microelectronics has not been engineered to remove a lot of **Si**. If we would use existing processes for etching cavities, it would take forever and probably would run out of producing the structures desired with increasing depth.
  - So let's modify those processes! Easier said than done. In fact there is one special process for deep etching with plasma, that we will look at below, Here we must now resort to the still rather **black art** of wet etching.
- Using isotropic chemical etchants thus as **HF** / **HNO_3** / **CH_3COOH** (very unhealthy!) would just produce a big hole with lateral dimensions no smaller than twice the etching depth and some semiglobular shape.
  - *Fortunately*, there is such a thing as **anisotropic chemical etching**; the most prominent chemical for this is **KOH** (at somewhat elevated temperatures like **80 ^oC**. *Unfortunately*, alkali metals like ***potassium*** (**K**) are ***anathema*** for **Si** microelectronics (and thus for MEMS with some electronics integrated on the chip).
  - The figures compare schematically isotropic and anisotropic etching.

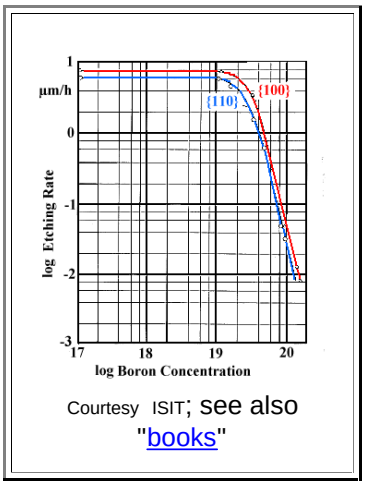


- We cannot use **KOH** for the reason given above, but its anisotropic etching properties gives us the idea to try other basic (instead of acidic) etchants.
  - What works and what we we can use is, for example, aqueous ***TMAH***(=tetra-methyl ammonium hydroxide; **$(CH_3)_4NOH$**) solutions in water (**5 % - 30 % TMAH**in **H_2O**), a harmless chemical from the viewpoint of easily-contaminated **Si**.
  - Often (but not always) the **{111}** planes act as "**stop planes**" for anisotropic etchants, meaning that the etching or dissolution rate on this plane is far slower than on other planes.
  - In fact, the etching rate with these anisotropic etchants on **{111}** planes can be about **1 000** times smaller than on the other low-indexed planes like **{100}** or **{110}**; i.e. the **selectivity** is very large.
  - If you start in a square opened in some chemistry-resistant mask on typical **(100) Si**, you will start to etch a perfect inverse truncated pyramid. If you etch long enough, a perfect pyramid with **{111}** sides will result. as shown below - we have already seen nice examples.
  - In the case of cantilever production as shown below, some more time would be needed to remove the "bridges" below the larger cantilevers. A time sequence would look like this:



- Electrochemical etching may also be highly anisotropic, but is not really used so far in mainstream MEMS technology.

- Why is anisotropic (electro)chemical etching of **Si** a black art? There is an easy simple answer, and a more general one. The simple answer is simple: Nobody, for example, can calculate the selectivity of the etchants named above (as a function of concentration, temperature, doping, and so on; of course) from "first" or even second principles.
  - Now let's be more general: If you consider the **parameter space** of an etching experiment - it is very large! You can have many compositions of the basic etchant (including those "magical" drops of this or that); the temperature and the doping of the **Si** are parameters of importance. If you actually do an *electro*chemical experiment, you have voltage and current as additional parameters. The results may even depend on the state of illumination of the the **Si**.
  - Let's assume you have found a set of parameters, i.e. one "point" in this large parameter space, that works. Now change one of the many parameters somewhat. Can you (or anybody else) predict theoretically (i.e. without trying out, or reverting to experiments already done before) what is going to happen? The answer is: Most likely not. Sometimes, predictions can be made with a high level of confidence, quite often unforeseen and amazing things happen. Understanding etching of semiconductors in general therefore is still a wide open field of research



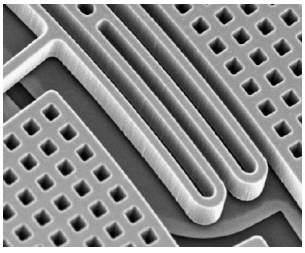
Courtesy ISIT; see also "books"

- Nevertheless, with anisotropic chemical etching we now have a way to make cavities in principle. This is great, but not good enough. We do not only need a large selectivity with respect to crystallographic planes, we also need selectivity with respect to the doping of the **Si**. We will see why in the next sub-chapter.
  - Fortunately (i.e. not predicted), for highly Boron (**B**) doped **Si**, i.e. **p^+-Si**, the etching rates in all direction become very small, i.e. **p^+-Si** does not dissolve anymore in basic etchants.
  - The effect is quite dramatic: As soon as some minimum **B**-concentration **[B]** has been reached, the etching rate decreases with the fourth power of **[B]** as shown in the figure.

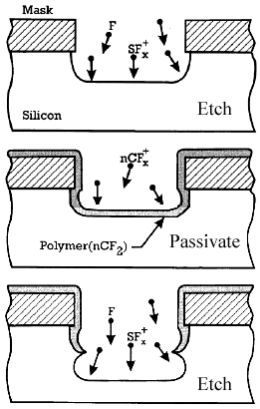
The Bosch Process

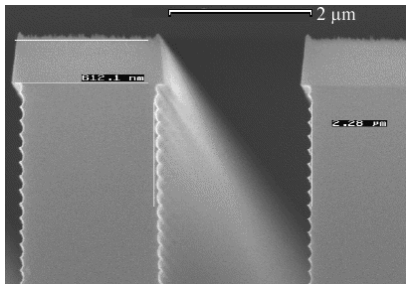
- As mentioned above, the large variety of **plasma etching** (also called **dry etching**) process developed for microelectronics are, of course, used in MEMS technology in many ways, but are simply not meeting all etching needs.

  - Look at the free moving part of a gyro in the picture. It consists of poly-**Si** about **10 µm** thick that was "etched out" of a solid poly-**Si** layer with very high precision. It has all those little square holes in it for reason we will encounter later - but they need to be made.
  - The solution to many of the dry etching or plasma etching problems encountered in MEMS technology is the **high-rate plasma etching** or the "**Bosch process**", patented in **1994** by Bosch
  - The Bosch process essentially decouples the two necessary ingredients for anisotropic etching: **1.** dissolving **Si** at the bottom, and **2.** passivating the side walls, i.e. protecting them against etching.


Courtesy of ISIT

- The process alternates every **10 s** or so between etching **Si** and depositing some polymer on the walls. The polymer deposited on the bottom of the area to be etches is taken off much faster than the coating of the side walls, allowing some net etching into the depth at rather high speeds and for (almost) arbitrary depth.
  - The etching process uses essentially SF_6 plasma chemistry.
  - Passivation uses a fluorocarbon process, using gases like C_4F_8 , C_3F_6 or CHF_3.
  - Etching rates of **> 10 µm/min** per minute are possible, allowing cost-effective processing even for deep etching, including all the way through a (thinned to about **300 µm**) wafer.
  - The frequent switching of gases and etching parameters makes the system quite complicated (and expensive). But the Bosch process is one of the key processes for MEMS and probably will ***spawn*** many variants in due time.



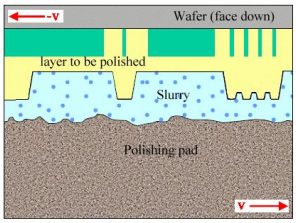


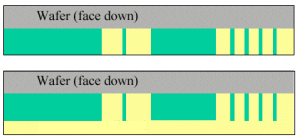
Courtesy of ISIT

The ***scalloped*** shape is typical for the Bosch process.

Chemical Mechanical Polishing

- We have encountered **chemical-mechanical polishing** or ***CMP*** before - in one sentence.
  - It is a key process for both microelectronics and MEMS. Microelectronics, however, thrived without **CMP** for more than **20** years, while many MEMS products would just not be possible without **CMP**.
  - How is it done? In principle, it is simply grinding down the surface with an abrasive. This is the *mechanical* part of polishing. The *chemical* part comes in because your ***slurry*** - the liquid or viscous goo that contains fine particles of the abrasive materials, also contains chemicals that etch off *only* the mechanically damaged **Si**. **KOH**, for an example, would be good at that but is not used for the reason given before.
  - The amazing (and at the time of its introduction quite unbelievable) thing is that you can use **CMP** with nanometer precision on huge **Si** wafers - if you know all the tricks of the trade, of course. Recipes for slurries, the kind of polishing pads to use for the kind of material you want to polish off, and everything else are closely guarded secrets.





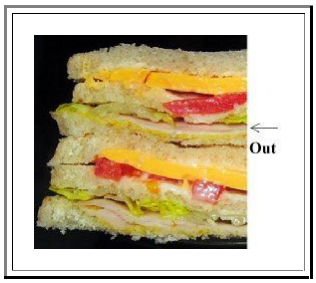
Possible results of planarization by **CMP**What you want to do is for example:
i) Just fill the "holes" in the green layer with something "yellow",
ii) Producing a flat yellow layer on top of the green layer (with holes filled).
Since your deposition technique produced a "rough" yellow layer, you need **CMP**.

- The figure shows the basic principle and what one can achieve.
  - **CMP** may also be counted to some extent among the "black arts". Not only does it have a chemical etching components; its parameter space is extremely large, too, and hard to consider in theory.

Other Typical MEMS Processes

There is much more, but her we can only give a few hints.

Etching of sacrificial layers.



- Free standing structures like in the gyro first are made on top of some *sacrificial layer* that is etched off at an opportune time. These sacrificial layers must meet a number of requirements, the most important one being that they must be compatible with whatever is under and above it, and that there must be a way of etching them off selectively without dissolving anything else.
- This is tough, man! If you just consider the etching process, it's like taking out the ham from your sandwich without opening it up and without doing *anything* to the bread, butter, mayonnaise, salad leaf etc., above and below. This means you can attack the ham only from the side, and that will take a long time during which a lot can go wrong.
- One common trick is to make the top layer (the one to be free standing after the process) full of small holes, so you can attack the sacrificial layer from more places. This is why we always see "perforated" structures, e.g. further up.
- There is no way, of course, to do the etching with plasma. You *must* use "black art" chemistry. On occasion, even gases are being used (e.g. **HF** vapor, an extremely dangerous chemical) because pulling filigree structures out of a liquid may make them stick together (think about your hair in the bathtub!) - even after drying.

Wafer bonding

- Consider an atomically flat **Si** wafer with a perfectly clean (non-oxidized) surface. Now put another wafer with an equally perfect surface on top. The two wafers would instantly bond into one piece of **Si** (with a grain boundary a the interface if they were not exactly at the same orientation).
- Real **Si** wafers are neither atomically flat nor perfectly clean - but they come rather close. If you join them, they will stick together by van-der-Waals bonding. Add a little heat and some mild pressure, and they will bond to on piece of **Si**. Now you have a well-established process that was developed for microelectronics, but is not much used (too expensive).
- In MEMS technology, wafer bonding might be a crucial process. Gyros or optical micro mirrors operating at their resonance frequency, for example, must be kept in a cavity with a precisely established low pressure because that determines the damping of the system. The tough part is to keep that pressure constant for **> 20** years.
- So bond another **Si** wafer (with a cavity) on top. Better hermetically sealing with no long-time stability problem at the interface is not possible, but there are many ways how to do worse.

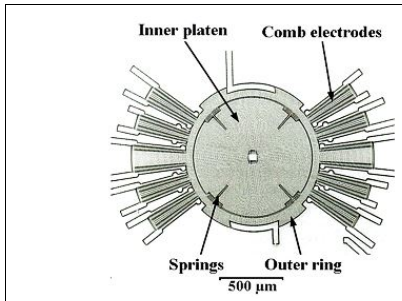
There are many tricks to achieve efficient bonding without having to be ultra-pure and ultra-precise. Moreover, in MEMS technology not only **Si**-wafer to **Si**-wafer bonding is used., but also, for example **Si**-wafer to glass-wafer bonding.

- This is easy to understand. Hermetically sealing your micro-mirrors with a **Si** wafer on top is fine as long as you only want to work only in the infrared. If you want to use *visible* light, your top better be transparent.
- Going through the various material combinations and the technologies used to bond them is far beyond the scope of this lecture course. Suffice it to state that bonding technologies are a major part of MEMS processing.

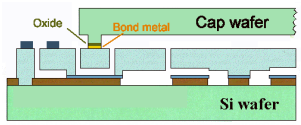
7.2.3 Example for Process Integration

The Goal

- Let's look at an advanced product, a **Ggyro** based on the Coriolis effect as described in some detail in modul 7.1.2.
  - Only a very cursory description will be possible; whatever follows is based on the description given in the **PhD** thesis of W. **Reinert** based on the work he performed at the **ISIT**. Of course, the finished product resulted from the work of many people who cannot all be credited in this context.
- Here is what we would like to produce:



Courtesy ISIT



Courtesy ISIT

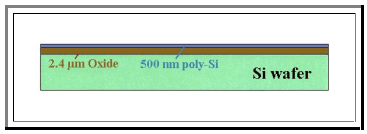
Not to scale, of course!

- Details, e.g. parts of the comb drive, you have seen before. Now let's look at the process sequence needed for manufacturing this MEMS device.

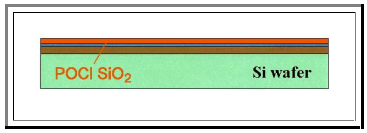
Process Sequence

- We start with some regular **Si** wafer. As far as the **MEMS** part is concerned, it is just a substrate and the only requirements are that it is (extremely) flat, can be oxidized, and is process compatible - all **Si** wafer meet these requirements.

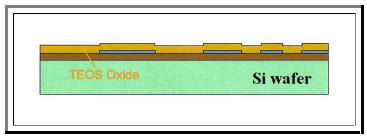
- If you want to have some electronic circuitry on the same substrate, you must now decide upon doping type and concentration, too. Here we take an **n**-type **1.5** Ω**cm** wafer.
- As a first process (always after cleaning etc.) we form a rather thick oxide - **2.4 µm** - that serves in particular as an electrical insulation even at rather high voltages.
  - Obviously (??) we do that with wet oxidation.
- On top of that oxide we deposit **500 nm** of poly-**Si** with a CVD process. It will serve as a buried conductor and therefore is called "Buried poly".



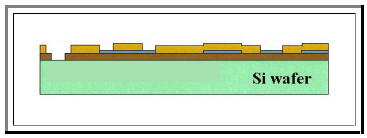
- The poly-**Si** needs to be highly doped if is is to be a conductor, for that we use the old-fashioned (but simple) **POCl process**.
  - **Ar** gas is bubbled through liquid **$POCl_3$**; on the hot wafer heavily **P**-doped **SiO_2** forms that is used as a diffusion source for diffusion of **P** into the poly-**Si**.
  - After the diffusion the **POCl-SiO_2** is removed in a so-called "de-glazing" process (wet chemistry).



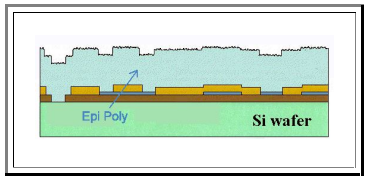
- In the next step the buried poly is structured (lithography and reactive ion etching (***RIE***).
- Then the structured buried poly is completely covered with a **1.6 µm** thick **SiO_2** layer, which will be the sacrificial layer for forming free-standing structures.
  - "Obviously" we use a TEOS CVD process for forming this oxide - fast, low temperature and stress-free.



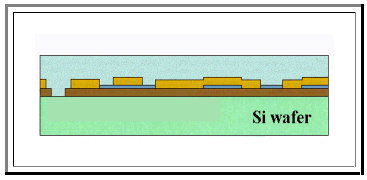
- Next, the sacrificial oxide layer is structured by a **2nd** lithography process, followed by wet chemical etching for removing the exposed oxide. Since the lateral structures are rather large this is good enough.
- The third lithography step follows right away, generating some openings in the first oxide needed for contacts to the substrate. All this contacts must be below the openings in the sacrificial oxide for obvious geometric reasons.



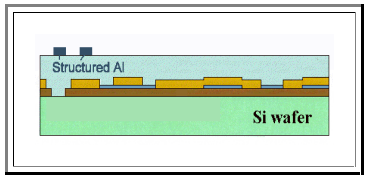
- Now we use a first process very special for **MEMS**. We deposit an *extremely* thick layer of poly-**Si** - **13.6 µm**.
  - In a regular **CVD** process this would take forever. So we do it at very high temperature - **1 050 °C**.
  - Since normal **CVD** reactors are not build for this kind of process, we do it in an epitaxial reactor and call the resulting layer, somewhat self-contradictory, "*Epi-poly*". The expression "epi-poly " is certainly a fine ***oxymoron*** but you better get used to things like that.



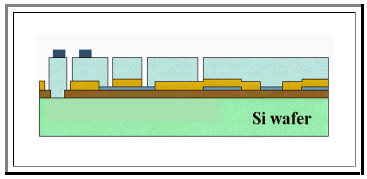
- The epi-poly layer is rather rough, so we now use ***CMP*** (chemical - mechanical polishing), one of the processes we had listed as quite **MEMS** specific, to make it perfectly flat.
  - The epi-poly will contain some mechanical stress - you just can't make a rather thick layer quickly without building up some stress. So we anneal everything ("stress anneal"), for somewhat obscure reasons in an oxygen atmosphere.
  - Since we will then oxidize some of the epi-poly, this oxide needs to be removed in another wet-chemical process step.



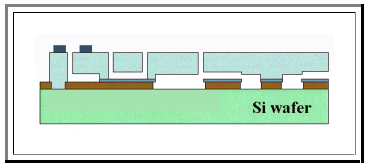
- Now it is time to produce the first metal layer; we take **Al** in this case.
  - It is deposited by sputtering **1.3 µm** on **Al** on top of the epi-poly followed by the **4th** lithography and wet chemical structuring.,



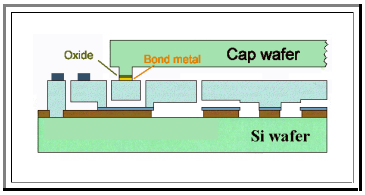
- At this point we need to use another **MEMS** specific process: the *Bosch process* or deep plasma etching. We need to structure the very thick epi-poly layer, and standard microelectronic processes just can't do this.
  - A fifth lithography defines the structures, and the alternating etching -side wall passivation process digs deeply into the layers, possibly creating some steps at interfaces



- A last very typical MEMS process follows: Removal of the sacrificial oxide by **HF** *gas* etching.
  - This is a process we would never use in microelectronics - far too messy and extremely dangerous! So why do we do it?
  - Think about what happens when you wash your hair and withdraw this long flexible fine structures from the liquid water - they stick together because of **capillary forces**.
  - Exactly the same thing happens when you withdraw the flexible fine structures released after dissolving the sacrificial oxide in some liquid - they stick together. But in contrast to your hair they will never ever come apart again - **stiction** takes over.
  - No capillary forces if you etch with gas - but lots of other problems that need to be solved.



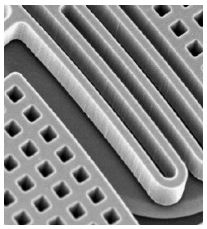
- All that remains to be done is to seal the gyro hermetically at a defined pressure that will not change in the next **20** years or so.
  - For this we bond a pre-structured **Si** wafer with some bond layers over the sensitive part, leaving the electrical connections free. We now have used thelast MEMS specific process: *Wafer Bonding*.
  - The total encapsulation process is a whole process sequence in its own right, but here we leave it at that..



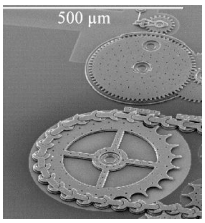
- Now we have a functioning gyro. If it works to specification, you now can sell it - for a few **€** a piece, if you are lucky.

7.2.4 Summary to: 7.2 Processes and Specialities

- MEMS uses all of "known" **Si** technology and has some specifics of its own.
  - Making cantilevers and membranes necessitates making "large" cavities.
  - Staying absolutely planar and stress-free is essential
  - Packaging can be far more demanding than for chips (e.g. transparent tops for **OMEMS**, keeping defined pressures for **> 10 a** in gyros).
- The ***bane*** of **MEMS** is **stiction**.
  - If you can't lubricate, it will stick sooner or later. Never bring moving parts in contact!
  - **MEMS** design therefore cannot just miniaturize exiting mechanical designs; it must look for new approaches.



Works



Sticks

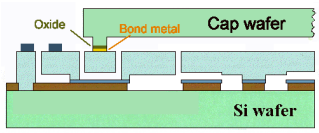
- **MEMS** employs some special processes and materials; they are the drivers of progress
  - Anisotropic chemical etching
  - High-rate plasma etching ("Bosch process")
  - Chemical-mechanical polishing
  - Sacrificial layers and removal (including chemical etching with "vapors")
  - Wafer bonding; in particular for packaging.

Making "large" cavities and extremely deep "holes"

Planarization

Free-standing structures

- Process integration looks simple if compared to an advanced **CMOS** process, but is actually rather involved due to the special processes needed and quality requirements



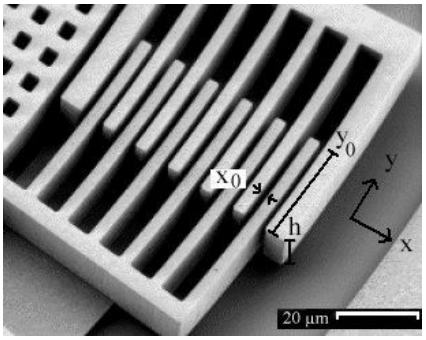
Exercise 7.2-1

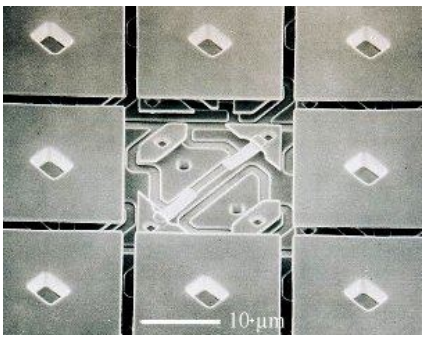
All Questiosn to 7.2

7.3 Summary

7.3.1 Summary to 7: MEMS

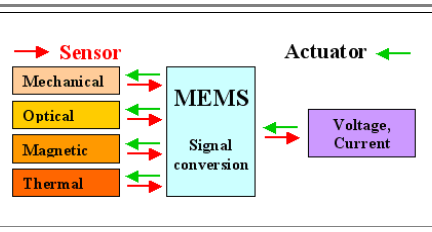
- **MEMS** are "Micro Electro Mechanical Systems" including also micro optics, micro fluidics and generally meaning micro systems.
  - **MEMS** uses **Si** substrates and technologies because "it is there and cheap" for the non-electronic part *and* because electronic components can be integrated on the same chip.
  - Examples of high-volume **MEMS** products are
    - (Pressure) sensors.
    - Accelerometers.
    - Gyros
    - "Beamer" chips (**DLP**)
  - More products are to come; **MEMS** is an *emerging* and often an *enabling* technology



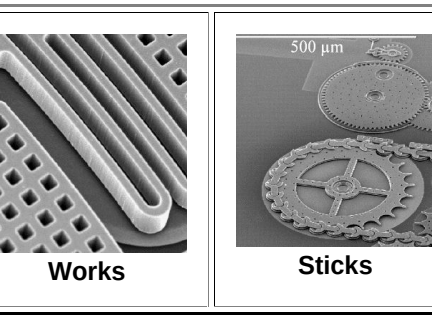


- **Gyros** are particular complex **MEMS** sensor products with a huge range of applications.
  - There must be a physical principle behind the sensor design; different approaches can be used.
  - One approach uses the **Coriolis force** causing detectable additional vibrations in an oscillator with two degrees of freedom if some rotation is experienced.

- Many **MEMS** devices are either sensors or actuators.
  - Looking only at mechanical **MEMS**, there is a need to couple mechanical movements to electrical signals and vice verse.
  - Ways to do this include.
    - Capacitive coupling
    - Piezoelectric and piezoresisitive coupling.
    - Thermal coupling (expansion, resistivity changes).
    - Magnetic coupling.
    - Optical coupling.
  - There is no "ideal" coupling; all methods suffer from certain problems.



- MEMS uses all of "known" **Si** technology and has some specifics of its own.
  - Making cantilevers and membranes necessitates making "large" cavities.
  - Staying absolutely planar and stress-free is essential
  - Packaging can be far more demanding than for chips (e.g. transparent tops for **OMEMS**, keeping defined pressures for **> 10 a** in gyros).
- The ***bane*** of **MEMS** is **stiction**.
  - If you can't lubricate, it will stick sooner or later. Never bring moving parts in contact!
  - **MEMS** design therefore cannot just miniaturize exiting mechanical designs; it must look for new approaches.



- **MEMS** employs some special processes and materials; they are the drivers of progress

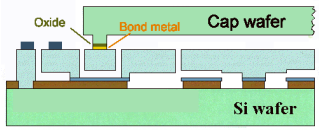
- Anisotropic chemical etching
- High-rate plasma etching ("Bosch process")
- Chemical-mechanical polishing
- Sacrificial layers and removal (including chemical etching with "vapors")
- Wafer bonding; in particular for packaging.

Making "large" cavities and extremely deep "holes"

Planarization

Free-standing structures

Process integration looks simple if compared to an advanced **CMOS** process, but is actually rather involved due to the special processes needed and quality requirements



Exercise 7.3-1

All Questions to 7.

8. Solar Cells

8.1 General Concerns

8.1.1 Basic Solar Cell Topics

8.1.2 Solar Cell Current-Voltage Characteristics and Equivalent Circuit Diagram

8.1.3 Summary to: 8.1 Solar Cells - General Concerns

8.2 Making Bulk Si Solar Cells

8.2.1 Production Necessities and Silicon Starting Material

8.2.2 Processes and Process Integration

8.2.3 Summary to: 8.2 Making Bulk Si Solar Cells

8.3 Making Thin Film Solar Cells

8.3.1 Types of Thin Film Solar Cells

8.3.2 Essentials of Producing Thin Film Solar Cells

8.3.3 Summary to: 8.3 Making Thin Film Solar Cells

8.4 Summary

8.4.1 Summary to: 8 Solar Cells

8. Solar Cells

8.1 General Concerns

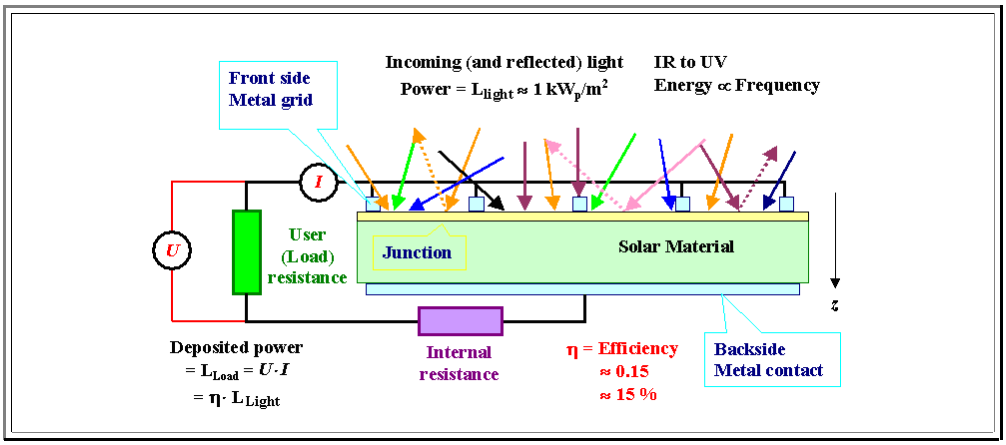
8.1.1 Basic Solar Cell Topics

Sun, Planet, and People

- At present, *we* have to make ends meet on *our* **planet** called **Earth** or **Terra**, and with *our* sun, called **Sun**. In all likelihood, this will not change in the foreseeable future.
  - "*We*" refers to *people* or **humans**, since it appears we are the only species on earth using energy *in excess* of what we need for just staying alive. We will see somewhat later or in this link just how much energy we need for staying alive (in a certain style).
  - What most certainly will change in the foreseeable future is *either* the way we supply ourselves with energy *or* the number of humans living (sort of) on Terra. Note that neither the planet nor the other expressions of live on Terra would give a damn about strongly declining numbers of Humans (they probably would be even very much in favor of that prospect).
- Since this Hyperscript is for that subgroup of the humans species that has a minimum of brain power *and* is willing to use it (this excludes, for example, at least one **American president**), we know that the future energy needs of humankind will have to be met by **solar energy** (in all forms: wind, "bio", solar heating, solar electricity, waves, tides, ...) and, if you like it or not, nuclear energy.
- So let's look at the most important hard facts first. They come from a host of numbers and relations, some of which are contained in the following links.
  - *World energy essentials*: Read the article Future Global Energy Prosperity: The Terawatt Challenge of Nobel prize winner **Richard Smalley** published in the MRS Bulletin **30, 2005**, and the article "Powering the Planet" of Nathan S. Lewis in the MRS Bulletin **32, 2007** to get a flavor of what it is all about.
  - Solar cell essentials
    This is the solar cell module from "Einführung in dieMaterialwissenschaft II", with which you should be thoroughly familiar.
  - Joules, Watts and kWh's
    This is a "basic" module within this Hyperscript, giving a few basic numbers and relations that you should get very well acquainted with. Your life will depend on it!
  - Silicon for solar cells.
    A module with a bit of the **30** year history around the efforts to generate cheap solar **Si**.
  - Poly Silicon solar cells.
    Some facts and stories around the "**multi crystalline**" solar cell.
  - Some (old) solar cell data.
    "Old" refers to the fact that solar cell research, development and production grows so fast that the above module from around **2004** is already quite outdated
  - ***FAQ***'s around solar cells.
    An "advanced" module of this Hyperscript with some more information about the applicability of solar cells for energy production. It is only "advanced" because our task here is to look *only* at the technology for making solar cells: not the *application* of solar cells.
  - Solarzellen und Materialwissenschaft A German Powerpoint Presentation to the general topic

Solar Cell Primer

- Now let's look at the basics of solar cells as shown in the figure below.

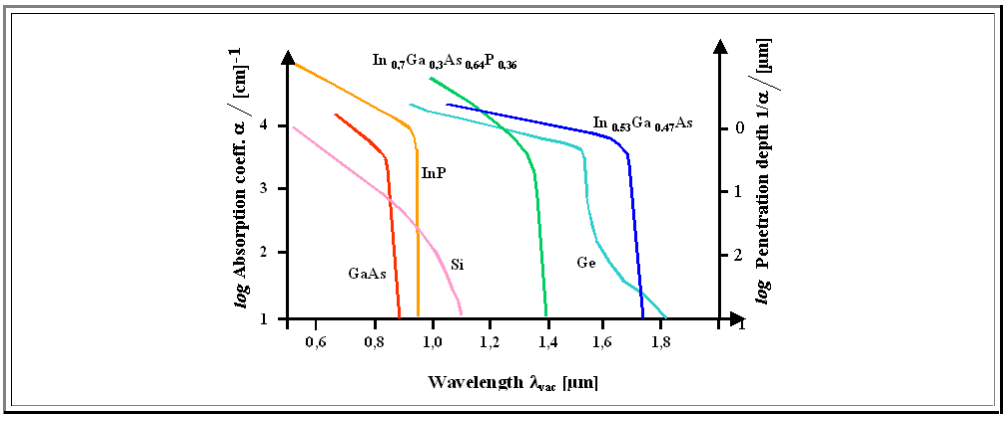


Light: We usually have a front side that absorbs the incoming *light* (there are a few solar cell concepts where the light can be absorbed on both sides that we will neglect here).

- The light has all kinds of frequencies according to its spectral distribution - this is pretty much a given.
- It is always coming from **all directions** because there is always scattering in the air. Only if you have **direct sun light**, it will come to a large extent from *one direction*, the sun. This has an important consequence: You can focus *direct* sunlight but not indirect sunlight. Turn a large parabolic mirror into the sun and put a relatively small solar cell in its focal point and you harvest the sun energy impinging on, for example, **1 m²** in a solar cell with an area of **1 cm²**. This looks attractive in terms of saving (expensive) solar cell area but the concept is not of much use in cloudy countries like Germany. Moreover, some of the savings from needing fewer solar cells must now go to pay for all kinds of other gadgets, and you have the non-trivial problem of keeping your solar cell cool.
- A large part of the light is reflected if you just take a piece of semiconductor - after all, **Si** is shiny and looks metallic because it reflects light quite nicely. An ideal solar cell should look pitch-black - the fact that you usually can *see* solar cells proves that they are not ideal in this context. Nevertheless, any solar cell must have some anti-reflection technology on its front side; and how to do this (cheaply) is a large part of solar cell technology.
- The light with the intensity $I_0(\delta, \nu)$ with δ= angle of incidence and ν= frequency at the surface, that is not lost by reflection or absorption in the grid, penetrates into the semiconductor and is absorbed according to

$$I(\delta, \nu, z) = I_0(\delta, \nu) \cdot \exp -(z/d_a)$$

- The parameter $d_a(\delta, \nu)$ is called the **penetration depth** of the light; it can be rather large for indirect semiconductors like **Si**, e.g. **10 µm**, but also quite small - a few **nm** - depending on the semiconductor type and the frequency of the light. Obviously, only light that is absorbed inside the semiconductor can contribute to energy conversion.
- Quite often a quantity called **absorption coefficient** $\alpha = 1/d_a$ is used; the figure shows some data (for perpendicular incidence).



Junction: Now we have generated exactly as may electron - hole pairs as we have absorped photons. We send some power through the load - always in the form of a simple resistor - by extracting *some* of the minority carriers produced by the light across a junction, which sweeps minority carriers that reach its **space charge region** (***SCR***) to the grid contact. This way we produce a **photo current density** j_P= number of minorities swept to the contact per **s** and **cm²** and a **photo potential** determined by the junction potential.

- Minority carriers generated by photons that do *not* reach the junction (because, for example, they *recombine* in the bulk or at some internal surface) obviously cannot contribute to the photo-induced current density j_P. Changes for loosing minorities to internal recombination grow if they have a long distance to cover (i.e. if they are produced far away from the junction) or if they find many opportunities to recombine (⇒ high defect density) on their way to the junction. That's why we need "clean" **Si** for good solar cells.

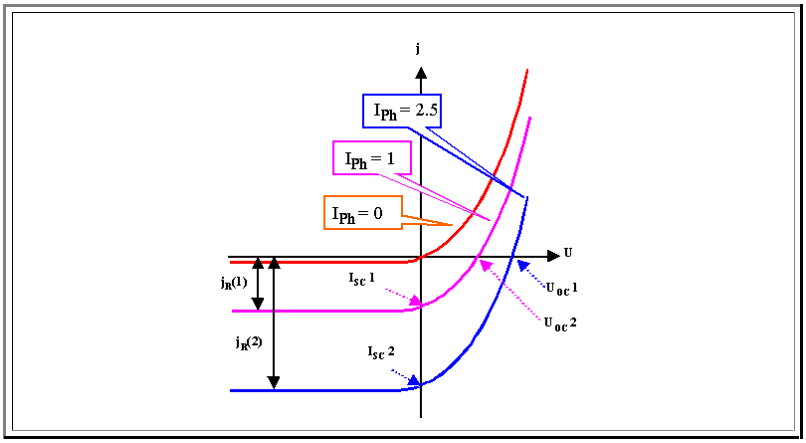
- We can make a suitable junction in many ways but usually we use a **pn-junction**. Schottky junctions, for example, could be (and are) used, too, but the metal part absorbs light. One kind of solar cell even relies on an electrolytic junction, but here we will not go into this.
- The current (density) - voltage ***j(U)*** relationship of an illuminated perfect **pn**-junction is described by our old **pn**-junction "master" equation

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

- ***L*** is the **diffusion length** of the semiconductor, n_{Min} the minority carrier concentration ($n_{Min}=n_i^2/N_{dop}$); n_i=intrinsic carrier concentration, N_{dop}= doping concentration); τ=life time ($\tau=L^2/D$); ***D***=diffusion coefficient of carriers), ***kT***=thermal energy, j_{Ph}=induced photo current.

The two expression in front of the brackets are the reverse current densities that come from the bulk (j_1) and the space charge region (j_2), respectively.

- While the j_2 term is often "negelcted" in text books, here we must consider the influence of the space charge region on the ***j(U)*** or, as it is often (sloppily) abbreviated, ***IV*** characteristics, if we want to be half-way serious about solar cells - it is the decisive term! If you activate the link you will see (once more?) that $j_2 >> j_2$ for **Si** and all other semiconductors with bandgayp **>≈1 eV.**
- The resulting curves for three illumination intensities I_o (including I_o**=0**=darkness) produce characteristics with short-circuit currents proportional to the illumination intensity (or photon flux) and an upper limit of **1** carrier per photon. They look schematically like this:



- We can measure the characteristics simply by changing the load resistor R_{load} from R_{load}**=0 Ω= short circuit** conditions to $R_{load} \Rightarrow \infty$=**open circuit** condition and keeping track of the current and voltage in the external circuit.
- From these measured ***IV***-characteristics we learn a lot about our solar cell; this will be treated in some detail in the next module.
- Here we need to do an exercise to acquaint us again to the diode equation from above and its relevance for solar cells.

Exercise 8_1_5

Diode equation and solar cells

- This is an important exercise! You should at least look at the solution, which will actually be given in an advanced module!

Contacts: We need electrical contacts to the front and back of the solar cell in order connect our *constant current source* "solar cell" with the load.

- On the front side this is either done by a **grid** of metal contacts, which then always absorbs some of the incoming light without generating power, or by a layer of some transparent conductor, which then will lead to problems with the resistivity.
- At this point we finally must realize that solar cell technology is the art of making **compromises** - you can never get one parameter at the best possible value without compromising others. This is true for purely technical parameters like in the task metioned above (find the best compromise between series resistance *and* the area needed for contacts), but in particular for all tasks where one of the parameters is **money**=costs.

- An always unavoidable internal **series resistance** R_{SE} is switched in series to the load and "eats up" a part of the voltage generated according to $U_S=R_{SE} \cdot I$ and thus reduces the efficiency.

Load: We have a load that "consumes" the power provided.

- The power provided by the solar cell in the external load (symbolized by a load resistor R_{load}) is simply given by $U_{load} \cdot I_{load}$.
- The current flowing through R_{load} is the same as the current I_{SC} (=I in the characteristics) flowing through the solar cell, but $U_{load} \cdot I_{load}$ is smaller than the voltage U_{SC} of the solar cell (=U (or V) in the characteristics) by the voltage drop in the serial resistor R_{SE}, i.e. $U_{load}=U - R_{SE} \cdot I$.
- Maximum power will only be delivered if the load is matched to the solar cell. R_{load} must have a value that adjusts U and I in such a way that the point of maximum power on the *IV* characteristics is met.
- In other words: We need some "**load management**" to always extract maximum power from our solar cell power plant. This looks more difficult than it really is; we therefore won't come back to this point again.

Efficiency: A solar cell is nothing but an energy transformer: Light energy goes in, electrical energy comes out. The question of efficiency then always comes up quickly.

- The relation between: input=light power to output=electrical power in the load depends on the **working point** of the solar cell, i.e. on which point exactly on the *IV* characteristics you run the system: solar cell plus external load.
- There is a maximum efficiency η, however, that obtains at the *working point of maximum efficiency*. It is easily extracted from the *IV*-characteristics (just multiply $I \cdot V$ of the curve and plot it), and this maximum efficiency is meant when we talk about the *efficiency* of a solar cell.
- There is a strict theoretical limit for η for any given solar cell; for **Si** solar cells $\eta \approx$ **25 %** obtains. There are also practical limits; at present commercial **Si** solar cells have $\eta \approx$ **15 %**.

Energy, Power, and Money A solar cell exposed to sunlight produces **power**=**energy/time** by converting the power contained in sunlight to electrical power with an efficiency η.

- The power contained in the sunlight varies with the daytime, the degree of cloudiness, and the latitude. Maximum sunlight power is obtainable on a perfect cloudless day on the equator at high noon; it is around **1 000 W/m^2** .
- At the latitude of Kiel (**54.12 ^{o} N**) you have to multiply by **cos(54.12 ^{o})=0,586**, so we get less sunlight than the people in the tropics (surprise!). But our solar cells stay cooler, and that is good for the efficiency η.
- When we characterize a solar cell (or a module of solar cells) in *one* number, we give its "Wattpeak" W_p nominal power, which, for a first guess, is $W_p=\eta \cdot$ **1 000 W/m^2** . It's a little bit more complicated than that (things like "**air masses**" ***AM 1.5*** or **AM 1** are involved).
- The average power delivered by a solar cell, considering that the sun isn't always shining, will be about **11%** of the Peak power. **1 m^2** of **η=15 %** solar cells will thus produce *per year* an amount of *energy* $E_y=1\,000 \cdot 0{,}15 \cdot 0{,}11 \cdot 24 \cdot 365$ **Wh=145 kWh**.
- At present rates you pay about **15 €** for this to your utility. If you count on a life time of **20** years for your solar cells, a square meter should cost no more than about **€ 300** to be competitive in the energy market. Of course, this kind of reckoning is naive, and the numbers have to be taken with a grain of salt, but you get the basic picture in this context. More about parts of this topic can be found in the link.

It is time for a little exercise:

Exercise 8_1_2

Optimal Working Point of Solar Cells

Modules

The typical **Si** solar cell - about **(12 x 12) cm^2** - is not what you find on your roof: up there are solar **modules**, typically about **(1.60 x 0.8) m^2** or **(1.60 x 1) m^2** or any other size in the **m^2** range.

- If the module contains **Si** solar cells, about **50 - 100** individual solar cells are connected (by ***soldering***) in series in a "string"; several strings are connected in parallel. The series connection is absolutely necessary to raise the voltage to acceptable levels (it is far more difficult in electrical engineering to deal with large currents and small voltages than vice verse) but causes a number of problemson its own.
- If the module contains thin film solar cells, long stripes of individual solar cells, about **1 cm** wide, are connected in series in-situ, i.e. during the production process of the whole module in one fell swoop. This causes a lot of problems on its own....

The total area of the solar cells in a module is always somewhat smaller than the module area, and the top layer of the module (usually glass) that protects the solar cells from rain, bird shit, and whatever else is in the air for a guaranteed **20** years, also reflects some of the incoming light. The efficiency of the module is therefore always somewhat smaller than the efficiency of the solar cells.

- Nevertheless, even if we count module technology under "semiconductor technology", it is not quite as challenging to solder solar cells together than to make them. It might be just as expensive, however. We will not go into this kind of module technology any more.
- But let's not kid ourselves: if you make thin film solar cells with $\mathbf{m^2}$ sizes, you must switch them in series even more urgently than with bulk **Si** solar. Typically, your individual solar cell has now a size of $\mathbf{(7 \cdot 1.000)\ mm^2}$, and switching these thin stripes in series must be done in-situ during manufacturing. There is no soldering involved anymore. We have a new, very complex, and very challenging facet of semiconductor technology instead.

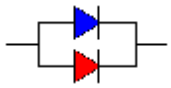
Typical module with **9 · 6=54** Si bulk solar cells

8.1.2 Solar Cell Current-Voltage Characteristics and Equivalent Circuit Diagram

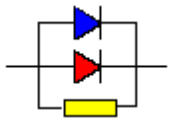
Basic Si Solar Cell

- It is important to look a bit more closely at the ***IV***-characteristics of a silicon **pn**-junction solar cell. The proper equation for that was already introduced before
  - In a kind of short-hand notation, and because it is what electrical engineers always do, we could symbolize that with the normal diode symbol:

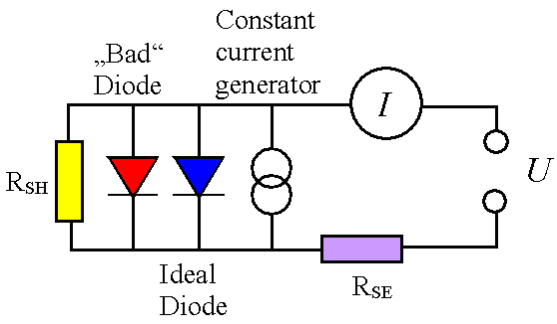
  - However, it is customary to use a **"two diode" model**, where the first diode describes just the simple **pn**-junction behavior without carrirer generation and recombination in the space charge region (**SCR**), and the second diode, which then must be switched in parallel, describes the "non-ideality" of the real **pn**-junction. The equivalent circuit diagram then would consist of two diodes in parallel:



  - Since our junctions area is **> 100 cm^2**, there might be regions where we have a *local* "**shunt**" in the junction. In the worst case this is a *local* short-circuit, caused for example because a metal particle was embedded in the junction (we obviously need to be clean if we make those junctions). The best we can do to model *globally* the many possibilties for shunts is to add a **shunt resistor** R_{Sh} in parallel to our diodes. "**Global**" in contrast to "**local**" means that we only care for whatever we can measure for the total solar cell, i.e. whatever we can get using only the two wires attached to it.



  - If we now add the internal **series resistance** that is always there *in series* to what we already have, and consider that the photocurrent flowing across the junctions(s) is simply a **constant current generator** in electrical engineering terms, we obtain the final (and most simple) **equivalent cuircuit diagram** of a solar cell and a somewhat modified master equation for the *equivalent circuit diagram*:



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

"Ideal" Diode — "Bad" Diode

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

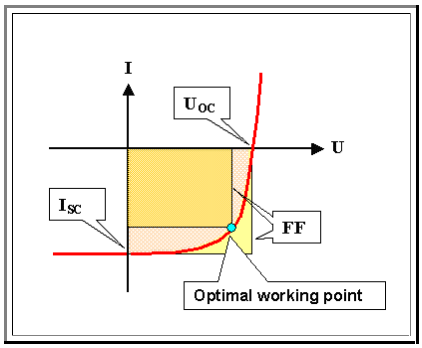
- It is easy to see how the master equation was changed.
  - First we switched from current densities ***j*** to currents ***I*** and simply used the terms I_1 and I_2 as abbreviations for the pre-exponential factors. I_1 and I_2 belong to the first (=ideal) and second (non-ideal) diode in the equivalent circuit diagram.
  - Next, we considered that the junction voltage U_{eff} is not identical to the externally measured voltage ***U*** but smaller by the voltage drop in the series resistor R_{SE}; i.e. $U_{eff}=U - I \cdot R_{SE}$.
  - Then we generalized the non-ideal diode a bit by not just having a factor of two in the denominator of the exponent, but an ideality factor ***n***. For ***n*=2** we would have the old equation, but since we arrived at that part by rather hand-waving approximations, real (non-ideal) diodes may be better described by some number other than **2**. Our ideal diode has ***n*=1**.
  - Finally we add the current lost in the shunt resistor, which is given by U_{eff} / R_{SH}. We have a positive sign because this current must be opposed to the photo current, which has a negative sign in the conventions chosen.
- Everything is quite clear - except that we can not solve this equation anymore analytically and plot **U(*I*)** curves with R_{SH}, R_{SE}, and ***n*** as parameters.

- We have reached a point where we need to do some exercise:

Exercise 8_1_3

IV characteristics of real solar cells

- It is absolutely essential that you do this exercise - at least ponder the question and look carefully at the solutions provided.
  - You will learn that the series and shunt resistance does not only matter very much but that we need to have extremely low values for the series resistance in the order of at most a few **mΩ** per **(10 × 10) cm²** solar cell.
  - **Class Exercise:** ***Cu*** *has a specific resistivity of about* **2μΩcm**. *Given a solid Cu wire with* **1 mm²** *cross section, what length will give you already* **10 mΩ***?*
- OK - coming back from the exercise (or just the solution) you learned that:
  - The **series resistance** really matters for high efficiency cells; i.e. for all solar cells with let's say **η > 10 %**. In fact, optimizing the series resistance of a standard commercial solar cells (now, in **2007**, with **η ≈ 15 %**) is one of the major tasks in solar cell **R&D**. We will look at some number in another exercise coming up in the next module.
  - The **shunt resistance**, if not extremely small, is less ***obnoxious*** for the efficieny η, but generates high leakage currents in reverse direction, increasing with the voltage. This is deadly in modules and thus cannot be tolerated either.
  - As far as the ideality factor is concerned, we do not notice major problems. Here you have to look at details to see that the impact of the second diode is ther and detrimental.
- The way to lump everything together *globally* is to describe the global ***IV***-characteristics with **3** numbers:



  - **1.** The **fill factor *FF***, defined as the relation between the area of the large yellow rectangle to the more orange area that is centered at the **optimal working point.**. That implies that the inner rectangle has the largest area of any rectangle (Area **=*U* · *I*=** power delivered by the solar cell). The fill factor measures the degree of "rectangularity" of the characteristics.
  - **2.** The **open-circuit voltage U_{OC}**, i.e. the maximum voltage the solar cell will produce for the load resistor
  - **3.** The **short-circuit current I_{SC}**.
  - The **efficiency** η then is directly proportional to U_{OC}, I_{SC}, and ***FF***, i.e.

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

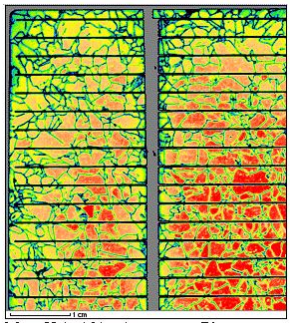
- How do we **measure** the ***IV***-characteristics of a real solar cell coming out of a production line?
  - Easy, you might think: Apply a voltage, measure the current, change the voltage, measure the current again... Do it automatically by using a voltage ramp and keeping track of the current. Yes - that would be perfectly OK except that solar cells from a real production come off the line at the rate of about ***1*** *solar cell per second*! Your measurement scheme is too slow for this because whenever the current changes the capacity of the **SCR** must be recharged and this takes some time.
  - The way it is done is simple but powerful: You impress a constant current on your solar cell by some external power source. At a given illumination intensity this takes a certain voltage as determined by the ***IV***-characteristics. If you change the illumination intensity, the voltage needed to drive the constant current changes, too. If you change your illumination intensity very quickly be just using a flash light, your voltage can follow just as quickly because you do not change the charge in the **SCR** capacitor.
  - The ***U*(*t*)** curve for a fixed current contains all the information you want - but now you can get it with flashing speed!
- You will actually do this in a Lab course; the link provides the details.

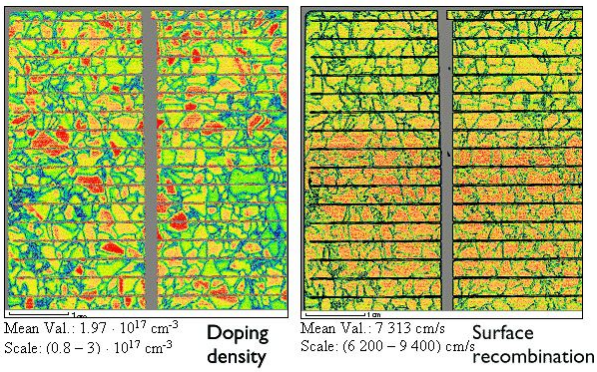
Making Things Really Complicated

- We are not yet done. Even if you have understood and committed to memory everything pointed out above with all its implications, at least two topics must be still be addressed:

  1. How about other types of solar cells?
  2. Connection between *local* and *global* parameters.

- The *first* question can be dealt with rather easily - in principle:
  - Whatever type of solar cell you have - **Si** bulk, **μ**-crystalline **Si** thin film type, amorphous **Si**, ***CIGS*** or **CdTe** thin films, ***dye***-based $\mathbf{TiO_2}$ electrolytic cells - to name just a few, they must have some characteristics similar to a diode, and you can always find a suitable equivalent circuit diagram for modelling its behaviour.
  - This equivalent circuit diagram might be somewhat different from the one given above, but you always have the problems associated with series and shunt resitances after you solved the basic junction problems.
  - Quite often your choices for contacts are severely restricted and your series resistances get so large that you only can get decent efficiencies by switching them in series in relatively small units. More to that will be found in the link.
- The second point is the difficult one. What he have discussed so far is the global behaviour of a solar cell - what you measure for the whole "global" **100 mm x 100 mm** cell, being uniformly illuminated.
  - Now take your virtual knife and cut your solar cell into **10 000 (1 x 1) mm^2** cells (without any damage and so on), an measure the ***IV***-characteristics of those **10 000** local solar cells. We assume that you know how to do this - it is possible ***albeit*** not easy.
  - If your global solar cell was not very uniform (rather likely for many solar cell types), your local solar cells may show wildly different behaviour. The ***IV*** characteristics of the, let's say, **8** local cells that contain serious shunts (=short circuits) will be far more affected by this than your global cell (where you have another **9 992** shunt-free cells switched in parallel).
- You get the point: Your major *global* parameters like short circuit current I_{SC}, open circuit voltage U_{OC}, fill factor **FF**, series and shunt resistances R_{SE} and R_{SH}, diode ideality factor ***n***, are *somehow* determined by the local values, but it is not immediately obvious how.
  - Besides the *global* short circuit current I_{SC}, which obviously is just the sum of the *local* short circuit currents, for all other parameters the relationship between local and global parameters is non-trivial or very complex. In the case of the *global* ideality factor ***n***, the number you extract from your global ***IV*** curves may not even have much to do with the junction properties.
- Do we need to know the local parameters? After all, what counts is the real=global solar cell. If it is as good as it should be, the local parameters might be of no interest.
  - Well - yes and no. Even if everything is OK globally, some small local defects may limit the life time of the solar cell, i.e. be a risk to its expected **20+** years of harvesting energy.
  - More important, if you don't know your local parameters, you have very little guidance in making your solar cell better. If, for example, the *global* short circuit current is too low, that may indicate that it is just too low everywhere or that it might be allright on most parts of the cell but really lousy in some small areas. If you don't know what is the reason, you can't fix the problem.
- To give an idea what we are talking about, look at the pictures of the "multi-crystalline" solar cell below:





- A perfect solar cell would show only one color - the optimal value. The colorful pictures thus demonstrate that there is room for improvement and that you have to have the information contained in the pictures to be able to make progress
- How are maps like these obtained? With the so-called CELLO-Plus technique, developed in Kiel. This is a rather sophisticated characterization technique that is described in some more detail in the link.

8.1.3 Summary to: 8.1 Solar Cells - General Concerns

- A solar cell converts light power into electrical power. It's overriding parameter is the over-all conversion efficiency η
  - Any solar cell is essentially a large -area junction, usually of the **pn**-type.
  - It's essential parameter are the short-circuit current I_{SC}, the open-circuit voltage U_{OC} and the fill factor **FF**
  - For optimal efficiency the bandgap E_g should be matched to the solar spectrum; we need around **1.5 eV**.
  - Maximum efficiency from the *semiconductor physics* point of view is achieved if all light with energy **>=**E_g produces minority carriers and all of these carrier are swept out as diode reverse current and
  - Maximum efficiency from the *module systems* point of view is achieved if the semiconductor part is **OK**, only very little light is reflected by the solar cell module, series resistances and shunt resistances can be neglected, and everything is uniform and homogeneous

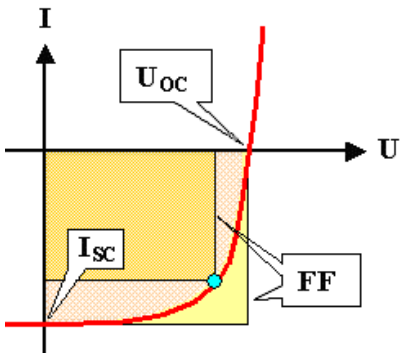
The equivalent circuit diagram with the basic equation has is all!

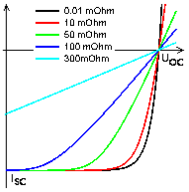
- Series and shunt resistances, unavoidable for large areas, are of overwhelming importance for solar cells with η **<≈ 10 %**

Important "raw" numbers.

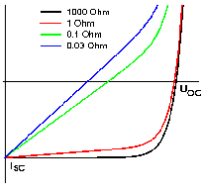
- **Maximum η Si solar cell ≈ 25 %**
- **Maximum sun power ≈ 1 kW / m².**
- **Maximum commercial solar cell power ≈ 200 W / m².**
- **Yearly average commercial solar cell power ≈25 W / m².**

Solar cell science and technology centers exclusively on *money*and *saving the earth*!

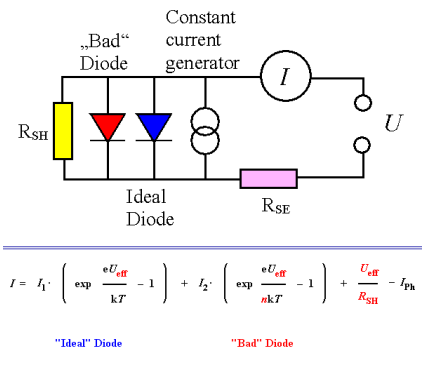




Effect of series resistance

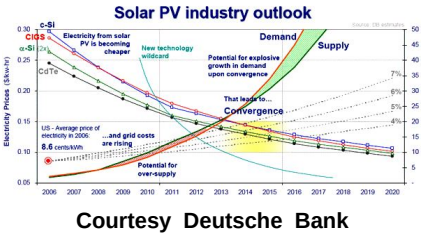


Effect of shunt resistance



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

- Switching solar cells with individual characteristics in series and / or in parallel causes all kinds of problems.
  - Worse: Any inhomogeneous solar cell (e.g. **mc-Si** solar cells) consists of *locally* different solar cells "somehow" connected internally
  - Optimizing solar cells with respect to "money" thus provides exciting science and engineering!



Courtesy Deutsche Bank

- There are many competing solar cell technologies and materials.
  - Bulk single-crystal and **mc Si** vs. thin film **Si (a-Si:H, µc-Si:H. ..**

- Other thin-film semiconductors:**CIGS, CdTe, ...**.
- Exotica: **TiO_2**- electrolyte ("Grätzel cell"), organic semiconductors, "Nano" materials, ...

Solar cells have a bright future!

Exercise 8.1-4
All Quick Questions to 8.1

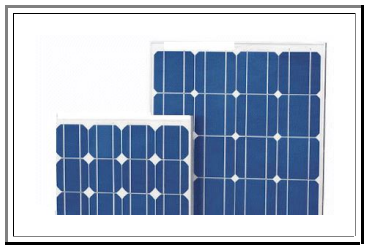
8.2 Making Bulk Si Solar Cells

8.2.1 Production Necessities and Silicon Starting Material

Boundary Conditions for Production

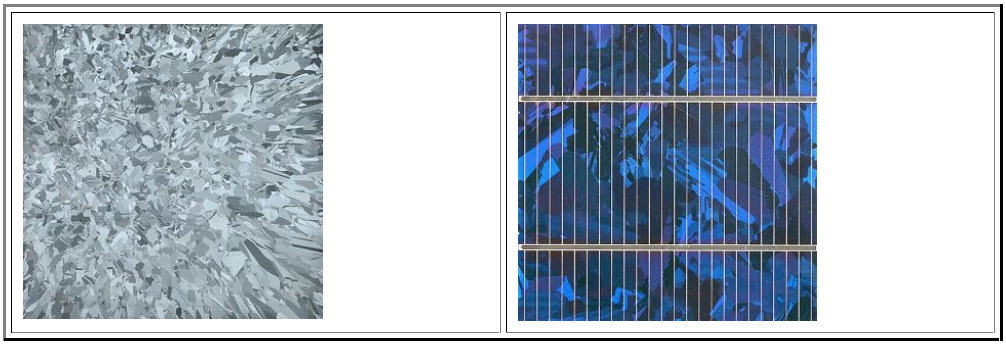
- It seems that a solar cell is just a **pn**-junction with a relatively large area, some defects and and some contacts, following relatively simple junction theory - as described before?
  - *No* - so far we have only looked at a large real **pn**-junction. A real *solar cell* that will end up on somebodies rooftop is a bit more than that. If it is a standard **2007 Si** bulk solar cell (***SC***), it must meet the following criteria on top of being just a large **pn**-junction:
- ***Processing Time***. A typical **12.5 cm x 12.5 cm** bulk Si solar cell comes off a production line at the rate of about **1 SC/s**.
  - If its size is **150 cm^2 = 1.5 · 10^{-2} m^2**, the factory produces **1.5 · 10^{-2} · 3 600 · 24 · 365 = 4.73 · 10^5 m^2** per year; good for about **7 · 10^7 W_p** or **70 MW_p**.
  - At at thickness of **300 µm** you used up **45 m^3** of **Si** or about **105 t**. At a **kg** price around **(10 - 40) €/kg**, this costs around **(1 - 4) Mio €**
  - Nevertheless, you still have a very small factory according to **2007** standards. Presently, the first **1 GW_p** factories are being built. Whichever way you look at it, is simply means that process time is even less than **1s** per cell, that you process in large batches, i.e. many cells in parallel like in a furnace, or you just have several factories in parallel. However way you look at a **1 GW_p/a** factory - this is tough, man!
- ***Cost Decrease*** The productions costs of a solar cell must go down about **5 % - 7 % /year** for the next **10 - 15** years.
  - If you can't meet this goal this, you close your factory and go bankrupt. There is no market for solar cells per se, there is only a market for power plants on roof tops (or other easily available cheap areas) with a well-defined price for the **kWh** produced.
  - The price in most countries is defined by politics and decreases with **5 % - 7 %** per year (in Germany it is the "**Energieeinspeisegesetz**" (***EEG***); presently emulated by many other countries). If there is such a well-defined price decrease for energy, people are only willing to pay for a solar module if its price decreases at the same rate. I you can't meet this goal, chances are that your competitors can - and that will be the end of *your* business while *they* will roll in money.
- ***Efficiency Increase*** The efficiency of your solar cells must go up by about **1 %** (absolute) every **12 - 18** month before it might level out around **η = 20 %**.
  - There are **η ≈ 20 %** solar cells on the market right now (Sunpower Corp.), proving that you can get efficiencies that high even in a (presently limited) mass production. If you competitor manages to get his efficiencies up and you don't - see above.
  - All serious companies therefore are working very hard on improving efficiencies -while still lowering costs.
- ***Key Material Supply*** Your solar cell will use up less of the key materials, especially **Si**, meaning it becomes thinner.
  - Not only does this save costs - there is simply not enough **Si** around at present (**2007**), to make all the solar cells the market would buy. Saving on **Si** thus makes a lot of sense.
  - If you can make your cell thinner, you can go farther with the **Si** you managed to buy. However, you cells are mechanically less stable and tend to break more easily.
- ***Business Growth Rates*** The solar cell you make is just one of many solar cell made world-wide per year - and their number increases with about **30 % - 40 %**.
  - You have exponential growth at an extremely high growth rate and what that means has been described very briefly in the context of microelectronics in chapter 5.3 and in somewhat more detail in the link. In fact, much of what has been stated (and meanwhile observed) in the microelectronic business will repeat itself with solar energy, just more so (there will be far more money involved as soon as solar energy can be measure in **%** of the total energy needs of the world).
- ***Summing Up*** Be prepared for some turbulent times in solar cell **R&D** and production! Watch some companies grow from nothing to Microsoft proportion and watch some people getting rich very fast
  - Maybe on of those winners will be *you*?

Silicon Starting Material



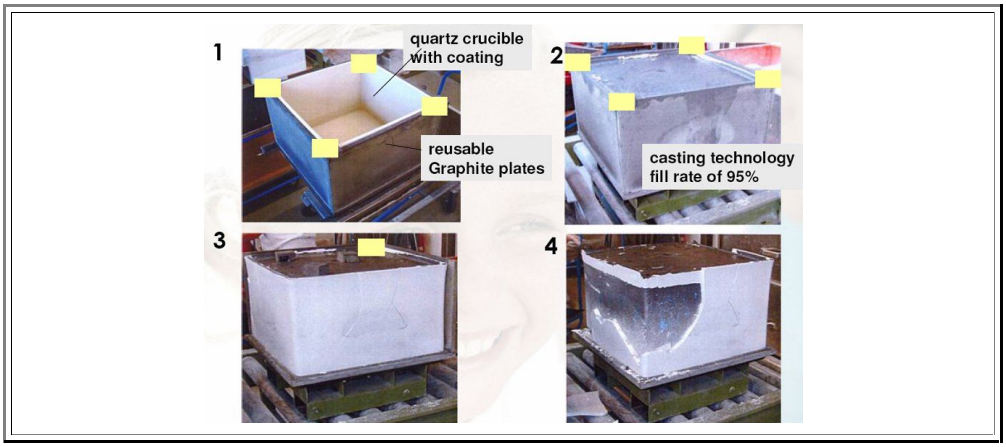
- You either start with single crystalline **Si**, i.e., more or less regular wafers, or with what is called **multi-crystalline Si**
  - If you go for single crystalline **Si**, you do everything pretty much as already described except for the following points.
    - You cut your crystal into an almost square shape because you loose to much area in your module if you "fill" it with round solar cells as made from regular wafers. So you must find the best compromise between wasting expensive single crystal **Si** or module area. The picture shows somewhat different solutions obviously because of different wafer size.
    - You try to stay cheap. You might use slightly dirtier **Si** for growing your crystal and you might grow them somewhat faster allowing a few dislocations or other defects not acceptable for microelectronics.
    - You cut your crystal into very thin wafers, as this as you can make it, say **250 µm** (instead of **(600 - 800) µm** used for microelectronics).
    - You do not produce wafers with a highly perfect and flat surface. You may just etch off the saw damage and leave it that.
  - There is nothing really new, so we will not go into any more detail here but take a slightly closer look at multi-crystalline **Si**.

- Multi-crystalline **Si** is simply coarse-grained poly-**Si** but since the name poly-**Si** is already taken, multi-crystalline **Si** is what it is called.
  - The picture below shows an example.



- Multi-crystalline **Si** is produce by **casting** and cutting. As raw **Si** you must use rather clean electronic grade **Si** coming straight from the Siemens process. Now melt this poly-**Si** it and cast it into a mold. That sounds easy but their are a few catches:
- **1. Melting point density anomaly**. **Si** belongs to the few material that do not contract upon crystallization but **expands** - by a whopping **10 %**!
  - Producing a sizeable lump of **Si** by simple casting would be equivalent of casting ice (solid $\mathbf{H_2O}$) at very low temperatures, say **80 K**. Pour you molten ice (= water) in a bucket kept at **80 K** and watch what will happen. A major explosion will happen!
  - If you just leave your bucket around at low temperature, the surfaces will cool down first and start to crystallize. Liquid water then is trapped inside a solid and tremendous forces will develop as soon as it crystallizes an tries to expand.
  - What we have to do is to keep the surface of the **Si** liquid and crystallize slowly from the bottom only. The crystal forming then can expand upwards and everything will be well behaved. The catch, of course, is that we need to control the temperature *and* the temperature gradients closely, taking into account that a lot of crystallization energy is released that needs to be taken out of the system. No is not a serious technical problem - except that solving it takes money and that you have to wait longer for your **Si** melt to crystallize and cool down (more money).
- **2. Reactivity of Si**. We have poured some liquid Si into a mold, cooled it down properly without encountering problems, and now want to take it out of the mold and reuse the mold for the next batch.
  - The problem now is that your **Si** is firmly bounded to the walls of your mold if didn't take proper precautions. In chemistry circles, **Si** is known as an **universal solvent** because everything dissolves in liquid **Si** and thus sticks to the solidified **Si** in a major way.
  - Of course, you can cut your **Si** ingot free, destroying the mold; all that takes is a lot money for molds. this is obviously not a good idea.

- A good idea is to use some "grease" that squeezes between the **Si** and the old (and covers the **Si** surface) because it lowers the surface tensions of the system.
- In other words, use a **liquid encapsulation technique** - we already encountered this trick when we looked at the growth of III-V single crystals.
- Now we encounter a new twist to an old technique: As far as casting **Si** is concerned, everything from now on is top secret. Exactly how is to done has never been fully disclosed, so we will not belabor this point anymore. The picture below shows how it is done in reality



- The yellow patches hiding something have been placed by the source. What you see is the solidified white "grease" and a sizeable chunk of **Si** in the **4th** picture. Now we can cast *and* reuse the mold, but we are still left with a problem.

3. Cleanliness and Structure. We don't just want any big chunk of **Si**, we want it to be *multi-crystalline* , meaning coarse- grained, and we certainly don't want it any dirtier that the raw **Si** we poured into the mold.

- So we have to make damned sure that the liquid **Si** - which, after all, is an universal solvent - doesn't pick up some dirt from its surrounding.
- We also have to make sure that crystallization produces big grains in some **columnar** arrangement (why we need this we will see right below). This brings us back to an optimized temperature control and some secret issues I can't tell you concerning casting **Si**.

Now we are done. If you actually could obaibn some poly-**Si**, that is. The solar energy business right now (**2007 /08**) is having a supply-side crisis; there just isn't enough electronic grade **Si** around for everybody. You should have some vague idea by now, why it took **>10** years and a lot of work and money to come up with an industrial casting process that meets all requirements (including costs).

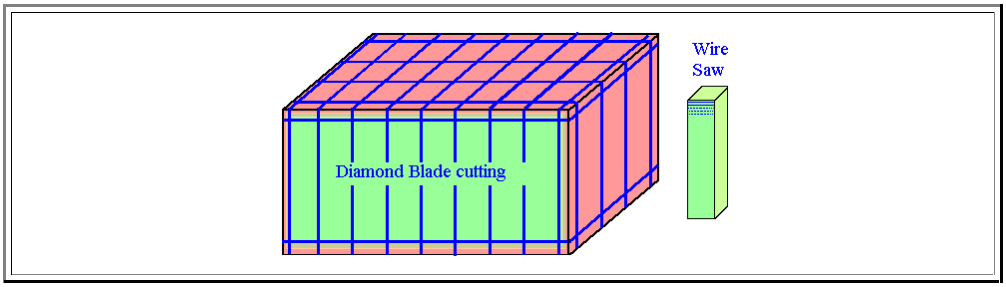
- All that's left to do is to cut the big block of **Si** into very thin, standard-sized, square, multi-crystalline **Si** wafers, taking care that grain boundaries are always at right angles to the surface.

Cutting Silicon Ingots into Wafers

Why is it important to have no grain boundaries running parallel to the wafer surface?

- Because grain boundaries are always efficient recombination centers and thus screen minorities "below" from making to the the front side side contacts and take out minority carriers from above that happen to come by.
- If the grain boundaries are at more or less right angles to the surface of the wafer, they are still bad but far less so then at **90°**.

So we first cut our huge **300 kg** ingot into square columns as indicated, taking care to cut off the "dirty" (= reddish) surface-near regions; especially the bottom regions, where grain nucleation started and we have fine-crystalline **Si**, and the top region, where all the dirt with low segregation coefficient will be found. This we do with a big diamond-blade saw.



- The columns left have the right wafer size and the grain boundaries inside all are more or less at right angles to the major axis if our casting process was OK.

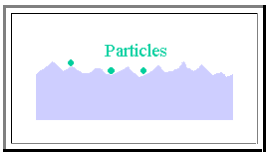
- All we need to do now is to cut the column into very thin wafers (about **250 µm** at present (**2008**)), making sure that the cutting grooves are very thin in order to minimize cutting losses.
  - Easy? - No, very difficult. In fact, a new process tool needed to be invented, the industrial-scale, extremely precise wire saw. A very fine but very strong closed wire loop runs around the **Si** column many times, moving fast, under high mechanical tension, and in an abrasive slurry.
  - In essence, its like a gigantic egg cutter with moving wires, just a bit more expensive. Cutting off the joke: the **Si** multi-wire saw, cutting fast, in parallel, and with very high precision, is an essential piece of equipment for multi-crystalline **Si** solar cells and you will thus not find much information about process details floating around.
- Just in case you didn't notice: We now have two *special* semiconductor technologies here that originated from solar cell technology: casting and wire-saw cutting.
  - In fact, cutting **Si** with wire saws is now moving into the wafer business for microelectronics, too.
- What we have no are square slices or wafers of multi-crystalline **Si**, about **250 µm** - **300 µm** thick and up to **(150 × 150) mm^2** in lateral size;.. And, yes, you're right - they fracture very easily.
  - Right after cutting, the surface of the multi-crystalline wafers is "destroyed"; or more precisely, on both sides are layers full of "saw-damage" that are electronically "dead". The thickness of these layers scales with the size of the hard grains in the abrasive slurry and thus will be in the range of several **µm**.
  - This surface damage must be removed before we can start to make a solar cell. Fortunately this is easy if, once more, something of a black art: Certain chemicals attack and dissolve "damaged **Si**" far faster than perfect **Si** (CMP relies on this!). So all we have to do is to put a batch of wafers in a chemical tank for a while.
  - Since we have a **batch process** with, say, **100** wafers, we can allow a process time of ≈ **100 s** and still meet our goal of **1** Solar cell per second.
- With the "over-etched" wafers we can now start solar cell production

8.2.2 Processes and Process Integration

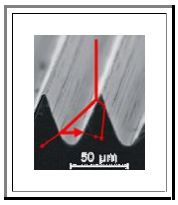
Overview

- Compared to the complexity of making an integrated circuit, making a solar cell is exceedingly simple. However, to belabor the point once more, making a *very good* and *very cheap* solar cell in a *1 s rhythm* is exceedingly difficult.
  - We will look at the necessary processes an the process integration in a very superficial way. There simply is not enough time to go into details; and the interesting details are kept very confidential at present anyway.
- What are the basic processes we need? Here is a list giving processes and the process integration sequence with short comments

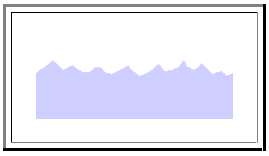
Surface roughening



- We must texture the surface in such a way that the reflection of light is reduced. This can be done by "grooving" or by any other way that produces surface textures that direct reflected light back at the **Si**
- The example on the right shows a light path that will still end up in the **Si** even after two reflections took place.
- Regular **Si**, as we know, looks like a metal, i.e. silvery-shiny - in other words it reflects light quite well. Measures for decreasing reflection are absolutely essential - already for low efficiency solar cells. Working on anti-reflection measures might be far more important for high-efficiency solar cells (i.e. **λ = 20 %**) than improving junction properties!

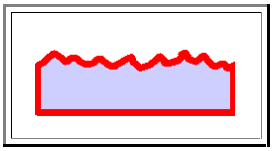


Surface Cleaning



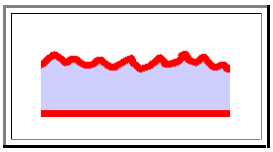
- Your mother is right! Cleanliness is next to Godliness, indeed. At least if you want to make solar cells (or any other semiconductor device).
- We had looked a the particle and contamination problem in the context of integrated circuits before; with solar cell production we have essentially the same problems.
- OK - solar cell structures might not be quite as sensitive to contamination than a **3 nm** gate oxide, but then we also don't want to run our production in an expensive super-cleanroom either. Finding efficient cleaning procedures is a good idea and essential for success.

Diffusion



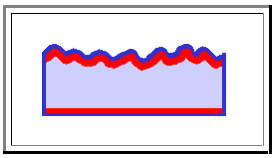
- Now we form the necessary **pn**-junction by diffusion. If our substrate wafer is **p**-type (the standard at present (2008)), we diffuse some phosphorous into the front side to a junction depth of **< 1 µm** (which is far thinner than the red line in the picture!). Note that we do *not* ion-implant the **P** to form the **emitter** of the solar cell (that's what we call this layer) as we would always do in microelectronics - far too expensive!
- We go back to an old-fashioned (and cheap) technique: We "somehow" smear some **P**-containing stuff on the **Si** surface and heat it up. **P** will out-diffuse from the source layer and penetrate into the **Si** if we do it right.
- One version for this is to use the so-called "**POCl**" process - we have encountered that before. But whatever - we will usually end up with a **pn**-junction all around the wafer; front-side, backside and edges, which is, of course, *not good*.

Edge Isolation



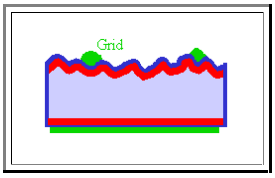
- We now remove the **pn**-junction around the edges, which otherwise would provide an electrical short-cut between front and back!
- Quite easy to draw. How about doing it? Can you come up with a viable process for this (good, quick, cheap, ...)? No! Well - it's not that easy. Obviously we have another quite solar-cell specific process here.
- One way to do is to stack a lot of wafers until you have a block. Then just etch off a few **µm** of the surface by plasma etching.
- One way for achieving edge isolation. There are other ways - and they are more or less confidential, once more

Passivation



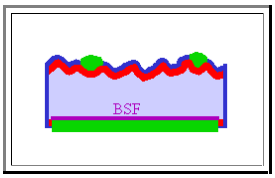
- We need to passivate the surfaces, i.e. remove surface states from the bandgap. Since we have a lot of surface it may act as very efficient "recombination center" where we loose our light-generated minority carriers.
- We can only have a high-efficiency solar cell by definition if pretty much all minority carriers generated by the light will reach the front contact. This means that we *must* prevent internal recombination at point defects, other lattice defects, internal interfaces like grain boundaries *and* at the front and backside surface by all means.
- The picture shows how its done: deposit some layer (like **SiO_2**?)! Yes - but! Again, we have a crucial process where you don't want to share your know-how with your competitors. At present it appears that the company Suntech has found a particularly efficient way of doing this, all other companies try to catch up, but are clearly behind.

Printing Contacts



- We now have to make the front side grid and the full-size backside contact. In other words, the tricky part is that we have to have a *structured* metal on the front side.
- No problem, you might think; we have done this before. Deposit the metal, structure by lithography and etching. Forget it. Far too expensive (not to mention too slow).
- We structure by **screen-printing** some goo or paste containing the desired metal particles on the solar cell. If you don't know how screen printing works in principle, look it up. We might do the same thing on the backside; there it is rather trivial because we do not need to structure.
- Sounds simple, is simple - but: It's a "rough" process, and if you damage the front just a tiny bit, your solar cell will be junk. What exactly do you use as the "paste"? Once more we have a special and crucial process with plenty of confidential know-how involved.

Sintering Contacts



- So far we only have some paste in the proper places. To get a metallic contact, we have to **sinter** the paste - in the usual way we make, e.g., ceramics out of paste.
- All that it takes is heat - we have a second high-temperature process. Instead of putting batches of wafers into a conventional furnace, we rather run them on some ***conveyor band*** through a long tunnel furnace (similar to making bricks). Once more we have a solar-cell specific process, quite tricky in reality, that is done with some special equipment.
- If we are really smart and thrifty we will diffuse some boron into the backside at the same time in order to produce a **back-surface field** (***BSF***). What we are producing then is a **p^+p** "junction" that carries a (relatively small) electrical field with it that will repel all minority carriers (i.e. electrons) that strayed to the backside in their random walk, giving them a change to reconsider and to go to the front side where we want to collect them. In addition, a **p^+** doped backside makes for an easier ohmic contact.

- Now we are almost done. We just put some anti-reflection coating on the surface that also serves as a protective layer, attach some leads ("wires") to strategic places so we can actually connect our solar cell to other, and, not to forget, measure everything we need to measure - within **1 s** - to characterize our individual solar cell electrically.
  - Since we have already covered how we can measure a complete current-voltage characteristic in a short time, we will not go into this here any more.

Famous Last Words

- What is described above are the bare essentials of making bulk **Si** solar cells. It actually doesn't matter much at this level of unsophistication if you start with multi-crystalline **Si** or single crystals, the general processes are the same.
  - A number of essential "tricks" have not been addressed at all. In particular, we must keep the impurities from being too harmful by using processes like "**gettering**" and "**hydrogen passivation**".
  - Nevertheless, if you can come up with the cash (about **50 Mio $**) *and* if you have a source of state-of-the art multi-crystalline **Si** (hard to find in 2008), you can buy a complete solar cell factory from some companies that will churn out decent solar cells almost automatically. The only problem is that if you do only this, you will go broke in a few years because your competitors, who actually understand what they are doing, will have better and cheaper solar cells on the market.
- So here is a prediction concerning the future:

Nobody knows *how* we will make solar cells in 10 - 15 years from now.
All we know is that we will make 'em better, cheaper and that we will make far, far more then we make today

Here is some sage advice: Buy stock from the right solar cell company now, and you will be rich in **20** years. (If you send me large amounts of money, I might tell you which company is "right").

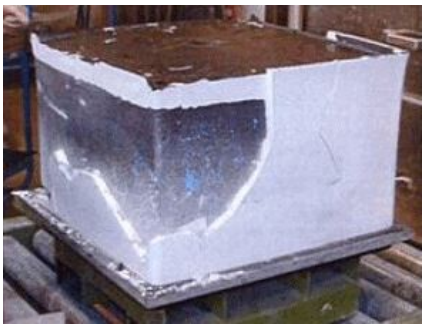
8.2.3 Summary to: 8.2 Making Bulk Si Solar Cells

- Bulk **Si** solar cells are made from (cheap) single crystalline wafers (cut squarish) or from square multicrystalline (***mc***) wafers. They account for about **85 %** of the installed solar power at present (**2008**).
  - A yearly production of **1 GW_{peak}** means about **10^7 m^2=10 km^2** **pn**-junction of good quality and much more
  - Consider ⇒

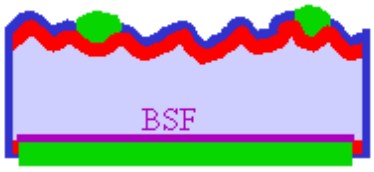
    A big problem is cranking up world wide **Si** production by **30 %** - **40 %** per year.

Processing Time	**1s / solar cell**
Cost Decrease	**5 % / a**
Efficiency Increase	**20 % in 2012 ?**
Key Material Supply	**30 % /a more Si**
Industry Growth Rates > 30 % for many years	**Supply capital and people**

- ***mc*** wafers are produced by **Si** casting. Problems are
  - Expansion upon crystallization.
  - Reaction with walls of mold
  - Columnar grain growth required
  - **300 kg** ingots are routinely cast in **2007**; liquid encapsulation and precise temperature control are essential
- Sawing the ingot into **mc-Si** wafers with as little losses as possible and with wafer thicknesses of **< 300 µm**, while straight-forward, is "high-tech".
  - Saw damage is removed by a chemical etch.



- Processing, simple in principle, has to meet the conditions above and is highly specialized. Essential processes are:
  - Diffusion, edge isolation, passivation, screen printing contacts and sintering contacts.
  - Essential device features are back surface field, gettering of impurity atoms, **H**-passivation of grain boundaries and other defects.



Exercise 8.2-1

Summing up Class Exercises

8.3 Making Thin Film Solar Cells

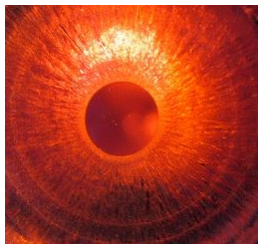
8.3.1 Types of Thin Film Solar Cells

General Remarks

- The **bulk-Si solar cell** was called "***bulk***" because its thickness of **> 200 µm** is far larger than the thickness of the space charge region in the necessary junction.
- There are two reasons why we make these bulk-**Si** solar cells:

  1. Infrared and even red light may penetrate **Si**to a depth of many **µm**. In fact, **Si** membranes with a thickness of just a few **µm** are somewhat transparent even in visible light as shown. If we want to absorb as much light as possible we must use "bulk" **Si**.
  2. If we use a thickness of **> 200 µm**, the solar cell is self-supporting (even so it breaks easily) and we do not need a substrate.

  - The catch is that we need a lot of expensive Silicon.
  - **Class Exercise:** *How much area in m^2 and how many **kg** of **Si** do you need for a **1 GW**power plant. Assume an efficiency **λ=15 %** and that the average power delivered is **12 %** of peak power.*



Backlit polycrystalline **Si**with a thickness around **5 µm**shows some transparency in the red.

- This gives us some (connected) major conditions for making **thin film solar cells**.

  1. The semiconductor should have a suitable direct band gap $E_g \approx$ **1.5 eV** since this ensures high absorption coefficients for all light with $h\nu > E_g$ and high efficiencies.
  2. There must be a process-compatible and cheap substrate on which we can deposit the thin layers.
  3. Since perfect single crystalline layers are impossible under the circumstances, "defects" in the layer and its interfaces must be harmless with respect to recombination.
  4. There must be some efficient (and cheap) way to produce a junction and decent ohmic contacts.

  - Those are tough requirements. While some (which ones?) are valid for *all* solar cells, they might be considered very special for a given thin film system.
  - We certainly have enough semiconductors to chose from - remember this link? So far, however, only a few semiconductors have made it - in the sense that you can find them in the form of solar modules on roofs. A few more can be found in laboratories, but all in all the list is rather short.

- Major *thin film* solar cells belong to one of the groups given below:

  1. Amorphous **Si**.
  2. Nanocrystalline thin film **Si**.
  3. Polycrystalline thin film **Si**.
  4. The **$CuIn_xGa_{1-x}Se_2$** or "**CIGS**" family.
  5. The **CdTe** solar cell.
  6. The **TiO_2** - dye based family.
  7. Multi-layer based **tandem cells** or multi-junction cells.
  8. Exotica.

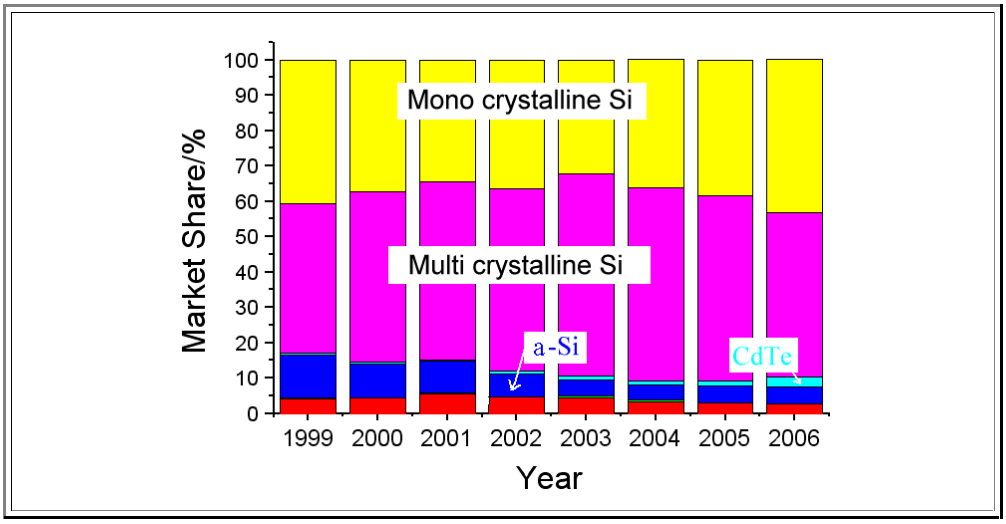
- If you feel that points **1 - 8** could be seen as the table of contents for a complete lecture course - you are right. Indeed, there would be a lot of ground to cover if we just were to look at the basic research, develeopment and - where already applicable - production of the thin film solar cells enumerated above. To give just one example of the complexity encountered:

  - Why is "**CIGS**" better than just "***CIS***" (=**$CuInGaSe_2$**)? Why are traces of **Na** important for making good **CIGS** solar cells? Why is **Mo** the best metal for the backside contact? Will there be enough **In** available for making **CIGS** cells in a big way? Why are properly made **CIGS** solar cells quite immune to radiation (as encountered by satellites in space)? Why are the recombination properties of the **CIGS** layer rather insensitive to its production method (sputtering, **CVD** in many variants, galvanic deposition, sintering, ...; everything seems to work)? What are the major limits to the efficiency? How stable is the combination of the many materials found in a **CIGS** cell over decades?

- All we can do here is to give a short summary of what is meant with the catch words in the list, and to look at a few essentials of production methods for thin film solar cells in the next sub-chapter. In time, advanced modules may spring into existence with additional information to each type of thin film solar cell.

Amorphous Si Solar Cells

- Amorphous **Si** or **a-Si** solar cells are almost as old as crystalline (or **c-Si**) solar cells. It is rather easy to produce thin film amorphous **Si** - just sputter from a **Si** target, or use **Si-CVD** at temperature **< ≈ 400 ^{0}C**. A layer thickness of **1 µm** or less is enough to absorb all light and thus we are truly talking "thin film" here.
  - However, pure amorphous **Si** is a good insulator and rather useless because it contains by necessity many "**dangling bonds**" that introduce electronic states in the band gap in a high density. Only if we "**passivate**" these dangling bonds with **hydrogen** (**H**) do we become a halfway decent semiconductor that behaves like a *direct* semiconductor with a band gap of about **1.7 eV**.
  - Now we have the first mystery about "amorphous **Si**". While the density of dangling bonds is large but still well below **1 %** of all bonds, we need to put **5 % - 30 %** hydrogen into the material to achieve decent semiconducting properties, i.e far more than we would need for just compensating the dangling bonds. In fact, what all and sundry calls "amorphous **Si**" is actually an amorphous **Si**-hydride; the abbreviation **a-Si:H** is often used.
- ***Eschewing*** these "details", we are left with a material that behaves in many respects roughly like **Si**, can be doped **n**-type and **p**-type, comes relatively cheap, and can be handled to a large extent by existing **Si** technology. It also can be deposited on large-area substrates.
  - We thus can produce solar cells, e.g. on a glass substrate. Contacts are a problem, but neglecting that for a moment, **a-Si:H** solar cells have been plagued from the very beginning by some specific problems that have not gone away even after decades of **R&D**:
    - The efficiency η is rather low (**<≈ 10 %** to give a number) for various reasons, and not easily (=cheaply) raised. For serious power application this is bad - being cheap at some point does no longer compensate for low efficiencies.
    - **a-Si:H** solar cells are prone to show the so-called **Staebler-Wronski effect**, i.e. their efficiency η degrades by up to **30 %** if the solar cell is exposed to light for some time!
    - The efficiency recovers upon heating the solar cell, but that is of little use. What exactly causes the Staebler-Wronski effect is a matter that has been intensively investigated for **> 20** years, but the final word is not yet in.
  - Nevertheless, **a-Si:H** solar cells do have a noticeable market share as shown below:



  - Note that a *constant* market share still means substantial growth in absolute numbers if there is substantial market *growth* - which we do haeve! Also note that beside **a-Si:H** solar cells, and lately **CdTe**solar cells, all other thin film cells hold only a miniscule market share (**CIGS**cells are shown in green; the red part denotes "***RGS***" - ribbon growth on substrate, a variant of **Si** solar cell technology).
- You probably own **a-Si:H** solar cells - in your pocket calculator, your watch, or in other small electronic devices where the efficiency (and durability) is not cricitical, but size, weight, price and easy production compatibility are important.
  - While there are also some large scale applications, it is not clear at present if **a-Si:H** solar cells will play a big role in the future.

Nanocrystalline Thin Film Si Solar Cell

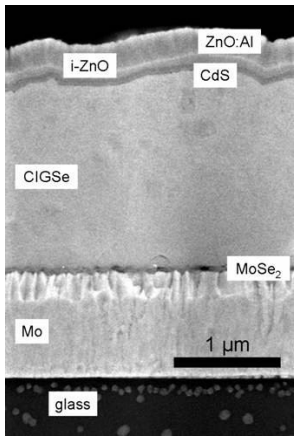
- While nowadays everything either just small or complex goes as "nano this-or-that", solar cells made from truly nanocrystalline thin fim **Si** is actually called "**micro-crystalline Si**" or **µc-Si:H** in most publications.
  - Be that as it may, the fact is that it became clear some **6** years ago that thin layers of **Si** that actually consist of **Si** nanocrystals (diameters in the **10 nm** region) embedded in a **α-Si:H** matrix (still containing plenty of hydrogen), if made exactly right, could be used to make solar cells with efficiencies of up to **14 %** and with practically no Staebler-Wronski effect.
  - That gave a big boost to the many researchers engaged with **a-Si:H** solar cells since it offered a new outlook to the future of these solar cells.
  - What exactly makes the **µc-Si:H** thin films so much better than just **a-Si:H**, and how far this can be carried, is not all that clear at present.
- To the best of my knowledge, there are no nanocrystalline thin film **Si** solar cells up on roofs in a major way at present. If this kind of thin film solar cell will make it into large scale production, is an open question at present, but it is seen as a major future contender for the presently firmly entrenched bulk **Si** solar cell.

Multi-crystalline "Thin" Film Si Solar Cell

- Make a bulk **mc-Si** solar cell with just about **(2 - 5) µm** thick multi-crystalline **Si**. It will not absorb all the light, so add a "mirror" at the back and other measures to keep the light inside the **Si**.
  - This would work quite well - except that you can neither make **5 µm** thin slices, nor handle them without fracturing them rather quickly.
  - So put your **5 µm**, or even just **2 µm**, of larged-grained or "multi-crystalline (**mc**)**Si** on a mechanically (and thermally) stable substrate. This is the concept of the "**thin film Si**" solar cell as it is usually called.
- There is no uncertainty about possible high efficiencies, long life times without degradation, etc. - it's good old and well-known **Si** technology after all. The problem, alluded to many times by now is: money.
  - In other words, you are competing aginst good, old, and well-known bulk **Si** technologies - and you must be just as good and considerably cheaper before you are being noticed. So you must solve the follwing problems:
    - Deposit rather thick layers of rather good **Si** (pure, large-grained, ...) *quickly* (at least **1 m^2/min**). Normal **Si CVD** processes are not that fast, and tend to produce small-grained films.
    - Find a cheap and process-compatible substrate.
    - Think about the temperature budget. While the **Si** part would be easier at high temperatures (deposition rates go up, grains grow larger, ...), you better stay at low temperatures all the time - or you will be very restricted in the choice of your substrate (heating also costs money).
    - What about the backside contact if that cannot be your substrate (because you picked glas)?
    - And what about the backside "mirror"?
    - And so on and so forth.
  - Impressive progress has been made with the thin film **Si** solar cell. If it will become competitive in the near future remains to be seen.

The CIGS Solar Cell

- Solar cells from the **CIGS** family have an "old" history - it goes back more than **20** years to the times around the first oil crisis (**1975**), when semiconductors of the metal**1**-metal**2**-chalgogenide family were studied with some effort..
  - One of the first promising materials contained mercury (**Hg**), and since mercury acquired a really bad name in those days, there were reservations about these materials, too. This might be about as smart as having reservations about **NaCl**because both **Na** and **Cl** do belong to the terrorist group in the periodic table.
  - Anyway, with lots of research it became clear that the "**CIGS**" 'family showed real promise. Serious production started around **2005**, and by now the first large factories are up and r unning.
- Some topics around **CIGS** solar cells have already been raised above. Here we note that having an (always **p**-doped) **CIGS** layer of about **(1 - 2) µm** in thickness is not enough, we also need a **pn**-junction and (low resistivity) contacts.
  - The **pn**-junction is actually a **hetero junction**; the **n**-part is supplied by a very thin layer of **CdS** and intrinsic (=high resistivity) **ZnO**. The backside metal is **Mo**, possibly reacting with the **CIGS** upon deposition to a thin intermediate layer of **$MoSe_2$** as shown above.



Cross-section through a **CIGS** cell

- Heavily **p**-doped (with **Al**) **ZnO** forms the transparent contact layer on top. The not-so-great resistivity of the top layers causes problems that will be dealt with in the next sub-chapter. Just take it for granted at that point that the simple technique of screen printing some metal contacts, as known from bulk **Si** solar cells, does not work here.

- Anyway, after many years of **R&D**, we have almost **20 %** efficiency **CIGS** cells in the laboratory and around **13 %** in mass production.
  - Also, after many years of **R&D**, we still have some "**black art**" involved in making **CIGS** solar cells. In the words of the Hahn-Meitner-Institut in Berlin, one of the leading **CIGS R&D** places:

    "*Despite the large CIGS progress made in the last 10 years, the material science (of CIGS) is still behind that of established semiconductors like, e.g., Si. Progress in increasing the efficiency by semi-empirical optimization of the physical and chemical processing steps was essential for success. In other words: 'It worked first and was explained later'*"
- Since **CIGS** solar cells have the potential for good efficiencies around **20 %** and cold be potentially cheap, they are seen as the major competitor to the **Si** bulk cell right now. Time will tell.

The CdTe Solar Cell

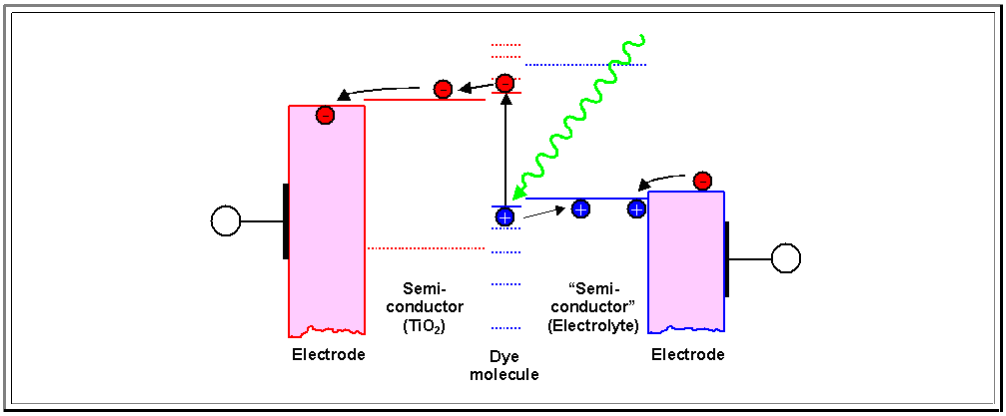
- **CdTe** solar cells have a bad image right now (around **2007**) because they use **Cd**, one of the more ***despicable*** elements in present mythology (possibly belonging to the terrorist group of the periodic table).
  - However, **CdTe** solar cells are relatively easy to make, not too bad, and rather cheap. The real problem comes from the fact that all of the solar cell industry must consider how to recycle their products eventually - and that might be not so cheap with **Cd** containing waste.
  - Beside image problems, a real problem might be that the efficiency of **CdTe** solar cells might be limited to values not competitive in future markets. There are, however, already large-scale installations as shown below. Again, time shall tell if **CdTe** solar cells are here to stay.

300 kW_p solar park with **CdTe**solar cells made by Antec; in Aurich, Germany

- **Feb. 2010**: Time did tell. First Solar, making exclusively **CdTe**solar cells, is now the biggest solar cell company on the planet.
  - Thin film **CdTe** solar cells enjoyed a tremendous success story in just a few years. The concern now is rather if there is enough **Te** around to allow many **GW** of installations.

The TiO_2 - Dye Based Family

The basic idea behind this solar cell concept is best illustrated in a schematic band diagram



- We use an optimized ***dye*** molecule that absorbs light extremely efficiently, and in doing so kicks an electron up from the **highest occupied molecular orbital** (***HOMO***-level; shown as solid blue line) to the **lowest unoccupied molecular orbital**(***LUMO***-level). In semiconductor terms an electron-hole pair is generated; in more general terms we have an exciton (=electron-hole pair at close range, feeling some interaction).
- The trick is to separate the two carriers. If we find suitable semiconductors that have their conduction band just right below the **LUMO** level (so the electron can jump over with some energy gain) or their valence band right above the **HOMO** level (so the hole can jump over with some energy gain), respectively, and some metal electrodes with suitable Fermi levels - we have a solar cell.
- Well, not quite, because at least one side needs to be transparent to light. In this case it will be the hole side, where Indium-Tin-Oxide commonly known as "***ITO***" works well enough. On the electron side we can use **TiO_2** and some metal with a lot of advantages as we will see.

The problem is the blue "hole" semiconductor. It must alos be transparetn to light - and presently there is no good candidate. The so-called "**Graetzel**" **cell**, the paradigm of this kind of cell, therfore uses a "redox" **electrolyte** that can transport the electrons from right to left (akin to transporting a hole from left to right) at the proper energy levels.

- The big advantage of the **Graetzel cell** (and its many variants) is that the **TiO_2** + dye side is very simple and dirt cheap: Take **TiO_2** nanoparticles, sinter them lightly onto a metal substrate, and coat the resulting nano-sponge with the dye. **TiO_2** nanoparticles dispersed in some solvent, in case you don't know, is what you buy as "white paint" in building supply stores - it is truly cheap.
- The dye might be a problem, but organic chemistry will solve it for you. It may not be dirt cheap but you only need tiny amounts.

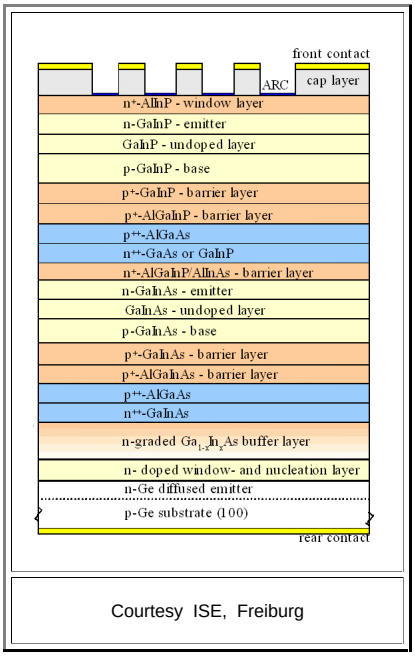
The real problem, of course, is the electrolyte side, in particular if it is supposed to work for **> 20** years without a problem.

- The Graetzel cell holds a lot of attraction for many **R&D** oriented people and organisations. It doesn't require expensive semiconductor equipment to start experiments, and it offers a chance for researchers from various disciplines other than semiconductor materials science to try their hand at solar energy. It is therefore small wonder that some "Graetzel cell" concept has been touted as the solution to the problems plaguing solar energy every few years since the eighties.
- However, the few efforts at large-area semi-commercial Graetzel cells invariably did not result in industrial production. One problem is that efficiencies are well below **10 %** at present - and that is just not good enough. Another problem is the required massive series connection, which is just not that easy if liquids are involved.
- Nevertheless, given the large number of researchers and institutions working on some variant of this cell concept, the final word is: Time will tell.

Tandem or Multi Junction Cells

- It was emphasized many times that with a given band gap the efficiency of a solar cell is limited, and that the "best" bandgap for sunlight would be about **1.5 eV**, giving a maximum efficiency of about **30 %**.
  - How about using solar cells made from semiconductors with different bandgaps, stacked on top of each other? The first one, facing the light should have a wide bandgap, taking out just the ultraviolet part of the spectrum. It will generate a relatively high voltage and some current.
  - Below the first layer then sits a medium bandgap solar cell, responding, let's say, to yellow and beyond. It would generate a medium voltage and some current.
  - Finally, in the **triple junction** solar cell we are generating, we have a small bandgap semiconductor absorbing the left-over red and infrared radiation; generating a small voltage and some current.
- We see the first problem of our **high-efficiency multi-junction** solar cell: we need **current matching**! The current generated in the individual cells must flow through all three cells since they are switched in series. If the three cells do not generate the same amount of current, we will have problems.
  - *Current matching* can be achieved if the thickness of the individual layers, the bandgaps, and so on, are matched to the solar spectrum. Tedious, but possible.

    Class Exercise: *What would happen if there is no current matching? It is sufficient to consider a tandem cell with two junctions.*
  - If we made sure that we haev current matching, the individual cell voltages will add up and we may have an efficiency exceeding the maximal **30 %** for single junction cells. For example, with **Ge** as the "bottom cell", and plenty of other cells on top (∞ many with incrementally increasing bandgap for "easy" theory), the theoretical limit is **> 50 %**.
- However, besides current matching, we must solve another problem. If we pile many different semiconductors on top of each other, the interfaces should better be "good" in the sense of not recombining carriers. In other words: no misfit dislocations.
  - That narrows done the possible combinations but we still can do quite a bit - as shown in the figure. The band gap goes up in steps from the **0.66 eV** of **Ge** to about **1.8 eV** of the **GaInP**.
  - The misfit is kept small, and experimental efficiencies in excess of **40 %** have been obtained



Courtesy ISE, Freiburg

- The catch is obvious: Multi-junction solar cells will not come very cheaply! There is no way to make **m²** of this structure in an inexpensive and fast way. This leaves two escape routes:
  - **1.** Make far simpler structures - e.g. a simple variant of a tandem cell, e.g. by just using **a-Si:H** on top of **a-Si/Ge:H**. Instead of using amorphous **p-Si** as absorber, alloy some **Ge** into this layer, lowering the band gap. On top put regular **n**-type amorphous **Si** with a larger band gap. This will not give you "high efficiency", but possibly more than just a simple **a-Si:H** solar cell. Or use **n**-type **a-Si:H** with is relatively wide band gap on top of crystalline bulk **p**-type **Si**. Or --- you get the idea.
  - **2.** Make the best multi-junction solar cell you can, and put it in the focus of some parabolic mirror, or any other cheap focussing device like Fresnel lenses. In other words, concentrate the sun light collected from a large area on a small area solar cell - which then can be expensive. This is the principle of the **concentrator solar cell** with the obvious catch that it will only work in direct sun light (i.e. not in my home state of Schleswig-Holstein, where we might not see the sun for weeks on end) *and and* if some mechanical contraption follows the sun.
- There is definitely a market for concentrator cells; how large it is remains to be seen.

Exotica

- Even the most stupid American (G.W. Bush?) realized by now (2007) that it is time to worry about the climate *and* about the finite resources of cheap energy carriers like oil. This means that the **USA** are just now (re)discovering solar energy. Since they almost completely missed the big economical boom of solar energy after the year **2000**, they are lacking behind in solar cell production but have now taking the lead in solar cell ***hyperbole***.
  - Check the Internet. Tip: Start with "nanosolar". You will find all kinds of new concepts for making cheap and very good solar cells. However, as the saying goes: "the proof of the pudding is in the eating" - or time will tell.
- There are, however, many more serious "exotic" solar cells concepts in the sense that they are either considered or actually used for very special applications (like supplying power to satellites), or objects of intensive **R&D**.
  - In the first category we have all kinds of **GaAs** based solar cells - optimal band gap for best efficiencies of single junction cells; used whenever price doesn't matter (in space).
- In the second category we have, for example, solar cells based on **organic semiconductors**.
  - Light emitting diodes (electric power in; light out) made from organic semiconductors are already on the market; so the inverse device (Light in; electric power out=solar cell) should be possible too as seen from a somewhat naive point of view.
  - Yes, organic solar cells are possible. Right now, however, they are subject to a host of problems. If those problems could be solved, leading to a potentially extremely cheap technology ("paint on your solar cell"), only time will tell.

Famous Last Words

- In the text above, we have a lot of "*remains to be seen*" or "*time will tell*". This is true enough - provided somebody does the work required.
  - This somebody will have to be someone like *you*! The matter is complex, transcending the ken of classical physicists or chemists. It is right at the heart of Materials Science and Engineering.

8.3.2 Essentials of Producing Thin Film Solar Cells

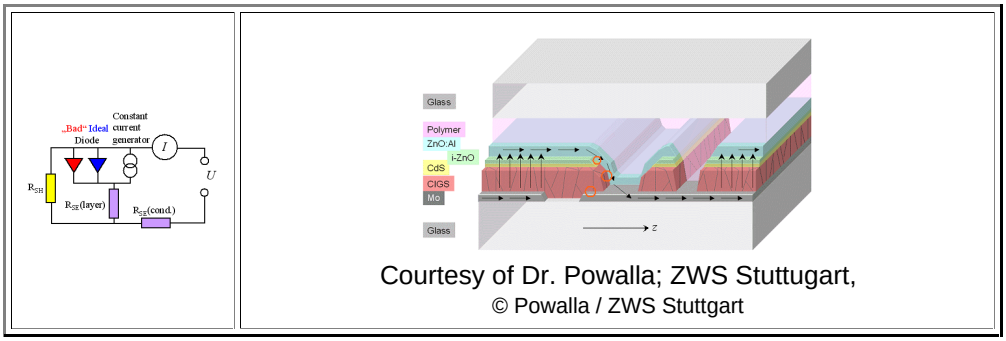
General Remarks

- If we look at the cross-section of the **CIGS** cell in the preceding subchapter, we see parts of what we are up to if want to mass produce this type of thin film solar cell.
  - We have to deposit various thin films on a glass substrate. In order of appearance we have
    - **Mo**
    - **CIGS**
    - **CdS**
    - Intrinsic-**ZnO**
    - **Al** doped **ZnO**
    - Some Polymer

    and then put the lid on - another glass plate.
  - Then we must "put the lid on" - another glass plate.
- The **CIGS** materials would be the "base" of the solar cell diode or the principal absorber material; it is always **p**-doped. The **CdS** and the **i-ZnO** form the **n**-doped or emitter part of the diode. They have bandgaps different from that of **CIGS**, so we actually have what is called a **hetero-junction**, something we will encounter extensively in Optoelectronics.
  - Hetero-junctions have basic properties similar to the **pn**-juntion we know, but a number of peculiarities, too, which we will not go into at present. The **CdS** layer, a rather crucial part of the **CIGS** solar cell, also serves as **"buffer" layer**, mediating the lattice mismatch between **CIGS** and **ZnO**, a necessary function to keep the interface properties in the tolerable range.
  - The heavily **p**-doped (by **Al**) top **ZnO** layer provides the transparent front side contact (and, perhaps, some anti-reflection coating).
  - The two top layers of polymer and glass are rather "trivial"; the essentially keep the rain out and just protect the whole cell for **20** years or so from the environment (including things like extremely corrosive birds shit).
- The first task thus is to deposit those layers - on a substrate typically **1 m × 1m** in size or, using some a role-to-role process, on a flexible substrate like polyimide or stainless steel that moves inside some piece of equipment at **1 m/min** or so and comes out at the other end with at least some of the relevant layers on it.
  - For layer deposition we can use some of the processes we learned about earlier, use some that have been around but that we have not encountered yet, or use the ones especially invented for exactly this task. In other words: We cannot go into this here at all.
  - Instead we look at some specialities of thin film layer solar cell production that goes beyond "merely" depositing thin films on a substrate

"In-situ" Series Connection

- As we have seen before, having very low series resistances is decisive if you want to have a solar cell with an efficiency η **> 10 %** or so.
  - Unfortunately, even heavily doped **ZnO** - a member of the important material class of "***TCO***" or **transparent conductive oxides** - has a lousy specific conductivity of at best **200 μΩcm** - two orders of magnitude above that of decent metals like **Ag** or **Cu**.
  - Worse, the intrinsic **ZnO** layer is almost an insulator, and the **CIGS** itself is also not overly conductive. We have a problem shown in the left half of the picture below; the right half shows the solution.



Courtesy of Dr. Powalla; ZWS Stuttugart,

- In some simple equivalence circuit diagram we must split the total series resistance somehow into an internal series resistance coming from the active layers and an "external" one coming from the "wiring", i.e. the conductors. The bad one is the internal one, because there is not much we can do about that for a single solar cell. This is one of the reasons why we do not use a low-resistance metal grid like for the **Si** bulk solar cell in this case: it wouldn't help much to bring the series resistance down.
- The only way to cope is to make the individual solar cells small in one direction (***z***-direction)- we have a stripe of about **1 cm** in width and up to **100 cm** long - and to switch these stripes in series. In other words, we produce little current in ***z*** direction but a lot of voltage. The voltage loss in the internal resistance is proportional to the current; if we keep the current low, we keep the internal losses down. In other words we use the same kind of reasoning that leads to high-voltage power lines in the face of finite line resistances.
- The problem, of course, is that we cannot manufacture individual stripes of thin film solar cells and then solder them together - we must do all this series connection "in-situ" while we deposit the various layers - fast and cheaply.

Sounds difficult - is difficult. We can also learn an important lesson:

- A potential disadvantage - large internal resistances - may be turned into an advantage. Just solve the problem of how to do "automatic" series connections, and you are no longer producing mediocre solar cells but large-area solar *modules* with a high-voltage output that not only alleviates the internal resistance problem but is just perfect for the power conversion electronics that must come after the module.
- As always, there are a few second-order problems with this approach too, but basically it is a sound approach that is used with practically all thin film solar cells.

The way it is done is shown in the schematic cross-section of a **CIGS** solar cell at the left-hand part of the picture above. If you follow the black arrows symbolizing current flow, you clearly see that the top of the left cell is connected to the bottom of the right cell; we have a series connection.

- If you look at the red circles, you see also that we three nominal short circuits in this structure:
  - The **ZnO:Al** layer, which is the front side "conductor" here, short circuits the junction formed by the **CIGS** and the **CdS/i-ZnO**.
  - The **ZnO:Al** layer also short circuits to the backside contact via the **CIGS** layer.
  - Both cells are short-circuited at the backside via the **Mo** layer overlap.
- This doesn't seem to make much sense. However, the short-circuits for the real thing are only nominal. In reality they have very high resistances (partially because the overlap areas are tiny) and therefore can be tolerated. If we accept this without further proof, a question still remains:
- Why? Why is this connection structure not made in a way that avoids these nominal short circuits? Even it they don't matter much, wouldn't it be better to stay on the safe side?
- The answer is: No! The way its done is the best - all things considered. One could easily draw cross-sections of smarter structures, but don't forget, you have to make them - cheaply and quickly. This leads to a little exercise:

Exercise 8_3_1

Making in-situ series connections

By now you should be utterly confused - and that is the point: What we see here (again) is that the simple issue of making solar cells, very simple devices after all, immediately blows up into your face and gets rather complicated if you look at any "detail".

- We will not go into more details here, but keep in mind that thin film solar cells *modules* are quite different from bulk **Si** solar cell modules.

Building and Running a Thin film Solar Module Factory

So your laboratory-scale process for a novel thin film solar cell has been very successful.- you can make something like **(5 × 5) cm^2** solar cells that have a decent efficiency of **η = 15 %** and can be made by what appears to be cheap processes.

- All that remains to be done is to built and equip a small (pilot) factory, capable of producing, let's say, solar cells good for **10 MW$_p$** per year. Having in mind that it takes about **8 m^2** for **1 kW$_p$**; we are talking about **8.000 m^2** solar cell area per year, or **22 m^2** per day. You want to process **1 m^2** at a time, i.e. make a whole module in one go.
- Where do you buy the layer deposition equipment and whatever else you need? The companies serving the microelectronic community can't offer anything for that kind of area. Companies making equipment for flat panel displays are more on target, but they don't know, for example, how to deposit **CdS**.
- And no company out there can help you with what ever it takes to do the in-situ series connections.

Your only choice left in the beginning it to make your own equipment, optimally in cooperation with some company that has a solid background in the general area.

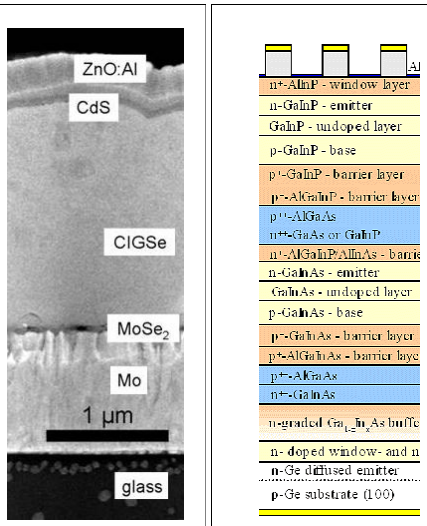
- After you did that, and it works, you do what? you keep it to yourself!

- Keeping this in mind, you understand why there is so little information around about how you actually make thin film solar cells on a large scale, and why there are some many variants of any process discussed in the literature.

What we are witnessing is an evolutionary process - trial and error - where competition on any level and survival of the fittest determines the outcome - and this is not necessarily what "pure" science would have chosen.

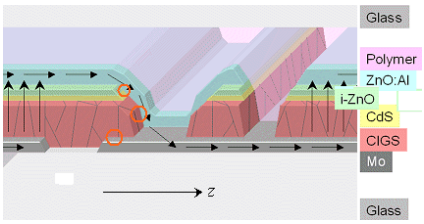
8.3.3 Summary to: 8.3 Making Thin Film Solar Cells

- Thin film solar cells need to meet some key requirements:
  - Process-compatible and cheap substrate ⇒ large area deposition.
  - Suitable direct band gap ⇒ high absorption coefficients f
  - Insensitivity to "defects"
  - Technology for junction and good ohmic contacts.
- Major contenders in (or close) to production are:
  - Amorphous **Si**.
  - Nanocrystalline thin film **Si**.
  - Polycrystalline thin film **Si**.
  - The **$CuIn_xGa_{1-x}Se_2$** or "**CIGS**" family.
  - The **CdTe** solar cell.
  - May others in **R&D**
  - The present "high potentials" are **CdTe** and **CIGS**.
- High-efficiency multi-junction solar cells may find applications as "concentrator cells" at the focus point of a large mirror or lens that tracks the sun.



CIGS | **Multi junction**

- **CIGS** and most other thin film solar cells have high internal resistances and need to be switches in series after about **1 cm** for high performance
  - This must be done automatically and in-situ as part of the production process.
  - A whole new technology needs to be developed for thin film solar cell mass production
- The race between bulk **Si** solar cells and thin film technologies is open in **2008**; the winning technologies are to be determined.



Exercise 8.3-2

All Quick Questions to 8.3

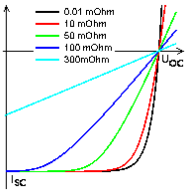
8.4 Summary

8.4.1 Summary to: 8 Solar Cells

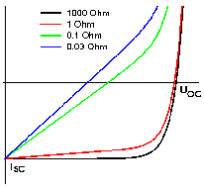
- A solar cell converts light power into electrical power. It's overriding parameter is the over-all conversion efficiency η
  - Any solar cell is essentially a large -area junction, usually of the **pn**-type.
  - It's essential parameter are the short-circuit current I_{SC}, the open-circuit voltage U_{OC} and the fill factor **FF**
  - For optimal efficiency the bandgap E_g should be matched to the solar spectrum; we need around **1.5 eV**.
  - Maximum efficiency from the *semiconductor physics* point of view is achieved if all light with energy **>=**E_g produces minority carriers and all of these carrier are swept out as diode reverse current and
  - Maximum efficiency from the *module systems* point of view is achieved if the semiconductor part is **OK**, only very little light is reflected by the solar cell module, series resistances and shunt resistances can be neglected, and everything is uniform and homogeneous

The equivalent circuit diagram with the basic equation has is all!

- Series and shunt resistances, unavoidable for large areas, are of overwhelming importance for solar cells with η **<≈ 10 %**



Effect of series resistance



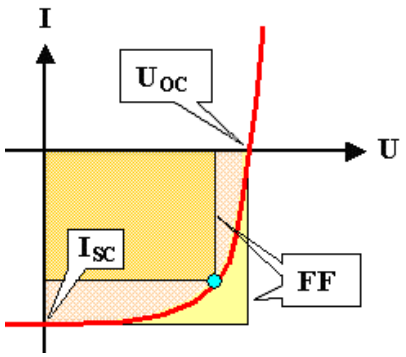
Effect of shunt resistance

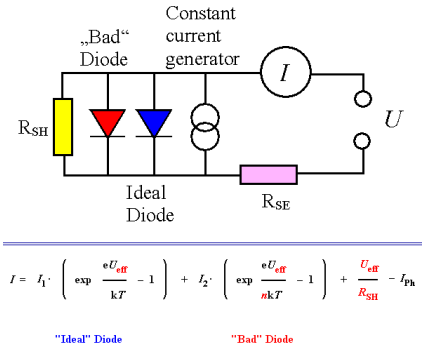
- Switching solar cells with individual characteristics in series and / or in parallel causes all kinds of problems.
  - Worse: Any inhomogeneous solar cell (e.g. **mc-Si** solar cells) consists of *locally* different solar cells "somehow" connected internally
  - Optimizing solar cells with respect to "money" thus provides exciting science and engineering!

Important "raw" numbers.

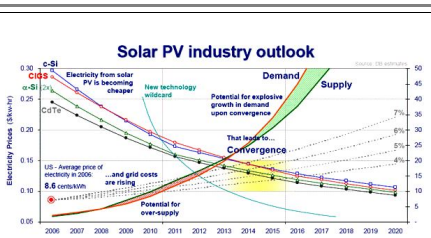
- **Maximum η Si solar cell ≈ 25 %**
- **Maximum sun power ≈ 1 kW / m².**
- **Maximum commercial solar cell power ≈ 200 W / m².**
- **Yearly average commercial solar cell power ≈ 25 W / m².**

Solar cell science and technology centers exclusively on *money* and *saving the earth*!





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



Courtesy Deutsche Bank

There are many competing solar cell technologies and materials.

- Bulk single-crystal and **mc Si** vs. thin film **Si (a-Si:H, µc-Si:H. ..**
- Other thin-film semiconductors:**CIGS, CdTe, ...**.
- Exotica: **TiO_2**- electrolyte ("Grätzel cell"), organic semiconductors, "Nano" materials, ...

Bulk **Si** solar cells are made from (cheap) single crystalline wafers (cut squarish) or from square multicrystalline (***mc***) wafers. They account for about **85 %** of the installed solar power at present (**2008**).

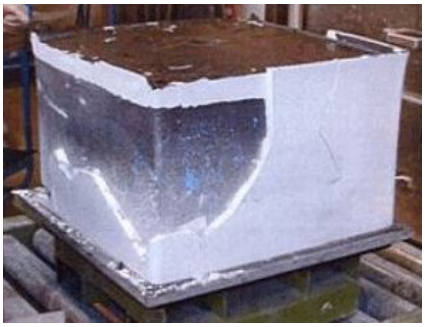
- A yearly production of **1 GW_{peak}** means about **10^7 m^2=10 km^2** **pn**-junction of good quality and much more
- Consider ⇒

  A big problem is cranking up world wide **Si** production by **30 %** - **40 %** per year.

Processing Time	**1s / solar cell**
Cost Decrease	**5 % / a**
Efficiency Increase	**20 % in 2012 ?**
Key Material Supply	**30 % /a more Si**
Industry Growth Rates > 30 % for many years	**Supply capital and people**

mc wafers are produced by **Si** casting. Problems are

- Expansion upon crystallization.
- Reaction with walls of mold
- Columnar grain growth required
- **300 kg** ingots are routinely cast in **2007**; liquid encapsulation and precise temperature control are essential

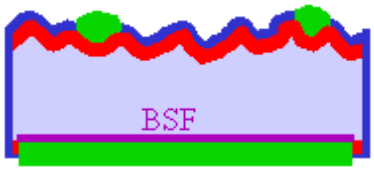


Sawing the ingot into **mc-Si** wafers with as little losses as possible and with wafer thicknesses of **< 300 µm**, while straight-forward, is "high-tech".

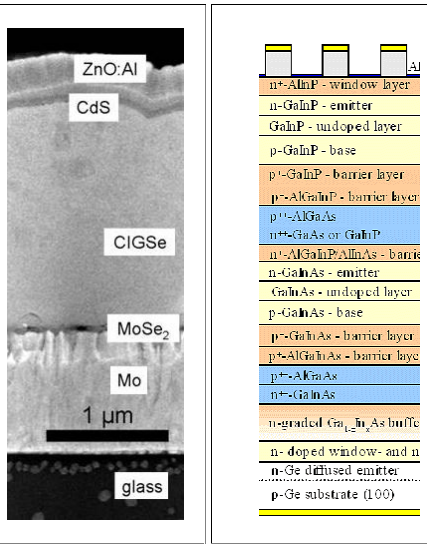
- Saw damage is removed by a chemical etch.

Processing, simple in principle, has to meet the conditions above and is highly specialized. Essential processes are:

- Diffusion, edge isolation, passivation, screen printing contacts and sintering contacts.
- Essential device features are back surface field, gettering of impurity atoms, **H**-passivation of grain boundaries and other defects.



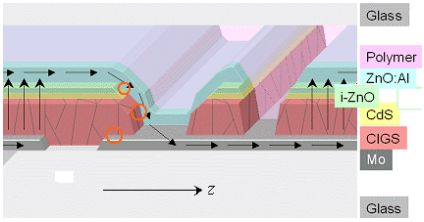
- Thin film solar cells need to meet some key requirements:
  - Process-compatible and cheap substrate ⇒ large area deposition.
  - Suitable direct band gap ⇒ high absorption coefficients f
  - Insensitivity to "defects"
  - Technology for junction and good ohmic contacts.
- Major contenders in (or close) to production are:
  - Amorphous **Si**.
  - Nanocrystalline thin film **Si**.
  - Polycrystalline thin film **Si**.
  - The **$CuIn_xGa_{1-x}Se_2$** or "**CIGS**" family.
  - The **CdTe** solar cell.
  - May others in **R&D**
  - The present "high potentials" are **CdTe** and **CIGS**.



CIGS | **Multi junction**

- High-efficiency multi-junction solar cells may find applications as "concentrator cells" at the focus point of a large mirror or lens that tracks the sun.

- **CIGS** and most other thin film solar cells have high internal resistances and need to be switches in series after about **1 cm** for high performance
  - This must be done automatically and in-situ as part of the production process.
  - A whole new technology needs to be developed for thin film solar cell mass production
- The race between bulk **Si** solar cells and thin film technologies is open in **2008**; the winning technologies are to be determined.



- Solar cells have a bright future!

Exercise 8.4-1

All Quick Questions to 8

9. Optoelectronics

9.1 General Concerns

9.1.1 Semiconductors and Light

9.1.2 Optoelectronic Devices

9.1.3 Summary to: 9.1 Optoelectronics - General Concerns

9.2 Important Principles and Technologies

9.2.1 Light Emitting Diodes

9.2.2 Laser Diodes

9.2.3 Summary to: 9.2 Optoelectronics - Important Principles and Technologies

9.3 Summary

9.3.1 Summary to: 9. Optoelectronics

9. Optoelectronics

9.1 General Concerns

9.1.1 Semiconductors and Light

Basic Questions

- **Optoelectronics** here means the coupling of *optics*, i.e. *light* (including all relevant wavelengths from the far infrared to the deep ultraviolet) and *electronics*, and that this is somehow done with semiconductors. We have two basic possibilities.
  - **1.** The input is light, the output is an electrical signal or electrical power.
    2. The input is electrical power and the output is light
  - If we go for less orthodox definitions, we may also consider
    3. Electrical input switches light on or off or, more generally, modulates light.
- Let's look at a few examples:
  - For the first point we have solar cells, if you like, although few people would count then among optoelectronics proper. The paradigm for this field is the **optical sensor**, sitting, e.g., at the end of a fibre optics cable and converting the signal coming in in the form of light being on or off to electrical signals. A more ubiquitous product example is the "***CCD***" chip in your digital **8** Megapixel camera, that contains no less then **8 Mio** optical sensors in a matrix.
  - For the second point we have *two* paradigmatic products. ***LED's*** - **light emitting diodes** and ***LASER* diodes**. Both devices are closely related technologically, but quite a bit apart from their working principles.
  - The third example could include **flat panel displays** in the form of "***LCD's***" and all the light manipulating **MEMS** devices we already considered in the MEMS context. Nobody would list those devices under optoelectronics proper, but we do have a connection between light and semiconductors.
- As far as this lecture course is concerned, we can just touch upon the basic principles of **LED's** and Lasers. So let's look at a few general points in the context of producing light in a semiconductor by running a current through it. This brings us to the first major point:
- **1.** Generating light takes *power* - current I times voltage U.
  - Just having almost powerless voltage swings, which are (almost) all you need for processing information, will not suffice. There is some energy $h \cdot \nu$ contained in the light generated that we want to flow *out* of the system, and thus some energy $I \cdot U \cdot t$ must flow *into* the system.
  - That means we always have some **efficiency** η (measured in %) associated with the process of light generation; and η, as always, should be as large as possible. In fact, with **LED's** we can achieve rather large efficiencies **> 50 %** (compare that to to a light bulb with $\approx$ **7 %**!), if everything is done just right.
- **2. LED's** generate light with just one wavelength λ (of course with some $\pm\Delta\lambda$).
  - The **wavelength** λ (or the frequency ν) of the light generated is more or less determined by the band gap. We already looked at the possibilities we have for **bandgap-engineering** in our "master diagram"; the basic relation is

$$\lambda = \frac{c}{\nu} = \frac{h \cdot c}{h \cdot \nu} = \frac{h \cdot c_0}{n \cdot E_g}$$

    with E_g=band gap energy, and c=speed of light in the medium the light is propagating, i.e. $c=c_0/n$ with c_0=speed of light in vacuum, and n=index of refraction $=(\epsilon_r)^{1/2}$; ϵ_r is the relative dielectric constant of the material.
  - There - you have elementary optics coming in as you should have suspected when the "*opto*"electronics came up. In other words, the **dielectric constant** and with it the **index of refraction** of our semiconductors enters the game here, a property we have hardly noticed so far.
  - Most semiconductors, actually, have a rather large index of refraction; some values are given in the table below. In case you wonder why materials that are not transparent to light should have an index of refraction, keep in mind

    1. All semiconductors are perfectly transparent to light with $h \cdot \nu < E_g$; so at least for small enough wavelengths an index of refraction makes perfect sense.
    2. In that part of optoelectronics where we want to produce light, we must work with wavelengths where the semiconductor is at least semitransparent. Otherwise no light could possible come out of the material.
    3. If we cool *any* semiconductor down to a sufficiently small temperature, it will be a good insulator for which we can easily define a dielectric constant and thus an index of refraction.
  - In fact, looking ahead to somewhat more advanced material science, everything contained in notions like *electric conductivity*, *dielectric constant*, *index of refraction*, or *absorption coefficient* are just special aspects of a so-called (complex) **dielectric function**, which is nothing but the (complex) dielectric "constant" as a function of frequency; more to that in the link.
- **3.** A third point of general interest is the *absolute intensity*, or even more specific, *the intensity density* we can produce. In other words, when we generate a certain "amount" of light as given by energy in times efficiency, from what kind of volume is it coming from.

- From a "point source" as we like it (in text books) for doing optics with lenses; from some finite volume like in a real light bulb, from some elongated structure like in a fluorescent tube, or from a large area like - ???.
- Well, with "normal" light sources you have a problem if you need two-dimensional light sources. You just can't throw a light switch and have a whole wall of your room start glowing uniformly - until you go for (future) **LED's**, that is.
- Whatever, we have an important point here that we need to address if we want to use **LED's** for making light.

If we take points **1 - 3** and remember that a photon in an **LED** will only be produced if an electron and a hole recombine in a process that transfers their energy to a photon (and not to the lattice or to something else) we have our task cut out:

1. Recombine as many electrons and holes as you need per second (⇒ *current*)
2. for the amount of light you want to produce (⇒ *intensity*)
3. with as many recombination events as possible producing light (⇒ *efficiency*)
4. in a defined volume (⇒ *intensity density*)
5. *and* get the light produced out of the semiconductor (⇒ *efficiency once more*).

- In simple and straight terms, we need to consider how we can handle *massive recombination* in a given volume. We need to do this for a number of different semiconductors because we need different wavelengths.
- This is new! So far we tried to either avoid recombination (e.g. in solar cells) or to ignore it (e.g. in the simple standard version of the diode equation). Now recombination is in the focus of what we need to optimize.
- What is also new to some extent is that we are now using (electrical) **power=*UI***. It you want to produce a lot of light (=energy), you must provide power. You can't get much light out of **1 W** light bulb and you can't get much light out of just a **1 W LED** either. This is new to some extent because in microelectronics (=communication and signal processing) we avoided power as much as we could - we had voltage swings, but no current if possible.

Material Issues

Color	Wavelength (nm)	Typical Semiconductor
Infrared	**880**	**GaAlAs/GaAs**
Red	**660 - 633**	**GaAlAs/GaAs**
Orange to **Yellow**	**612 - 585**	**AlGaInP** **GaAsP/GaP** **GaAsP/GaP**
Green	**555**	**GaP** **AlInGaP**
Blue to **Ultraviolet**	**470 - 395**	**AlInGaPN****GaN/SiC** **GaN/SiC** **InGaN/SiC**

As already pointed out above, we need several suitable semiconductors to cover all aspects of optoelectronics.

- It also became clear that we have to look at more properties than we did so far : For example, *dielectric constants* and *recombination mechanisms* are now prime parameters we must consider.
- The value of the band gap, too, is now of *prime* importance (In contrast, we wouldn't have cared much about the *precise* value of the bandgap for **Si** technologies!).

Besides appreciating the material overview given in the table at the right, let's look a bit closer at major semiconductors and their properties in detail: The same table but with more entries concerning properties we have not yet encountered can be found in the link.

- Interesting points that can be found in this table are
  - *Indirect* semiconductors like **GaP** are listed!
  - A *strange recombination* mechanism ("exciton - band") is listed
  - A *strange semiconductor* (**$In_{0,53}Ga_{0,47}As$**) is listed

Here is a table with a lot (but not all by far) interesting properties of some semiconductors:

Properties	Si	Ge	GaAs	InP	InSb	$In_{0,53}Ga_{0,47}As$	GaP	GaN	SiC	Diamond	Remarks
Crystal											
Unit weight [mol]	29,09		144,63	145,79		168,545					
Density [g/cm^3]	2,33		5,32	5,49		5,49			3.166 (cubic) 3.211 (hex)	3,51	

Crystal structure	Diamond	Diamond	Sphalerite	Sphalerite	Sphalerite	Sphalerite			Many variants cubic, hex, rhombohedral	Diamond	
Lattice constant [nm]	0,5431	0,565	0,565	0,587	0,648	0,5867			a=0,30 c many values		
Transport properties											
Band gap [eV]	1,12	0,66	1,42	1,35	0,17	0,75	2,26		2.39 - 3.26	5.47	
Type	Indirect	Indirect	Direct	direct		direct	indirect		indirect		
Effective e^- mass $[m^*/m_0]$	0,98								0,24 - 0,7		
Effective h^+ mass $[m^*/m_0]$ light heavy	0,16 0,49		0,082 0,45	0,12 0,56	7,3	0,051 0,50			0.9		
N_{eff} in C $[10^{18}cm^{-3}]$	28 (32)	10,4	0,47	0,54	0,042	0,21					
N_{eff} in V $[10^{18}cm^{-3}]$	10 (18)	6	7	2,9		7,4					
n_i $[10^6cm^{-3}]$	6 600 13.000		2,2	5,7		63 000					
Mobility (undoped) $[cm^2/Vs]$ μ_n μ_h	1 500 450	1 900 3 900	8500 450	5 000 200	80 000 1 250	14 000 400	300 150		500 - 1 000 20 - 50	200 - 2 200 1.800 - 2 100	
Lifetime (general) [µs]	2500		0,01	0,005		0,02					
Mecha-nism of lumines-cence	None		band-band	band-band		band-band	exciton-band if doped				
Dielectric properties											
Dielectric constant	11,9	16	13,1	12,4	17,7	13,7	11,1		9.7 - 10	5.5	
Break through field strength [kV/cm]	300		350	400		100					
Specific intrinsic resistance [MΩcm]	0,2		310	11		0,0008					
Electron affinity [eV]	4,0	4,05	4,07	4,4	4,59	4,63	4,3				
Thermal Properties											
Expansion coefficient $[10^{-6}\ ^oC^{-1}]$	2,6		6,86	4,75		5,66	5,3			1	
Therm. conductivity [W/cmK]	1,5		0,45	0,68		0,05			5.0	22	Cu: 4.01
Specific heat $[J/g^oC]$	0,7		0,35	0,31		0,29			0.671	0.428	Cu: 0.38

Melting point [^{o}C]	1 412	937	1 238	1 062	527	970					

- **Si**, **Ge** and diamond are included as references; if there are several numbers, they are from different sources. We find expected properties, but also, perhaps, some unexpected ones:
  - *Dielectric constants* are relatively large even at the very high optical frequencies. That means we have also large *indexes of refraction* $n=(\epsilon)^{½}$.
  - **Thermal conductivity** can be rather large (**SiC** and diamond!). That's good because we have to get a lot of heat out of our optoelectronic "power" devices.
  - There is at least one recombination mechanism not mentioned before: *Exciton - band* recombination.
- This tells us that we now need to look a bit more closely at radiant *recombination*, the key to light generation.

Radiant and Non-radiant Recombination

- Let's start from the typical situation of a working **LED** or **Laser**. 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***.
  - What they *want* to do is to recombine because that would be the right step towards equilibrium.
  - What they are going to do depends on the conditions. They must be "just right" otherwise recombination will not take place. Essentially three conditions must be met:
    1. The electron and the hole must be in the *same place* at the *same time*. The probability for that happening is certainly proportional to the respective concentrations n_e and n_h.
    2. We must *preserve energy*. The energy released by the recombination ($=E_g$) must appear somewhere else; in the world of quantum mechanics always in some *particle* like a freshly generated photon.
    3. We must conserve (crystal) momentum.
  - In direct semiconductors we know we don't have to worry about point **2** and **3**. Whenever boy meets girl in this case, things can happen and they can recombine, emitting a photon. We have what we now will call **direct recombination**.
- This is still true and will remain true. But now we have to look a bit more closely and realize that other things beside the direct recombination *might* happen too.
  - For instance, 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 is exactly how recombination happens with the help of defects in indirect semiconductors. There is no reason in the world why this must not happen in direct semiconductors, too. It does, and the problem is that you can meet conditions **2.** and **3.** from above without emitting a photon, The third partner to the process - the defect - takes care of energy and momentum conservation.
- We thus have recombination via defects as a second **recombination channel**, to use the proper term. We don't like this recombination channel because it takes electrons and holes out of circulation without producing light, and therefore decreases the efficiency of our **LED**.
- Now we must make a leap in imagination. Third partners to the recombination process, that much is clear, may somehow influence what happens. The question is thus: What kind of third partners besides defects do we find in semiconductors and what, exactly, are they doing to recombination?
  - A simple first answer is: Electrons in the conduction band. What happens is that an electron meets an attractive hole, recombines, and the energy released goes right to a second electron that happens to be close by. This process is called "**Auger recombination**" and we don't like it because it does not produce light.
  - A not-so-simple-answer is: Bound **excitons**! So an electron gets very close to nice hole, but can't recombine because they are in an indirect semiconductor, and their momentums just don't fit. However, there is still some attraction (Coulomb attraction) as long as they stay real close because they have opposite charges. So they form a bound electron-hole pair, encircling each other (always exactly opposite momentum!), and run about as a pair called "*exciton*" that would like to do it but can't quite without a little help from a friend. Then they encounter a friend in the form of a special defect (**N** atoms in **GaP**, for example) that traps the couple, i.e. *localizes* it in space. Great - because now Heisenbergs uncertainty relation kicks in: $\Delta x \cdot \Delta p > h$. If Δx is small, Δp, the uncertainty of the momentum p, is large. If your momentum is uncertain, what exactly needs to be preserved? Right - happy recombination, you have made it now. And don't forget to ~~use a condom~~ emit a photon while you're doing it because we still must have energy conservation!
  - Wow! We now have a somewhat strange recombination channel that produces radiation, which is good!. It is actually the only reason why the indirect semiconductor **GaP**, if "doped" with **N** to catch the excitons (in addition to the doping that produces the electrons and holes), will be a quite efficient material for green **LED's**.
  - This little excursion into more advanced semiconductor technology just serves to demonstrate that there are indeed more recombination channels than we might have thought of. Moreover, the "blabla-ons", they many so-called "**quasi-particles**" of solid state physics, often considered to be exotic curiosities with no conceivable uses, are quite useful, after all. At least some of them, including the exciton. The link will go somewhat deeper into these subjects.

- We aren't even done yet. There are even more recombination channels but what we have learned so far suffice to make the point:
  - If you want a high-efficiency **LEDs** and Lasers, you must optimize your recombination channels - and this may not be easy.

9.1.2 Optoelectronic Devices

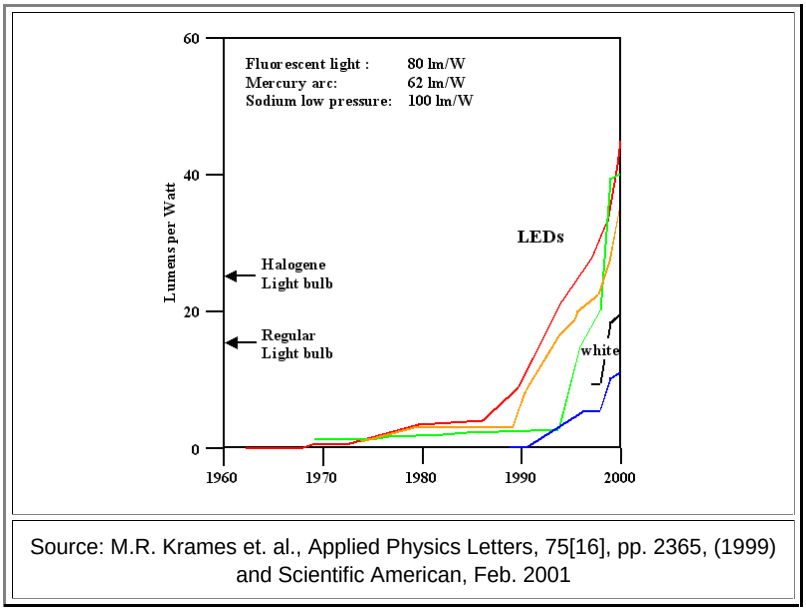
Remarks

- This module will start in a very superficial way. It will be neither complete in the sense that all important devices are mentioned, nor exhaustive in the sense that the most important specifics of the devices will be included
  - In time, it is hoped, this module will become more volume and depth from the work of the students in the accompanying seminar.

Light Emitting Diodes

- Light Emitting Diodes or **LED's** are the cornerstone of optoelectronic products. Very roughly they are used for two main product lines and for one speciality:
- **1. Signal lights** All those little red, yellow, green or blue lights (sometimes *blinking* annoyingly), mostly used for indicating that something has been turned on, is in a certain mode, or simply is just there (e.g. blinking red bicycle lights).
  - The main requirement for signal light **LED's** is that they come in many *colors* (so **designers** of dashboards etc. are not limited in their creativity; an important condition considering that designers appear to be limited in many other respects) and that they are *cheap*.
  - Nowadays (**2008**) we do have all colors - including "white" - but that was not always so. For generating a specific "color" including white you can go three routes:

    1. Take a semiconductor with the appropriate bandgap. This generates a "true" color, i.e. light around one wavelength as given by our "master diagram" from before.
    2. Take semiconductors that generates **UV** or at least high-energy light and use it to excite some fluorescent material - exactly like in fluorescent light tubes. That can produce white light or any color you find a fluorescent material for.
    3. Take three semiconductors that produce "**RBG**", i.e. red - green - blue, in such an intensity mixture as to produce the color wanted - exactly at it is done by any screen or display.
  - Obviously the first way is potentially the cheapest as long as you don't require white light. That's why you mostly find **LED's** belonging to the **GaAlAs** family (red), **GaP** family (green) or the nitride family **GaAlN** (blue to **UV**). Going from red to blue / **UV** also mirrors the history of **LED** development.



Source: M.R. Krames et. al., Applied Physics Letters, 75[16], pp. 2365, (1999) and Scientific American, Feb. 2001

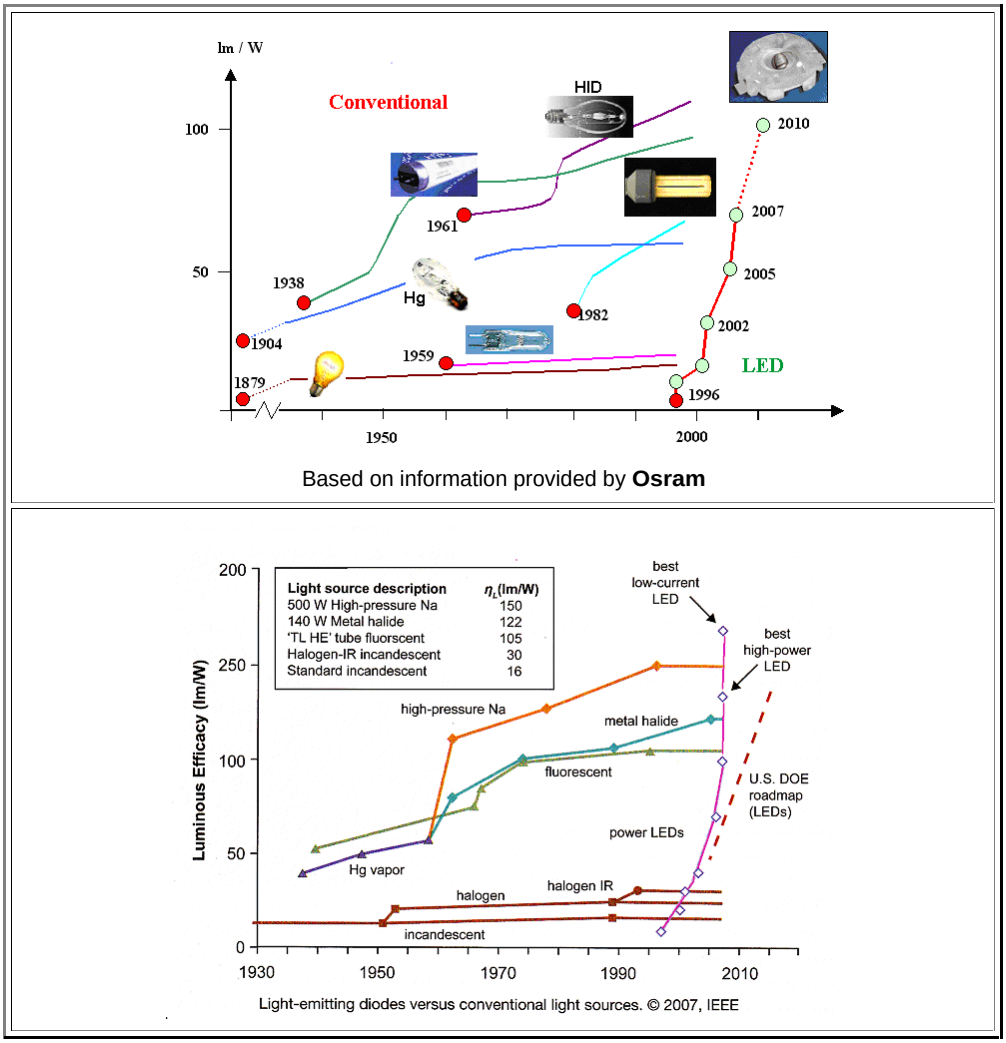
  - The orange **LED** dscribed in the article mentioned above already had an efficiency of **> 100 lm/W**, and today (Feb. 2008) **150 lm/w** white **LED's** can be purchased, for example from **Nichia** (the company that pioneered the blue / white **LED**). But now we moved already into the second topic:
- **2. Light** With the advent of the blue / **UV GaN**-based **LED** in **1993** (that comes with a quite interesting story around its inventor, Shuji **Nakamura** and the company he was then working for (Nichia)), making white light with **LED's** was possible for the first time.
  - For general lighting purposes - your room, the street, a congress hall, the street in front of you bicycle or car, - you name it - you need first of all *white light*. After you can do that with **LED's**, you need:

    - High **efficiency** as measured by lumen / Watt (**lm/W**) or by total "plug efficiency" in **%**, meaning the ratio of

light energy out to electrical energy ***Ult*** in.

- High *intensity*. It is not good enough to have a high-efficiency light source if the best you can offer produces the same intensity as, let's say, a **20 W** conventional light bulb.
- Large **life time**. You do not want to change your light fixtures too often. For something in the better quality region, you want several years of operation time at the least.
- Low *price*. The price you can get for your white light **LED** depends on what you offer. If it is much better than a regular light bulb, it does not have to be "cheap" - but it still must be worth its price.

The picture below shows the efficiency of white **LED's** vs. existing light sources:



The insets symbolize the type of "light bulb" sufficiently; the one belonging to the purple line (the top performer of the "classical" light sources) is the metal halide "bulb", belonging the the "high intensity discharge" (***HID***) type of light source. If you compare the development of the white **LED** to all the other light sources, you get a first impression why everybody in the lighting business is so excited about **LED's** as the light source of the future. Potential energy savings are enormous!

A quick word to the unit "**lumen**" and to absolute efficiencies:

The "*lumen*" (***lm***) is the **SI** unit of the *perceived* power of light; it measures the luminous *flux*. Now the natural way to measure the flux of light would be to measure the *energy flux* of the light. Since the eye, however, is not equally sensitive to the various wave-lengths it can see, the lumen corrects for this.

As far as converting **lm/W** into absolute efficiency goes, we have the following approximate relations

- **(10 - 15) lm/ W ⇔ (5 - 9) %**
- **(70 - 100) lm/ W ⇔ (25 - 35) %**

With respect to efficiencies, **LED's** do have a bright future, indeed. What about the other criteria?

Product life time is not a problem but an asset. Well-made **LED's** will last for **>10** years, outperforming more or less all other light sources.

The problem is: *intensity*. The light is typically produced in a small volume (great for focussing etc.), and if you put a power of let's say **100 W** into a volume of <**1 mm^3**, you better have some concept of keeping the temperature down.

Closely related is the problem of plug compatibility, meaning that you want to use **230 V 50 Hz AC**, **110 V 60 Hz AC**, or whatever your country has as its consumer electrical energy standard. A **LED**, however, is a forwardly biased **pn**-junction, running at something like **3 V DC** (and then **33 A** if you want **100 W**). While this is an electrical engineering problem, it is still a big problem.

The (inorganic) materials used for both applications (always as thin layers) are once again:

- Aluminium gallium arsenide (**AlGaAs**) — red and infrared
- Aluminium gallium indium phosphide (**AlGaInP**) — high-brightness orange-red, orange, yellow, and green
- Gallium phosphide (**GaP**) and Aluminium gallium phosphide (**AlGaP**) — green
- Gallium arsenide phosphide (**GaAsP**) — red, orange-red, orange, and yellow
- Gallium nitride (**GaN**) and Indium gallium nitride (**InGaN**) — near ultraviolet, bluish-green and blue

Laser Diodes

Semiconductor Lasers will be treated in some more detail in module 9.2.2. Here we simply note that the theory of (semiconductor) Lasers is rather complex, but the technology is not.

- In principle, many **LED's** "automatically" become a Laser if you run a very high current through them (producing a lot of light) without destroying them first. It is therefore not too difficult to produce a Laser diode in the Lab - all you need (haha) is very efficient cooling of your experimental **LED** device.
- You may know already that Lasers always need some kind of optical feed back, usually provided for by mirrors, and ask yourself where the mirrors are if we use a simple **LED** as Laser. The answer is that plan-parallel surfaces of the semiconductor might be already sufficient for that because the interface semiconductor - air does act as a "semi"-transparent mirror and that might be good enough.

Laser and optics of CD player.

The truth, however, is that it often takes many years after a certain new **LED** has been marketed, before the long-lived, reliable and cheap Laser diode follows.

- The first **GaN**-based blue **LED's** were on the market around **1993**, whereas the blue **GaN** based Laser had to await **2005** or so (in **2006** SONY, for example, still had major production problems).

Using semiconductors for making a Laser is just one way for making Lasers - you can use other solids, liquids and gases for that. This brings up the question of pro and cons - what are the advantages and disadvantages of semiconductor Lasers?

- Look at **CD** and **DVD** or now **blue ray** disc drives - the are the major market for semiconductor Lasers besides the very pedestrian "Laser pointer". The advantages are obvious:
  - Very cheap.
  - Very small.
  - Electric energy supply at low voltage.
- The major disadvantages are
  - Low power at decent quality (around **1 W** maximum).
  - Limitations as to color (= frequency).

The cheap and reliable semiconductor Lasers are actually the **enabling** devices for all this memories! No suitable Laser - no discs.

Sorry - your Laser *pointer* just can't be turned into the Laser *gun* you might fancy for fighting those aliens if you are a male (or, in your adult life, for cutting metal or other materials). For this you need other Laser types which can deliver real *power* - far heavier, bulkier and far more expensive than your semiconductor Laser.

- Of course there are constant optimizations and new developments - newer and better semiconductor Lasers are frequently announced. We haven't seen the last of this yet.

Displays and *OLED's*

Take a million or so **LED's**, arrange them in a matrix, and make sure they can be individually addressed - you now have a **display**.

- If you make your display with individual **LED's**, soldered together somehow, you will get an expensive big display with lousy resolution - the kind of boards you see on Times Square or other places that have been taken over by the evil advertising people.
- If you *could* make your **LED's** tiny in the lateral direction *and* all of them close together, i.e. on *one* substrate, *and* for all three **RGB** colors *and* individually addressable, you would have a **flat panel display** that would be great for **TV**, computers, cell phones and laptops because it could have a high energy efficiency.

OLED based **TV**
Courtesy of **Samsung**

- Unfortunately, the possible substrates for inorganic semiconductor **LED's** are far too small (we have only the **III-V** single crystals, essentially **GaAs**, **SiC** and perhaps **Al_2O_3** (= Sapphire)) and those potentiial substrates do not come even close to what would be required.
- Fortunately, an unexpected discovery made accidentially in the **70ties** by **Shirakawa** in Japan (and to some extent by others before him) has helped:
  - There is such a thing as an **organic conductor**, and an **organic semiconductor** leading to an **organic light emitting diode** - an **OLED**.
  - **OLED's** have only be around for less then **5** years, but we already have flat panel displays based on **OLED's** in cell phones and the first ones for **TV** are announced right now.
- Let's be clear about one thing: *Organic semiconductors* right now are still lousy semiconductors (and extremely sensitive to oxygen).
  - They have a tremendous advantage over inorganic semiconductors, however: they can be very cheap and, far more important, they can be deposited at low temperatures by rather simple techniques on cheap and very large substrates and they are easy to pattern.
  - In other words: **OLED** displays are easy to make, the light emission is **OK**, and the product life time is **OK** at present for consumer items where demands are a bit more relaxed.
- What we are witnessing right now is the beginning of a completely new field of *semiconductor materials science and technology*. Who knows where it will end!

9.1.3 Summary to: 9.1 Optoelectronics - General Concerns

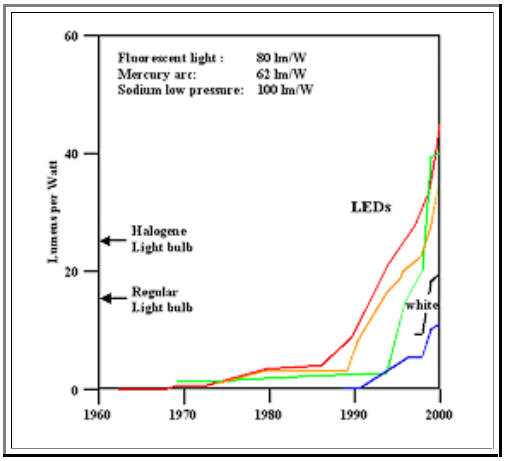
- Optoelectronics has *two* basic branches:
  - **1.** Light in ⇒ electrical signal out:
    - **Optical sensors** as single elements
    - "**CCD**" chips in "megapixel" matrices.
  - **2.** Electricity in ⇒ light out; in two paradigmatic versions:
    - **LED's**
    - **Laser** diodes
- Here we only look at the second branch.
  - The semiconductors of choice are mostly the **III-V's**, usually in single-crystalline perfect thin films.
  - The present day (**2008**) range of wavelength covers the **IR** to near **UV**.
  - Indirect semiconductors like **GaP** can be used too, if some "tricks" are used.

	Wavelength (nm)	**Typical Semiconductor**
Infrared	**880**	**GaAlAs/GaAs**
Red	**660 - 633**	**GaAlAs/GaAs**
Orange to **Yellow**	**612 - 585**	**AlGaInP GaAsP/GaP GaAsP/GaP**
Green	**555**	**GaP**
Blue to **Ultraviolet**	**470 - 395**	**GaN/SiC GaN/SiC InGaN/SiC**

- The index of refraction $n=(\epsilon)^{1/2}$ and thus the dielectric constant ϵ become important
  - Semiconductors have a relatively large index of refraction at photon energies below the bandgap of $n \approx$ **3 - 4**.
  - Diamond has the highest n in the visible region
- The *thermal conductivity* becomes important because for generating light one needs *power* (which we avoided as much as possible for signal processing with **Si**!)
  - Again, diamond has the highest thermal conductivity of all known materials - **5** times better than **Cu**!

Typical Semiconductor	**Dielectric constant**	**Thermal conductivity [W/cm · K]**
Si	11.9	1.5
GaAs	13.1	0,45
GaP	11.1	1.1
GaN	8.9	1.3
SiC	10	5
C (Diamond)	5.8	22

- **LED's** come as cheap little "indicator" lights and recently also as replacement for "light bulbs".
  - Intense white light from **LED's** becomes possible, Advantages: High efficiencies and long life time
  - The key was the "taming" of the **GaN** material system for blue and **UV LED's**.
- **LED's** based on organic semiconductors (**OLED**) are rapidly appearing in **OLED** based displays.
  - *Advantage:* High efficiencies because of active light generation.
    Problem: Product life time; sensitivity to air.



- Semiconductor "Diode" **Lasers** are high-power" **LED's** plus "mirrors"
  - *Advantage:* Small and cheap.
    Problems: Low power, "Quality".

Enabling technology for CD / DVD / Blue ray / ... memory technologies!

Exercise 9.1-1

All Quick Questions to 9.1

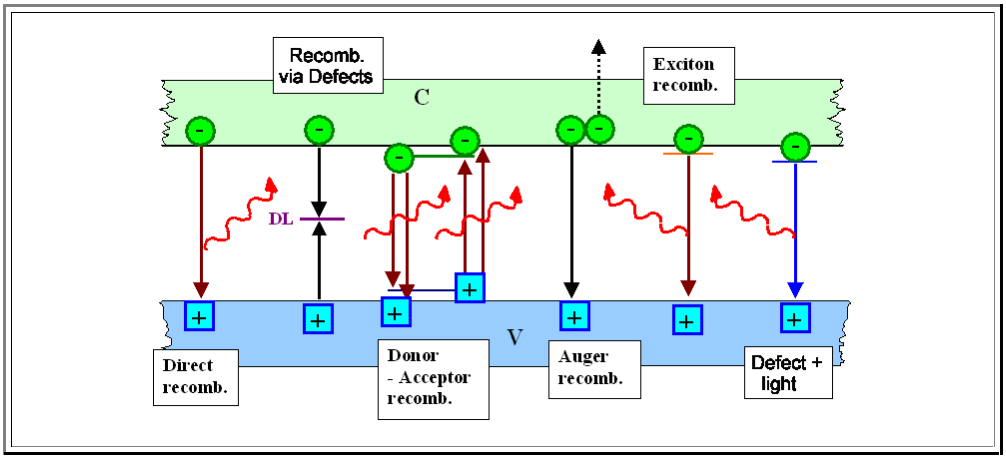
9.2 Important Principles and Technologies

9.2.1 Light Emitting Diodes

Quantum Efficiency

- First thing we need for efficient **LED's** and Lasers is a large **quantum efficiency**, that is a high percentage of radiative recombinations. Since this is so important, let's repeat some of what we have already covered in the sub-chapter before this one.
  - In more elementary semiconductor materials science we learned that in *direct* semiconductors a recombination event happens if an electrons in the conduction band meets a holes from the valence band; the pair recombines and emits a photon. Let's call that process now **direct recombination** or **band-band recombination**- and yes, a photon will still be emitted.
- There is one other recombination channel that you have already learned about: recombination via defects in indirect semiconductors or **defect recombination**. We might distinguish point defects in the bulk acting as **recombination center**, dislocations or interfaces - they all have one thing in common: They introduce some defect levels usually deep in the bandgap (so-called "deep levels").
  - Recombination happens because these *deep levels* can be occupied by electrons and holes. In a simple visualization as shown below, the deep level captures an electron and within time constraint a hole, too - the net effect is the disappearance of an electron-hole pair, i.e. recombination.
- Now we must look at even more more recombination mechanisms, or **recombination channels** as we will call it now, then you have not encountered so far (don't overburden the freshmen and -women!).
  - We have, for example, recombination via defects not only for *indirect* semiconductors. There is no reason why it should not also happen in *direct* semiconductors. This recombination channel then is in competition to the direct recombination but will *not* produce photons.
- The picture below shows a lot more possible recombination channels, all applicable to direct semiconductors, and we cannot possibly understand all of them in detail at present (more to this topic can be found in this link). However, we can try to appreciate a few basic points:

  - In *direct* semiconductors there are a number of recombination channels that do *not* produce light - in particular defect based non-radiative recombination (prominent if there are many defects) and **"Auger" recombination** (prominent at very high carrier concentrations). *That is bad*.
  - In *indirect* semiconductors we may find radiative recombination channels - in particular via "**excitons**" or some special defects. The former is responsible for the use of **GaP** - an *indirect* semiconductor - as material for green **LED's**, the latter for light coming out of **SiC**; also an *indirect* semiconductor. *This is good*.

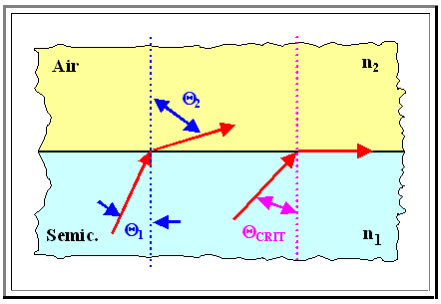


- A high quantum efficiency calls for having mostly *radiating* recombination channels, and that calls for optimizing the material you have according to its peculiar specifics, your technological potentials, and perhaps your budget.
  - It is not necessarily an easy task. That it can be done at an impressive level for some semiconductor materials is demonstrated by the fact that we have working and affordable high-efficiency **LED's** right now.
  - How far this can go, in particular with completely new materials like organic semiconductors out there, remains to be seen.

Getting the Light Out

Unfortunately, a **LED** generates light *inside* the semiconductor where we don't need it. We must get the light out in the air, preferably with some defined radiation characteristics.

- Look at you electronic gadget from somewhere in the room. If you want to see if it is "on", the (usually) green light generated inside the **GaP LED** must come out. Preferably in the shape of a rather large cone as measured with some spherical angle. The first priority, however, is to get the light across the *semiconductor - air interface* at all.
- Remembering **Snell's law** from basic optics, and knowing that n_1 = index of refraction of air = **1**, we have with respect to the situation shown in the picture

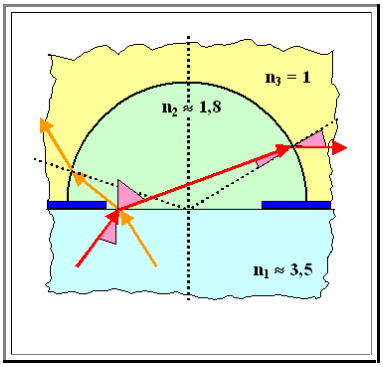


$$\frac{\sin\Theta_2}{\sin\Theta_1} = n_2$$

As also shown in the picture, total (internal) reflection will happen if $\sin\Theta_1 = 1/n_2$. Considering that n_2 = index of refraction of the semiconductor ≈ **3 - 4**, only light emitted towards the surface into a relatively small cone will be able to "come out" at all.

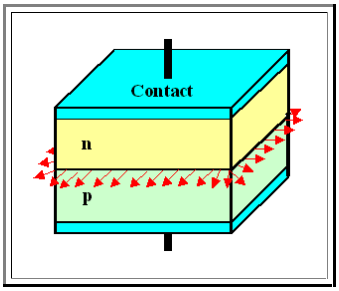
This is not good and causes a serious problem for high efficiencies. A lot of the light produced will simply not be able to come out into the air, and some of the light that makes it will be in the wrong direction (e.g. "backwards"). We must employ some *optical engineering* at this point.

- As always, there are simple and more complex ways to solve the problem.
- A very simple and cheap solution is shown in the picture. The idea is not to allow large jumps in the refractive index across a boundary but to use **index grating**.
- Just one intermediate layer with an index n_3 between n_1 and n_2 may already provide a big improvement. In a simple version with wide-spread application, just adding a "dome" of some epoxy as shown will not only get more light out of the semiconductor, but will also distribute it more or less evenly in a large steric angle - exactly what you want for the little signal lights of your gadgets.



There is one more problem, however. While we may now be able to get most of the light out that reaches the semiconductor surface, we still have the problem of getting the light there in the first place. Considering that light might be generated deep inside the semiconductor, we must take into account that it will be absorbed again before it reaches the surface.

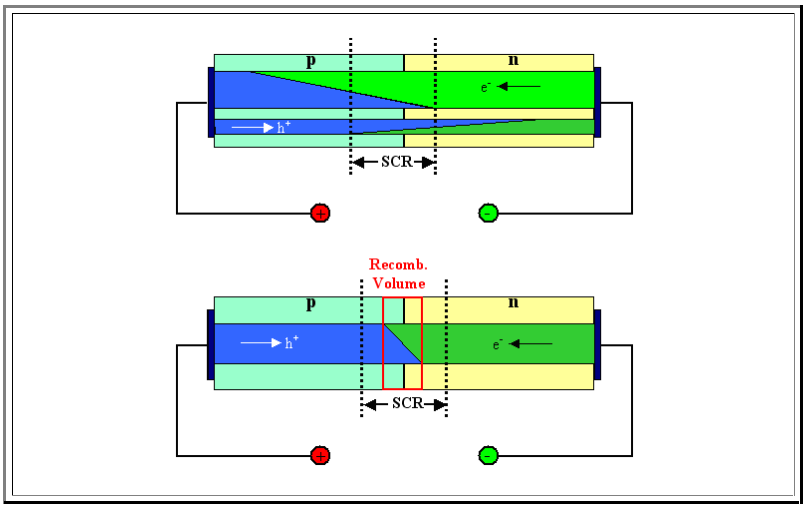
- After all, the energy of the light generated comes from the recombination of an electron and an hole, the energy of the light thus by definition is just what is needed to generate an electron-hole pair.
- In other words: A run-of-the-mill digital camera, operating with a **Si** "***CCD***" chip that detects light by generating electron-hole pairs in **Si**, will also detect the (little bit) of light produced by very good solar cells under open circuit condition after a lot of electron hole pairs have been generated by a high intensity light flash. Even so **Si** is an indirect semiconductor, if you block its usual defect recombination channels, which you do by definition in a very good solar cell, some radiative recombination will occur and detecting it will tell you a lot of what is going on inside your solar cell.



- In yet other words: if we just make a **pn**-junction as shown in the picture in order to produce a simple **LED**, we get some light from around the edges (in all directions; not shown), but most of the light generated in the interior will never make it to the outside world.
- There seems to be no easy fix for this problem. Indeed - there isn't, as long as we use only *one* type of semiconductor. However, if we now turn to *hetero junctions*, we can solve this and other problems very elegantly - and acquire new problems in the process.

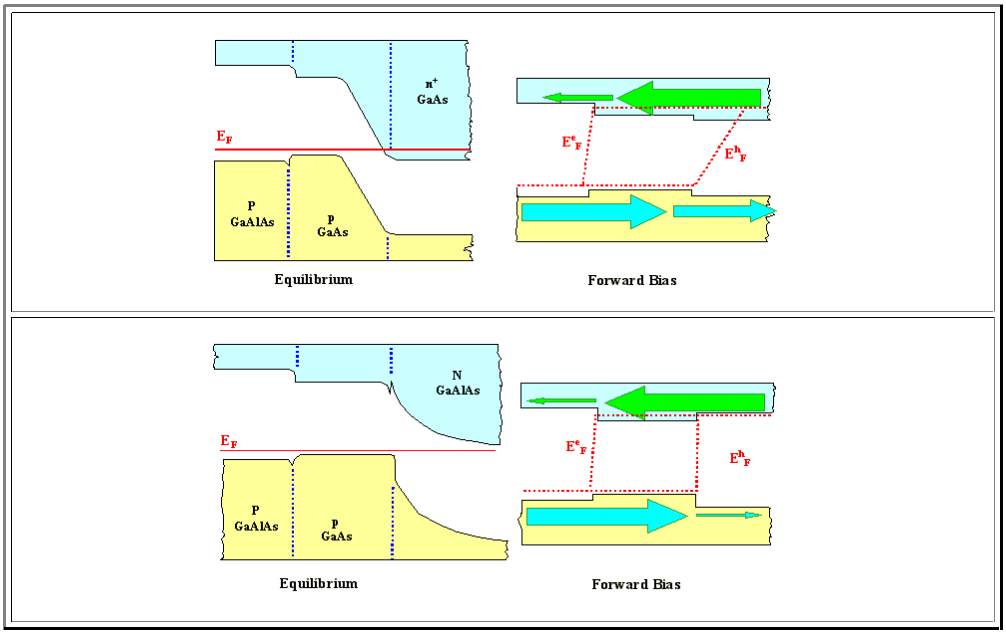
Recombination Volume and Hetero Junctions

- The problems pointed out above together with some other considerations make it clear that we must have control over where *exactly* the light is generated.
  - If we make just a regular **pn**-junction and run it under forward bias, we inject a large forward current of electrons and holes from the **n**- part and the **p**-part of the junction into the other side, respectively . As soon as our majority carriers cross the junction, they become minority carriers and start to recombine with majorities.
  - This is shown schematically in the upper half of the picture below. The small reverse currents have been omitted, and it was assumed that the **n**-part is more heavily doped than the **p**-part (can you tell how?).
  - Recombination happens inside and outside the space charge region (***SCR***) as shown; the exact geometry is given by the life time, the corresponding diffusion length, and God (actually rather Shockley, Read and Hall) knows what else. We have already looked at recombination inside the **SCR** in a simple-minded way, found that it might be substantial, and that it may even be dominating in semiconductors with bandgaps **> 1 eV** or so.
- We might expect therefore that most of the light will be generated in the **SCR**, but we want to do better than that. We want to determine exactly where the light is produced - in a **recombination zone** that we want to *make* to specifications.
  - What that might look like is shown in the lower part of the picture. Again we inject large currents of electrons and holes across a junction with some **SCR**, but they all recombine within a defined recombination zone that is smaller than the **SCR**. The two injection currents are shown as just one current in this case.



- As already mentioned, we can do all this by using *hetero junctions*. We then also enter a rather difficult part of semiconductor physics, and we can not go into much details here - activate the link if you want to learn more.
  - We will just ask two simple questions and look at the answers without doing the reasoning necessary to derive those answers.
- **1.** How do we construct a **pn**-junction between two semiconductors with different bandgaps?
  - The answer is: Exactly as before for a **homo junction**. However, we encounter a problem: There must be a **discontinuity** of the conduction and valence band right at the (metallurgical) junction - and we don't know how to deal with that.
- **2.** What happens if we form a junction between two semiconductors with different bandgaps but of the same doping type - a **Pp** or **nN** junction (with capital **P** and **N** denoting the semiconductor with the larger band gap)?
  - The answer is: same procedure as before: Align the Fermi energies, consider that far away from the junction nothing has changed, move electrons and holes at the junction to the part where their energy is lowered, consider the charge and the associated band bending ... . But again, there must be a *discontinuity* now, and we don't know how to deal with that.
- Cutting a long (and involved) problem short, below are some band diagrams of typical hetero junctions. The necessary discontinuity in the energy has been split into a ΔE_C and a ΔE_V "somehow".

- Shown are both cases: a **Ppn** - single hetero junction and a **PpN** double hetero junction in equilibrium and under extreme forward bias with electron and hole currents indicated.
- For the forward bias case the Fermi energies in a strict sense do no longer exist (they are equilibrium properties, after all); instead "**Quasi Fermi energies**" are drawn in that need not concern us here, but are explained in this link.



- What we are interested in is the **PpN** device with two junctions, biased in forward direction. In this case we inject a large electron current from the **N** part into the **p** part, and a large hole current from the **P** part into the **p** part - just as in the case of the single hetero junction. The difference is that these currents cannot easily get out on the other side of the small bandgap semiconductor because we have an additional energy barrier now
- If the electrons and holes injected into the small bandgap semiconductor from the two wide bandgap sides cannot get out, their only way to oblivion is to recombine in the small band gap semiconductor.

Let's see what we have achieved:

- All the electrons and holes injected via forward bias must recombine in the volume of the small bandgap semiconductor - they can get in, but not out! We now have a defined *recombination volume*.
- The light generated in the recombination volume can easily move into the large bandgap semiconductors because there is no large difference of the index of refraction. And in these semiconductors it will *not be absorbed* at all for obvious reasons.

- We have now solved all the problems mentioned above - and generated the following *new* problems:

  - We now have **phase boundaries** between different crystals, which we have to keep free of defects, in particular we must avoid misfit dislocations. That severely limits the choice of materials.
  - If we use a semiconductor like **GaN** to generate **UV** light as the "small bandgap semiconductor", there is not much choice anymore for the "large" bandgap semiconductors.
  - Ohmic contacts to large bandgap semiconductors (that we still need, of course) are notoriously more difficult to make than to small bandgap semiconductors.
  - More involved technology increases costs.

All things considered, we see that there is far more to making a decent **LED** from some direct (inorganic) semiconductor than just making a **pn**-junction. The field is still progressing rapidly, and only time will tell what we are going to "see" in a few years from now.

Molecular Beam Epitaxy

So far we have already encountered two examples - one, two - of devices where lots of thin layers of different semiconductors were stacked on top of each other; another one will come up

- The question is - how is it done? How do we make complex successions of thin layers from different materials with various thicknesses and doping? Are we going to use variants of the **CVD** process that we used for epitaxial **Si**?
- No - we won't. it would be too cumbersome and slow to grow one layer, then to change gases and whatnot for the next layer, always hoping that all parameters are just right and that no cross-contamination occurs. We use molecular beam epitaxy or ***MBE***, the universal tool developed for just this task

The principle of **MBE** is simple. Suppose you want to make **84 nm** thick layer of $\mathbf{Ga_{0.8}Al_{0.s}As}$ doped with $\mathbf{5 \cdot 10^{17}}$ of **Si** on a plain **GaAs** substrate. All you have to do is:

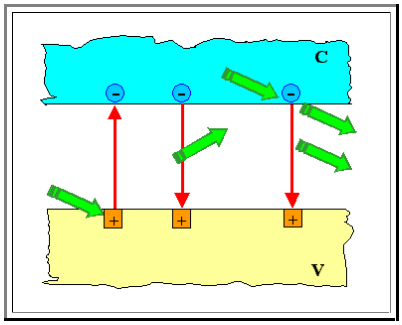
1. Keep your substrate with an atomically clear surface at some suitable temperature.
2. Direct a low-energy flow of all the atoms (or small molecules) needed with just the right rate - atoms per second and **cm^2** - on your substrate.
3. Let 'em react right on your substrate to what you want. They actually have no choice in this example. Each atom will occupy just right position as soon as it is chemisorbed on the substrate.
4. Stop the process suddenly as soon as the required thickness is reached.

- So the principle is easy, the technology, unfortunately, is not. Let's see what we need and how we can meet these needs.
  - The *first point* clearly calls for **ultra-high vacuum** (***UHV***), a heater with temperature control, and some way to clean the surface after the substrate is in place. This can be done, for example, by blasting off (ever so gentle) the surface oxide or whatever you have with an **Ar** ion beam
  - There is no real problem here - provided you have the cash to pay for **UHV**, a notoriously expensive undertaking
  - The *second point* first of all calls for very low flow rates because you want each atom to go only to the substrate - no interaction with all the other atoms going there on the way. This means you must keep the concentration low or, in other words, the mean free path length very large. Nevertheless, the low flow rates still need to be precisely controlled - and now we have a problem: how do we achieve that for all kinds of atoms?
  - By designing sources called "**effusion cells**" or "**Knudsen effusion cells**" The word "effusion" is reminiscent of "diffusion" but the meaning is different. ***Effusion*** means "pouring forth" or, according to the *Webster*: "the escape of gases through minute apertures into a vacuum. That is exactly what an effusion cell does: Provide and keep a very well defined vapor pressure of the material in question inside a "container" and let a defined amount of the vapor escape through a suitable orifice. The key word is "defined" because **MBE** needs to be a high precision technique. Obviously, out **MBE** machine needs several effusion cells; each one optimized for, e.g. **Ga**, **As**, **Al** and **Si**.
  - The *third point* is simple now. All we have to do is to make sure that we have the same rate of incoming atoms everywhere on the substrate (*not* easy for large wafers!) and that the substrate has the optimal temperature to enable the various bonding processes that have to occur on its surface.
  - The *fourth point* is not so simple. We need to incorporate some in-situ technique that monitors the thickness of the growing layer without interfering with the growth process, and we need some way to stop the process. The latter requirement is easy: Just close the orifices of the effusion cell by some mechanical shutter - except that mechanical movements in an **UHV** are not all that simple.
  - The in-situ measurement of choice is mostly "***RHEED***" or "Reflection High Energy Electron Diffraction". The name says it all, and we won't elaborate on this
- Taking everything together, a **MBE** system is a fairly complex piece of equipment but central to many branches of semiconductor technology. The link provides some pictures and a few more details
  - Of course, for serious production you try to find solutions that are less complex and expensive and geared to one process only.

9.2.2 Laser Diodes

Stimulated Emission and Inversion

- In principle, anything that emits electromagnetic radiation can be turned into a "**LASER**" - but what *is* a Laser?
  - The word "*Laser*" was (and of course still is) an ***acronym***, it stands for "**L**ight **A**mplification by **S**timulated **E**mission of **R**adiation" By now, however, it is generally perceived as a standard word *in any language* meaning something that is more than the acronym suggests (and we will no longer write it with capital letters)!
  - A Laser in the direct meaning of the acronym is a black box that emits (=outputs) more light of the same frequency than what you shine (=input) on it - that is the *amplifier* part. But the "*stimulated emission*" part, besides being the reason for amplification, has a second indirect meaning, too: The light emitted is exactly in phase or **phase coherent** (or simply *coherent*) to the light in the input. Coherent then also means that the waves are all traveling in exactly the same direction.
  - Unfortunately, Lasers in this broad sense do not really exist. *Real*Lasers only amplify light with a very specific frequency i.e. **monochromatic** light - its like electronic amplifiers for *one frequency only*.
- A Laser in the *general* meaning of the acronym thus produces intense monochromatic electromagnetic radiation in the wavelength region of light (including infrared and a little ultra violet; there are no sharp definitions) that it is coherent to the (monochromatic) input. If you "input" light containing all kinds of frequencies, only *one* frequency becomes amplified.
  - A Laser in the *specific* meaning of everyday usage of the word, however, is more special. It is a device that produces a coherent beam (=only in direction) of monochromatic light and, at least for semiconductor Lasers, *without some input light* (but with a "battery" or power source hooked up to it). It is akin to an electronic oscillator that works by internally **feeding back** parts of the output of an amplifier to the input for a certain frequency only.
  - Before the advent of hardware Lasers in the sixties, there were already "**Masers**" - just take the "**M**" for "microwave" and you know what it is.
  - And even before that, there was the basic insight or idea behind Masers and Lasers and, as ever so often, it was **A. Einstein** who described the "**S**timulated **E**mission part in **1917/1924**. More to the history of Lasers can be found in the link.
- Obviously, for understanding Lasers, we have to consider *stimulated emission* first, and then we must look at some *feedback* mechanism.



- Understanding stimulated emission is relatively easy; all we have to do is to introduce one more process for the interaction between light and electrons. So far we considered two basic processes, to which we will now give their proper names:
  - **1. Fundamental absorption**, i.e. the interaction of a photon with an *electron* in the *valence band* resulting in a electron - hole pair. This kind of absorption is fundamental because that is the way light is always fundamentally absorbed - not just in semicondcutors but in all materials: an electron is kicked to a higher energy levely a photon that then ceases to exist, i.e. gets absorbed.
  - **2. Spontaneous emission** of a photon by the recombination of an electron-hole pair via direct band-band recombination. The word "spontaneous" refers to a certain stochastic component of this process. This is just the normal radiative recombination we have discussed before and, in a more general sense, the process that *always* generates light, e.g. in "excited" atoms or molecules.
- This is clear enough and we have learned a lot of details about these processes without delving too deeply into quantum mechanics or quantum optics. This is no longer true for the third process concerning the interaction between light and electrons and holes - *stimulated emission*. So let's just describe it:
  - **3. Stimulated emission**is simply a special kind of interaction of a photon with an electron on an "excited" energy level (=electron in the *conduction band* for semiconductors). A photon that happens to "fly by" then *stimulates* the recombination and thus the emission of a *second photon*. The way this happens is always such that the emitted photon is "identical" to the photon stimulating its generation - same energy, same direction, same phase.

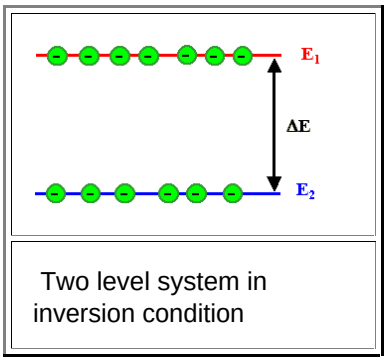
- All three processes are schematically shown in the band diagram on the right. Stimulated emission, in a way of speaking, just takes the randomness out of the spontaneous emission.
- Looking at this picture, you should wonder why one obvious process is missing? How about an electron in the conduction band simply absorbing a photon and getting kicked up the energy scale in the conduction band?
  - In other words: An electron in the conduction band absorbs a photon, moves up the amount $\mathbf{h \cdot \nu}$ in the conduction band, and comes back to the band edge by tranferring its surplus energy to phonons, i.e. heating the lattice.
  - The answer is simple: this process does take place, but is not very strong if we do not have many electrons in the conduction band. It is also not necessary for "lasing", but rather detrimental and we will not consider it any more.
- Stimulated emission is a *resonant process*. It only works if the photons have exactly the right energy and momentum (=wave vector), corresponding to the energy that is released if the electron makes a transition to some allowed lower level. Of course, you have the usual quantum mechanical paradox that the incoming photon is in resonance with something it has not yet created; it is just, so to speak, probing possible outcomes of possible processes.
  - From the resonance argument it follows that two photons have the same wavelenth and are *exactly in phase* with each other. For semiconductors, the photon energy must be pretty much the energy of the band gap, because all conduction band electrons are sitting at the conduction band edge (with some small $\Delta \boldsymbol{E}$, of course), and the only available lower energy level are the free positions occupied by holes at the valence band edge, restricting photons that can be released (and be in resonance with incoming ones) to $\boldsymbol{E_{photon}=h\nu \approx E_g}$.
  - Stimulated emission thus may be seen as a competing process to the fundamental band-band absorption process described above. But while *all* photons with an energy $\mathbf{h\nu > E_g}$ may cause fundamental absorption because there are many unoccupied levels above $\boldsymbol{E_g}$, *only* photons with $\mathbf{h\nu = E_g}$ (give or take some small $\Delta \boldsymbol{E}$) may cause stimulated emission.
- Einstein showed that under "normal" conditions (meaning conditions not too far from thermal equilibrium), *fundamental absorption by far exceeds stimulated emission*. Of course, Einstein did not show that for semiconductors, but for systems with well defined energy levels - atoms, molecules, whatever.
  - However, for the special case that *more* electrons occupy an excited energy state than the related ground state - this condition we will call **inversion** - stimulated emission may dominate the electron-photon interaction processes. Then *two* photons of identical energy and being exactly in phase come out of the system for *one* photon going into the system.
  - The kind of *inversion* we are discussing here should not be mixed up with the *inversion* that turns **n**-type **Si** into **p**-type or vice versa that we encountered before. Same word, but different phenomena!
  - These two photons may cause more stimulated emission - yielding **4**, **8**, **16**, ... photons, i.e. an avalanche of photons will be produced until the excited electron states are sufficiently depopulated.
  - In other words: One photon $\mathbf{h\nu}$ impinging on a material that is in a state of *inversion* (with the right energy difference $\mathbf{h\nu}$ between the excited state and the ground state) may, by stimulated emission, cause a lot of photons to come out of the material. Moreover, these photons are all in phase (and thus travelling in the same direction), i.e. we have now a strong and coherent beam of light as output.
- We are now stuck with two basic questions:
  - **1.** What exactly do we mean with "inversion", particularly with respect to semiconductors?
  - **2.** How do we induce a state of "inversion" in semiconductors?
- Let's look at these questions separately

Obtaining Inversion in Semiconductors

- If you shine **10** input photons on a crystal, **6** of which disappear by fundamental absorption, leaving **4** for stimulated emission, you now have **8** output photons. In the next round you have $\mathbf{2 \cdot (8 \cdot 0{,}4)=6{,}4}$ and pretty soon you have none.
  - Now, if you reverse the fractions, you will get **12** photons in the first round, $\mathbf{2 \cdot (12 \cdot 0{,}6)=14{,}4}$ the next round - you get the idea.
- In other words, the *coherent* amplification of the input light only occurs for a *specific condition*:
  - There must be *more* stimulated emission processes than fundamental absorption processes if we shine light with $\boldsymbol{E=h\nu=E_g}$ on a direct semiconductor - this condition defines "*inversion*" in the sense that we are going to use it.
  - We only look at *direct* semiconductors, because radiant recombination is always unlikely in indirect semiconductors, and while stimulated emission is generally possible, it also needs to be assisted by phonons and thus is unlikely, too.
- Lets first consider some basic situations for defining *inversion* in full generality.For the most simple system, we might have two energy levels $\boldsymbol{E_1}$ and $\boldsymbol{E_2}$, with the lower one ($\boldsymbol{E_1}$) mostly occupied by electrons, the upper one ($\boldsymbol{E_2}$) relatively empty. $\boldsymbol{E_2}$ could be highest energy level still occupied in some atom or molecule, or whatever.

- *Inversion* then means that the number of electrons on the upper level, n_2, is *larger* or at least equal to n_1.
- In *equilibrium*, however, we would simply have

$$\frac{n_1}{n_2} = \frac{D_1}{D_2} \cdot \exp -\frac{\Delta E}{kT}$$



Two level system in inversion condition

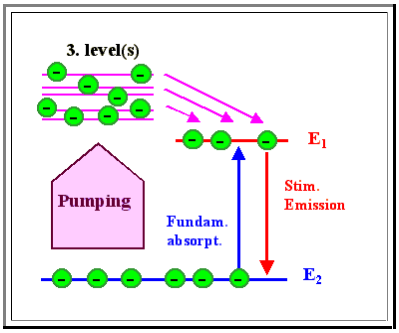
- With $\Delta E = E_1 - E_2$, and $D_{1,2}$= the maximum number of electrons allowed on $E_{1,2}$ (the "density of states").
- In words: In equilibrium we have far more electrons at E_2 than at E_1.

For inversion to occur, we must be *very far from equilibrium* if ΔE is on the order of **1 eV** as needed for visible light. In fact, the systems would have to have a *negative temperature* for such a distribution if we keep the concept of "temperature" that is only properly defined in equilibrium. This is something you should figure out by yourself.

- If we now shine some light on our two-levels system that we brought into inversion, stimulated emission would quickly depopulate the E_2 levels, while fundamental absorption would kick some electrons back. Nevertheless, after some (short) time we would be back to equilibrium having produced a short Laser light pulse at best.
- To keep stimulated emission going, we must move electrons from E_1 to E_2 *by some outside energy source* all the tine - we must "**pump**" the system. Doing this with some other light source providing photons of the only usable energy $\Delta E = E_1 - \Delta E = E_2$ would defeat the purpose of the game; after all that is just the light we want to generate. In semiconductors now, we could inject electrons from some other part of the device but a two-level system is not a semiconductor, so that is not possible here.
- *In short*: Two level systems are no good for practical uses of stimulated emission.

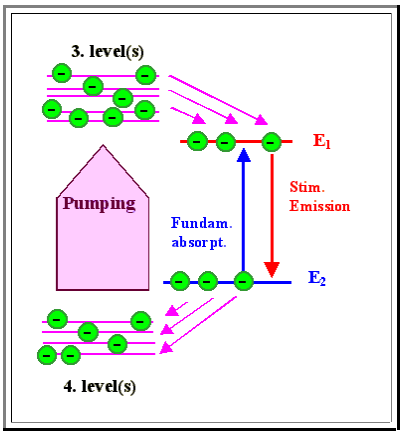
What we need is an *easy* way to move a lot of electrons to the energy E_2. This can be achieved in a **three level system** as shown below (and this was the way it was done with the first ruby Laser).

- The essential trick is to have a whole system of levels - ideally a band - *above* E_2, from which the electrons can descend efficiently to our single level E_1 but not easily back to E_2 where they came from. Schematically, this looks like in the figure on the right.
- The advantage is obvious. We now can take light with a whole range of energies - always larger than ΔE- to "*pump*" electrons up to E_2 via the reservoir provided by the third level(s).
- The only disadvantage is that we have to take the electrons from E_1. And no matter how hard we **pump** (this is the word used for this process), the probability that a quantum of the energy we pour into the system by pumping will find an electron to act upon, will always be proportional to the number (or density) of electrons available for kicking up to E_2. In the three level system this is still at most D_1. If we sustain the inversion, it is at most $0{,}5 \cdot D_1$, because by definition we have at least one-half of the available electrons already on E_2.

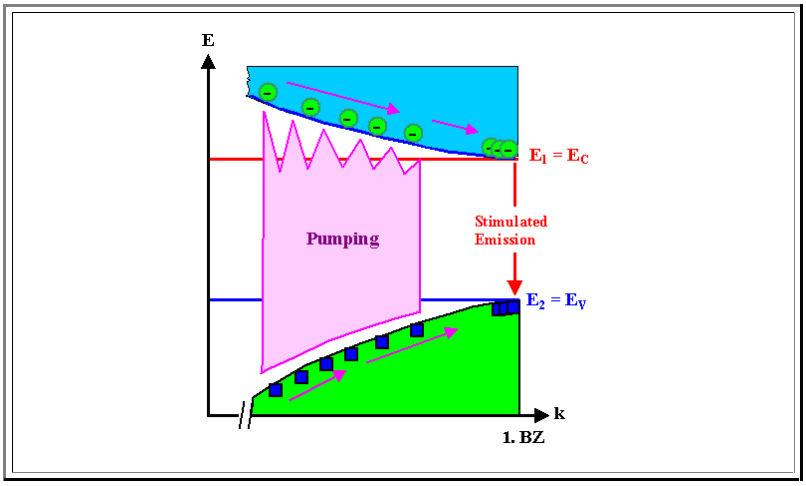


It is clear what we have to do: Provide a *fourth energy level* (even better: a band of energy levels) *below* E_1, where you have a lot of electrons that can be kicked up to E_2 via the third level(s). It is clear that we are talking semiconductors now, but lets first see the basic system:

- We simply introduce a system of energy states below E_1 in the picture from above. We now have a large reservoir to pump from, and a large reservoir to pump to.
- All we have to do is to make sure that pumping is a one-way road, i.e. that there are no (or very few) transitions from the levels **3** to levels **4**.
- This is not so easy to achieve with atoms or molecules, but, as you should have perceived by now, this is exactly the situation that we have in many direct band gap semiconductors. All we have to do to see this, is to redraw the **4**- level diagram at the right as a band diagram. To include additional information, we do this in *k-space*.



We have the following general situation for producing inversion in direct semiconductors:



Electrons may be pumped up from anywhere in the valence band to anywhere in the conduction band - always provided the transition goes vertically upwards in the reduced band diagram.

- The electrons in the conduction band as well as the holes in the valence band will quickly move to the extrema of the bands - corresponding to the levels E_2 and E_1 in the general four level system.
- "*Quickly*" means within a time scale of typicall 10^{-13} **s**. This time scale is so small indeed that it introduces some uncertainties in the energies via the **uncertainty relation** but that need not bother us here.

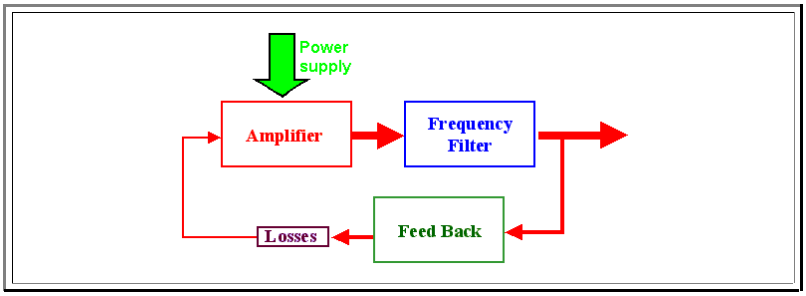
So all we have to do for a semiconductor Laser made from a fancy **LED** is to take the hetero junction configuration introduced before and run enough electrons and holes into the recombination zone defined by the small band gap semiconductor.

- This will raise the concentration of electrons at the conduction band edge of the small energy gap semiconductor and lower the concentration of electrons at the valence band edge (by rasing the concetrations of holes). If the current is large enough, we will always have more electrons at the conduction band edge than at the valence band edge, and this is what we called inversion above.
- In other words: run enough current into your **LED** (without heating it up too much!) and you achieve inversion and thus meet what is called the "**first Laser condition**".
- In orher words: We now can *amplify* light. All we need to do now for or making a Laser is to provide some feed back mechansims in the sense mentioned above.

A General Look at Feedback and Oscillations

So far we have the the amplification of light by stimulated emission. Making a Laser in the conventional sense of the word still requires to produce a light beam with a "battery" and withoutsome "input light".

- This is the same task as to produce an *oscillator* from an electronic amplifier, and the solution of this task for a Laser is achieved in the saem basic way.
- **Feed back** *one* frequency from the output of the amplifier to the input, and make sure it is in phase (or as we say for light, "coherent"). This frequency will be amplified, the feed back increases, it will become more amplified, ..., pretty soon your system is now an oscillator for the frequency chosen. You just need to feed back a large enough part of the output to account for losses that may be occurring in the feed back loop. The essential parts are shown in the drawing



- If you think about this, you will discover a problem. If there is enough signal at the input, the output will go up forever or until a fuse blows - there is no stability in the system
- We need some kind of servo mechanism that adjusts the amplification factor to a value where only the losses are recovered by amplification, so that a stable, preferably adjustable output amplitude is obtained.

This is clear enough for electrical signals, but how do we do this with light? Well, we do everything with **mirrors**:

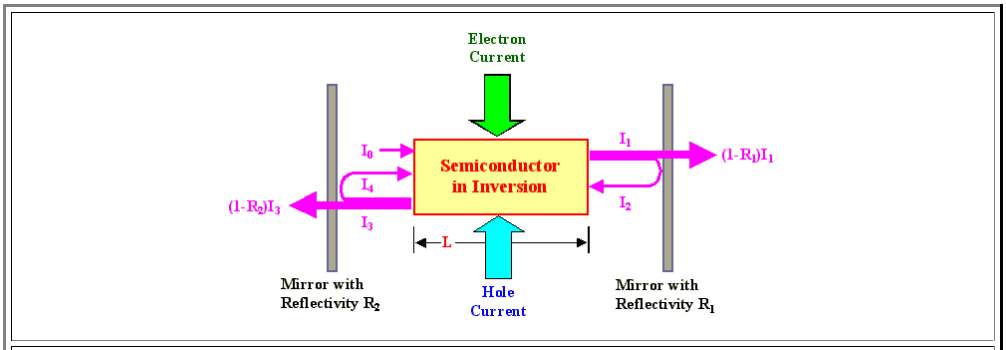
- **1**. The **feed back** part.
- **2.**The **coherency** requirement.
- **3.** The **selection of the frequencies**.
- **4.**The **guidance of the light** including the "*beam shaping*" of the output.

The **4th** point is new - after all, electrical signals go to wherever the wires go, but with light we have to make sure we get a *single* beam coming out in the right direction. We will only look at the general principle of light feedback without worrying much about the three other points other points..

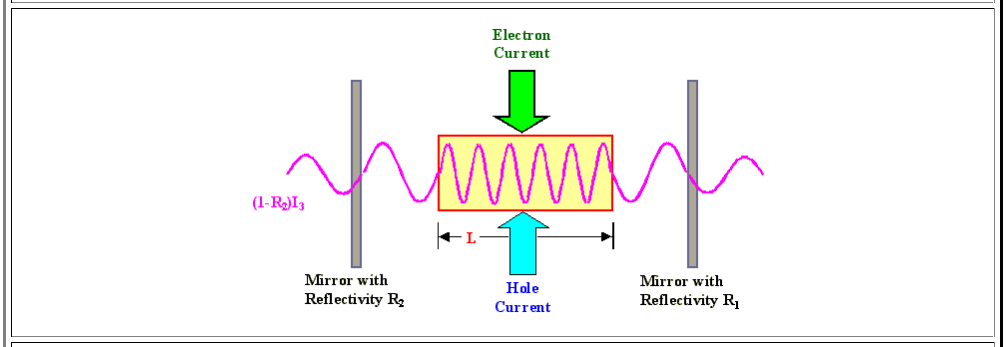
Feedback with Mirrors

All we have to do is to put the piece of semiconductor that is supposed to amplify the light by stimulated emission between two *partially transparent* mirrors

- The whole system then looks like this:



Principle of optical feed back



Formation of a standing wave as a result

Assuming some amplification, i.e. $I_1 = a \cdot I_0$ with $a > 0$, and for reasons of symmetry, $I_3 = a \cdot I_2$, we could now start some calculations.

- We won't, however, because all we could get at that level of simplicity is that the system is either quiet (too little feed back) or "explodes". What will really happen (if no meltdown occurs or a fuse blows first) is is that with increasing output intensity some losses will go up, too and the amplification factor ***a*** comes down until some balance is achieved.

- We have feed back now, but how do we achieve monochromatic light (i.e. what provides the filter) and coherency, not to mention the guidance of the light; i.e. the other points mentioned above?
  - Well, we have already started that we do that with mirrors, too - and we even use the same mirrors we used for the feed back.
- All we have to do is to chose the (optical) distance between the mirrors to be a *multiple of half the wave length* we want. Only in this case a **standing wave** in between the mirrors can be formed. Waves with wavelengths other than the "fitting "ones would simply cancel by interference.
  - This is best seen by looking at what would happen to light with a "wrong" wavelength. Every time it travels through the system, its phase is shifted to some extent, and pretty soon you have a wave with the phase $-\Phi$ for any wave with the phase Φ and destructive interference will cancel everything except waves with phases that fit.
  - This is the same old principle that governs diffraction of electrons or **X**-ray beams in crystals, all musical instruments, and, if you believe Richard **Feynman**, just about everything else, too.

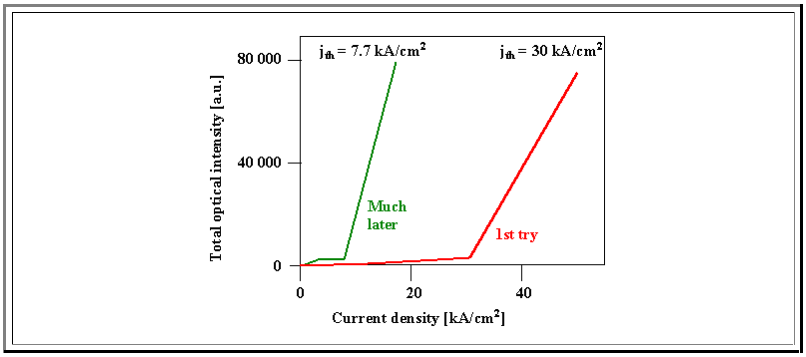
Making a Simple Laser Diode

- If we are not too particular about certain aspects of our semiconductor Laser, we do not even *need* external mirrors. We may now use the effect that annoyed as in the context of getting light out of a semiconductor: that the surfaces of the crystal already act as a partially transparent mirror (with a reflectivity of roughly **30 %** for perpendicular incidence)..
  - The surface mirrors will even be extremely plan-parallel (which is important for obvious reasons) if we obtain them by ***cleaving*** the crystals because with a bit of care they will fracture on the same lattice plane type (usually **{110}** or **{111}**); you can't do much better than this.
- The length crystal ***L***=distance between the "mirrors" then determines the allowed wavelengths as follows:

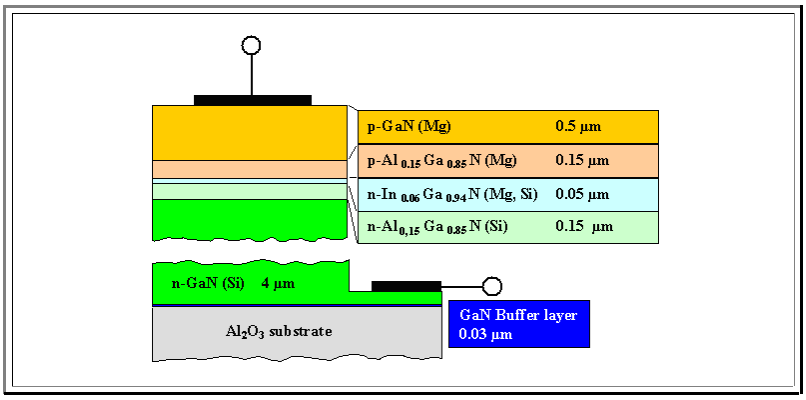
$$L = \frac{m \cdot \lambda_{air}}{2} = \frac{m \cdot \lambda}{2n_{ref}} \qquad m=1, 2, 3, 4,...,$$

$$\lambda = \frac{2 \cdot L \cdot n_{ref}}{m}$$

  - With λ_{air}= wave length in air, λ= wave length in the crystal, n_{ref}= refractive index of the crystal.
- However, with a typical Laser diode that is not extremely small, ***L*** is far larger than the the wavelength λ of visible light and we thus have many allowed wavelengths at this point. The question is how we make sure that we amplify just *one* of the many allowed ones.
  - Easy. Amplification depends on stimulated emission, and stimulated emission *only*works for $h \cdot \nu_{ref} = h \cdot c/\lambda = E_g$. From all the allowed waves that can live within our *Fabry-Perot resonator*, only the one with the proper energy will actually occur.
- If we have made a Laser and now turn it on we will typically encounter the following results:



  - We take the device and run current into it. For low current densities, we have a **LED** producing some light in all directions. Increasing the current produces more light and at some threshold current density j_{th} the losses in the optical feed back provided by our surface mirrors are overcome and lasing starts. We will notice that because now the total amount of light produced goes up sharply with thecurrent density and we also get the quitessential Laser beam
  - If we are lucky and get lasing at all, we probably will encounter a high $j_{th}$and our Laser will cease to operate after a few minutes or hours because all kinds of tricky processes (not always well understood) will kill it.
- Our "simple" Laser may have looked like that (after the basic inventor and investigator Nakamura)

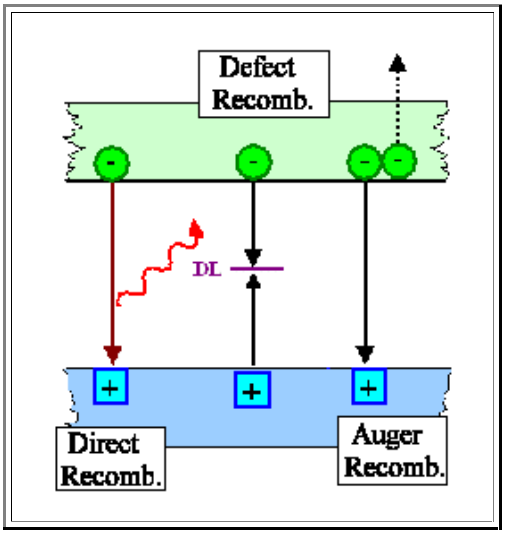


- Now it's time for serious work. Later - and with far more complicated strucures than the one shown above, you may get the second curve in the diagram above with far lower j_{th} and hopefully a Laser life time of many years. Now you may start mass production, ebéventuyll selling your blue Laser diode for a few **€** a piece.
- All in all, we see that the claim from before is correct: the theory of (semiconductor) Lasers is rather complex, but the technology is not. However, that is only true as long as you don't care all that much about the quality of your Laser. In other words as long as you make mostly cheap low-power Lasers or exactly what is needed the mass market for **CD's**, **CVD's**, and watever comes after that.
  - If you think a minute about that: only the advent of **CD's** made progress in **PC's**and Laptops possible, i.e enabled microelectronics, the hard core of semicondutor technology to go on. **CD's** are unthinkable without cheap small Lasers with little power consumption, so semiconductor technolog*ies* are all intertwined by no and move on together, rapidly gaining momentum and complexity.

9.2.3 Summary to: 9.2 Optoelectronics - Important Principles and Technologies

There are always several recombination channels active in parallel

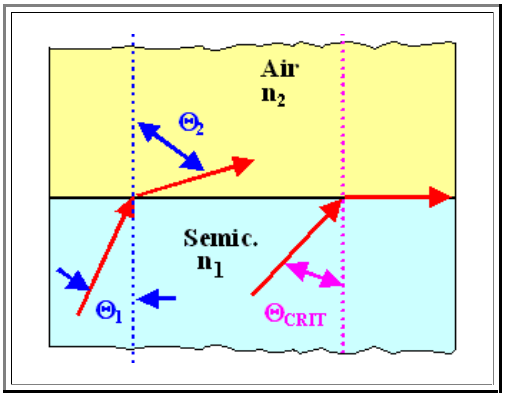
- Direct **band-band recombination**; *producing light*.
- **Defect recombination**; *not* producing light.
- **Auger recombination**; *not* producing light.
- "Exotic" mechanisms like **exciton recombination**; producing light in *indirect* semiconductors like **GaP**



- High efficiency **LED's** need optimized recombination.

Without "tricks" only a fraction of the light produced gets out of the semiconductor

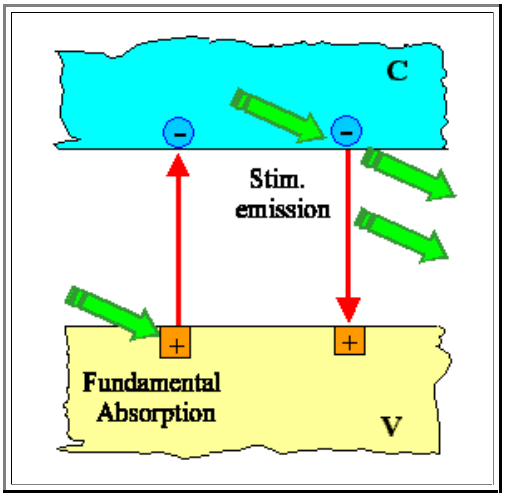
- Index grating is essential
- Avoiding re-absorption is essential
- Defined recombination volumes are important



Hetero junctions of the **NnP** or **NpP** type are the solution, but create problems of their own

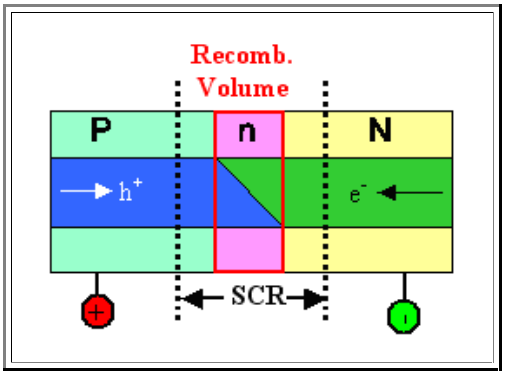
- Hetero-interfaces must be defect free ⇒ Avoid misfit dislocations!

Laser diodes are similar to **LED's** but need to meet two additional conditions



1. The rate of **Stimulated emission**, a new process predicted by A. Einstein concerning the interaction of light and electrons in the conduction band, must be at least as large as the rate of **fundamental absorption**

- *Stimulated emission* results in *two* **fully coherent** photons for *one* incoming photon and thus allows optical *amplification*.
- Strong stimulated emission his requires large non-equilibrium electron concentrations in the conduction band. ⇒strong "pumping" is necessary, moving electrons up to the conduction band just as fast as they disappear by recombination.
- In semiconductor junctions pumping can be "easily" achieved by very large injection currents across a forwardly biased (hetero) junction.⇒ cooling problem!

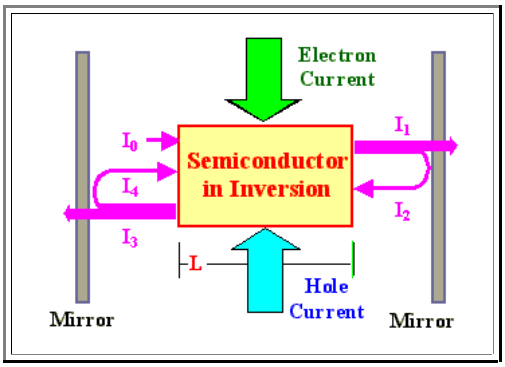


2. There must be some feed-back that turns an (optical) amplifier into an oscillator for one frequency

- *Feed-back* is achieved by partially transparent mirrors.

- Monochromatic output is achieved by the optical resonator forme by two exactly plan-parallel mirrors

- Only wavelengths $\lambda = 2L/i$ (i = integer) that "fit" into the cavity will be able to exist. Together with the condition
  $h\nu = hc/\lambda = E_g$
  the Laser wavelength is given



- Semiconductor Lasers now span the range from **IR** to **UV**; essential materials are all **III-V's**, in particular the **GaN** family.

- **Molecular beam epitaxy** is the deposition method of choice for epitaxial multilayer structures

Exercise 9.2-1

All Quick Questions to 9.

9.3 Summary

9.3.1 Summary to: 9. Optoelectronics

- Optoelectronics has *two* basic branches:
  - **1.** Light in ⇒ electrical signal out:
    - **Optical sensors** as single elements
    - "**CCD**" chips in "megapixel" matrices.
  - **2.** Electricity in ⇒ light out; in two paradigmatic versions:
    - **LED's**
    - **Laser** diodes
- Here we only look at the second branch.
  - The semiconductors of choice are mostly the **III-V's**, usually in single-crystalline perfect thin films.
  - The present day (**2008**) range of wavelength covers the **IR** to near **UV**.
  - Indirect semiconductors like **GaP** can be used too, if some "tricks" are used.

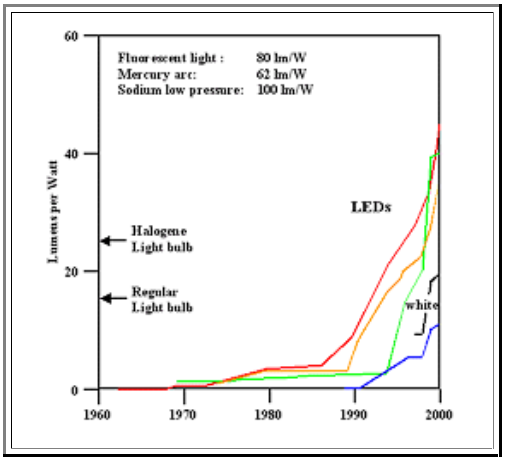
	Wavelength (nm)	**Typical Semiconductor**
Infrared	**880**	**GaAlAs/GaAs**
Red	**660 - 633**	**GaAlAs/GaAs**
Orange to **Yellow**	**612 - 585**	**AlGaInP GaAsP/GaP GaAsP/GaP**
Green	**555**	**GaP**
Blue to **Ultraviolet**	**470 - 395**	**GaN/SiC GaN/SiC InGaN/SiC**

- The index of refraction $n=(\epsilon)^{½}$ and thus the dielectric constant ϵ become important
  - Semiconductors have a relatively large index of refraction at photon energies below the bandgap of $n \approx 3 - 4$.
  - Diamond has the highest n in the visible region

Typical Semiconductor	**Dielectric constant**	**Thermal conductivity [W/cm · K]**
Si	11.9	1.5
GaAs	13.1	0,45
GaP	11.1	1.1
GaN	8.9	1.3
SiC	10	5
C (Diamond)	5.8	22

- The *thermal conductivity* becomes important because for generating light one needs *power* (which we avoided as much as possible for signal processing with **Si**!)
  - Again, diamond has the highest thermal conductivity of all known materials - **5** times better than **Cu**!

- **LED's** come as cheap little "indicator" lights and recently also as replacement for "light bulbs".
  - Intense white light from **LED's** becomes possible, Advantages: High efficiencies and long life time
  - The key was the "taming" of the **GaN** material system for blue and **UV LED's**.
- **LED's** based on organic semiconductors (**OLED**) are rapidly appearing in **OLED** based displays.
  - *Advantage:* High efficiencies because of active light generation.
    Problem: Product life time; sensitivity to air.

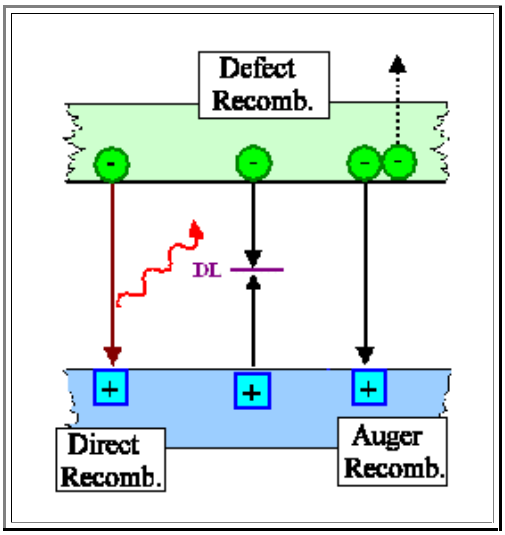


- Semiconductor "Diode" **Lasers** are high-power" **LED's** plus "mirrors"
  - *Advantage:* Small and cheap.
    Problems: Low power, "Quality".

Enabling technology for CD / DVD / Blue ray / ... memory technologies!

There are always several recombination channels active in parallel

- Direct **band-band recombination**; *producing light*.
- **Defect recombination**; *not* producing light.
- **Auger recombination**; *not* producing light.
- "Exotic" mechanisms like **exciton recombination**; producing light in *indirect* semiconductors like **GaP**



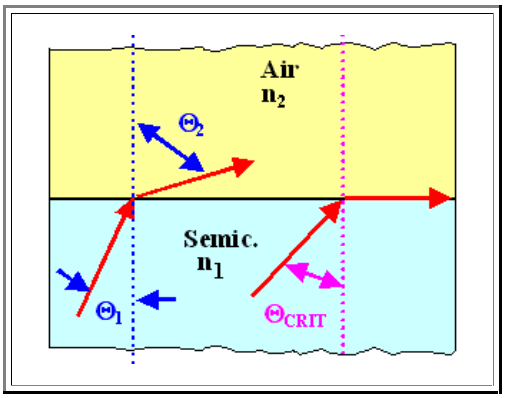
High efficiency **LED's** need optimized recombination.

Without "tricks" only a fraction of the light produced gets out of the semiconductor

- Index grating is essential
- Avoiding re-absorption is essential
- Defined recombination volumes are important

Hetero junctions of the **NnP** or **NpP** type are the solution, but create problems of their own

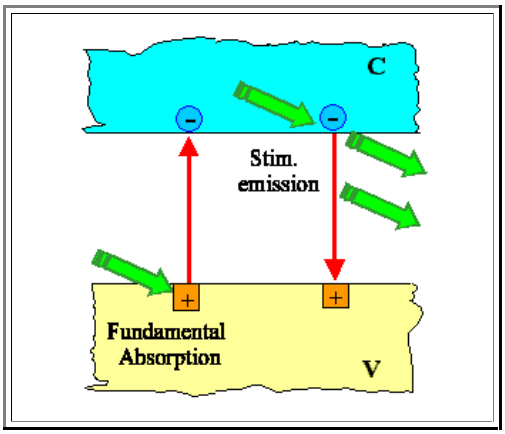
- Hetero-interfaces must be defect free ⇒ Avoid misfit dislocations!



Laser diodes are similar to **LED's** but need to meet two additional conditions

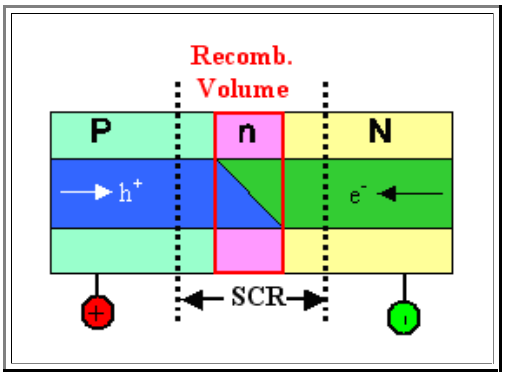
1. The rate of **Stimulated emission**, a new process predicted by A. Einstein concerning the interaction of light and electrons in the conduction band, must be at least as large as the rate of **fundamental absorption**

- *Stimulated emission* results in *two* **fully coherent** photons for *one* incoming photon and thus allows optical *amplification*.
- Strong stimulated emission his requires large non-equilibrium electron concentrations in the conduction band. ⇒strong "pumping" is necessary, moving electrons up to the conduction band just as fast as they disappear by recombination.
- In semiconductor junctions pumping can be "easily" achieved by very large injection currents across a forwardly biased (hetero) junction.⇒ cooling problem!



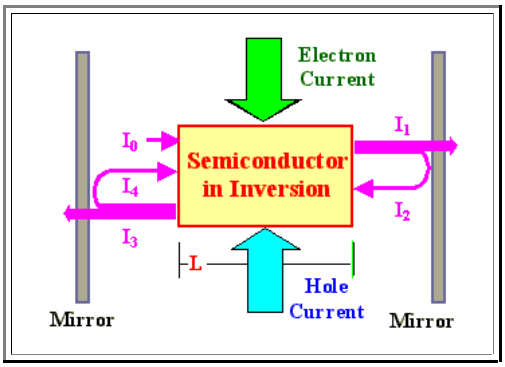
2. There must be some feed-back that turns an (optical) amplifier into an oscillator for one frequency

- *Feed-back* is achieved by partially transparent mirrors.



- Monochromatic output is achieved by the optical resonator forme by two exactly plan-parallel mirrors

- Only wavelengths $\lambda = 2L/i$ (i=integer) that "fit" into the cavity will be able to exist. Together with the condition
  $h\nu = hc/\lambda = E_g$
  the Laser wavelength is given



- Semiconductor Lasers now span the range from **IR** to **UV**; essential materials are all **III-V's**, in particular the **GaN** family.

- **Molecular beam epitaxy** is the deposition method of choice for epitaxial multilayer structures

Exercise 9.3-1

All Quick Questions to 9.